


## Dynamical Information Synergy in Biochemical Signaling Networks

Lauritz Hahn<sup>1</sup>, Aleksandra M. Walczak<sup>2,\*</sup> and Thierry Mora<sup>1,3\*</sup>

*Laboratoire de Physique de l'École normale supérieure, CNRS, PSL University,  
Sorbonne Université, and Université Paris Cité, Paris, France*

 (Received 10 January 2023; accepted 28 July 2023; published 20 September 2023)

Biological cells encode information about their environment through biochemical signaling networks that control their internal state and response. This information is often encoded in the dynamical patterns of the signaling molecules, rather than just their instantaneous concentrations. Here, we analytically calculate the information contained in these dynamics for a number of paradigmatic cases in the linear regime, for both static and time-dependent input signals. When considering oscillatory output dynamics, we report on the emergence of synergy between successive measurements, meaning that the joint information in two measurements exceeds the sum of the individual information. We extend our analysis numerically beyond the scope of linear input encoding to reveal synergetic effects in the cases of frequency or damping modulation, both of which are relevant to classical biochemical signaling systems.

DOI: 10.1103/PhysRevLett.131.128401

To react and adapt to varying internal and external conditions, cells use networks of signaling proteins to convey and process information. Recent experiments have shown that these networks often show complex dynamical behavior, such as relaxation to a steady state, pulses, oscillations, or bistable switches [1–3]. Given a specific regulatory network topology, different stimuli can produce distinct dynamical responses by messenger molecules. For example, the transcription factor nuclear factor kappa-B (NF- $\kappa$ B) exhibits damped oscillations in cells stimulated with tumor necrosis factor- $\alpha$  (TNF $\alpha$ ) [4,5] whereas stimulation with bacterial lipopolysaccharide (LPS) leads to a single, prolonged wave [5]. These distinct responses help to explain how NF- $\kappa$ B can be involved in such diverse processes as inflammatory response, cell differentiation, cell proliferation, apoptosis, and more [4,6]. This has led to the hypothesis that cells use the temporal dynamics of signaling molecules to transmit information about both identity and intensity of stimuli [1,2].

Information theory has been a useful tool for quantifying the information flow in biochemical networks [7–10]. While its application is often restricted to static measurements of the output signal, recent experimental studies have used information theory to quantify the reliability of signal transmission in biochemical networks by estimating the mutual information (MI) between input stimuli and the dynamical responses of signaling molecules such as NF- $\kappa$ B [9,11,12], extracellular signal-regulated kinase (ERK) [11,13], calcium ions (Ca<sup>2+</sup>) [3,11], or nuclear translocation of transcription factors [14]. These results indicate that information transmission is increased when considering the temporal dynamics as compared to a static scenario. Concomitantly, progress has been made in calculating

analytically the mutual information between input stimuli and dynamical responses for time traces of infinite lengths in simple linearized models of biochemical signaling [15,16]. However, computations of information for trajectories of finite lengths have mostly been limited to numerical investigation [17,18] (sometimes aided by analytical approximations [19]), despite their relevance for cells that make quick decisions in response to external cues. Here we develop a framework for computing analytically the mutual information between the input and the time trace of the output signal, using a linear approximation. Focusing on the onset of a constant input, we demonstrate the existence of regimes in which information transmission is synergetic, i.e., the information contained in two time points jointly is larger than the sum of the information contained in the individual time points.

We consider a biochemical network with one input species  $X$  and one output species  $Y$  [Fig. 1(a)]. Our goal is to calculate the mutual information between a time trace of the input concentration, denoted by vector  $\mathbf{x} \equiv (x_{t_1}, \dots, x_{t_m})$ , and a set of measurements of the output concentration  $\mathbf{y} \equiv (y_{t_1}, \dots, y_{t_n})$ :  $I(\mathbf{x}; \mathbf{y}) = \int d\mathbf{x}d\mathbf{y} p(\mathbf{x}, \mathbf{y}) \times \log p(\mathbf{x}, \mathbf{y}) / p(\mathbf{x})p(\mathbf{y})$  [Fig. 1(b)].

In general, changes in the output species  $Y$  are determined by the history of  $x$  and  $y$  themselves. We start by assuming that this dependency is instantaneous, with no delays, so that increments of  $y$  happen with a rate depending on  $x_t$  and  $y_t$  only (we will relax that assumption later). We also assume  $y \gg 1$ , so that we can use the small-noise approximation and describe the evolution of  $y$  through the stochastic differential equation:

$$\dot{y}_t = f(x_t, y_t) + \sqrt{2D(x_t, y_t)}\eta_t, \quad (1)$$

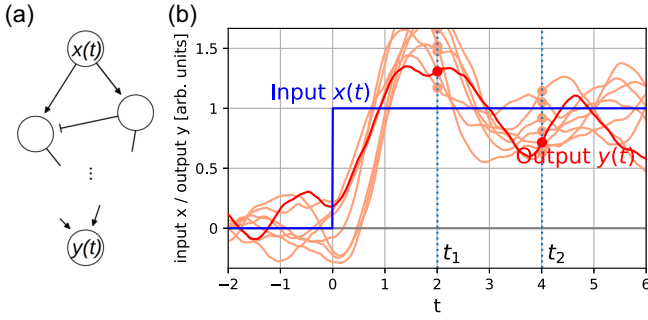


FIG. 1. Calculating mutual information in time traces. (a) We consider an input signal, either static or dynamic, that is transmitted through a biochemical network and thus produces a time-dependent output signal whose dynamic patterns may encode information about the input. (b) To evaluate the information transmission, we sample the input-output trajectories at a given number of time points and compute the mutual information. In certain cases this is possible analytically; otherwise numerical estimators can be used.

where  $f$  is an arbitrary regulation function that subsumes all the details of the regulation network between  $Y$  and  $X$ , and  $\eta_t$  a unitary Gaussian white noise. Linearizing these dynamics around  $(x^*, y^*)$ , we can write

$$\dot{y}_t = \tau^{-1}(x - y_t) + \sqrt{2D}\eta_t, \quad (2)$$

where  $x$  and  $y$  were shifted and rescaled without losing generality so that  $(x^*, y^*) = (0, 0)$ , and  $\partial_y f = -\partial_x f$ .

We start by considering the situation in which the input changes from 0 to a value  $X \sim \mathcal{N}(0, \sigma_x)$  at  $t = 0$ , corresponding to the sudden activation of the signaling pathway, e.g., due to some environmental change. Equation (2) may be integrated exactly, so that  $y_t$  conditioned on  $x$  is normally distributed:

$$p(y_t|x) = \frac{1}{Z} \exp \left[ -\frac{[y_t - x(1 - e^{-t/\tau})]^2}{2D\tau(1 - e^{-2t/\tau})} \right], \quad (3)$$

where we have assumed  $y(0) = 0$ . Since  $x$  is also normally distributed,  $(x, \mathbf{y})$  is distributed as a multivariate Gaussian, and all mutual information values may be calculated exactly (see SM [20]). The information  $I(X; Y_{t_1})$  carried by a single measurement of  $y$  at time  $t_1$  can be written as a function of the signal-to-noise (SNR) ratio  $S(X; Y_{t_1})$  (SM [20]):

$$I = \frac{1}{2} \log(1 + S), \quad (4)$$

$$S(X; Y_{t_1}) = \frac{\sigma_x^2}{D\tau} \frac{1 - e^{-t_1/\tau}}{1 + e^{-t_1/\tau}}. \quad (5)$$

Typically,  $I(X; Y_{t_1})$  is reported in the  $t_1 \rightarrow \infty$  limit, where it is maximal, yet cells can rarely wait that long for a readout.

However, a cell is not limited to one measurement of  $y$ : it can make multiple measurements, or exploit the information contained in the whole output trajectory.

To calculate the mutual information between a static input  $X$  and an interval of the output trajectory  $\mathbf{y} = \{y_t\}_{t \in [t_1, t_1+T]}$ , we use Bayes's law to write the posterior probability of  $x$  given the history of  $y$  as

$$p(x|\mathbf{y}) = \frac{p(x)p(y_{t_1}|x)}{p(\mathbf{y})} \prod_{t=t_1, t_1+\delta t, \dots} p(y_{t+\delta t}|y_t, x). \quad (6)$$

The logarithm of each term of this product is quadratic in  $x$ , meaning that the posterior is Gaussian. Collecting the terms in  $x^2$  and taking the  $\delta t \rightarrow 0$  limit gives us the inverse of the posterior variance  $\text{Var}(X|\mathbf{y})$ , from which we deduce the mutual information  $I(X; \mathbf{Y})$  contained in the entire trajectory as Eq. (4) with

$$S(X; \mathbf{Y}) = \frac{\sigma_x^2}{D\tau} \frac{1 - e^{-T/\tau}}{1 + e^{-T/\tau}} + \frac{\sigma_x^2 T}{2D\tau^2}. \quad (7)$$

This SNR is the sum of the SNR given by a single measurement [Eq. (5)], and that provided by an effective number  $T/2\tau$  of additional *independent* measurements that the trace provides. As usual when combining several measurements, the SNR grows linearly and the MI logarithmically according to a law of diminishing returns, meaning that these measurements are redundant, making synergy between them impossible with these memoryless dynamics.

The model of Eq. (2) cannot describe oscillatory behavior, which is observed in several well-studied systems such as the above-mentioned NF- $\kappa$ B, ERK, and  $\text{Ca}^{2+}$ , but also other transcription factors such as p53 [23,24], Crz1 [25] or Msn2 [26,27]. To account for this behaviour, we can consider linearized second-order dynamics, which take the form of an underdamped oscillator under external forcing:

$$\ddot{y}_t = -\gamma \dot{y}_t - \omega_0^2(y_t - x) + \gamma \sqrt{2D}\eta_t, \quad (8)$$

where  $\gamma$  is the damping coefficient,  $\omega_0^2 = \gamma/\tau$ , and  $\Omega = \sqrt{\omega_0^2 - \gamma^2/4}$  the natural frequency of the oscillator. Model (2) corresponds to the overdamped limit  $\gamma \rightarrow \infty$ .

Using the same approach as above (see SM [20]), we can calculate  $I(X; Y_{t_1})$  as Eq. (4) with

$$S(X; Y_{t_1}) = \frac{\sigma_x^2 [1 - e^{-\gamma t_1/2} (\cos \Omega t_1 + \frac{\gamma}{2\Omega} \sin \Omega t_1)]^2}{\text{Var}(Y_{t_1}|X)}, \quad (9)$$

with  $\text{Var}(Y_t|x) = (D\gamma/4\Omega^2\omega_0^2)[4\Omega^2(1 - e^{-\gamma t}) + \gamma^2 e^{-\gamma t} \times (\cos 2\Omega t - 1) - 2\gamma\Omega \sin 2\Omega t]$ . Unlike in the overdamped case, the MI does not necessarily increase with  $t_1$ , but instead is itself subject to oscillations, and is maximal for  $t_1 = \pi/\Omega$  [Fig. 2(b)].

We can next calculate the mutual information between  $x$  and the trajectory of  $y$  by introducing the auxiliary variable  $z = \dot{y}$  to make the system Markovian.  $I(X; Y) = I(X; Y, Z)$  is given by Eq. (4) with

$$S(X; Y) = \frac{\sigma_x^2 [1 - e^{-\frac{\gamma t_1}{2}} (\cos \Omega t_1 + \frac{\gamma}{2\Omega} \sin \Omega t_1)]^2}{\text{Var}(Y_{t_1}|X)} + \frac{\sigma_x^2 \omega_0^4 e^{-\gamma t_1} \sin^2 \Omega t_1}{\text{Var}(Z_{t_1}|X) \Omega^2} + \frac{\omega_0^4 \sigma_x^2}{2\gamma^2 D} T, \quad (10)$$

with  $\text{Var}(Z_t|x) = (D\gamma/4\Omega^2)[4\Omega^2(1 - e^{-\gamma t}) + \gamma^2 e^{-\gamma t} \times (\cos 2\Omega t - 1) + 2\gamma\Omega \sin 2\Omega t]$ . Note that the  $T \rightarrow 0$  limit corresponds to the information given by an instantaneous measurement of  $Y$  and its derivative  $Z$ , which gives more information than  $Y$  alone [Eq. (9)].

The scaling with observation time  $T$  in Eq. (10) has the same property of diminishing return as the overdamped case. However, synergy can emerge if we consider two measurements  $y_{t_1}$  and  $y_{t_2}$ . The corresponding mutual information  $I(X; Y_{t_1}, Y_{t_2})$  may be calculated analytically using the Markovian propagator  $p(y_{t_2}|y_{t_1}, z_{t_1}, x)$  (see SM [20]). At steady state ( $t_1 \rightarrow \infty$ ), it simplifies to Eq. (4) with

$$S = \frac{\omega_0^2 \sigma_x^2}{\gamma D} + \frac{\sigma_x^2 (1 - e^{-\frac{\gamma \Delta t}{2}} (\cos \Omega \Delta t + \frac{\gamma}{2\Omega} \sin \Omega \Delta t))^2}{\text{Var}[Y(\Delta t)|X] + \frac{D}{\gamma^2} e^{-\gamma \Delta t} \sin^2 \Omega \Delta t}, \quad (11)$$

with  $\Delta t = t_2 - t_1$ , which is plotted in Fig. 2(a). We observe that the joint information is maximal when the second measurement is done  $\Delta t = \pi/\Omega$  after the first, i.e., at opposite phase. The resting position of the oscillator is then approximately the average of the two measurements, irrespective of the phase of the first measurement. In that case, the two measurements are *synergistic*,  $\text{Synergy} = I(X; (Y_{t_1}, Y_{t_2})) - I(X; Y_{t_1}) - I(X; Y_{t_2}) > 0$ . They provide more information together than the sum of each, which are confounded by lack of phase information. Figure 2(b) shows the more realistic case of a first measurement in finite time, and optimal delay  $\Delta t = \pi/\Omega$ , confirming that synergy is a generic outcome. Optimal readout strategies are usually assumed to maximize information. The corresponding optimal measurement times are then  $t_1^* < \pi/\Omega$  and  $t_2^* - t_1^* = \pi/\Omega$ , versus  $t^* = \pi/\Omega$  for a single measurement. While in this case maximizing information and synergy are different, the most informative measurements are synergistic. The phase diagram of the steady-state synergy as a function of the dimensionless parameters of the dynamics,  $\omega_0/\gamma$ , and the steady-state SNR  $\equiv \omega_0^2 \sigma_x^2 / D\gamma$  [Fig. 2(c)] shows that synergy emerges when damping is weak and noise is large, which is the regime in which resonant effects are strong.

We can apply our formulas to experimental measurements of the response of the ERK pathway activated by an optogenetic actuator [13] in the oscillatory regime. From

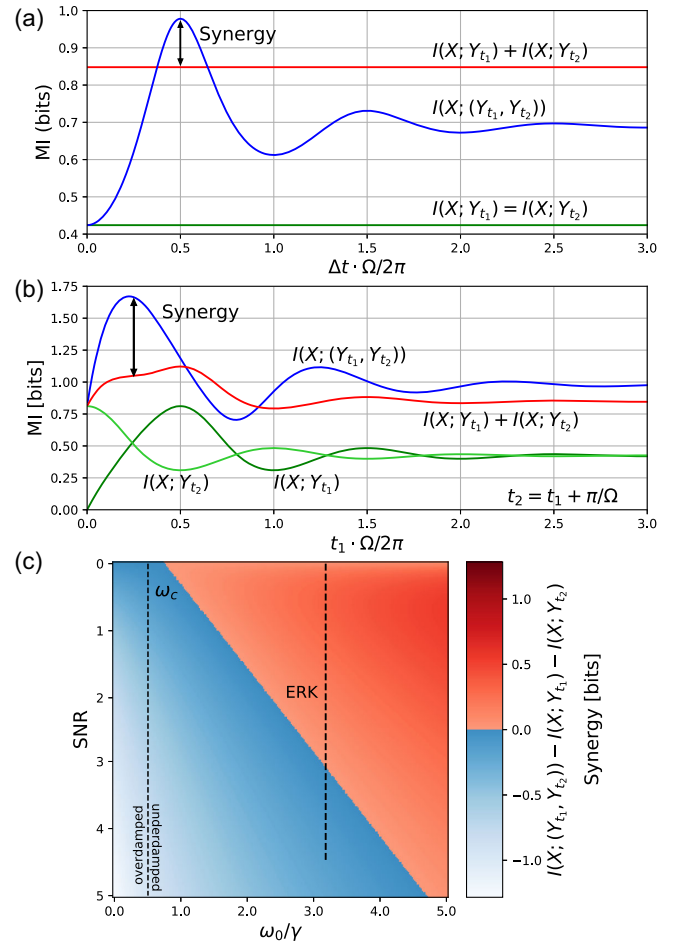


FIG. 2. Synergy in information transmission between successive measurements. (a) Information contained jointly in two successive measurements as a function of the delay  $\Delta t$  between them, for the oscillatory dynamics of Eq. (8) at steady state. Synergy is observed when  $I(X; (Y_{t_1}, Y_{t_2})) > I(X; Y_{t_1}) + I(X; Y_{t_2})$ . (b) Information of two measurements as a function of the time of the first measurement,  $t_1$ , for fixed delay  $\Delta t = t_2 - t_1 = \pi/\Omega$ . (c) Phase diagram of the synergy in steady state as a function of SNR  $= \omega_0^2 \sigma_x^2 / D\gamma$  and  $\omega_0/\gamma$ , with  $\Delta t = t_2 - t_1 = \pi/\omega$ . Dashed line shows the range of experimental values estimated from Ref. [13] where ERK is activated by optogenetics. All data are for  $\sigma_x^2 = 1$ ,  $\gamma = 1$ , and in (a) and (b),  $D = 5$ ,  $\omega_0 = 2$ .

Fig. 2(e) of Ref. [13] we estimate  $\Omega^{-1} \approx 1.43$  min and  $\gamma^{-1} = 4.5$  min, so that  $\omega_0/\gamma \approx 3.2$ , and  $\text{Var}(Y_{t \rightarrow \infty}|X) \approx 0.013$  in the experiment's arbitrary units of normalized fluorescence. The output variance depends on the dynamic range of inputs, but is lower than 0.06 in those units, giving a SNR varying between 0 and  $\approx 4.5$ . In this experimental regime, the total information of two time points may be as large as 2.6 bits, with synergy appearing for  $\text{SNR} \lesssim 3.3$  [Fig. 2(c), dashed line]. This suggests that synergy may be relevant in the physiological regime of those experiments, and could be exploited by cells in downstream signaling.

So far we have considered the case of an input affecting the resting position of the output  $y$ . However, other

encodings are also common in biological systems, such as frequency modulation, proposed for Msn2 [27], Crz1 [25], and  $\text{Ca}^{2+}$  [28,29], or damping modulation, similar to the distinct dynamical responses of NF- $\kappa$ B when stimulated with TNF $\alpha$  or LPS [5]. These encodings are no longer Gaussian, so we must turn to numerical methods to estimate mutual information. We first generate solutions to Eq. (8) for a large number ( $N \sim 10^4$ ) of sampled inputs. We then either calculate the empirical mean of  $-\log[P(x|\mathbf{y})/P(x)]$  using the Markovian expression [Eq. (6)], or use the  $k$ -nearest-neighbor (knn) estimators developed by Kraskov, Stögbauer, and Grassberger (KSG) [30], and Selimkhanov *et al.* [11]. While these estimators are very flexible and widely used, they have been criticized for not taking information encoded in the temporal order of successive measurements into account [12]. We benchmark the estimators using the exact results derived in the previous section (see SM [20] for more details).

We first study frequency modulation by considering an underdamped system [Eq. (8)] with null resting position, and two equiprobable input frequencies  $\{\omega_{0,1}, \omega_{0,2}\}$ , corresponding to two discrete stimuli that the cell is trying to distinguish, and which control the response frequency. We initialize the system either at a random value drawn from the steady state, or from  $y(0) = 0$ , and compute the MI in the relaxation dynamics, i.e., we ask how well the two frequencies can be distinguished. When starting from a random value, a single measurement should contain very little information, but two measurements should together allow for a good estimate of  $\omega_0$ , so we expect to find synergy. These predictions are confirmed by Fig. 3(a), which shows that information carried by several measurements is larger than the sum of individual ones (global synergy,  $I_{1:n} > I_1 + \dots + I_n$ ). In the case of a fixed initial condition, one measurement can already distinguish frequencies since the initial phase is fixed, unless the noise has had time to randomize the phase.

Next, we consider damping modulation by computing the mutual information between an output evolving according to Eq. (8) with null resting position and equiprobable binary input damping coefficients  $\{\gamma_1, \gamma_2\}$ , chosen on both sides of the critical damping transition  $\gamma_1 > \gamma_c = 2\omega_0 > \gamma_2$ . Once again, synergy is observed, especially when the initial condition is drawn from the steady state [Fig. 3(b)]. Since all other parameters are equal, single measurements give almost no information on whether the dynamics are overdamped or underdamped.

In this Letter we have derived analytical and numerical solutions for the information carried by an output signal in response to a constant input, which corresponds to the typical experiments of Refs. [3,11,12,14]. In contrast, previous theoretical work on information from temporal trajectories has focused on the case of Gaussian fluctuating inputs at steady state [15,16], where information can be decomposed in the frequency domain [31,32]. For

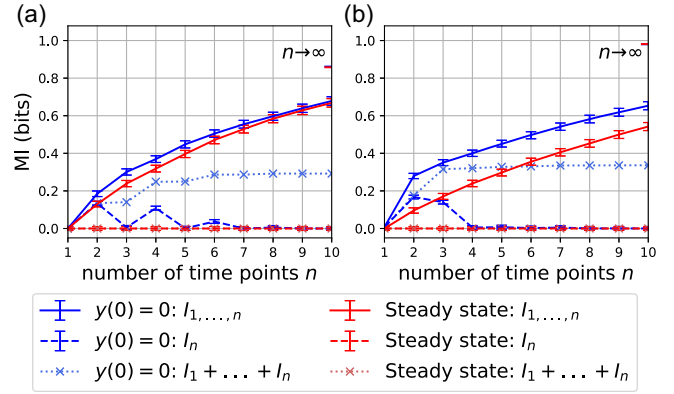


FIG. 3. Frequency and damping modulation. Information contained in  $n$  successive equidistant measurements, with fixed ( $y(0) = 0$ ) or random (steady-state) initial conditions. The binary input consists of (a) two different frequencies in the underdamped regime; and (b) two different damping coefficients in the overdamped and underdamped regimes respectively. Solid curves show the joint MI of  $n$  measurements, dashed curves the MI of the  $n$ th measurement alone, and the dotted curves the sum of the individual MIs up to that point. Synergy is observed when the solid line is above the dotted line. Note that since the input is binary, all  $\text{MI} \leq 1$  bit. The curves were obtained using  $T = 8$ ,  $D = 4$ ,  $\sigma_x^2 = 1$ , and (a)  $\omega_0 = 2$ ,  $\gamma_1 = 1$ ,  $\gamma_2 = 6$ ,  $\Delta t = \pi/\Omega$ , (b)  $\gamma = 1$ ,  $\omega_{0,1} = 2$ ,  $\omega_{0,2} = 3$ ,  $\Delta t = 2\pi/(\Omega_1 + \Omega_2)$ .

completeness, here we present exact results for a fluctuating input  $x$  [following an Ornstein-Uhlenbeck process  $\dot{x}_t = -x_t/\tau_x + \sqrt{2D_x}\eta_x(t)$  with  $\langle \eta_x(t)\eta_x(t') \rangle = \delta(t-t')$ ] but with a finite observation window  $(0, T)$  at steady state. When  $y$  responds to  $x$  as before, Eqs. (2) or (8), the joint distributions of  $\mathbf{x} = (x_t)_{t \in (0, T)}$  and  $(\mathbf{y}, \mathbf{z}) = (y_t, z_t)_{t \in (0, T)}$  are multivariate Gaussians whose covariance matrices have a tridiagonal structure. After calculating their determinants we obtain exact but lengthy expressions for  $I(\mathbf{X}; \mathbf{Y})$ , which are given in SM [20]. Taking the large time limit gives back the classical information rate of [16]

$$\lim_{T \rightarrow \infty} \frac{I(\mathbf{X}; \mathbf{Y})}{T} = \frac{1}{2\tau_x} \left( \sqrt{1 + \frac{D_x \tau_x^2}{D\tau^2}} - 1 \right), \quad (12)$$

for both the overdamped and underdamped cases.

Information synergy had been previously discussed in neuroscience as a property of groups of neurons [33], or between spikes of the same neuron [34]. Our models demonstrate that synergy could also be relevant in cellular signaling in the physiological regime when the response function is dynamic and nonlinear, allowing cells to extract more information from stimuli and to make faster decisions. Potential candidates for synergetic signaling include Msn2, Crz1, or  $\text{Ca}^{2+}$ , which have been found to use frequency modulation to encode information, and NF- $\kappa$ B in response to TNF $\alpha$  and LPS, which can show both overdamped and underdamped dynamics depending on

the stimulus [5]. Frequency modulation has also been hypothesized to be used in some pathways to encode information [27–29]. However, it is not clear whether such synergistic signaling designs are optimal or evolutionary adaptive, and how the nature and statistics of the input modulate synergy. Another outcome of our work is exact solutions for the information content in finite trajectories, which we used to benchmark the performance of mutual information estimators that are frequently used in experimental research. Our results provide analytical foundations for a better understanding of dynamical information in biochemical signaling, and suggest new directions for both experimental and theoretical research.

We thank M. Kramar and Huy Tran for fruitful discussions.

---

\*These authors contributed equally.

- [1] B. N. Kholodenko, *Nat. Rev. Mol. Cell Biol.* **7**, 165 (2006).
- [2] J. E. Purvis and G. Lahav, *Cell* **152**, 945 (2013).
- [3] G. D. Potter, T. A. Byrd, A. Mugler, and B. Sun, *Biophys. J.* **112**, 795 (2017).
- [4] A. Hoffmann, A. Levchenko, M. L. Scott, and D. Baltimore, *Science* **298**, 1241 (2002).
- [5] M. W. Covert, T. H. Leung, J. E. Gaston, and D. Baltimore, *Science* **309**, 1854 (2005).
- [6] S. L. Werner, D. Barken, and A. Hoffmann, *Science* **309**, 1857 (2005).
- [7] A. A. Margolin, I. Nemenman, K. Basso, C. Wiggins, G. Stolovitzky, R. Dalla Favera, and A. Califano, *BMC Bioinform.* **7 Suppl 1**, S7 (2006).
- [8] G. Tkacik, C. G. Callan, and W. Bialek, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 12265 (2008).
- [9] R. Cheong, A. Rhee, C. J. Wang, I. Nemenman, and A. Levchenko, *Science* **334**, 354 (2011).
- [10] A. M. Walczak and G. Tkačik, *J. Phys. Condens. Matter* **23**, 153102 (2011).
- [11] J. Selimkhanov, B. Taylor, J. Yao, A. Pilko, J. Albeck, A. Hoffmann, L. Tsimring, and R. Wollman, *Science* **346**, 1370 (2014).
- [12] Y. Tang, A. Adelaja, F. X. Ye, E. Deeds, R. Wollman, and A. Hoffmann, *Nat. Commun.* **12**, 1 (2021).
- [13] C. Dessauges, J. Mikelson, M. Dobrzyński, M.-A. Jacques, A. Frismantiene, P. A. Gagliardi, M. Khammash, and O. Pertz, *Mol. Syst. Biol.* **18**, e10670 (2022).
- [14] A. A. Granados, J. M. J. Pietsch, S. A. Cepeda-Humerez, I. L. Farquhar, G. Tkačik, and P. S. Swain, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 6088 (2018).
- [15] F. Tostevin and P. R. Ten Wolde, *Phys. Rev. Lett.* **102**, 218101 (2009).
- [16] F. Tostevin and P. R. Ten Wolde, *Phys. Rev. E* **81**, 061917 (2010).
- [17] L. Duso and C. Zechner, in *2019 IEEE 58th Conference on Decision and Control (CDC)* (IEEE, Nice, France, 2019), pp. 6610–6615.
- [18] M. Reinhardt, G. Tkačik, and P. R. ten Wolde, *arXiv*: 2203.03461.
- [19] A.-L. Moor and C. Zechner, *Phys. Rev. Res.* **5**, 013032 (2023).
- [20] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevLett.131.128401> for mathematical details, which includes Refs. [21,22].
- [21] C. M. Holmes and I. Nemenman, *Phys. Rev. E* **100**, 022404 (2019).
- [22] S. Gao, G. V. Steeg, and A. Galstyan, *Proc. Mach. Learn. Res.* **38**, 277 (2015), <https://proceedings.mlr.press/v38/gao15.html>.
- [23] D. A. Hamstra, M. S. Bhojani, L. B. Griffin, B. Laxman, B. D. Ross, and A. Rehemtulla, *Cancer Res.* **66**, 7482 (2006).
- [24] E. Batchelor, A. Loewer, C. Mock, and G. Lahav, *Mol. Syst. Biol.* **7**, 488 (2011).
- [25] L. Cai, C. K. Dalal, and M. B. Elowitz, *Nature (London)* **455**, 485 (2008).
- [26] C. Garmendia-Torres, A. Goldbeter, and M. Jacquet, *Curr. Biol.* **17**, 1044 (2007).
- [27] N. Hao and E. K. O’Shea, *Nat. Struct. Mol. Biol.* **19**, 31 (2012).
- [28] M. Berridge, *Nature (London)* **386**, 759 (1998).
- [29] M. J. Boulware and J. S. Marchant, *Curr. Biol.* **18**, 769 (2008).
- [30] A. Kraskov, H. Stögbauer, and P. Grassberger, *Phys. Rev. E* **69**, 066138 (2004).
- [31] R. M. Fano, *Transmission of Information: A Statistical Theory of Communication* (MIT Press, Cambridge, MA, 1961).
- [32] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes* (Holden-Day, San Francisco, 1964).
- [33] E. Schneidman, W. Bialek, and M. J. Berry II, *J. Neurosci.* **23**, 11539 (2003).
- [34] N. Brenner, S. P. Strong, R. Koberle, W. Bialek, and R. R. d. R. van Steveninck, *Neural Comput.* **12**, 1531 (2000).