Maximized redundant and synergistic information transfers predict the rise in the output gene expression noise in a generic class of coherent type-1 feed-forward loop networks

Md Sorique Aziz Momin* and Ayan Biswas[†]

*Department of Chemistry, Bose Institute,
93/1 A P C Road, Kolkata 700009, India*

(Dated: September 12, 2021)

Abstract

We apply the partial information decomposition principle to a generic coherent type-1 feed-forward loop (C1-FFL) motif with tunable direct and indirect transcriptional regulations of the output gene product and quantify the redundant, synergistic, and unique information transfers from the regulators to their target output species. Our results which are obtained within the small-noise regime of a Gaussian framework reveal that the redundant and synergistic information transfers are antagonistically related to the output noise. Most importantly, these two information flavors are maximized prior to the minimization and subsequent growth of the output noise. Therefore, we hypothesize that the dynamic information redundancy and synergy maxima may possibly be utilized as efficient statistical predictors to forecast the increasing trend of the fluctuations associated with the output gene expression dynamics in the C1-FFL class of network motifs. Our core analytical finding is supported by exact stochastic simulation data and furthermore validated for a diversified repertoire of biologically plausible parameters. Since, the output gene product serves essential physiological purposes in the cell, a predictive estimate of its noise level is supposed to be of considerable biophysical utility.

I. INTRODUCTION

Cells maintain an intricate relationship with their environments where they come across different stimuli such as usable nutrients, detrimental molecules, external stresses (e.g., heat-shock and osmotic-pressure), etc. In order to survive in these convoluted circumstances, cells respond to the incoming signals by producing appropriate proteins/enzymes at the right time and in economic amounts to achieve desired physiological goals with appreciable precision. A typical *E. coli* and yeast cells contain roughly 4×10^6 and 4×10^9 proteins, respectively, with individual types of protein numbered in $\mathcal{O}(100)$ copies [1–3]. The instructions for synthesizing this large number of different proteins come from different genes: the coding regions of DNA. The information processing in between environmental signals and genes is mediated by a special type of catalytic protein, namely, transcription factor (TF). At first, the inducer signal activates certain TFs which, upon binding to the regulatory promoter regions

* soriqueaziz@gmail.com

† Corresponding author: ayanbiswas19@gmail.com

of their enslaved genes, increase (activation) or decrease (repression) the transcription rates. This is done by altering the binding probabilities of the target genes with the RNA polymerase. Thus, the production rates of mRNAs and consequently that of the end-products (proteins/enzymes) are modified at par with the cellular necessities. Some of these gene products can act as TFs thereby regulating other genes both up and downstream along the signaling pathway. This TF-gene interaction which forms an arrayed structure is generally known as the gene transcription regulatory network (GTRN) [2].

The GTRN of an organism, in terms of its size and activities, is very complex in nature and contains different architectural patterns, e.g., cascades, feed-forward and feed-back loops, etc. Some of these network patterns, often of small sizes, are over-represented in the GTRN of a particular organism than in an equivalent randomized network. These recurring small networks are known as the network motifs [4]. Associating three genes/gene products as different nodes and their activatory interactions as the inter-node edges, there appears an ensemble of thirteen three-node random subgraphs. Out of these different possible network configurations, the C1-FFL is found to appear significantly in the GTRNs of *E. coli*, yeast, and many other higher organisms. Hence, for these GTRNs, the C1-FFL is qualified as the only network motif. Within this motif, TF “S” activates another TF “X” whereas both of them activate the production from the output gene “Y”. Here, for simplicity, we use the same notations for the biochemical species (genes and gene products) and their population levels. Their meanings should be inferred depending upon the context of their usage. Therefore, the production of gene product Y is accomplished through two parallel regulatory paths: direct regulation ($S \rightarrow Y$) and another indirect regulation ($S \rightarrow X \rightarrow Y$). In the terminology of networks, the former regulatory process forms a one-step cascade whereas the latter constitutes a two-step cascade. In the absence of an edge between X and Y, one obtains another additional one-step process: $S \rightarrow X$ apart from the previously mentioned $S \rightarrow Y$. The resulting pattern out of these two one-step branches is: $Y \leftarrow S \rightarrow X$ otherwise known as the fan-out network. For a generic C1-FFL with tunable direct and indirect regulatory controls over its output gene, the two-step cascade and fan-out structure may be considered as two of its extreme realizations where the direct and indirect decoding rate, respectively, are zero! Therefore, except for these two special patterns, a typical C1-FFL motif is capable of decoding the upstream signal through two parallel regulatory pathways. The decoding process is perfect when their individual shares in the output gene expression level are equal

in amounts.

This paper intends to explore the information processing capabilities of a generic C1-FFL motif by keeping the interplay of its direct and indirect decoding strengths in perspective. To be specific, our objective is to build a multivariate information-theoretic framework to compare different types of the intergenetic information transfer and their relationships with the output gene expression noise for the generalized C1-FFL class of motifs. According to our network modification scheme which will be formally introduced in Sec. II C, each of these C1-FFL configurations are genotypically different by means of different direct and indirect production rates (decoding strengths) of the parallel regulatory pathways. In contrast, our biochemical construct envisages a situation in which all the inter-convertible structures of the C1-FFL type may manage to maintain fixed expression levels of the three constituent genes. This constraint on the gene products' abundances implies that different C1-FFLs are equitable on phenotypic footing although being distinct in the underlying genotypic spaces. To add a biological perspective to our current information-theoretic question, it is imperative to mention that the C1-FFL motif performs essential physiological functions in living systems. To exemplify, C1-FFL employing a SUM signal integration logic is operational in making flagellar motor proteins in *E. coli* [5]. The network modification scheme in our work is somewhat broadly inspired, albeit on a theoretical front, by the foundational experiments performed on the *arabinose* utilization system in *E. coli*. There, Mangan and colleagues have compared a C1-FFL module with one of its extreme network instances, namely, the simple regulation which is otherwise known as the fan-in network. Within the fan-in network, the two regulators do not interact transcriptionally with each other [6] and interestingly, just like the two-step cascade and fan-out network, the fan-in pattern is also a member of the previously mentioned ensemble of thirteen three-node random networks. Although, the biological impact of our hypothesized construct is further subject to experimental scrutiny, the prescribed theoretical model is a rich ground to explore the information processing implications of plausible modular topologies within the broad C1-FFL family.

The biochemical kinetics associated with the synthesis and degradation of different gene products are afflicted with noise in primarily two ways: (i) due to the low copy numbers of the biomolecules participating in various chemical reactions and (ii) fluctuations of global origin flowing downstream to the enslaved genes. Therefore, noise is an inherent element to living systems, specifically in the context of a single cell, and thus becomes an indispensable

feature of cellular functionalities at both local and global levels [7–11]. To coherently describe the noisy gene expression dynamics at the single-cell level, one needs to bring in the tools and techniques of stochastic processes. An information-theoretic framework [12–14] is one of such routes which coarse-grains the biochemical signaling process as a noisy input/output communication channel. The application of information theory is diversified as it spans the domain of neuroscience, gene transcription regulation as well as intercellular communication, to name a few [15–26]. In this paper, we study the effects of interspecies interactions within an ensemble of C1-FFLs with tunable direct and indirect decoding strengths on the efficient maintenance of the directional information flow. Transfer entropy (TE) offers us the mathematical machinery to quantify this information dynamics which detect the asymmetric statistical dependencies between different biochemical species in C1-FFL patterns [27, 28]. In recent years, applications of the TE-based measures have successfully generated new insights in a number of physical and physiological systems, e.g., Ising models, epileptic brains, *HeLa* cell-cultures, etc. [29–31]. On the other side, the scope of measuring entropy transfer in the GTRN motifs is still unexplored to a significant extent. The current paper is intended to make progress in this uncharted domain of research. Such an information-theoretic construct fundamentally involves measuring correlations between gene expression levels from two nearby time points at steady state in contrast with single-time-point measurements. The metric of TE goes beyond measuring the nonlinear correlation between different gene products and quantifies the extent to which the information from the past states of a set of source species can predict the future state of the target species. Thus, TE provides us with a mechanistic understanding of the control over the fluctuations in the target gene expression using the information sources. For our model C1-FFL, as will be demonstrated in Sec. III A, the information transferred from the regulators to their enslaved gene product often comes from unique sources. Other types of information flow may involve either the master- or the co-regulator depending upon the relative values of their information transfer capacities. Additionally, the regulators may even complement each other to transfer information. Here, we are concerned with the question: How different types of information transfer are interrelated with the output noise? The dynamical C1-FFL model will be introduced in Sec. II A. To establish the required analytical framework in discrete-time domain, we apply a stationary multivariate vector autoregressive (MVAR) process [31–33] as deliberated in the Appendix which also contains the details of the stochastic simulation

method. The information-theoretic measures will be formally introduced in Sec. II B. The scheme for the architectural modification and the rationale behind our choices of parameters are contained in Sec. II C. Sec. III A provides our main analytical results supported by exact stochastic simulation data. Secs. (III B-III E) provide further analytical support for our core findings within a broad regime of parametric realization. Finally, we conclude our article with a summary of our findings in Sec. IV.

II. THE MODEL

A. Gene expression dynamics in a coarse-grained C1-FFL motif

Using the Langevin formalism [24, 34, 35], we model the dynamics of the populations (random variables) of the three gene products (S, X, and Y) which constitute the C1-FFL motif. These population (gene expression) levels are expressed in the unit of copy numbers in an effective cellular volume of magnitude unity. The set of dynamical equations is as follows:

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

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

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

f_s , $f_x(S)$, and $f_y(S, X)$ are the production/synthesis rates of S, X, and Y, respectively. These input rate functions may depend upon the regulators' populations in a (non)linear fashion. Usage of the Hill function to model transcriptional regulation is a standard practice supported by experimental data [2, 16, 36–38]. To exemplify in the present model, the synthesis rate for $S \rightarrow X$ is taken: $f_x(S) = k_{sx}[S/(K_{sx} + S)]$ where k_{sx} is the synthesis rate parameter interpreted as the maximal production rate from the gene promoter of X. The remaining part in f_x describes the promoter occupancy probability of gene X by the activated TF S. The activation coefficient K_{sx} is numerically equivalent to the copy number of the activated TF S which is needed to reach the half-maximal expression of X [2]. The production rate for the reaction channel: $\phi \rightarrow S$ is taken to be independent of any TF population external to the three-node subgraph and hence, we have $f_s = k_s$. The combined output production rate considering an additive logic operating upon the di-

rect and indirect regulatory pathways is $k_{sy}[S/(K_{sy} + S)] + k_{xy}[X/(K_{xy} + X)]$. k_{sy} (k_{xy}) is the production rate parameter for $S \rightarrow Y$ ($X \rightarrow Y$) whereas K_{sy} (K_{xy}) is the corresponding activation coefficient. The degradation rates of S, X, and Y are modelled as linearly proportional to their corresponding populations, i.e., $\mu_s S$, $\mu_x X$, and $\mu_y Y$, respectively, where μ_s , μ_x , and μ_y are the related degradation rate parameters. To represent the inherently stochastic synthesis and degradation reactions, the Langevin description includes the Gaussian-distributed white noise terms $\xi_s(t)$, $\xi_x(t)$, and $\xi_y(t)$ for species S, X, and Y, respectively. Their statistical properties are as follows [39–42]: $\langle \xi_z(t) \rangle = 0$, $\langle \xi_z(t) \xi_{z'}(t') \rangle = \langle |\xi_z(t)|^2 \rangle \delta_{zz'} \delta(t - t')$ with $\{z, z'\} \in \{s, x, y\}$ being the respective indices for gene products S, X, and Y. $\delta_{zz'}$ and $\delta(t - t')$ represent the Kronecker- and Dirac-delta functions, respectively. $\langle \dots \rangle$ indicates steady-state ensemble-averages whereas t and t' stand for two different time points. The steady-state ensemble-averaged noise strength is defined as $\langle |\xi_z(t)|^2 \rangle = \langle f_z(Z) \rangle + \mu_z \langle Z \rangle = 2\mu_z \langle Z \rangle$ where, $Z \in \{S, X, Y\}$ are the populations of different gene products. The last equality comes from the steady-state ensemble-averaged versions of Eqs. (1a-1c). The steady-state expressions of the population levels of the gene products are: $\langle S \rangle = k_s/\mu_s$, $\langle X \rangle = (k_{sx}/\mu_x) \langle f_x(S) \rangle \approx (k_{sx}/\mu_x) f_x(\langle S \rangle) = (k_{sx}/\mu_x) [\langle S \rangle / (K_{sx} + \langle S \rangle)]$. The approximation is reasonable within the small-noise regime [24]. Similarly, $\langle Y \rangle = \langle Y(S) \rangle + \langle Y(X) \rangle$ where $\langle Y(S) \rangle \approx (k_{sy}/\mu_y) [\langle S \rangle / (K_{sy} + \langle S \rangle)]$ and $\langle Y(X) \rangle \approx (k_{xy}/\mu_y) [\langle X \rangle / (K_{xy} + \langle X \rangle)]$. Evidently, $\langle Y(S) \rangle$ and $\langle Y(X) \rangle$ are the steady-state ensemble-averaged populations of the output gene product which are synthesized by the master- and co-regulator, respectively. The units for the degradation rate parameters are min^{-1} . The synthesis rate parameters are expressed in the units of $\text{c.n.} \times \text{min}^{-1}$ where c.n. stands for copy number (per an effective cellular volume of magnitude unity). The activation coefficients have the units of c.n. as in the case for population levels. Coming back to the random forcing terms ($\xi_z(t)$), we have used Gaussian noise with zero means and zero cross-species correlations as well as zero temporal correlations. These are reasonable assumptions for coarse-grained transcription networks which incorporate birth-death types of reactions like our present model [24]. At steady state, noise strengths are conceived to originate equally from the stochastic synthesis and degradation events of each of the biochemical species. Besides, usage of constant steady-state noise strengths is an approximation adopted to ease the analytical tractability of the forthcoming computations of the (co)variances of various gene products [see the Appendix]. There, we will also discuss the effect of the time-discretization of Eqs. (1a-1c) on the noise

strengths.

B. Information-theoretic measures

While explicitly dealing with the temporal structure of gene expression data, the one-time-point/static mutual information (MI) can be upgraded to evaluate the association between two random variables from two different time points, e.g., Y_t and Y_{t+1} . This enables us to quantify the dynamics of MI. In both the static and dynamic cases, MI may involve more than two random variables corresponding to distinct biochemical species. By judiciously combining two- and three-variable two-time-point MIs together, we can alleviate the non-directional aspect of the static information. The resulting metric of TE quantifies the information transfer from the input to the output level of an information channel. The joint effect of two source species in transferring information to a single target species is also possible to compute using the multivariate TE. In our present case, TE may be used to predict the future state of the output (i.e., Y_{t+1}) using the combined knowledge of its present state (Y_t) and that of another source species (e.g., S_t) beyond what can be predicted by Y_t alone [27, 28, 43, 44]. Therefore, TE from S to Y is expressed as:

$$T_{s \rightarrow y} \equiv I(Y_{t+1}; S_t | Y_t) \stackrel{\text{def}}{=} I(Y_{t+1}; Y_t, S_t) - I(Y_{t+1}; Y_t) = \frac{1}{2} \log_2 \left[\frac{\Sigma(Y_{t+1} | Y_t)}{\Sigma(Y_{t+1} | S_t, Y_t)} \right]. \quad (2)$$

Here, $I(\dots; \dots)$ are various dynamic MIs whereas $I(\dots | \dots)$ represent different conditional two-time-point MIs. The partial variances $\Sigma(\dots | \dots)$ are computed using Eqs. (A.10a, A.10b) given in the Appendix. The last equality in terms of the time-lagged partial variances is naturally established for the present Gaussian system [31, 32, 45]. Similarly, TE from X to Y is computed using:

$$T_{x \rightarrow y} \equiv I(Y_{t+1}; X_t | Y_t) \stackrel{\text{def}}{=} I(Y_{t+1}; Y_t, X_t) - I(Y_{t+1}; Y_t) = \frac{1}{2} \log_2 \left[\frac{\Sigma(Y_{t+1} | Y_t)}{\Sigma(Y_{t+1} | X_t, Y_t)} \right]. \quad (3)$$

Lastly, TE due to the joint action of S and X targeting Y is the following:

$$T_{sx \rightarrow y} \equiv I(Y_{t+1}; S_t, X_t | Y_t) \stackrel{\text{def}}{=} I(Y_{t+1}; Y_t, X_t, S_t) - I(Y_{t+1}; Y_t) = \frac{1}{2} \log_2 \left[\frac{\Sigma(Y_{t+1} | Y_t)}{\Sigma(Y_{t+1} | S_t, X_t, Y_t)} \right]. \quad (4)$$

To reiterate, the TE metrics: $T_{s \rightarrow y}$, $T_{x \rightarrow y}$, and $T_{sx \rightarrow y}$ measure the information contents transferred from S to Y, from X to Y, and jointly from S and X to Y, respectively.

According to the prescription of *Partial information decomposition* (PID), $T_{sx \rightarrow y}$ is dissected into four distinct/independent non-negative and non-overlapping information transfer elements. Unique information $U_{s \rightarrow y}$ is transferred to Y from S alone. Similarly, unique information $U_{x \rightarrow y}$ is directed to Y only from X. Redundant information transfer $R_{sx \rightarrow y}$ is due to that regulator which transmits a lesser amount of information in comparison with the other information source. On the other hand, $S_{sx \rightarrow y}$ is the synergistic information transmitted complementarily by the pair of sources. As the name suggests, synergistic information is the amount of information processed by the source variables/species operating in unison while predicting their target. The set of PID elements are related to the TE metrics by the following set of equations [31, 32, 46–48]:

$$T_{s \rightarrow y} = U_{s \rightarrow y} + R_{sx \rightarrow y}, \quad (5a)$$

$$T_{x \rightarrow y} = U_{x \rightarrow y} + R_{sx \rightarrow y}, \quad (5b)$$

$$T_{sx \rightarrow y} = U_{s \rightarrow y} + U_{x \rightarrow y} + R_{sx \rightarrow y} + S_{sx \rightarrow y}. \quad (5c)$$

In the transfer entropy space, $T_{s \rightarrow y}$ and $T_{x \rightarrow y}$ are embedded within $T_{sx \rightarrow y}$. The non-overlapping portions of $T_{s \rightarrow y}$ and $T_{x \rightarrow y}$ contribute the unique information transfers, i.e., $U_{s \rightarrow y}$ and $U_{x \rightarrow y}$ whereas the common (overlapping) space provides $R_{sx \rightarrow y}$. $S_{sx \rightarrow y}$ lies beyond the combined space of $T_{s \rightarrow y}$ and $T_{x \rightarrow y}$ and is contained within the space of $T_{sx \rightarrow y}$ [31]. To completely specify the PID elements in Eqs. (5a-5c), the redundant information transfer is defined as the minimum of the single-source-species TEs considering a unique target, i.e., $T_{s \rightarrow y}$ and $T_{x \rightarrow y}$. Thus, the formula to evaluate the redundant information transfer stands as [32]:

$$R_{sx \rightarrow y} \stackrel{\text{def}}{=} \min\{T_{s \rightarrow y}, T_{x \rightarrow y}\}. \quad (6)$$

Finally, using the formalism of the *minimum mutual information* PID (MMI PID), i.e., Eqs. (5a-5c, 6), we can quantify the unique, synergistic, and redundant information transfers at steady state for the C1-FFL class of motifs. These working formulae are common for our (approximate) analytical and (exact) stochastic simulation-based data. Due to the usage of base “2” for the logarithm functions, the TEs and the PID elements are measured in the units of “bits”. Besides, we compute the noise level in the output gene expression using the concept of *coefficient of variation* (CV). The metric of interest is $CV_y^2 \stackrel{\text{def}}{=} \Sigma(Y_t)/\langle Y_t \rangle^2$. We use Eq. A.7f to analytically evaluate the output noise which is also measured exactly from the stochastic simulation. The output noise has the units of c.n.⁻¹. Sec. III A presents

these datasets which form the basis of our conjecture regarding the connection between the i/o information transfers and the output gene expression noise in the C1-FFL modular configurations.

C. Architectural modification scheme for the C1-FFL class of networks

In one of the extreme configurations of the generic C1-FFL where the direct transcriptional regulation is absent ($k_{sy} = 0 \Rightarrow \langle Y(S) \rangle = 0$), Y is entirely produced by the TF X (i.e., $\langle Y \rangle = \langle Y(X) \rangle$). Now, increasing $\langle Y(S) \rangle$ by one copy at a time by means of an appropriate increment in k_{sy} and simultaneously decreasing $\langle Y(X) \rangle$ by the same amount through lowering k_{xy} , we reach a set of production rates where all of Y is produced by S alone (i.e., $\langle Y \rangle = \langle Y(S) \rangle$ with k_{xy} and $\langle Y(X) \rangle = 0$). In this other extreme realization of the C1-FFL motif, TF X does not share an edge with the output species Y. Hence, in between these two extreme points in the C1-FFL configuration space, schematized in Fig. 1, we have a range of intermediate structures with non-zero $\langle Y(S) \rangle$ and $\langle Y(X) \rangle$ ($\langle Y(S) \rangle = 1 \rightarrow 99$ and $\langle Y(X) \rangle = 99 \rightarrow 1$ subject to the constraint: $\langle Y \rangle = \langle Y(S) \rangle + \langle Y(X) \rangle = 100$). Hence, the ratio of the direct production level to the total production level of the output gene product serves as an ideal normalized metric to indicate the inter-transmutations of different varieties of the C1-FFL network modules. In Sec. III, we will analyze the PID and noise elements with respect to this transformed tuning parameter: $\langle Y(S) \rangle^f =: \langle Y(S) \rangle / \langle Y \rangle$. $\langle Y(S) \rangle = \langle Y(X) \rangle = \langle Y \rangle / 2$, or $\langle Y(S) \rangle^f = 0.5$ corresponds to a perfect C1-FFL structure with equally strong direct ($S \rightarrow Y$) and indirect ($X \rightarrow Y$) decoding paths.

The strength of interaction between each of the target promoters of S and X on the output gene Y and the corresponding RNA polymerase is designated by the maximal production rates, i.e., k_{sy} and k_{xy} , respectively. These synthesis rate parameters form part of the genotypic space of the generic C1-FFL whereas the output abundance level ($\langle Y \rangle$) is its phenotypic characteristic. The latter is essentially interlinked with the benefits derived from this motif by a single cell. Therefore, a fixed value of $\langle Y \rangle$ signifies a many-to-one mapping from the genotypic to the phenotypic space. By maintaining fixed steady-state population levels: $\langle S \rangle = \langle X \rangle = \langle Y \rangle = 100$, we ensure that different configurations of the generic C1-FFL type are compared on an equal footing [2]. The adoption of steady states to analyze the information transfer and noise management is motivated by the fact that the

FIG. 1. A schematic wiring diagram of a generic C1-FFL motif enabled with an additive logic (denoted by “+”) and tunable direct and indirect regulatory controls over the output gene product. The blue and red arrows designate synthesis (due to transcriptional activation) and degradation reactions of gene products, respectively. $k_{..}$ and $\mu_{..}$ are the corresponding reaction rate parameters [see Sec. II(A)]. The extreme left and right configurations are special types of a generic C1-FFL motif where one of the regulatory branches leading up to the output species is absent. These non-existent interactions are marked by dashed blue lines with k_{sy} ($\langle Y(S) \rangle$) or k_{xy} ($\langle Y(X) \rangle$) = 0. The middle topology depicts a perfect C1-FFL with $\langle Y(S) \rangle = \langle Y(X) \rangle$. k_{sy} (k_{xy}) is increased (decreased) in a way so that $\langle Y(S) \rangle$ ($\langle Y(X) \rangle$) increases (decreases) by one copy at a time thereby keeping the total population $\langle Y \rangle = \langle Y(S) \rangle + \langle Y(X) \rangle$ fixed. Hence, the entire network modification scheme reveals a variety of C1-FFLs each of which is characterized by unique sets of $\{k_{sy}, k_{xy}\}$ or equivalently $\{\langle Y(S) \rangle, \langle Y(X) \rangle\}$.

steady-state populations of the biochemical species are responsible for their optimal functioning in the single cell [2]. Fixed regulator abundances also imply that the TF-promoter interactions with the output gene (strengths of the Hill functions) are fixed. If we relax this constraint of population constancy and instead increase $\langle S \rangle$ and $\langle X \rangle$, the regulator-level intrinsic variability/noise ($\{\langle S \rangle^{-1}, \langle X \rangle^{-1}\}$) which act like extrinsic noise for their enslaved gene products will drop. The reduction in flow of fluctuations will automatically result in reduced magnitudes of the entropy transfers and PID metrics. Since, the TE-based measures typically have low numerical values, we remain cautious about further lowering of their magnitudes. Keeping in mind the translocation of TF proteins from the cytoplasm to the nucleus in eukaryotic cells, the population levels of all the gene products are expressed in the

units of molecules per effective cellular volume of magnitude unity. The degradation rate parameters considered for our main set of results are: $\mu_s = 0.1$, $\mu_x = 0.5$, and $\mu_y = 5.0$ all in the units of min^{-1} . With upstream regulators more stable than their downstream targets ($\mu_s < \mu_x < \mu_y$), we ensure sufficient flow of fluctuations and avoid the risk of insignificant PID values. In Sec. III C, we will examine the effects of a tunable input degradation rate parameter on our metrics of interest. The Hill coefficient $n = 1$ is used to avoid strong nonlinear effects. Activation coefficients are set at $K_{sx} = K_{sy} = \langle S \rangle$ and $K_{xy} = \langle X \rangle$ in units identical with that of the population levels. This type of choice denotes a particular operating point on the activation curves of X and Y where their promoter regions are half-filled by their respective regulators. Quantitatively speaking, the approximated steady-state values of the Hill functions thus become 0.5 and hence this type of regulatory state is known as the half-maximal activation [2]. If the activation coefficients are taken to be lower than the previously mentioned values, the input functions start to get steeper/nonlinear. The alternative choice, i.e., activation coefficients being higher than the population levels, make the input functions more linear but experimental wisdom suggests that the transcriptional control starts to mount with sufficient efficacy at the half-maximal point on the activation curve [2]. Considering these trade-offs, we present our main analytical and stochastic simulation results by taking into account the half-maximal activation. In Sec. III D, we will consider variations of activation coefficients around their half-maximal points. Since, the principle of PID is applicable for either linear or linearized systems, the values of n and $K_{\dots}$ should be chosen in a way so that the input functions become sufficiently linearized. The activation coefficient is represented as the ratio of the off and on rate parameters of TF-promoter interaction, i.e., the ratio of rate parameters corresponding to the backward and forward reactions between a certain TF and its target promoter, forming the TF-promoter complex. For $K_{\dots} = (K_{\dots}^{\text{off}}/K_{\dots}^{\text{on}}) > 1$, it is evident that $K_{\dots}^{\text{off}} > K_{\dots}^{\text{on}}$. On the other side, the binding energy is quantified as $E_b = -k_B T \ln K_{\dots}$ [3, 17]. Therefore, $K_{\dots} > 1$ suggests that $E_b \neq 0$ which is an indirect signature of the underlying out-of-equilibrium dynamics of the system. It is to be pointed out that there exists a clear separation of timescales between different stages of gene transcription regulation. For instance, the inducer-dependent activations of the TFs and the formations of the TF-promoter complexes occur at a relatively faster timescales with respect to the time taken by the gene products to achieve their steady-state population levels. Taking this into consideration, our present coarse-grained dynamical

model portrays a suitably approximated equilibrium picture of the actual nonequilibrium character of the biochemical kinetics [2]. The synthesis rate parameters k_s , k_{sx} , k_{sy} , and k_{xy} are determined using the steady-state versions of Eqs. (1a-1c) along with the steady-state approximate forms of the Hill functions and the fixed steady-state population constraints. In Sec. III E, we will loosen this constraint of identical population levels of the regulators and the output gene product and examine the resulting effects on the PID measures and the output noise.

III. RESULTS AND DISCUSSION

A. Maxima of the synergistic and redundant information transfers indicate the impending growth of the output noise

In Fig. 2(a), we observe that with increasing direct production strength or $\langle Y(S) \rangle^f$, both the redundant and synergistic information transfers from the regulators increase and maximize for a nearly balanced C1-FFL ($\langle Y(S) \rangle^f = 0.53$). Since, the quantification of $R_{sx \rightarrow y}$ and $S_{sx \rightarrow y}$ involves both of the regulators, one can interpret their simultaneous maximization in terms of approximately equal amounts of $\langle Y(S) \rangle$ and $\langle Y(X) \rangle$. It is the direct and indirect productions of the output gene product through which the information processed at the interregulator level flows to the terminal species. We also notice that for this near-perfect configuration of the C1-FFL, $U_{s \rightarrow y}$ takes its first non-zero value whereas $U_{x \rightarrow y}$ turns zero. This further demonstrates that the two types of the unique information transfers can only exist independent of each other. For this C1-FFL topology, the regulators predominantly convey the redundant and synergistic information flavors to their predicted biochemical species. The decline in $\langle Y(X) \rangle$ level during the growth of redundancy and synergy explains the diminishing trend of information uniquely transferred by the co-regulator X. Similar logic gives out meaning to the trend of $U_{s \rightarrow y}$. The dominance of $R_{sx \rightarrow y}$ over $S_{sx \rightarrow y}$ as documented in Fig. 2(a) is also significantly observed in other small networks under the control of different tunable biochemical parameters [34]. Interestingly, the regulators continue to transfer the redundant information even when one of the regulatory pathways that control the output gene expression is absent but this is not the case with the synergistic information transfer. The numerical values of the PID elements have specific physical

FIG. 2. Panel (a) portrays the four PID elements representing different types of information transferred from the regulators to the output gene product, measured in the units of bits. Panel (b) shows the output gene expression noise measured using the squared coefficient of variation (CV) in the unit of c.n.⁻¹. The solid lines are the approximate analytical results derived within the small-noise domain. The symbols are the steady-state average datasets from an ensemble of 5×10^4 independent Gaussian time series generated from the exact Langevin stochastic simulation. The parameters common to both treatments are listed in Sec. II C. Besides, $\Delta t = 10^{-1}$ unit of time is used to generate the analytical and stochastic simulation-based results. The dashed gridlines demarcate the values of $\langle Y(S) \rangle^f$ at which $R_{sx \rightarrow y}$ and $S_{sx \rightarrow y}$ are maximized in panel (a) and CV_y^2 is minimized in panel (b). The mismatch between the analytical and the stochastic simulation-based datasets is due to the presence of intrinsic noise which is significantly generated from the discreteness of copy numbers of the gene products in the Langevin stochastic simulation.

meanings attached to them and may be understood by taking into account the physical interpretation attributed to the concept of information, in general [49, 50]. Regulator S (X) transferring $U_{s \rightarrow y}$ ($U_{x \rightarrow y}$) bits of information to Y signifies that the predictor state S_t (X_t) uniquely reduces the uncertainty of the output gene expression state Y_{t+1} by an extra factor of $2^{U_{s \rightarrow y}}$ ($2^{U_{x \rightarrow y}}$), on average, with respect to the datum line contribution from Y_t to Y_{t+1} . Similarly, the predictor states (S_t and X_t) reduce the entropy of Y_{t+1} on a shared (complementary) basis by an extra factor of $2^{R_{sx \rightarrow y}}$ ($2^{S_{sx \rightarrow y}}$), with respect to the reference entropy reduction done by Y_t .

Fig. 2(b) documents the response of the output gene expression noise in terms of CV_y^2 . The concave-up profile has an opposite trend with respect to that of $R_{sx \rightarrow y}$ and $S_{sx \rightarrow y}$.

As the master-regulator-influenced output production grows, the output noise tends to decrease until it becomes minimized for an approximately balanced C1-FFL ($\langle Y(S) \rangle^f = 0.57$). Thereafter, the output noise begins to rise. For equal target production levels ($\langle Y(S) \rangle$ and $\langle Y(X) \rangle$), the direct and indirect decoding processes of the input signal become equally strong. This establishes maximum control over the output variance leading to the minimization of the output noise. In the absence of such a balanced regulatory control, the output gene expression shows an increased noise level in either direction away from $\langle Y(S) \rangle^f = 0.5$. The exact positions of the redundant and synergistic information transfer maxima and the output noise minimum depend upon the intricate interplay among different systemic parameters. Consulting Figs. 2(a, b) together, we can appreciate the fact that the redundant and synergistic information transfers are appropriate predictors for the output gene expression noise. Looking at the values of $\langle Y(S) \rangle^f$ at which $R_{sx \rightarrow y}$ and $S_{sx \rightarrow y}$ attain their peak values and output noise is minimized, it appears that the maximization of $R_{sx \rightarrow y}$ and $S_{sx \rightarrow y}$ can forecast subsequent minimization of the output noise and most importantly its forthcoming growth. Since, the output noise is an important quantifier of the corresponding gene expression stability which is invariably interlinked to the physiological fitness of a single cell, the redundant and synergistic information transfers may be considered as metrics of considerable biophysical utility. Due to the complex functional dependencies of the TE metrics and the output variance on the biochemical parameters, it is difficult to mathematically account for the relative positions of the redundant and synergistic information transfer maxima and the output noise minimum and the phenomenon demands further investigations. Nevertheless, we will test our current hypothesis in a reasonably broad and biochemically plausible parametric regime in the upcoming segments of our analysis. For ease of comparison, we will present the new datasets from the analytical computations due to the revised parameter choices alongside our control analytical data from Fig. 2.

B. Effects of the temporal step Δt on the dynamic information redundancy, synergy, and the output noise

The numerical value of the temporal step Δt is of great significance in our analysis of the information dynamics. Eqs. (A.1a-A.1c) show that the synthesis and degradation reaction propensities at time point t , in general, are $f_z(Z_t)\Delta t$ and $\mu_z Z_t \Delta t$, respectively. On the other

FIG. 3. Panels [(a), (b)] compare $\{R_{sx \rightarrow y}, S_{sx \rightarrow y}\}$ and CV_y^2 , respectively, for two values of the temporal step $\Delta t = 10^{-1}$ (solid lines) and $\Delta t = 10^{-2}$ (symbols). All the datasets are obtained from the analytical calculations using the small-noise approximation. Differently colored vertical axes correspond to the reference scales for the statistical metrics obtained from different values of Δt . The biochemical parameters used are identical to the case in Fig. 2.

hand, the Gaussian noise processes, at time point t are $\epsilon_z(t) = \sqrt{[f_z(Z_t) + \mu_z Z_t] \Delta t}$. For the analytical purpose, we use the approximate steady-state average values ($f_z(\langle Z_t \rangle)$, $\mu_z \langle Z_t \rangle$) for the propensities and noise, a judicious choice for the computational tractability in sync with the small-noise approximation [24]. For the purpose of stochastic simulation, their exact time-dependent values are used. Changing the temporal step size affects the reaction probabilities and the random forcings linearly and sublinearly, respectively. Since, the analytical computations are performed within the small-noise regime, we choose a different value of Δt which is smaller than the control value in Fig. 2 and by doing so ensure that the noise strengths in the dynamics decrease. It is to be reiterated that Δt is also the temporal distance between the predictor and target variables. Hence, a decreased Δt also decreases the numerical values of the predictor PID elements.

When we compare the PID elements analytically obtained using $\Delta t = 10^{-1}$ and 10^{-2} units of time in Fig. 3(a), we observe that their overall variational ranges in the former case are roughly 10 times the corresponding ranges in the latter case. Surprisingly, their variational trends with respect to $\langle Y(S) \rangle^f$ is preserved in the Δt transformation. When we examine the analytical profiles of the output noise in Fig. 3(b), it is found that $CV_y^2(\Delta t = 10^{-1}) > CV_y^2(\Delta t = 10^{-2})$, although in contrast with the PID metrics, the output noise does

not undergo numerical change by any order of magnitude. We have compared the positions of the maxima and minima of the redundant and synergistic information transfers and the output noise, respectively, in Table I. From these results, we conclude that a reasonable alternative choice for Δt does not alter the fact that the appearances of maxima of the redundant and synergistic information transfers precede the appearance of minimum of the output noise profile. Hence, the maximized redundant and synergistic information transfers can safely predict the minimization and the subsequent growth in the output noise level even when the temporal interval available for prediction (Δt) is reasonably small.

TABLE I. The list compares the positions of the redundant and synergistic information transfer peaks and the output noise minima on the $\langle Y(S) \rangle^f$ axis for different values of Δt in Fig. 3.

Δt (unit of time)	$\{R_{sx \rightarrow y}, S_{sx \rightarrow y}\}_{\text{Max}}$ at $\langle Y(S) \rangle^f =$	$CV_y^2 _{\text{Min}}$ at $\langle Y(S) \rangle^f =$
10^{-1}	0.53	0.57
10^{-2}	0.52	0.56

C. Effects of the input stability on the redundant and synergistic information transfers and the output noise

FIG. 4. Each of the panels [(a), (b), (c)] shows variations in $R_{sx \rightarrow y}$, $S_{sx \rightarrow y}$, and CV_y^2 , respectively, with respect to $\langle Y(S) \rangle^f$ for $\mu_s = 0.1, 1.0$, and 5.0 min^{-1} . The datasets shown are obtained analytically using the small-noise approximation. All other independent parameters are kept fixed at their control values in Fig. 2.

The set of degradation rate parameters $\{\mu_s, \mu_x, \mu_y\}$ are the independently tunable parameters given the constraint of constant steady-state gene expression levels. These rate

parameters are connected to the lifetimes of respective gene products through the relation $\tau_z = \ln 2 / \mu_z$ and thus are the characteristic inverse timescales of the dynamics in Eqs. (1a-1c). As we increase the input relaxation rate parameter from the reference value $\mu_s = 0.1$ to 1.0 and finally 5.0 min^{-1} , Fig. 4(a) shows a decreasing $R_{sx \rightarrow y}$. The trend is opposite for $S_{sx \rightarrow y}$ in Fig. 4(b) and needs further in-depth mathematical explanation but for our present practical purpose, is irrelevant. It is also observed that with decreasing stability of the master-regulator, the redundant and synergistic information transfers are maximized at the cost of lesser amounts of the direct output production. In Fig. 4(c), the output noise, similar to the redundant information transfer, decreases with increasing μ_s but the positions of its minima do not follow any specific trend. Table II collates the positions of the maxima and minima on the $\langle Y(S) \rangle^f$ axis for different values of μ_s . The one important message conveyed by Table II is that the output noise minimization occurs after the redundant and synergistic information transfers are maximized in the system. Therefore, our hypothesis proposing the maximized synergistic and redundant information transfers as suitable predictors for the rise of the output gene expression noise is valid irrespective of the stability of the input signaling species that the C1-FFL motif may encounter in a fluctuating environment.

TABLE II. Comparison table for the positions of the maximum and minimum points of the redundant and synergistic information transfers and the output noise, respectively, for different values of μ_s in Fig. 4.

μ_s (min^{-1})	$\{R_{sx \rightarrow y}, S_{sx \rightarrow y}\} _{\text{Max}}$ at $\langle Y(S) \rangle^f =$	$CV_y^2 _{\text{Min}}$ at $\langle Y(S) \rangle^f =$
0.1	0.53	0.57
1.0	0.50	0.54
5.0	0.47	0.57

D. Effects of the activation coefficients on the redundant and synergistic information transfer elements and the output gene expression noise

Next, we investigate the validity of our core hypothesis when the activation coefficients: K_{sx} , K_{sy} , and K_{xy} are set at an operating point other than the half-maximal point as in the control data in Fig. 2. The TF-promoter binding strength is measured by the inverse

FIG. 5. Panels [(a), (c), (e)] show the profiles of $R_{sx \rightarrow y}$, $S_{sx \rightarrow y}$, and CV_y^2 , respectively, for different values of K_{sx} . Panels [(b), (d), (f)] contain the plots for these three metrics in the same order but for different values of K_{sy} and K_{xy} . To showcase a reasonable number of interesting parametric realizations, the constraint of $K_{sy} = K_{xy}$ is maintained. Each of the three types of activation coefficients is assigned values such that two points around the respective half-maximal points on different activation curves may be accessed. Apart from the half-maximal case, the other two values for each of the activation coefficients are assigned according to the prescription: $K_{\dots} = \langle Z \rangle \pm 20\% \langle Z \rangle$. Other independent parameters and population levels are kept unchanged from Fig. 2. All the datasets are analytical results obtained within the small-noise domain.

of the activation coefficient and hence an increased K_{sx} signifies that the transcriptional link is weakened. Depending upon the value of the activation coefficient with respect to the associated TF copy number, the (non)linear nature of the input Hill function may change. To exemplify, with $K_{sx} \gg S$, the synthesis rate of X becomes $f_x(S) = k_{sx}[S/(K_{sx} + S)] \approx (k_{sx}/K_{sx})S = k'_{sx}S$. Hence, the transcriptional regulation becomes linear in S (first-order kinetics) with a reduced maximum production rate $k'_{sx} = k_{sx}/K_{sx}$. On the other hand, with $K_{sx} \ll S$, an approximate zero-order kinetics follow as $f_x(S) \approx k_{sx}$. The half-maximal activation point due to $K_{sx} = S$ retains a sufficient amount of (non)linearity which suits the requirements of PID. Notably, the MMI PID is valid when the dynamics under investigation is either linear or linearized. Taking this into our consideration, we intend to vary various K_{sx} applying reasonably small changes around their half-maximal values. The activation coefficient can be changed by means of altering the gene promoter sequences so that the respective promoter's chemical affinity towards its input TF is changed [2]. For $K_{sx} = \langle Z \rangle \pm 20\% \langle Z \rangle = 1.2\langle Z \rangle, 0.8\langle Z \rangle$, the various approximate steady-state values of the Hill functions associated with the independent reaction channels become $\approx 0.455, 0.556$, respectively.

When we change K_{sx} taking $K_{sx} = \langle S \rangle = 100$ as our datum line and keeping $K_{sy} = \langle S \rangle$ and $K_{xy} = \langle X \rangle$, Fig. 5(a) shows that an increased (decreased) K_{sx} lifts (drops) the redundant information transfer. Similar trends are also visible in the output noise level portrayed in Fig. 5(e) whereas the synergistic information transfer is robust to the altered K_{sx} as shown in Fig. 5(c). When the direct and indirect transcriptional links regulating the output gene expression are subjected to similar types of change while keeping $K_{sx} = \langle S \rangle$, the data projects that increasing (decreasing) $\{K_{sy}, K_{xy}\}$ results in increased (decreased) levels of $\{R_{sx \rightarrow y}, S_{sx \rightarrow y}, CV_y^2\}$ as profiled in Figs. 5(b, d, f), respectively. These results may be explained by taking into consideration the fact that with reduced activation coefficients, the binding affinities become higher or in other words, the input TFs bind strongly with their target promoters and such tight binding events lower the fluctuations in the transcriptional activation reactions. Therefore, the levels of the output noise and most of the redundant and synergistic information transfers drop in magnitudes. The changes in relevant statistical metrics around the reference half-maximal data (red solid lines) in Figs. 5(b, d, f) are larger in magnitudes in comparison with their counterparts in the variable K_{sx} case, i.e., Figs. 5(a, c, e), respectively. These datasets show that the transferred informations and the output

noise are less susceptible to the change in the binding affinity of $S \rightarrow X$ than in $S \rightarrow Y$ and $X \rightarrow Y$. This is possibly due to the fact that the former link is static in terms of fixed production level $\langle X \rangle$ whereas the latter two edges are sufficiently dynamic because of varying $\langle Y(S) \rangle$ and $\langle Y(X) \rangle$ levels. The data in Fig. 5(c) showing an almost identical response of $S_{sx \rightarrow y}$ towards different values of K_{sx} need further interpretation in terms of detailed changes in the analytical expressions of the single- and two-time-point (co)variances and also possibly in the TE metrics. But due to the presence of the nonlinear logarithm function, such fine-grained mathematical explanation is difficult to achieve. Fortunately, our main concerns in these measurements involve the antagonistic trends of $\{R_{sx \rightarrow y}, S_{sx \rightarrow y}\}$ with respect to CV_y^2 and the appearances of the information maxima followed by the noise minima. The values of $\langle Y(S) \rangle^f$ for which the redundant and synergistic information transfers are maximized and the output noise is minimized for different numerical values of the activation coefficients, are presented in Table III. These datasets categorically suggest that even when the operating

TABLE III. The values of $\langle Y(S) \rangle^f$ at which the redundant and synergistic information transfers are maximized and the output noise is minimized for different values of the activation coefficients are listed. Only those activation coefficients, whose variations are investigated, have been explicitly mentioned and the table should be consulted along with Fig. 5. Other activation coefficients are kept fixed at their respective half-maximal values. The activation coefficients are expressed in terms of the population levels: $\langle S \rangle = \langle X \rangle = \langle Y \rangle = 100$.

Activation coefficients	$\{R_{sx \rightarrow y}, S_{sx \rightarrow y}\} _{\text{Max}}$ at $\langle Y(S) \rangle^f =$	$CV_y^2 _{\text{Min}}$ at $\langle Y(S) \rangle^f =$
$K_{sx} = \langle S \rangle$	0.53	0.57
$K_{sx} = 1.2\langle S \rangle$	0.53	0.58
$K_{sx} = 0.8\langle S \rangle$	0.52	0.55
$\{K_{sy}, K_{xy}\} = \{\langle S \rangle, \langle X \rangle\}$	0.53	0.57
$\{K_{sy}, K_{xy}\} = 1.2\{\langle S \rangle, \langle X \rangle\}$	0.53	0.57
$\{K_{sy}, K_{xy}\} = 0.8\{\langle S \rangle, \langle X \rangle\}$	0.53	0.57

points on various activation curves are shifted from the respective half-maximal points, $R_{sx \rightarrow y}$ and $S_{sx \rightarrow y}$ attain their peaks before CV_y^2 reaches its lowest value. Hence, these two PID flavors again become enabled to forecast the growth in the output gene expression fluctuations.

E. Effects of differential gene expressions of the regulators and the output species on the redundant and synergistic information transfers and the output noise

FIG. 6. Panels [(a), (c)] compare $\{R_{sx \rightarrow y}, S_{sx \rightarrow y}\}$ and CV_y^2 , respectively, for $\langle Y \rangle = 200$ (symbols) with the reference data generated for $\langle Y \rangle = 100$ (solid lines). Panels [(b), (d)] do similar comparisons for the data due to $\langle Y \rangle = 50$ (symbols). The regulator abundances are kept identical to the previous cases, i.e., $\langle S \rangle = \langle X \rangle = 100$. The constraint of $\langle Y(X) \rangle = \langle Y \rangle - \langle Y(S) \rangle$ is still in place so are other independent choices of different parameters. The datasets are derived from analytical computations under the aegis of small-noise approximation.

The gene expression levels of the predictor and target species are important factors determining how much redundant and synergistic informations are transferred from the input to the output level in the motif and the extent to which the target's fluctuations are altered? We bring in significant asymmetry between the populations of the information sources and target to demonstrate this point while keeping the parametric realization in Fig. 2 as our reference. Since, both S and X are the information providers, we retain their populations identical to each other ($\langle S \rangle = \langle X \rangle = 100$) while that of their target is either increased or decreased significantly.

Fig. 6(a) suggests that when the regulators predict the target species with a population pool twice as strong as the sources ($\langle Y \rangle = 200$), both $R_{sx \rightarrow y}$ and $S_{sx \rightarrow y}$ increase with respect to the reference datasets of Fig. 2(a). On the output noise side in Fig. 6(c), due to its population-normalized definition ($CV_y^2 = \Sigma(Y_t)/\langle Y_t \rangle^2$), increased $\langle Y \rangle$ (or equivalently $\langle Y_t \rangle$) invariably results in decreased output fluctuations. In the opposite scenario, i.e., regulators predicting their target with a population half of that of the former species ($\langle Y \rangle = 50$), the redundant and synergistic information transfers and the output noise level are decreased and increased, respectively, with regards to the designated control datasets [see Figs. 6(b, d)]. The opposite trend of the output noise profile with respect to the redundant and synergistic information transfers reappears in the present parameterization and moreover ensures that the relevant information transfer elements maximize before the output fluctuations are minimized. The specific positions of different maxima and minima are listed in Table IV. Therefore, even with the asymmetrical production machineries of the regulators and their

TABLE IV. The values of $\langle Y(S) \rangle^f$ corresponding to the maxima of the redundant and synergistic information transfers and the minima of the output noise for different output gene expression levels are tabulated. The regulator population levels are kept at $\langle S \rangle = \langle X \rangle = 100$. The profiles presented in Fig. 6 should be consulted along with these datasets for various extrema.

Output gene expression level ($\langle Y \rangle$)	$\{R_{sx \rightarrow y}, S_{sx \rightarrow y}\} _{\text{Max}}$ at $\langle Y(S) \rangle^f =$	$CV_y^2 _{\text{Min}}$ at $\langle Y(S) \rangle^f =$
100	0.53	0.57
200	0.525	0.565
50	0.52	0.56

enslaved gene product, the maximized redundant and synergistic information transfers can efficiently predict the growth of the output noise.

IV. CONCLUSION

In summary, we have studied the information transfer and output noise management capacities of various manifestations of a generic C1-FFL motif. For fixed steady-state abundances of the three constituent gene products, we have invoked the small-noise approximation within a Gaussian framework and applied the principle of MMI PID to the time-discrete

MVAR dynamics. This has helped us to quantify the redundant, synergistic, and unique flavors of information transfer from the upstream species to the final gene product. For an approximately balanced C1-FFL (equally strong direct and indirect decoding strengths), these two types of dynamic information dominate their unique counterparts which become vanishingly small as the redundant and synergistic information transfers are maximized in the system. Hence, with appreciably strong direct and indirect regulations of the target (output) species, the regulators cease to independently predict their target and only act together.

We have demonstrated that the redundant and synergistic information transfers are antagonistically related to the output noise level. Moreover, these information transfer elements are maximized before the minimization and subsequent growth of the output noise. Depending upon these findings, we hypothesize that the redundant and synergistic information transfer elements may be referred to as potential predictors (indicators) for the rise in the fluctuation level associated with the output gene expression. Since, the output gene product may serve as an essential enzyme/protein/TF for various crucial physiological purposes of a single cell, a predictive control over its fluctuation is of utmost importance for the cell's survival in a changing environment. We have supported our main piece of the analytical result with data generated from the exact procedure of Langevin stochastic simulation coupled with sufficient ensemble averaging over a large number of independently generated Gaussian time series.

The order in which the redundant and synergistic information transfer maxima and the output noise minimum appear with respect to changing $\langle Y(S) \rangle^f$ remains a puzzle both at the mathematical and biochemical levels of understanding. We hope that further research will be able to demystify the phenomenon in its entirety. Nevertheless, we have tested our core hypothesis under sufficiently varied choices for different parameters. We have found out that our main result stands firmly when: (i) the time duration available to the regulators for predicting their target is reduced by an order-of-magnitude, (ii) the input species' stability is significantly modified, (iii) various activation coefficients are tuned in and around their individual half-maximal operating points, and (iv) the population level of the predicted output is made to be double/half of the individual populations of its two upstream predictor species. Thus, our main finding regarding the predictive capacities of the redundant and synergistic information transfers targeting the output gene expression noise is pretty robust

with respect to reasonable variations of the relevant biochemical parameters conforming to the small-noise approximation.

In the absence of a standardized MMI PID technique for more than three stochastic variables, the generic C1-FFL network with tunable transcriptional control over the output gene product is an ideal model system to systematically investigate the response of different flavors of information transfer to various important biochemical parameters. Previously, the principle of MMI PID has been applied to the two-dimensional Ising model and the findings have categorically demonstrated that the redundant and synergistic information transfer terms may act as warning signals for the transition from a disordered to an ordered state [29]. Our present hypothesis should be further tested for its generality, e.g., when the biochemical system involves transitions between different steady states. Since, the prescription of MMI PID is valid for either linear or reasonably linearized systems, it remains an open question whether the present theoretical approach can be safely applied to such nonlinear biochemical transitions. The situation is further complicated due to the presence of large fluctuations in and around the point of transition and therefore the present analytical treatment using the small-noise approximation will be defunct in tackling such problems. Explorations along these lines of research are supposed to generate new insights about the real-world applications of the multivariate information theory.

ACKNOWLEDGMENTS

Md SAM is supported by DST, Govt. of India, through INSPIRE research fellowship (DST/INSPIRE Fellowship/2018/IF180056). Md SAM and AB acknowledge Bose Institute, Kolkata for research support.

Appendix: Computing the time-discrete variances and covariances

- **Analytical expressions using the small-noise approximation:** The Langevin equations (1a-1c) are in the continuous-time domain. In this section, we use the MVAR process to calculate TEs between different nodes of the reference C1-FFL network [31–33]. To do so, we discretize Eqs. (1a-1c) considering the gene expression dynamics with a finite temporal step/resolution (Δt). In this time interval, changes in the population

level of species S is $S_{t-1} \rightarrow S_t$ where $t-1$ symbolically denotes the immediate past (i.e., Δt time behind) time point of t . The consequent rate of change in copy number state is $\frac{\Delta S}{\Delta t} = \frac{S_t - S_{t-1}}{\Delta t}$. The equivalence between the continuous and discrete versions is through the relation: $\frac{dS}{dt} = \lim_{\Delta t \rightarrow 0} \frac{\Delta S}{\Delta t}$. Hence, the time-discrete Langevin equations are approximations to their continuous counterparts. The resulting discretized dynamical system of equations is as follows:

$$S_t \approx f_s(S_{t-1})\Delta t + (1 - \beta_s)S_{t-1} + \epsilon_s(t), \quad (\text{A.1a})$$

$$X_t \approx f_x(S_{t-1})\Delta t + (1 - \beta_x)X_{t-1} + \epsilon_x(t), \quad (\text{A.1b})$$

$$Y_t \approx f_y(S_{t-1}, X_{t-1})\Delta t + (1 - \beta_y)Y_{t-1} + \epsilon_y(t). \quad (\text{A.1c})$$

Here, we have simplified our analytical treatment by assuming $\epsilon_z(t-1) \approx \epsilon_z(t)$. Since, our forthcoming analysis is intended to be at steady state, approximating noise terms from neighbouring time points to be equal to each other is a useful simplification for the analytical purpose. This is based upon the assumption that the steady-state copy numbers do not change significantly between two neighbouring time points, i.e., $\langle Z_{t-1} \rangle \approx \langle Z_t \rangle$. Since, the noise strengths are functions of the copy numbers, we can safely use approximately equal noise strengths for time points t and $t-1$. However, while executing the exact stochastic simulation, such an approximation is not tenable because the system's simulated dynamics also involve transient states. Additionally, within a finite time-span Δt , depending upon the number of reactions happening, $Z_{t-1} \approx Z_t$ may not hold for all the time points (even at steady state) each time an independent Gaussian stochastic time series is generated in repeated time-course simulation experiments. For fixed Δt , $\beta_z = \mu_z \Delta t$ represent the inverse timescales in dimensionless forms for our time-discrete dynamical system. Therefore, the rescaled random forcing terms for this discretized system are $\epsilon_z(t) = \xi_z(t)\sqrt{\Delta t}$. Using the properties of $\xi_z(t)$ from Sec. II A, we note that $\langle \epsilon_z(t) \rangle = 0$ and $\langle \epsilon_z(t)\epsilon_{z'}(t') \rangle = \langle |\xi_z(t)|^2 \rangle \Delta t \delta_{zz'} \delta(t-t')$. The steady-state ensemble-averaged noise strengths in Eqs. (A.1a-A.1c) are related to their counterparts in Eqs. (1a-1c) through the relation $\langle |\epsilon_z(t)|^2 \rangle = \langle |\xi_z(t)|^2 \rangle \Delta t$. Ensemble-averaging Eqs. (A.1a-A.1c) at steady state, we obtain:

$$\langle S_t \rangle = f_s(\langle S_{t-1} \rangle)\Delta t + (1 - \beta_s)\langle S_{t-1} \rangle,$$

$$\langle X_t \rangle = f_x(\langle S_{t-1} \rangle)\Delta t + (1 - \beta_x)\langle X_{t-1} \rangle,$$

$$\langle Y_t \rangle = f_y(\langle S_{t-1} \rangle, \langle X_{t-1} \rangle)\Delta t + (1 - \beta_y)\langle Y_{t-1} \rangle.$$

Where, we have utilized the fact that $\langle \epsilon_z(t) \rangle = 0$. Now, at time point t , the copy number fluctuations around the respective steady-state mean values are:

$$\delta S_t \approx S_t - \langle S_t \rangle = [f_s(S_{t-1}) - f_s(\langle S_{t-1} \rangle)]\Delta t + (1 - \beta_s)\delta S_{t-1} + \epsilon_s(t), \quad (\text{A.2a})$$

$$\delta X_t \approx X_t - \langle X_t \rangle = [f_x(S_{t-1}) - f_x(\langle S_{t-1} \rangle)]\Delta t + (1 - \beta_x)\delta X_{t-1} + \epsilon_x(t), \quad (\text{A.2b})$$

$$\delta Y_t \approx Y_t - \langle Y_t \rangle = [f_y(S_{t-1}, X_{t-1}) - f_y(\langle S_{t-1} \rangle, \langle X_{t-1} \rangle)]\Delta t + (1 - \beta_y)\delta Y_{t-1} + \epsilon_y(t) \quad (\text{A.2c})$$

Here, we have invoked the small-noise approximation [40, 41, 51, 52], according to which, at steady state, the copy number fluctuations of order higher than the linear one is appreciably smaller and can be neglected when compared to the steady-state mean copy numbers. To paraphrase mathematically, this approximation implies that $\delta Z_t / \langle Z_t \rangle \ll 1$ where $Z_t \in \{S_t, X_t, Y_t\}$. The condition is similarly valid for the immediate past time point. Hence, we can rewrite the steady-state perturbations in the input functions due to the infinitesimal changes, i.e., $\{\delta S_{t-1}, \delta X_{t-1}\} \rightarrow 0$ in the following manner:

$$\begin{aligned} f_s(S_{t-1}) - f_s(\langle S_{t-1} \rangle) &\equiv \delta f_s(S_{t-1}) \approx f'_s(\langle S \rangle) \delta S_{t-1}, \\ f_x(S_{t-1}) - f_x(\langle S_{t-1} \rangle) &\equiv \delta f_x(S_{t-1}) \approx f'_{x,s}(\langle S \rangle) \delta S_{t-1}, \\ f_y(S_{t-1}, X_{t-1}) - f_y(\langle S_{t-1} \rangle, \langle X_{t-1} \rangle) &\equiv \delta f_y(S_{t-1}, X_{t-1}) \approx f'_{y,s}(\langle S \rangle, \langle X \rangle) \delta S_{t-1} \\ &\quad + f'_{y,x}(\langle S \rangle, \langle X \rangle) \delta X_{t-1}. \end{aligned}$$

Here, we have taken recourse to: $\langle S_{t-1} \rangle \approx \langle S_t \rangle \equiv \langle S \rangle$ and $\langle X_{t-1} \rangle \approx \langle X_t \rangle \equiv \langle X \rangle$. Here, $\langle f'_{x,s}(S) \rangle =: (df_x/dS)|_{S=\langle S \rangle} \approx f'_{x,s}(\langle S \rangle)$ and so on for other derivatives of the input functions. In our analytical calculations, performed within the small-noise domain, we have used a couple of judicious steady-state approximations, e.g., $\langle f_x(S) \rangle \approx f_x(\langle S \rangle)$ and $\langle f_y(S, X) \rangle \approx f_y(\langle S \rangle, \langle X \rangle)$. Extending this approximation to the derivatives, we have: $\langle f'_{x,s}(S) \rangle \approx f'_{x,s}(\langle S \rangle) = k_{sx}[K_{sx}/(K_{sx} + \langle S \rangle)^2]$, $\langle f'_{y,s}(S, X) \rangle \approx f'_{y,s}(\langle S \rangle, \langle X \rangle) = k_{sy}[K_{sy}/(K_{sy} + \langle S \rangle)^2]$, and $\langle f'_{y,x}(S, X) \rangle \approx f'_{y,x}(\langle S \rangle, \langle X \rangle) = k_{xy}[K_{xy}/(K_{xy} + \langle X \rangle)^2]$. Nevertheless, while executing the stochastic Langevin simulation, we have used actual copy numbers instead of their steady-state values in the input functions/reaction propensities [24]. Similar approximations at the previous time point lead us to $\langle f_s(S_{t-1}) \rangle \approx f_s(\langle S_{t-1} \rangle)$, $\langle f_x(S_{t-1}) \rangle \approx f_x(\langle S_{t-1} \rangle)$, $\langle f_y(S_{t-1}, X_{t-1}) \rangle \approx f_y(\langle S_{t-1} \rangle, \langle X_{t-1} \rangle)$. For the purpose of stochastic simulation, we have used dynamic input functions, i.e., $f_z(Z_{t-1})$ instead of their steady-state values: $\langle f_z(Z_{t-1}) \rangle$ or, as in our analytics, $f_z(\langle Z_{t-1} \rangle)$. For the details

of the stochastic simulation method, see the last part of this section. Rewriting the set of Eqs. (A.2a-A.2c), we get:

$$\delta S_t \approx [f'_s(\langle S \rangle)\Delta t + (1 - \beta_s)]\delta S_{t-1} + \epsilon_s(t), \quad (\text{A.3a})$$

$$\delta X_t \approx f'_{x,s}(\langle S \rangle)\Delta t\delta S_{t-1} + (1 - \beta_x)\delta X_{t-1} + \epsilon_x(t), \quad (\text{A.3b})$$

$$\delta Y_t \approx f'_{y,s}(\langle S \rangle, \langle X \rangle)\Delta t\delta S_{t-1} + f'_{y,x}(\langle S \rangle, \langle X \rangle)\Delta t\delta X_{t-1} + (1 - \beta_y)\delta Y_{t-1} + \epsilon_y(t). \quad (\text{A.3c})$$

Eqs. (A.3a-A.3c) in matrix form stand as:

$$\delta \mathbf{C}_t = \mathbf{B}\delta \mathbf{C}_{t-1} + \mathbf{N}_t, \quad (\text{A.4})$$

where,

$$\delta \mathbf{C}_t = \begin{pmatrix} \delta S_t \\ \delta X_t \\ \delta Y_t \end{pmatrix}, \quad \delta \mathbf{C}_{t-1} = \begin{pmatrix} \delta S_{t-1} \\ \delta X_{t-1} \\ \delta Y_{t-1} \end{pmatrix},$$

$$\mathbf{B} = \begin{pmatrix} f'_s(\langle S \rangle)\Delta t + 1 - \beta_s & 0 & 0 \\ f'_{x,s}(\langle S \rangle)\Delta t & 1 - \beta_x & 0 \\ f'_{y,s}(\langle S \rangle, \langle X \rangle)\Delta t & f'_{y,x}(\langle S \rangle, \langle X \rangle)\Delta t & 1 - \beta_y \end{pmatrix}, \quad \text{and } \mathbf{N}_t = \begin{pmatrix} \epsilon_s(t) \\ \epsilon_x(t) \\ \epsilon_y(t) \end{pmatrix}.$$

By multiplying Eq. A.4 by its transposed counterpart and then performing the steady-state ensemble-averaging, we arrive at:

$$\langle \delta \mathbf{C}_t \delta \mathbf{C}_t^T \rangle = \mathbf{B} \langle \delta \mathbf{C}_{t-1} \delta \mathbf{C}_{t-1}^T \rangle \mathbf{B}^T + \langle \mathbf{N}_t \mathbf{N}_t^T \rangle. \quad (\text{A.5})$$

Since, the mean squared deviations of the copy numbers are approximately stationary at steady state, they are approximately equal to each other at successive steady-state time points. Hence, we can write:

$$\mathbf{A}_t \equiv \langle \delta \mathbf{C}_{t-1} \delta \mathbf{C}_{t-1}^T \rangle \approx \langle \delta \mathbf{C}_t \delta \mathbf{C}_t^T \rangle = \begin{pmatrix} \Sigma(S_t) & \Sigma(S_t, X_t) & \Sigma(S_t, Y_t) \\ \Sigma(S_t, X_t) & \Sigma(X_t) & \Sigma(X_t, Y_t) \\ \Sigma(S_t, Y_t) & \Sigma(X_t, Y_t) & \Sigma(Y_t) \end{pmatrix}.$$

We denote the matrix incorporating the steady-state ensemble-averaged noise strengths as $\mathbf{M}_t \equiv \langle \mathbf{N}_t \mathbf{N}_t^T \rangle$. In the absence of cross-species noise-correlations, $\mathbf{M}_t$ is a diagonal

matrix with elements $\langle |\epsilon_z(t)|^2 \rangle$. Finally, we have a steady-state fluctuation-dissipation type of equation:

$$\mathbf{A}_t = \mathbf{B}\mathbf{A}_t\mathbf{B}^T + \mathbf{M}_t. \quad (\text{A.6})$$

Solving Eq. A.6, we arrive at the closed-form analytic expressions of the steady-state copy number variances and covariances. Here for notational simplicity, we have denoted the steady-state derivatives of different input functions as: $f'_s(\langle S \rangle) \equiv \langle f'_s \rangle$, $f'_{x,s}(\langle S \rangle) \equiv \langle f'_{x,s} \rangle$, $f'_{y,s}(\langle S \rangle, \langle X \rangle) \equiv \langle f'_{y,s} \rangle$, and $f'_{y,x}(\langle S \rangle, \langle X \rangle) \equiv \langle f'_{y,x} \rangle$.

$$\Sigma(S_t) = \frac{2\beta_s \langle S \rangle}{1 - (\langle f'_s \rangle \Delta t + 1 - \beta_s)^2}, \quad (\text{A.7a})$$

$$\Sigma(S_t, X_t) = \frac{(\langle f'_s \rangle \Delta t + 1 - \beta_s)(\langle f'_{x,s} \rangle \Delta t) \Sigma(S_t)}{1 - (\langle f'_s \rangle \Delta t + 1 - \beta_s)(1 - \beta_x)}, \quad (\text{A.7b})$$

$$\Sigma(S_t, Y_t) = \frac{(\langle f'_s \rangle \Delta t + 1 - \beta_s)[\langle f'_{y,s} \rangle \Delta t \Sigma(S_t) + \langle f'_{y,x} \rangle \Delta t \Sigma(S_t, X_t)]}{1 - (\langle f'_s \rangle \Delta t + 1 - \beta_s)(1 - \beta_y)}, \quad (\text{A.7c})$$

$$\Sigma(X_t) = \frac{2\beta_x \langle X \rangle + (\langle f'_{x,s} \rangle \Delta t)^2 \Sigma(S_t) + 2(\langle f'_{x,s} \rangle \Delta t)(1 - \beta_x) \Sigma(S_t, X_t)}{1 - (1 - \beta_x)^2}, \quad (\text{A.7d})$$

$$\begin{aligned} \Sigma(X_t, Y_t) = & [(\langle f'_{x,s} \rangle \Delta t)(\langle f'_{y,s} \rangle \Delta t) \Sigma(S_t) + [(\langle f'_{y,s} \rangle \Delta t)(1 - \beta_x) + (\langle f'_{x,s} \rangle \Delta t)(\langle f'_{y,x} \rangle \Delta t)] \Sigma(S_t, X_t) \\ & + (\langle f'_{y,x} \rangle \Delta t)(1 - \beta_x) \Sigma(X_t) + (\langle f'_{x,s} \rangle \Delta t)(1 - \beta_y) \Sigma(S_t, Y_t)] \times [1 - (1 - \beta_x)(1 - \beta_y)]^{-1}, \end{aligned} \quad (\text{A.7e})$$

$$\begin{aligned} \Sigma(Y_t) = & [2\beta_y \langle Y \rangle + (\langle f'_{y,s} \rangle \Delta t)^2 \Sigma(S_t) + 2(\langle f'_{y,x} \rangle \Delta t)(\langle f'_{y,s} \rangle \Delta t) \Sigma(S_t, X_t) + 2(\langle f'_{y,s} \rangle \Delta t)(1 - \beta_y) \Sigma(S_t, Y_t) \\ & + 2(\langle f'_{y,x} \rangle \Delta t)(1 - \beta_y) \Sigma(X_t, Y_t) + (\langle f'_{y,x} \rangle \Delta t)^2 \Sigma(X_t)] \times [1 - (1 - \beta_y)^2]^{-1}. \end{aligned} \quad (\text{A.7f})$$

In the next step, we compute the one-time-lagged covariances by replacing t by $t + 1$ in Eq. A.4. We have again used the simplifying assumption $\mathbf{N}_{t+1} \approx \mathbf{N}_t$ considering our steady-state analytical requirements. This leads us to the following:

$$\delta \mathbf{C}_{t+1} = \mathbf{B} \delta \mathbf{C}_t + \mathbf{N}_t$$

By multiplying the above equation with $\delta \mathbf{C}_t^T$ and upon subsequent steady-state ensemble-averaging, we get:

$$\mathbf{A}_{t+1,t} = \mathbf{B} \mathbf{A}_t \quad (\text{A.8})$$

Here, we have used the fact that $\langle \mathbf{N}_t \delta \mathbf{C}_t^T \rangle = \mathbf{0}$. $\mathbf{A}_{t+1,t} \equiv \langle \delta \mathbf{C}_{t+1} \delta \mathbf{C}_t^T \rangle$ is the one-time-lagged covariance matrix, individual elements of which, can be obtained by solving Eq. A.8.

The closed-form analytic expressions are listed below:

$$\Sigma(S_{t+1}, S_t) = (\langle f'_s \rangle \Delta t + 1 - \beta_s) \Sigma(S_t), \quad (\text{A.9a})$$

$$\Sigma(S_{t+1}, X_t) = (\langle f'_s \rangle \Delta t + 1 - \beta_s) \Sigma(S_t, X_t), \quad (\text{A.9b})$$

$$\Sigma(S_{t+1}, Y_t) = (\langle f'_s \rangle \Delta t + 1 - \beta_s) \Sigma(S_t, Y_t), \quad (\text{A.9c})$$

$$\Sigma(X_{t+1}, S_t) = \langle f'_{x,s} \rangle \Delta t \Sigma(S_t) + (1 - \beta_x) \Sigma(S_t, X_t), \quad (\text{A.9d})$$

$$\Sigma(X_{t+1}, X_t) = \langle f'_{x,s} \rangle \Delta t \Sigma(S_t, X_t) + (1 - \beta_x) \Sigma(X_t), \quad (\text{A.9e})$$

$$\Sigma(X_{t+1}, Y_t) = \langle f'_{x,s} \rangle \Delta t \Sigma(S_t, Y_t) + (1 - \beta_x) \Sigma(X_t, Y_t), \quad (\text{A.9f})$$

$$\Sigma(Y_{t+1}, S_t) = \langle f'_{y,s} \rangle \Delta t \Sigma(S_t) + \langle f'_{y,x} \rangle \Delta t \Sigma(S_t, X_t) + (1 - \beta_y) \Sigma(S_t, Y_t), \quad (\text{A.9g})$$

$$\Sigma(Y_{t+1}, X_t) = \langle f'_{y,s} \rangle \Delta t \Sigma(S_t, X_t) + \langle f'_{y,x} \rangle \Delta t \Sigma(X_t) + (1 - \beta_y) \Sigma(X_t, Y_t), \quad (\text{A.9h})$$

$$\Sigma(Y_{t+1}, Y_t) = \langle f'_{y,s} \rangle \Delta t \Sigma(S_t, Y_t) + \langle f'_{y,x} \rangle \Delta t \Sigma(X_t, Y_t) + (1 - \beta_y) \Sigma(Y_t). \quad (\text{A.9i})$$

Here, $\langle f'_s \rangle = 0$ ($\because f_s = k_s = \text{Constant}$). According to the prescription in Ref. [32], the partial variances can be written as:

$$\Sigma(Y_{t+1}|Y_t) = \Sigma(Y_t) - \frac{\Sigma^2(Y_{t+1}, Y_t)}{\Sigma(Y_t)}, \quad (\text{A.10a})$$

$$\begin{aligned} \Sigma(Y_{t+1}|S_t, Y_t) &= \Sigma(Y_t) - \left(\Sigma(Y_{t+1}, S_t) \quad \Sigma(Y_{t+1}, Y_t) \right) \\ &\quad \times \begin{pmatrix} \Sigma(S_t) & \Sigma(S_t, Y_t) \\ \Sigma(S_t, Y_t) & \Sigma(Y_t) \end{pmatrix}^{-1} \begin{pmatrix} \Sigma(Y_{t+1}, S_t) \\ \Sigma(Y_{t+1}, Y_t) \end{pmatrix}, \end{aligned} \quad (\text{A.10b})$$

$$\begin{aligned} \Sigma(Y_{t+1}|S_t, X_t, Y_t) &= \Sigma(Y_t) - \left(\Sigma(Y_{t+1}, S_t) \quad \Sigma(Y_{t+1}, X_t) \quad \Sigma(Y_{t+1}, Y_t) \right) \\ &\quad \times \begin{pmatrix} \Sigma(S_t) & \Sigma(S_t, X_t) & \Sigma(S_t, Y_t) \\ \Sigma(S_t, X_t) & \Sigma(X_t) & \Sigma(X_t, Y_t) \\ \Sigma(S_t, Y_t) & \Sigma(X_t, Y_t) & \Sigma(Y_t) \end{pmatrix}^{-1} \begin{pmatrix} \Sigma(Y_{t+1}, S_t) \\ \Sigma(Y_{t+1}, X_t) \\ \Sigma(Y_{t+1}, Y_t) \end{pmatrix}. \end{aligned} \quad (\text{A.10c})$$

The working formula for $\Sigma(Y_{t+1}|X_t, Y_t)$ can be deduced similarly.

• **Computations using the exact Langevin stochastic simulation method:** Expressions (A.7a-A.7f) and (A.9a-A.9i) are essential ingredients to analytically evaluate the PID elements and output noise within the small-noise regime. To validate the approximate results, we further execute stochastic simulation using Eqs. (A.1a-A.1c). Unlike the analytical case, we use exact copy numbers (Z_{t-1}) instead of their steady-state average values ($\langle Z_{t-1} \rangle$) in the input functions $f_z(\dots)$ as well as the degradation terms in the right-hand-side of these equations. Similarly, the noise strengths are utilized as functions of actual copy

numbers and not their steady-state average values. As mentioned earlier, for the purpose of stochastic simulation, we maintain $\epsilon_z(t-1) \neq \epsilon_z(t)$. Therefore, for the stochastic trajectories, we have $|\epsilon_z(t-1)|^2 = |\xi_z(t-1)|^2 \Delta t = [f_z(Z_{t-1}) + \mu_z Z_{t-1}] \Delta t$ as the noise strength operating at time point $t-1$. For each of the three species, the associated Gaussian noise terms with zero means and unit variances are generated from the random number generating function *randn* in MATLAB. Hence, the modified random forcing in the simulated version of Eqs. (A.1a-A.1c) appears as $\epsilon_z(t-1) = \sqrt{[f_z(Z_{t-1}) + \mu_z Z_{t-1}] \Delta t} \times \text{randn}$. The choice for the numerical value of Δt is crucial in the stochastic simulation for a couple of reasons. As mentioned before, the modified noise strength at time point t , i.e., $|\epsilon_z(t-1)|^2 \propto \Delta t$. Hence, a larger Δt will evidently increase the stochasticity in the system and the simulated values of our metrics of interest will cease to be in sync with their analytical counterparts which are evaluated within the small-noise regime. On the other hand, the reaction propensities for different synthesis and degradation events happening in between time points $t-1$ and t are $f_z(Z_{t-1}) \Delta t$ and $\mu_z Z_{t-1} \Delta t$, respectively. Thus, a smaller Δt will make these reactions less probable in each time step and the predictability of future target states by the present source states will be hampered. As a result, the PID measures will be significantly reduced in their magnitudes. In fact, it can be analytically shown that with $\Delta t \rightarrow 0$, the variances and covariances obtained in Eqs. (A.7a-A.7f) reduce to their corresponding continuous-time versions derived in Ref. [35]. Therefore, one has to choose a reasonable value for Δt keeping in mind both of these trade-offs/constraints. For the present case, we simulate the exact versions of the discretized Langevin Eqs. (A.1a-A.1c) for sufficiently long time t_f (200 units of time) so that (i) the system reaches a steady state and (ii) using an appropriate Δt (10^{-1} unit of time), we can ensure that the system evolves via a sufficiently large number of iterations ($t_f/\Delta t = 2000$). Sec. IIIB analytically demonstrate that even with a smaller value of Δt and consequently reduced values of the PID elements, our core hypothesis remains intact.

In this way, for each of the values of k_{sy} (k_{xy}), we obtain a large number of independent Gaussian time series (5×10^4) for the dynamical variables. The rationale/constraint through which the direct and indirect synthesis rate parameters of the output species are tuned is specified in Sec. IIC. In the next stage, the necessary steady-state first and second moments (both single-time-point and one-time-lagged) of the copy number distributions are obtained considering the time points t_f and t_f-1 by averaging over the previously mentioned

ensemble of time series. These are subsequently used to compute the single- and two-time-point (co)variances for the framework of PID. The averaged imprint of noise at steady state $t = \{t_f, t_f - 1\}$ from this large collection of independent time series corresponds to the mean-field usage of $\langle |\epsilon_z(t)|^2 \rangle$ in our approximated analytical calculations [see $\mathbf{M}_t$ in Eq. A.6]. For an individual stochastic variable, one can conceptualize these independent stochastic time series as the gene expression profiles from individual single cells in a clonal population. Otherwise, these time traces may be interpreted as gene expression profiles gathered from a single cell in repeated time-course experiments with identical initial conditions and other constraints. Due to the probabilistic nature of biochemical reactions in a single cell, these gene expression profiles are all different from each other, a phenomenon which is popularly known as *isogenic heterogeneity* [9]. The Langevin stochastic simulation, in essence, mimics this biological reality. Since, this process involves none of the simplifying steps applied in our previous analytical treatment, the obtained statistical metrics are exact here. Sec. III A presents our main results using both analytical and stochastically simulated data for the PID elements and the output noise so that a meaningful and predictive comparison may be done.

-
- [1] R. Milo and R. Phillips, *Cell Biology by the Numbers* (2015), ISBN 9781317230694.
 - [2] U. Alon, *An Introduction to Systems Biology: Design Principles of Biological Circuits*, 2nd ed. (CRC press, 2020).
 - [3] G. Tkačik, A. M. Walczak, and W. Bialek, Phys. Rev. E **80**, 031920 (2009).
 - [4] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, Science **298**, 824 (2002).
 - [5] S. Kalir, S. Mangan, and U. Alon, Mol. Syst. Biol. **1**, 2005.0006 (2005).
 - [6] S. Mangan, A. Zaslaver, and U. Alon, J. Mol. Biol. **334**, 197 (2003).
 - [7] D. T. Gillespie, J. Phys. Chem. **81**, 2340 (1977).
 - [8] E. M. Ozbudak, M. Thattai, I. Kurtser, A. D. Grossman, and A. van Oudenaarden, Nat. Genet. **31**, 69 (2002).
 - [9] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, Science **297**, 1183 (2002).
 - [10] P. S. Swain, M. B. Elowitz, and E. D. Siggia, Proc. Natl. Acad. Sci. **99**, 12795 (2002).

- [11] L. S. Tsimring, Rep. Prog. Phys. **77**, 026601 (2014).
- [12] C. E. Shannon, Bell. Syst. Tech. J. **27**, 379 (1948).
- [13] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. (Wiley-Interscience, New York, 2006).
- [14] D. J. C. MacKay, *Information Theory, Inference & Learning Algorithms* (Cambridge University Press, New York, 2002).
- [15] S. F. Taylor, N. Tishby, and W. Bialek (2007), arXiv:0712.4382v1 [q-bio.PE].
- [16] W. Bialek and S. Setayeshgar, Phys. Rev. Lett. **100**, 258101 (2008).
- [17] A. M. Walczak, G. Tkačik, and W. Bialek, Phys. Rev. E **81**, 041905 (2010).
- [18] W. H. de Ronde, F. Tostevin, and P. R. ten Wolde, Phys. Rev. E **82**, 031914 (2010).
- [19] R. Cheong, A. Rhee, C. J. Wang, I. Nemenman, and A. Levchenko, Science **334**, 354 (2011).
- [20] W. Bialek, *Biophysics: Searching for Principles* (Princeton University Press, Princeton, 2012).
- [21] J. Selimkhanov, B. Taylor, J. Yao, A. Pilko, J. Albeck, A. Hoffmann, L. Tsimring, and R. Wollman, Science **346**, 1370 (2014).
- [22] F. Tostevin and P. R. ten Wolde, Phys. Rev. Lett. **102**, 218101 (2009).
- [23] P. Mehta, S. Goyal, T. Long, B. L. Bassler, and N. S. Wingreen, Mol. Syst. Biol. **5**, 325 (2009).
- [24] W. H. de Ronde, F. Tostevin, and P. R. ten Wolde, Phys. Rev. E **86**, 021913 (2012).
- [25] N. B. Becker, A. Mugler, and P. R. Ten Wolde, Phys. Rev. Lett. **115**, 258103 (2015).
- [26] A. Erez, T. A. Byrd, M. Vennettilli, and A. Mugler, Phys. Rev. Lett. **125**, 048103 (2020).
- [27] T. Schreiber, Phys. Rev. Lett. **85**, 461 (2000).
- [28] T. Bossomaier, L. Barnett, M. Harré, and J. Lizier, *An Introduction to Transfer Entropy* (2016), ISBN 978-3-319-43221-2.
- [29] D. Marinazzo, L. Angelini, M. Pellicoro, and S. Stramaglia, Phys. Rev. E (RC) **99**, 040101 (2019).
- [30] S. Stramaglia, J. M. Cortes, and D. Marinazzo, New J. Phys. **16**, 105003 (2014).
- [31] L. Faes, D. Marinazzo, and S. Stramaglia, Entropy **19**, 408 (2017).
- [32] A. B. Barrett, Phys. Rev. E **91**, 052802 (2015).
- [33] L. Novelli, F. M. Atay, J. Jost, and J. T. Lizier, Proc. R. Soc. A **476**, 20190779 (2020).
- [34] A. Biswas, Chaos **29**, 063108 (2019).
- [35] M. S. A. Momin and A. Biswas, Phys. Rev. E **101**, 052411 (2020).
- [36] L. Bintu, N. E. Buchler, H. G. Garcia, U. Gerland, T. Hwa, J. Kondev, and R. Phillips, Curr.

- Opin. Genet. Dev. **15**, 116 (2005).
- [37] E. Ziv, I. Nemenman, and C. H. Wiggins, PLoS ONE **2**, e1077 (2007).
 - [38] G. Tkačik, T. Gregor, and W. Bialek, PLoS ONE **3**, e2774 (2008).
 - [39] S. Tănase-Nicola, P. B. Warren, and P. R. ten Wolde, Phys. Rev. Lett. **97**, 068102 (2006).
 - [40] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry, 3rd ed.* (North-Holland, Amsterdam, 2007).
 - [41] J. Elf and M. Ehrenberg, Genome Res. **13**, 2475 (2003).
 - [42] J. Paulsson, Nature **427**, 415 (2004).
 - [43] J. T. Lizier, F. M. Atay, and J. Jost, Phys. Rev. E **86**, 026110 (2012).
 - [44] M. Wibral, N. Pampu, V. Priesemann, F. Siebenhühner, H. Seiwert, M. Lindner, J. T. Lizier, and R. Vicente, PLoS One **8**, e55809 (2013).
 - [45] L. Barnett, A. B. Barrett, and A. K. Seth, Phys. Rev. Lett. **103**, 238701 (2009).
 - [46] P. L. Williams and R. D. Beer (2010), arXiv: cs.IT/1004.2515.
 - [47] N. Timme, W. Alford, B. Flecker, and J. M. Beggs, J. Comput. Neurosci. **36**, 119 (2014).
 - [48] J. T. Lizier, N. Bertschinger, J. Jost, and M. Wibral, Entropy **20** (2018).
 - [49] A. Rhee, R. Cheong, and A. Levchenko, Phys. Biol. **9**, 045011 (2012).
 - [50] A. Levchenko and I. Nemenman, Curr. Opin. Biotechnol. **28**, 156 (2014).
 - [51] J. Keizer, *Statistical Thermodynamics of Nonequilibrium Processes* (Springer-Verlag, Berlin, 1987).
 - [52] R. Grima, Phys. Rev. E **92**, 042124 (2015).