The strength of protein-protein interactions controls the information capacity and dynamical response of signaling networks

Ching-Hao Wang* and Pankaj Mehta†
Department of Physics and Biological Design Center, Boston University, Boston, MA 02215, USA

Caleb J. Bashor‡
Department of Bioengineering, Rice University, Huston, TX 77030, USA
(Dated: November 13, 2018)

Eukaryotic cells transmit information by signaling through complex networks of interacting proteins. Here we develop a theoretical and computational framework that relates the biophysics of protein-protein interactions (PPIs) within a signaling network to its information processing properties. To do so, we generalize statistical physics-inspired models for protein binding to account for interactions that depend on post-translational state (e.g. phosphorylation). By combining these models with information-theoretic methods, we find that PPIs are a key determinant of information transmission within a signaling network, with weak interactions giving rise to "noise" that diminishes information transmission. While noise can be mitigated by increasing interaction strength, the accompanying increase in transmission comes at the expense of a slower dynamical response. This suggests that the biophysics of signaling protein interactions give rise to a fundamental "speed-information" trade-off. Surprisingly, we find that cross-talk between pathways in complex signaling networks do not significantly alter information capacity–an observation that may partially explain the promiscuity and ubiquity of weak PPIs in heavily interconnected networks. We conclude by showing how our framework can be used to design synthetic biochemical networks that maximize information transmission, a procedure we dub "InfoMax" design.

PACS numbers: 75.50.Pp, 75.30.Et, 72.25.Rb, 75.70.Cn

Introduction

Cells have evolved complex protein signaling networks to process information about their living environments [1–4]. These networks play a central role in cellular decision-making, development, growth, and migration [5–7]. In eukaryotic cells, signaling pathways such as Wnt/β-Catenin[8, 9] and TGF-β pathways[10] have important homeostatic functions (e.g., cell proliferation, differentiation, and fate determination), with disruptions in their signaling leading to tumorigenesis and drive metastasis [11, 12].

Information transfer in signaling networks occurs via the addition of covalent chemical groups that alter the regulatory state of a signaling protein (e.g. phosphorlyation of a Tyrosine residue). Addition and removal these post-translational modifications (PTMs) are respectively catalyzed by "writer" (e.g. a kinase) and "eraser" (e.g. a phosphatase) enzyme activities. Information transfer occurs when the ratio of these opposing activities is altered by an upstream input (e.g. ligand binding to receptor), and becomes rapidly and reversibly encoded in the PTM state of the downstream substrate (e.g. phosphorylated or non-phosphorylated). An important breakthrough in the understanding of signaling network connectivity came with the discovery of protein-protein interaction (PPI) domains that specifically bind to PTM-modified motifs, effectively "decoding" the PTM state of a substrate[6, 13]. By linking an activity to a substrate through binding, PTM-mediated PPI interactions serve as signaling network links by interconnecting writer/eraser cycles (Fig. 1A). One of the best-known examples of a PTM-binding domain is the Src homology 2 (SH2) domain, which specifically docks to motifs containing phosphorylated tyrosine. For example, SH2 recognition plays a central role in the EGF pathway, connecting initial receptor autophosphorylation to downstream signaling events via recruitment of SH2 domain-containing enzymes to their substrates (Fig. 1A,C).

Given their role in mediating information transfer between signaling proteins, the question naturally arises as to how the biophysical features of PTM-PPI interactions relate to a pathway's emergent, network-level information processing properties. Here, we create a theoretical framework for exploring this relationship using a thermodynamically-inspired statistical model in which biochemical partition functions relate the probability of finding the system in a given state (e.g. bound, unbound, etc.) to relevant biophysical features like interaction affinity and species concentration [14–16]. Models of this class have been successfully used to understand the biophysics of promoter regulation in transcriptional networks[16–19]. Here, we extend this approach to signaling networks by introducing variables representing PTM-dependent PPIs, thereby accounting for the

*Electronic address: chinghao@bu.edu
†Electronic address: pankajm@bu.edu
‡Electronic address: caleb.bashor@rice.edu

non-equilibrium nature of reversible, enzyme-catalyzed phosphorylation. We combine this statistical physics approach with information theory[20, 21], which has seen widespread recent application in biology[22]. Examples include the modeling of information processing in gene networks[23–26], enzyme cascades [27], and bacterial signaling networks[28, 29], as well as calculating information capacity in canonical eukaryotic signaling networks from single cell measurement of input-output relationships[30, 31].

Our joint framework allows us to investigate the relationship between the biophysics of PPI-PTM interactions and signaling network information processing. We chose to model a simple, idealized signaling pathway in order to more directly probe this relationship. Here, our approach is inspired by synthetic biology, where a principle goal is engineering synthetic regulatory circuits capable of executing designed regulatory function, typically through direct experimental manipulating features like protein expression level and PPI strength. Thus, in contrast to previous approaches that investigate the information capacity of pre-existing, native networks, our goal with this work is to ask how we can manipulate the biophysics of PPIs to engineer new networks that optimize information transmission. Information processing circuits must necessarily balance three competing requirements that are often in tension: i) minimizing unwanted "noise" that corrupts the true signal, ii) ensuring that the circuits can respond quickly to dynamical perturbations, and iii) maximizing the dynamic range of inputs. In signaling networks, it has been argued significant noise is introduced by weak, promiscuous PPIs[32, 33], often in combination with low levels of background kinase and phosphatase activity[34, 35].Thus, we hypothesize in the current work that while increasing the strength of PTM-PPI interactions may reduce noise, it may also involve inherent tradeoffs in response times and dynamic range.

Motivated by these considerations, we focus in this article on a series of interrelated conceptual questions: How can we quantify noise due to promiscuous PPIs? How does the strength of PPIs affect information transmission and dynamic response times in signaling networks? How do network architecture and cross-talk affect information transmission[33, 36–39]? Can we rationally choose PPIs in synthetic biochemical networks that maximize information transmission? We begin by discussing how to generalize thermodynamic models to binding that include PTMs. We then discuss how basic elements of these models can serve as an input into information theoretic calculations. Using this framework, we quantitatively show how weak PPIs give rise to non-specific binding, resulting in "noise" that reduces information transmission. We then show that while noise can be diminished by increasing PPI strength, increased information transmission comes results in a slower dynamical response—a biophysical manifestation of what in engineering is often called the "gain-bandwidth" tradeoff. We then show that cross-talk between pathways in highly interconnected signaling networks does not significantly alter information capacity. We conclude by discussing "InfoMax", a new procedure for designing synthetic biochemical networks that optimize gain-bandwidth tradeoff.

Including post-translational modifications in thermodynamic models

To construct a thermodynamic model, we consider an idealized post-translational signaling network with phosphorylation as the only PTM. Each node in the network represents a distinct kinase activity, and linkages between nodes are mediated by PTM-dependent PPIs (Figure 1A). Here, phosphorylation of a kinase node by an upstream activity renders it 'active' and competent to engage with and phosphorylate (and subsequently activate) a downstream kinase. We sought to create a generalizable thermodynamic expression for describing such a network.

For a given multi-state molecular system, thermodynamics provide a concise description of the statistical weight of each state, and therefore the probability of observing a state when the system is at steady state. At thermal equilibrium the statistical weight of a given microscopic configuration is proportional to its Boltzmann factor defined as $e^{-\beta E}$, where E is the energy of this microstate and $\beta = 1/(k_B T)$ is the inverse temperature with k_B being the Boltzmann constant. As we noted, conventional thermodynamic prescription based on transcriptional regulation[16–19] does not include PTMs and PTM-dependent bindings. Here we introduce a new set of variables to account for PTMs.

For brevity, we consider a simplified signaling network (Figure 1C); a linear pathway consisting of a membrane-spanning receptor kinase R, a single freely-diffusing protein kinases K_1, and target transcription factor TF. We treat $R, K_1, TF \in \{0, 1\}$ with value 1 indicating a phoshphorylated state (transcribed state for TF) and 0 otherwise. Pathway activation (input) is initiated by ligand (L) binding to the receptor at the cell surface, leading to receptor autophosphorylation (i.e. $R = 1$). This results in phosphorylation-dependent recruitment and phosphorylation of K_1 (i.e. $K_1 = 1$). Phosphorylated K_1 then translocates into the nucleus where it binds to and phosphorylates TF (i.e. $TF = 1$), activating transcription.

Within the context of this simplified signaling system, we begin to describe the thermodynamics of the interactions involved, breaking down the network depicted in Figure 1C into three parts and enumerating the possible states within each. As depicted in Figure 1A,C, the receptor kinase only has two possible PTM states (phosphorylated or not). We label the probability of phosphorylated receptor kinase as $\mathbb{P}(R = 1) = q$, where $q \in [0, 1]$ is the parameter that encapsulates ligand ac-

tivation. The probability of the complementary configuration is therefore given by $\mathbb{P}(R = 0) = 1 - q$. (ii) Based on our discussion above, the interaction between R and K depends crucially on the value of R. Simple enumeration reveals that there are four possible scenarios: $(K_1, R) \in \{(0,0),(1,0),(0,1),(1,1)\}$, as shown in Figure 1D. The first two involve the interaction between unphosphorylated receptor (i.e. $R = 0$) and K_1 while the last two involve that between phosphorylated receptor (i.e. $R = 1$) and K_1. Thermodynamics dictates that when a system reaches equilibrium, the steady-state distribution of a microscopic state is given by the Boltzmann factor of that state divided by the sum of the Boltzmann factor of all possible states (i.e., partition function). It's worth noting that although enzymatic reactions (i.e. phosphorylation) are involved in signaling can drive a system out of equilibrium, we show in Appendix that the steady state distribution of a given state takes the Boltzmann form. In other words, the probability of having a phosphorylated PK given that the receptor kinase is phosphorylated, *viz.*

$$\mathbb{P}(K_1 = 1|R = 1) = \frac{e^{-\beta\theta_{R,K_1}}}{1 + e^{-\beta\theta_{R,K_1}}}, \tag{1}$$

where the numerator is the Boltzmann factor associated with this configuration while the denominator is the sum of this factor and that associated with $(K_1, R) = (0,1)$ (i.e., factor 1). Here we denote θ_{R,K_1} as the binding affinity (BA) of R to K_1 (i.e. $\theta_{R,K_1} = \Delta F - \tilde{\mu}$, where ΔF is the free energy difference between the bound and unbound state and $\tilde{\mu}$ is the chemical potential of phosphorylated R; see SI Section for its expression in terms of kinetic parameters.) By conservation, the complementary configuration has probability $\mathbb{P}(K_1 = 0|R = 1) = 1 - \mathbb{P}(K_1 = 1|R = 1) = 1/(1 + \exp(-\beta\theta_{R,K_1}))$. The cases where receptor is not phosphorylated (i.e. $R = 0$) are similar except that the binding affinity is parameterized by W. Since W controls the amount of low-probability, non-specific binding, we assume W is positive and large (i.e. $\beta W \gtrsim 1$) so that the probability of having a phosphorylated K_1 given that there's no signal input is almost zero, *viz.*

$$\mathbb{P}(K_1 = 1|R = 0) = \frac{e^{-\beta W}}{1 + e^{-\beta W}} \approx 0, \tag{2}$$

which implies $\mathbb{P}(K_1 = 0|R = 0) = 1 - \mathbb{P}(K_1 = 1|R = 0) \approx 1$. Note that practically this would require $\beta W \geq 4.60$ in order to achieve $e^{-\beta W} \leq 0.01$. With all these defined, one can summarize all four configurations and their statistical weights by the phosphorylation probability of K_1 conditioned on the state of R, $P(K_1|R)$ (see Fig. 1D). (iii) Finally, since the thermodynamic description of the interaction between K_1 and TF is the same as that between R and K_1, one can write down $\mathbb{P}(TF|K_1)$ in a similar fashion by relating θ_{R,K_1} to $\theta_{K_1,TF}$ (see Fig. 1D.)

Mutual information and PPIs

Mutual information between two random variables measures how much knowing one tells us about the other, usually measured in units of bits[20, 21]. In biology, it has been widely used to characterize the information transfer by biochemical systems [22–25, 27, 28, 30, 31]. Here we focus on defining this information-theoretic quantity in terms of PPIs for a given PK signaling network.

The mutual information of interest is that between the receptor kinase and TF output, $I(R;TF)$, since it quantifies how many input states cell can distinguish solely by examining its TF readout. Mathematically,

$$I(R;TF) = \sum_R \sum_{TF} \mathbb{P}(R)\mathbb{P}(TF|R) \log_2 \left[\frac{\mathbb{P}(TF|R)}{\mathbb{P}(TF)}\right]. \tag{3}$$

Note that since the summations in Eq.(3) are over $\{0,1\}$, this signaling network represents a discrete (binary) channel[21]. Physically speaking, $\mathbb{P}(TF)$ quantifies the transcriptional readout, $\mathbb{P}(TF|R)$ defines the input-output relation (i.e., channel transfer function), and $\mathbb{P}(R)$ measures the input, all at steady-states. Note that the state of PK, K_1, is absent from this expression since it is embedded in the input-output relation. Within the thermodynamic framework defined based on Fig. 1C and detailed in Fig. 1D, all quantities in Eq.(3) can be explicitly calculated: signal input $\mathbb{P}(R)$ is given in Fig. 1D while the channel input-output relation (i.e. transfer function), $\mathbb{P}(TF|R)$, is obtained by first invoking the conditional independence of TF and R on K_1, then marginalizing contributions from K_1, *viz.* $\mathbb{P}(TF|R) = \sum_{K_1} \mathbb{P}(TF|K_1)\mathbb{P}(K_1|R)$. Finally, the output is simply given by $\mathbb{P}(TF) = \sum_R \mathbb{P}(TF|R)\mathbb{P}(R)$. Explicitly, the transfer function is given by:

$$\begin{aligned}
\mathbb{P}(TF = 1|R = 1) &= \mathbb{P}(TF = 1|K_1 = 1)\mathbb{P}(K_1 = 1|R = 1) + \mathbb{P}(TF = 1|K_1 = 0)\mathbb{P}(K_1 = 0|R = 1) \\
&= \left(\frac{e^{-\beta\theta_{K_1,TF}}}{1 + e^{-\beta\theta_{K_1,TF}}}\right)\left(\frac{e^{-\beta\theta_{R,K_1}}}{1 + e^{-\beta\theta_{R,K_1}}}\right) + \left(\frac{e^{-\beta W}}{1 + e^{-\beta W}}\right)\left(\frac{1}{1 + e^{-\beta\theta_{R,K_1}}}\right) \qquad (4) \\
&\approx \left(\frac{e^{-\beta\theta_{K_1,TF}}}{1 + e^{-\beta\theta_{K_1,TF}}}\right)\left(\frac{e^{-\beta\theta_{R,K_1}}}{1 + e^{-\beta\theta_{R,K_1}}}\right) \qquad (5)
\end{aligned}$$

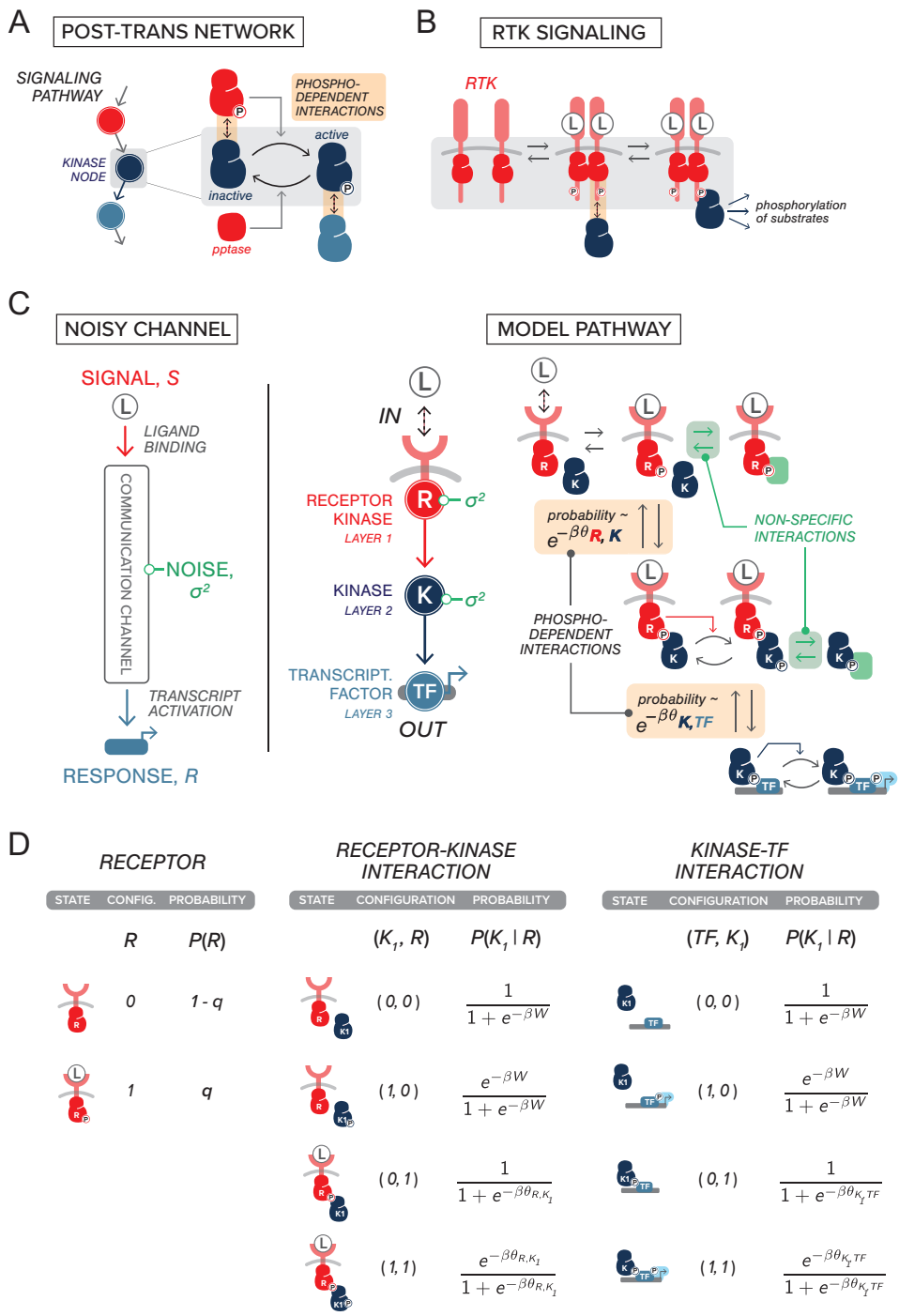

FIG. 1: Relating the biophysics of post-translational modification (PTMs) and protein-protein interactions (PPIs) to network-level information properties (*A*) A model signaling network that consists of three nodes is shown. The PPIs and PTMs that involve a single node (e.g. kinase node, colored dark blue) is illustrated on the right. In a PTM network, edges between nodes encodes both the phosphorylation dependent PPIs and the resulting change in enzymatic activities (e.g. active/inactive). (*B*) Naturally occurring pathways that can be conceptualized as the network shown in *A*. In receptor tyrosine kinase (RTK) signaling, binding of ligand (e.g. epidermal growth factor, EGF) to the extracellular domain of EGF receptor (EGFR, pink lobe) leads to the its dimerization and the phoshorylation of its kinase domain. This triggers signaling through phosphoryaltion-dependent interactions. (*C*) A signaling pathway can be viewed as a noisy communication channel (left). The input to this pathway is a ligand (L) that binds to the receptor kinase (R) which, through allosteric interactions, leads to receptor kinase phosphorylation. The phosphorylated receptor kinase then specifically binds its cognate kinase (K) which in term translocates into the nucleus to activate transcription. A pictorial summary of these events are shown on the right. The steady-state phosphorylation probability is annotated. Non-specific interactions (i.e. those highlighted in green) serve as noise in the network representation. All species are colored according to the nodes they correspond to (left). (*D*) Probability of PTM states in the thermodynamic model. Species are labeled with reference to *A*. As in the main text, binary variables $R, K_1, TF \in \{0, 1\}$ are used to indicate the PTM states of these species with value 1 indicating a phosphorylated state (transcribed state for TF) and 0 otherwise. Panels are organized according to the binding interactions involved and are indicted at the top. In C, D, $\beta = 1/(k_B T)$ is the inverse temperature and $\theta_{i,j}$ is the binding affinity of i to j (see main text for definition).

$$\begin{aligned}
\mathbb{P}(TF=1|R=0) &= \mathbb{P}(TF=1|K_1=1)\mathbb{P}(K_1=1|R=0)+\mathbb{P}(TF=1|K_1=0)\mathbb{P}(K_1=0|R=0) \\
&= \left(\frac{e^{-\beta\theta_{K_1,TF}}}{1+e^{-\beta\theta_{K_1,TF}}}\right)\left(\frac{e^{-\beta W}}{1+e^{-\beta W}}\right)+\left(\frac{e^{-\beta W}}{1+e^{-\beta W}}\right)\left(\frac{1}{1+e^{-\beta W}}\right) \quad (6) \\
&\approx 0 \quad (7)
\end{aligned}$$

$$\begin{aligned}
\mathbb{P}(TF=0|R=1) &= \mathbb{P}(TF=0|K_1=1)\mathbb{P}(K_1=1|R=1)+\mathbb{P}(TF=0|K_1=0)\mathbb{P}(K_1=0|R=1) \\
&= \left(\frac{1}{1+e^{-\beta\theta_{K_1,TF}}}\right)\left(\frac{e^{-\beta\theta_{R,K_1}}}{1+e^{-\beta\theta_{R,K_1}}}\right)+\left(\frac{1}{1+e^{-\beta W}}\right)\left(\frac{1}{1+e^{-\beta\theta_{R,K_1}}}\right) \quad (8) \\
&\approx \left(\frac{1}{1+e^{-\beta\theta_{K_1,TF}}}\right)\left(\frac{e^{-\beta\theta_{R,K_1}}}{1+e^{-\beta\theta_{R,K_1}}}\right)+\left(\frac{1}{1+e^{-\beta\theta_{R,K_1}}}\right) \quad (9)
\end{aligned}$$

$$\begin{aligned}
\mathbb{P}(TF=0|R=0) &= \mathbb{P}(TF=0|K_1=1)\mathbb{P}(K_1=1|R=0)+\mathbb{P}(TF=0|K_1=0)\mathbb{P}(K_1=0|R=0) \\
&= \left(\frac{1}{1+e^{-\beta\theta_{K_1,TF}}}\right)\left(\frac{e^{-\beta W}}{1+e^{-\beta W}}\right)+\left(\frac{1}{1+e^{-\beta W}}\right)\left(\frac{1}{1+e^{-\beta W}}\right) \quad (10) \\
&\approx 1, \quad (11)
\end{aligned}$$

where the approximation in the last line of these expressions indicates the limit where $\beta W \gtrsim 1$ so that $e^{-\beta W} \to 0$. In this limit, the output is simply

$$\mathbb{P}(TF=1) \approx \left(\frac{e^{-\beta\theta_{K_1,TF}}}{1+e^{-\beta\theta_{K_1,TF}}}\right)\left(\frac{e^{-\beta\theta_{R,K_1}}}{1+e^{-\beta\theta_{R,K_1}}}\right)q \quad (12)$$

$$\mathbb{P}(TF=0) \approx \left[\left(\frac{1}{1+e^{-\beta\theta_{K_1,TF}}}\right)\left(\frac{e^{-\beta\theta_{R,K_1}}}{1+e^{-\beta\theta_{R,K_1}}}\right)+\left(\frac{1}{1+e^{-\beta\theta_{R,K_1}}}\right)\right]q+(1-q) \quad (13)$$

With all these at hand, we can express Eq. (3) as a function of BAs $\theta_{i,j}$. In SI Section 2, we provide the analytic expression of mutual information Eq.(3) in terms of BAs. We have thus established an explicit functional relation between mutual information and PPIs.

Results

Weak binding affinities result in noise that limit the signal-to-noise ratio and information capacity

A key biophysical quantity that controls the network level properties is the binding affinity – or equivalently the binding energy – between proteins. When the binding affinity is large, proteins stay tightly bound to there targets. Small binding affinities allow proteins to quickly bind and unbind from targets but can give rise to transient binding. Here we examine how these considerations affect information transmission through a signaling network. To understand this tradeoff quantitatively, we consider a family of single-input, single-output signaling networks consisting of a receptor kinase R that phosphorylates a variable size intermediate layer consisting of n kinases K_i ($i = 1, \cdots, n$), and a transcriptional output TF (see Fig. 2A). The binary variables $R, K_i, TF \in \{0, 1\}$ encode the PTM-state of the protein with the value 1 indicating a phosphorylated state and 0 an unphosphorlyated state. We assume that the output transcription factor is active if and only if it is phosphorylated and the that the circuit is designed to activate the TF in the presence of a ligand at concentration L. We focus on information transmission at steady-state and neglect information encoded in the temporal dynamics.

A fundamental measure of noise in signaling networks is the signal-to-noise ratio (SNR) [21, 27]. To define the SNR, we make use of the probability that the output TF is active in the presence of the ligand $Q(L) \equiv P(TF = 1|L)$. In general, this input-output function is probabilistic. The stochasticity in Q stems from the probabilistic nature of protein-protein binding that is in-

herent in our thermodynamically-inspired models. And as in all thermodynamic models the more negative the binding affinities ($\theta_{k,j}$ where $k, j \in \{R, K_i, TF\}$), the smaller the effect of thermal fluctuations. In terms of $Q(L)$, the output obtained under a high input, $L = 1$, (e.g. large number of phosphorylated receptor kinase) defines the best "signal" one can obtain for a given realization of BAs. On the other hand, there can still output signals even when the input is absent (i.e. $L = 0$) due to thermal "noise" inherent in PPIs (i.e. contributions from W, see SI Section 1 for details). We therefore define the signal-to-noise ratio (SNR) of a given network/channel as the ratio between $Q(L = 1)$ and $Q(L = 0)$, averaged over realizations of BAs.

To understand the effect of the strength of PPI on the SNR, we consider drawing the binding affinities for the interactions in our network from a normal distribution $\theta_{i,j} \sim \mathcal{N}(\mu, \sigma)$ with mean binding affinity $\mu \equiv \langle\theta\rangle$ and variance σ^2, where $\langle\cdot\rangle$ refers to average over different realizations of BAs. The PPIs involving W, which sets the time scale of unbinding between unphosphorylated kinase to its substrate, is varied in the following analysis. This allows us to probe the effect of both the mean binding strength as well as the thermal noise resulting from W. Under these assumptions, we can analytically derive a formula for the SNR (see SI Section 1 for full derivations). When proteins bind tightly (i.e. large negative binding energies $\beta\mu \lesssim -1$), the SNR for the simplest signaling network $L \to R \to TF$ reduces to the following simple expression:

$$\text{SNR} \equiv \frac{\langle Q(L=1)\rangle}{\langle Q(L=0)\rangle} = e^{\beta W}\left[1 - e^{\beta\left(\mu + \frac{\sigma^2}{2}\right)}\right] \tag{14}$$

For networks with n-layers of kinase between input R and output TF, as depicted in Figure 2A, we plot the color map of their log-SNR at different level of specific and non-specific PPIs in Figure 2B. Regardless of the depth of network, n, strong specificity in PPIs, namely, tighter binding, always leads to higher SNR. This suggests that BA is an important source of "noise" that limits the resolution of output signal. To further explore this idea, we calculate the corresponding input-output relation (i.e. $\mathbb{P}(TF = 1)$ as a function of $\mathbb{P}(R = 1) \equiv q$) in Figure 2C both at tight- and weak-binding. As shown, networks with strong BAs always have a larger gain, implying a higher information capacity[21, 27]. Note that the activation of the receptor kinase, $R = 1$, depends on whether it is bound to ligand, and thus q is implicitly a function of ligand concentration L. In SI Section 2, we explicitly calculate input-output mutual information, $I(R; TF)$, for networks of varying depth at both binding scenarios. We also examined the effect of input distributions on mutual information (see SI Section 2 for details). As expected, the mutual information is zero when input is completely certain, *viz.* $q = 0, 1$. When binding is tight (i.e., $\beta\mu \lesssim -1$), the optimal input distribution $q^\star$ that maximizes the mutual information is $q^\star = 0.5$ – the input distribution with highest entropy (see SI Section 2 Figure S2). Surprisingly, for weak binding we find numerically and analytically that $q^\star \leq 0.5$ (see SI Section 2 Figure S5).

To summarize, we have found that the binding affinity of interactions can be directly related to the information transmission and the signal-to-noise ratio. We find that weak binding affinities give rise to noise stemming from thermal fluctuations and that this noise can always be reduced by increasing binding affinities and making binding more deterministic.

Noise due to non-specific PPIs mediates the "information-speed" trade-off

The previous observations are hard to reconcile with the observation that many PTM-recognition domains such as SH2 and SH3 have only moderately strong binding affinities [40, 41]. For this reason, we investigated tradeoffs that arise from having strong PPIs. One common requirement of eukaryotic signaling pathways is that they should be able to quickly respond to changes in the environmental conditions. This led us to ask how the strength of PPIs affects kinetics. Stronger binding affinities make it harder for proteins to disassociate, suggesting that there maybe a trade-off between reducing noise and responding quickly in the biophysics of PPIs.

To test these ideas, we 'translated' our thermodynamic model for the cascade studied in Figure 3A into a kinetic model (see SI Section 3 for details). Note that the thermodynamic model presented in Fig. 1D can be explicitly derived from the kinetic formulation. Here we invoke this duality to investigate both the signaling dynamics through kinetic formulation as well as the steady-state information capacity by the thermodynamic calculation presented in the previous section. Based on our thermodynamic framework, we first calculated input-output mutual information, *viz.* $I(R; TF)$ in Fig. 1C, with BAs drawn from distributions with different means $\langle\theta\rangle$. Due to the interplay between the kinetic and thermodynamic picture which we explicitly derived in SI Section 3, we mapped these mean BAs $\langle\theta\rangle$ to their corresponding kinetic rates. The key idea behind this mapping is that the steady state solution of the kinetic model with these rate constants is equivalent to its probabilistic counterpart in the thermodynamic model presented above. For example, the fraction of phosphorylated PK i at steady state is the same as $\mathbb{P}(K_i = 1)$ in the thermodynamic model. The BAs in the thermodynamic picture, $\theta_{i,j}$, is related to the Michaelis constant of kinase j phosphorylation reaction by i, K_m, via $\theta_{i,j} = k_B T \ln(K_m / X_i^{SS})$, where X_i^{SS} is the steady state concentration of phosphorylated kinase i.

We performed simulations to measure dynamic response of the signaling circuit to an abrupt perturbation where the input signal was suddenly removed (see Figure 3A). We characterized the response times by measuring the time τ it took the output to reach a new steady-

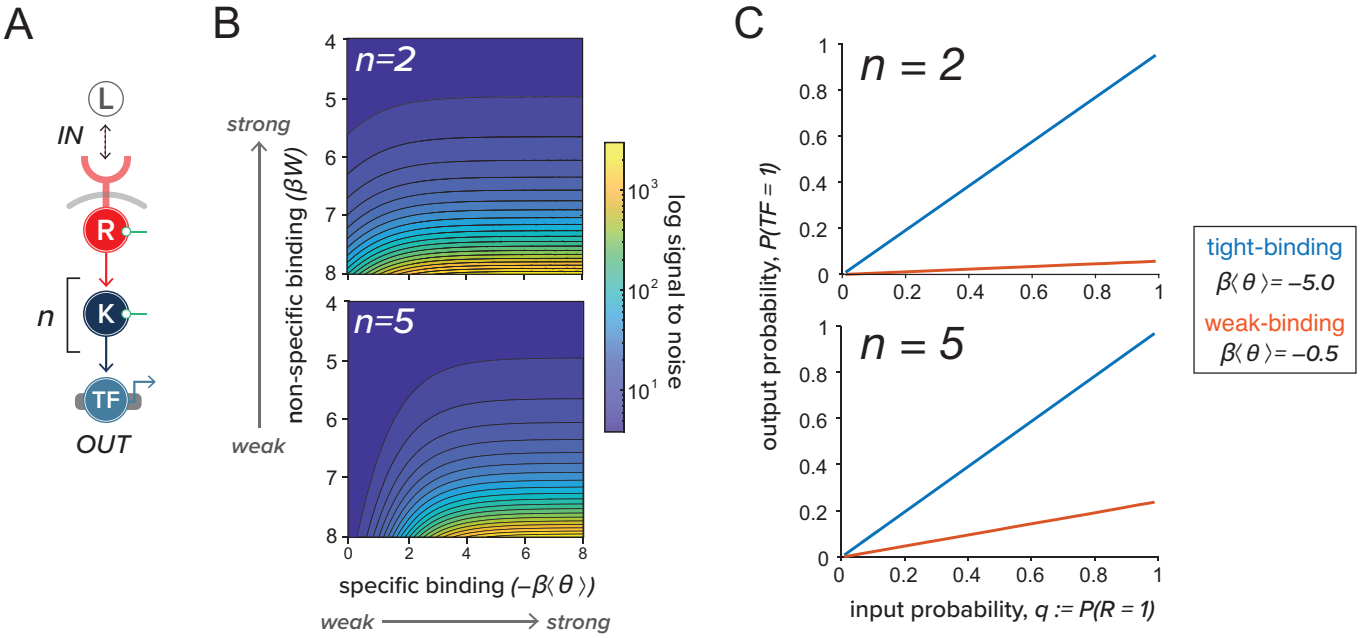

FIG. 2: Noise due to non-specific protein-protein interactions (PPIs) limits the quality of information transmission. (*A*) A simple linear network that mediates information of input (L) through a n-layer kinase cascade (K_i, $i = 1, 2, \cdots, n$), to an output transcription factor (TF) which is active when phosphorylated. As in Figure 1 *C*, green circles indicate noise. (*B*) Color map shows the numerically simulated log-signal-to-noise ratio (log-SNR), defined in Eq.(14), of the network shown in *A* at different level of specific and non-specific interactions. Binding affinities $\beta\theta_{i,j}$ is drawn from a normal distribution with mean $\langle\theta\rangle$ and variance $\sigma^2 = 0.01$ (see main text for simulation details). This quantity can also be obtained by solving Eq.(S9)(S10) (see SI Section I). In this panel, we show the result for $n = 2, 5$. (*C*) Input-output relation of the n-layer kinase cascade ($n = 2, 5$) for tight- and weak-binding with $\beta W = 2.0$.

state. We repeated this procedure for binding affinities drawn from distributions with different means $\langle\theta\rangle$. In Figure 3B, we plot both mutual information and the response speed, defined as the inverse of the response time τ^{-1}, against $\beta\langle\theta\rangle$. This plot shows that response speed and mutual information change in opposite ways as the binding affinity is decreased. Tight-binding (specific PPIs, more negative $\beta\langle\theta\rangle$) allows the network to transmit more information at the expense of a slower dynamical response (see Figure 3C).

This "speed-information" trade-off can be viewed as a biophysical manifestation of the gain-bandwidth trade-off [27]. Intuitively, tighter binding means that the binding off-rate is fairly small compared to the on-rate which is dictated by diffusion. This implies once proteins are bound through specific interactions, the lifetime of the bound complex is long.

Information loss in signaling 'can' be mitigated by cross-talks when inputs are correlated

Thus far, we have considered discreet, linearly connected pathways with a single input and output. However, native eukaryotic signaling networks are highly interconnected, with multiple inputs and outputs that cross-talk through PPIs. For this reason, we wanted to better understand how information transduction capacity in multi-input, multi-output (MIMO) networks depended on both the strength of PPIs and the structure of the input signal (i.e. the correlation between inputs). To do so, we studied two parallel pathways, each consisting of an input receptor kinase R and output TF (see Figure 4A). In this scheme, cross-talk refers to interactions where proteins in one pathway activate those in the other (i.e., dashed lines in Figure 4A). We varied the binding affinity and correlation between two inputs, R_1 and R_2 – defined as the connected correlation function (covariance) between the inputs $c \equiv \langle R_1 R_2\rangle - \langle R_1\rangle\langle R_2\rangle$ with $\langle\cdot\rangle$ indicating an average over the joint input distribution $\mathbb{P}(R_1, R_2)$ – and calculated the mutual information, $I(\{R_1, R_2\}; \{TF_1, TF_2\})$ between all the inputs outputs, (see Figure 4B for examples). We found that, regardless of the degree of correlation between inputs, pathway cross-talk is always detrimental to information transmission when noise from non-specific binding is small (i.e., tight-binding). However, for weak binding and positively correlated inputs, cross-talk can confer a slight benefit, actually increasing information transmission (see Figure 4C and SI Section 6 Figure S8 for full statistics under the distribution of correlations). This can be rationalized by noting that cross-talk allows cells to reduce noise by "averaging" the two input signals. This averaging is of course only possible if the signals

are correlated and contain redundant information.

Our results show that while inter-pathway cross-talk usually degrades information, it may actually provide a benefit when input signals are correlated by reducing noise due to weak PPIs. Simulations on larger pathways confirm this qualitative trend (though it becomes more difficult to define cross-talk for more complex circuits). Finally, we note that our presentation has been limited to the case where cross-talk involves cross-activation between pathways (i.e., pathway 1 activates the pathway 2 intermediate and vice versa). This is reasonable since we have restricted ourselves to considering networks consisting of kinases and some background phosphatase activity. If instead, we had allowed for cross-inhibition between pathways (i.e., pathway 1 inhibits the pathway 2 intermediate and vice versa), information capacity would be slightly increased for negatively correlated signals (results not shown) and diminished for correlated inputs.

Information maximization for complex multi-input, multi-output circuits

System-wide studies of phosphorylation-based signaling networks have revealed underlying PPI networks to be highly interconnected [42, 43]. Here we asked how interconnectivity within signaling network can affect its information capacity. To explore this question, we developed a new algorithm we dub "InfoMax", which identifies the binding affinities and protein concentrations that maximize information transmission for a given network topology. InfoMax, which stands for information maximization, begins with an initial random guess of binding affinities. It then utilizes the thermodynamic framework we developed to calculate the input-output mutual information using these affinities. Optimization is then performed on these affinities to maximize mutual information. Since the explicit functional dependence of mutual information on binding affinity is known (c.f. Eq.(3) and Figure 1D), this procedure can be done through a combination of analytic and iterative schemes. To make our approach more generalizable and agnostic to topology, we opted to use simulated annealing to conduct optimization (see Algorithm 1 in *Materials and Methods*). A Python implementation of InfoMax is freely available at the author's Github repository: `https://github.com/chinghao0703/InfomaxDesign`.

In order to test the utility of InfoMax, we constructed a library of one-input-one-output networks where we systematically varied network depth (n in Figure 5A) and two-input-two-output networks where we varied the width (n_w in Figure 5B). PPI affinities in these networks were optimized using the InfoMax algorithm, allowing us to identify the PPIs configuration with the highest maximum mutual information, subject to the constraints that BAs are bounded within a given range. In the one-input-one-output networks, we found that increasing network depth always decreased information transmission. This can be understood by noting that additional signaling layers increase non-specific PPI-mediated noise, without ever increasing the strength of the input signal. This observation is a manifestation of the data-processing inequality (DPI), which states that information is never gained by addition more layers when transmitting across noisy channels [21, 44](see Figure 5A and SI Section 4 Figure S7). For the optimal solution found, mutual information saturates around 1 bit. As expected, introducing perturbations to this optimal solution by using a sup-optimal binding affinity at an intermediate layer in the kinase cascade substantially diminish information transmission (see Figure 5A).

In contrast, for the two-input-two-output networks we found that increasing the width of the intermediate network can increase information transmission modestly for small widths. As seen in Figure 5B, these gains quickly saturate after the network reaches the 2-4-2 topology ($n_w = 4$). This suggests that modestly widening networks can alleviate bottlenecks in information transmission by reducing noise from weak PPIs. Interestingly, InfoMax also reveals that the optimal PPI-design strategy for more complicated networks can be quite different from the one-input-one-output case where it is always advantageous to have tight-binding between all proteins. In general, we find that the binding affinities that maximize information transmission for multi-input-multi-output networks take on a wide variety of values (see Figure 5B).

Discussions

The ability of cells to reliably transduce environmental signals is critical for their survival, growth, and proliferation. In this article, we developed a theoretical framework for relating the biophysics of post-translational modifications (PTMs) and protein-protein interactions (PPIs) to information processing in eukaryotic signaling networks. We showed that PPIs with moderate binding affinities necessarily result in thermal noise that limits information transmission within a signaling pathway. While noise can be reduced by increasing binding affinities, this comes with the expense of sluggish dynamic responses, highlighting a fundamental trade-off between information and signaling pathway response dynamics. Although extensive pathway cross-talk is relatively common in signaling networks, we found that it confers little or no advantage to a signaling networks information capacity.

Our results are consistent with other theoretical works that implicate noise as a major source of information transmission error in signaling [23, 24, 27–31]. What is novel about in this work is the ability to directly trace the origin of noise in eukaryotic signaling networks to the strength of PTM-mediated protein-protein interactions. Our results on cross talk also agree with those

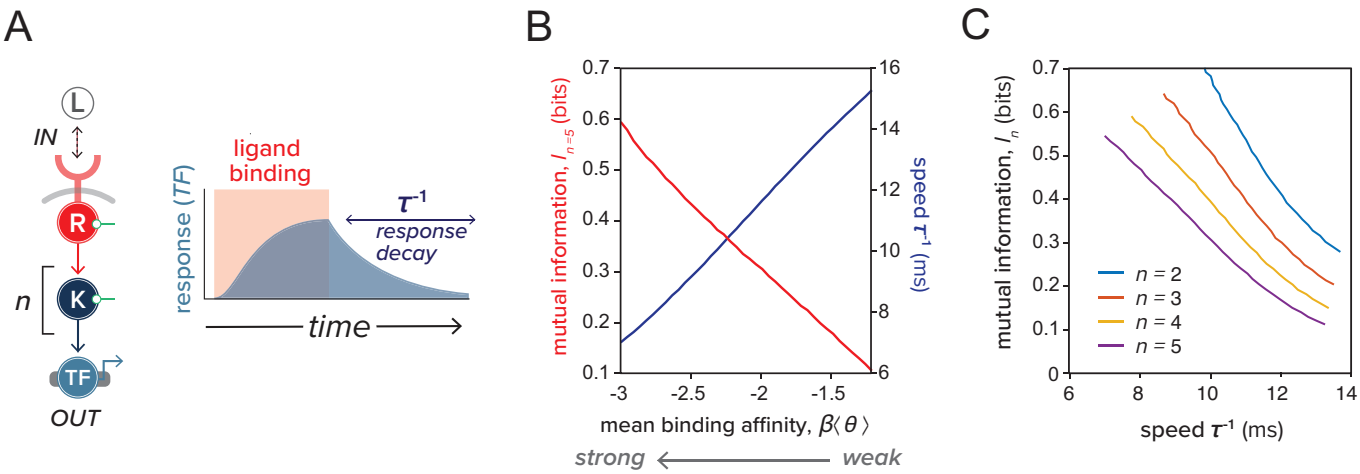

FIG. 3: Specificity in PPIs mediates the information-speed tradeoff. (*A*) A non-zero constant input (ligand binding of duration indicated in red) is administered to the network shown on the left. This signal turns on the output (TF) to its steady-state before switching off. The speed of response is defined as τ^{-1}, where τ is the time for output TF to reach a new steady-state after the input is turned off (indicated as response decay in blue). (*B*) Mutual information and response speed as a functions of mean binding affinity $\beta\langle\theta\rangle$ for the network shown in *A* with $n = 5$. (*C*) Mutual information versus response speed as $\beta\langle\theta\rangle$ is varied. Different colors correspond to networks of different depth n.

obtained in [45] that cross-talks degrades information for channels where input noise can be neglected and inputs are uncorrelated. In addition, our information-theoretic analysis reveals the disadvantages of a deep signaling network, particularly in the face of high non-specific binding (see SI Section 4 Figure S7). This is consistent with previous work on MAP kinase cascade [27], where the authors argued that maintaining fast response times requires a smaller number of steps with a higher gain per node in order to overcome molecular shot noise. Our simulations also show that information transmission quickly degrades for depths larger than three, which potentially explains the ubiquity of MAP-kinase cascades.

Our work has interesting implications for both natural and synthetic circuits. A recent study of the kinase-phosphatase interaction network in budding yeast identified 1844 interactions in budding yeast. Somewhat surprisingly, the binding affinities of many of the identified interactions fell into a narrow affinity window[43]. Binding affinity clustering was particularly pronounced for the kinase/phosphatase catalytic domains that mediate phosphorylation-dependent binding[46]. Our work on the information-speed tradeoff outlined above suggests such an optimized affinity range could be a common feature of networks that need to transmit signals reliably yet quickly in response to noisy environments [32].

Another intriguing observation from the yeast kinase-phosphatase interactions is the existence of extensive cross-talk between signaling pathways [43] that the authors describe a 'collaborative network of interactions' – a topology that suggests a distributed cellular decision-making strategy [42]. In this article, we show that cross-talk, while unlikely to increase information transmission, is also not particularly detrimental for signaling. Thus, widespread experimental observations of cross-talk in yeast signaling networks likely has an alternative origin. An intriguing hypothesis is that cross-talk arises because of evolutionary selection for signaling robustness[42]. Distributing information processing tasks to many interacting proteins may allow cells to maintaining reliable information transmission even when proteins are deleted or modified.

Our study is directly inspired by synthetic biology, where a long-standing engineering goal is to create cell-based therapies by reprogramming the way in which cells interact with their environment[47]. Creating synthetic kinase-based signaling circuitry that enables user-customized sense and respond function will necessarily involve information processing considerations, and may favor circuit designs that maximize mutual information between receptor-mediated input and transcriptional output. The potential design space for signaling circuits is vast-unlike genetic circuits, signaling circuits consists of freely diffusible molecular components and thus possess many more tunable parameters that have to be accounted for during design, including circuit topology, intracellular species concentrations, lifetimes, interaction affinities, and intrinsic catalytic rates [48]. Conclusions from our work suggest some general rules that could be used to constrain the search for productive circuit configurations. For example, focusing on engineering high interaction specificity for parts that mediate PTM-mediated PPIs could potentially mitigate noise, while using Infomax could be used to maximize the information capacity for a given circuit architecture.

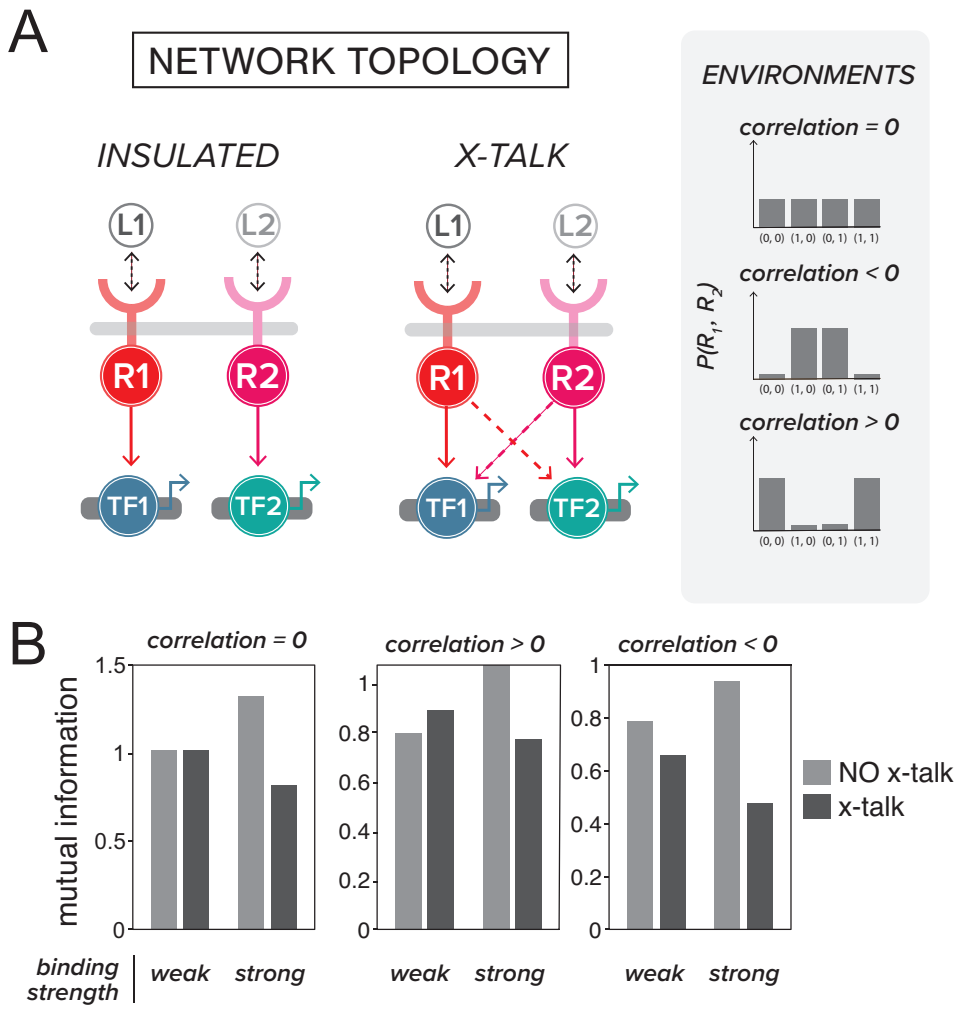

FIG. 4: Effect of cross-talks between pathways on information capacity. (*A*) Schematic of insulated and cross-talked networks are shown. Dashed connections represent cross-talks between pathways. Three environmental conditions that differ in terms of the correlation between the two signals they provide to the network (left) are illustrated: inputs with zero, negative, and positive correlation (see *Materials and Methods* for details). (*B*) Mutual information between inputs (R_1, R_2) and outputs (two TFs) is calculated for different input correlations, different mean binding affinities, and different cross-talk levels. Columns are arranged based on the sign of correlation, bars are grouped according to the strength of binding affinities for protein-protein interactions (weak/tight-binding corresponds to $\beta\langle\theta\rangle = -5/-1$), and colors indicate the presence or absence of cross-talks. For networks with (without) crosstalk, the binding affinity of the cross-talk interactions (i.e. dashed lines in the cross-talked networks shown in *A*) is set to $\beta\langle\eta\rangle = -5/0$ (see SI Section 5 for more details.)

Materials and Methods

Optimizing mutual information with simulated annealing

Let G be the network given and $\{\theta^{(0)}\}$ be the initial BAs (parameters) which we sample uniformly from $[-8, 2]$. For each time step t ($t = 0, \cdots, T = 10^6$), we either add to each elements of $\{\theta^{(t)}\}$ a fixed finite amount $\pm\delta\theta = \pm 10^{-4}$ or leave it un-perturbed, completely at random. The perturbed parameter $\{\theta^{(t)\star}\}$is accepted with probability $p = \min\{1, \exp\{\alpha[I(\{\theta^{(t)\star}\}) - I(\{\theta^{(t)}\})]\}\}$, where $\alpha = \log t$. If accepted, set $\{\theta^{(t+1)}\} \leftarrow \{\theta^{(t)\star}\}$; otherwise, set $\{\theta^{(t+1)}\} \leftarrow \{\theta^{(t)}\}$. This procedure continues until any elements in $\{\theta^{(t)}\}$ falls beyond $[-8, 2]$ or $t = T$, whichever comes earlier. Python codes for such implementation is available at the author's Github repository: https://github.com/chinghao0703/InfomaxDesign. In Figure 5, we perform 100 simulated annealing routines with un-correlated inputs and report the realization that gives maximum mutual information denoted as $I_{\max}$. The pseudo-code of the InfoMax procedure is given in Algorithm 1.

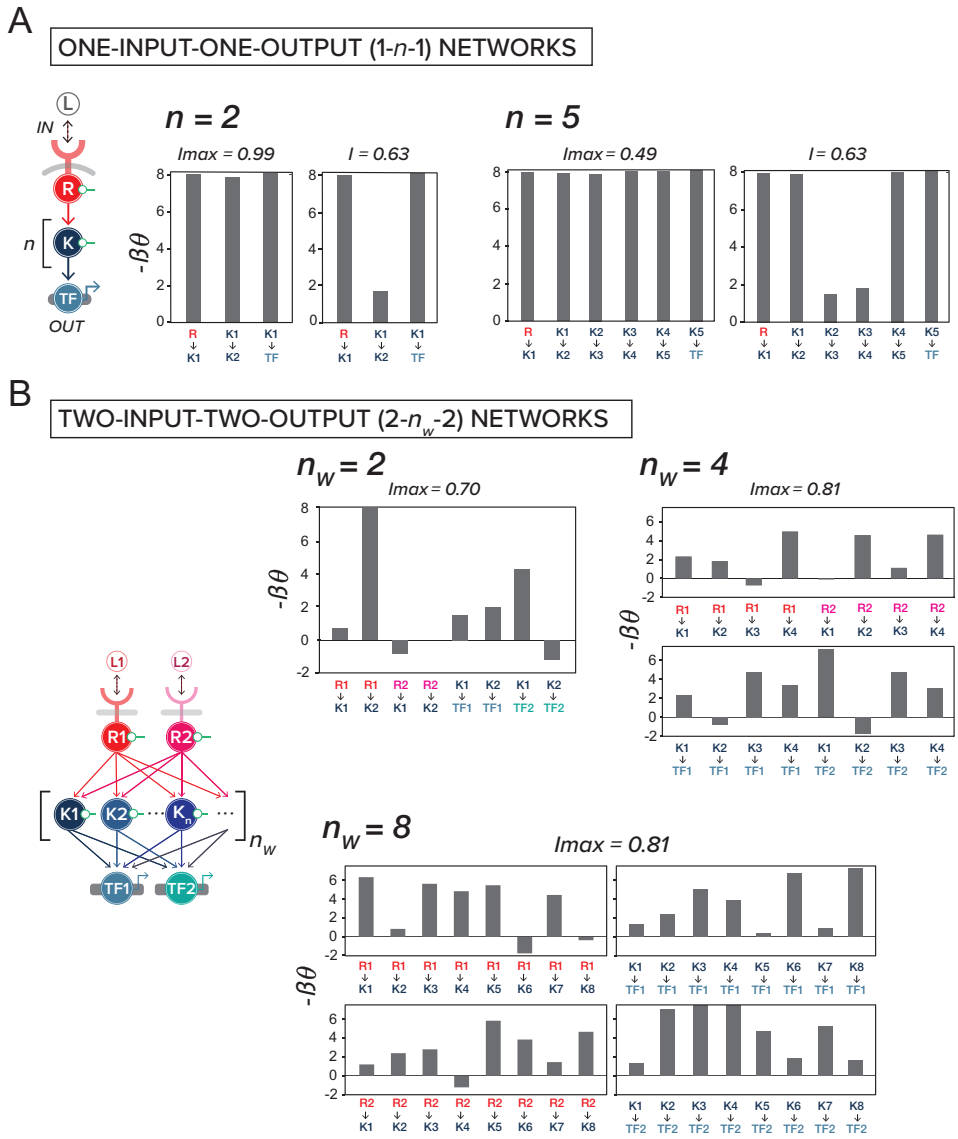

FIG. 5: InfoMax design finds the PPIs that maximize information transmission. (*A*) InfoMax is applied to networks with one input, n layers of kinase, and one output. For $n = 2$ and 5, the bar graph on the left shows the binding affinities that give the maximum mutual information (indicated as I_{max} on top), as opposed to a non-optimal solution with mutual information $I < I_{max}$ shown on the right, all measured in bits. Bars indicate binding affinities between proteins. For example, bar labeled as $R \to K_1$ is the binding affinity of receptor kinase R to kinase K_1. (*B*) InfoMax applied to 2-n_w-2 networks. This nomenclature refers to all-to-all connected networks with two nodes in the input layer, n_w in the hidden, and two at the output layer. Binding affinities of all networks are optimized to achieve maximum mutual information using simulated annealing (see *Materials and Methods*). Bar charts show the optimized binding affinities with I_{max} indicated on top. In this panel, networks are subject to inputs with zero correlation. In all panels, $-\beta\theta$ is constrained to be within $[-2, 8]$.

Calculating correlation between inputs

Here we consider two-input-two-output all-to-all connected network. As before, let x_1, x_2 be the inputs while x_3, x_4 be the outputs. Let's also define the connected correlation functions for the input as:

$$c = \mathbb{E}[x_1 x_2] - \mathbb{E}[x_1]\mathbb{E}[x_2], \tag{15}$$

where $\mathbb{E}[\cdots]$ is taken with respect to the joint distribution $\mathbb{P}(x_1, x_2)$ given by

(x_1, x_2)	$\mathbb{P}(x_1, x_2)$
$(0,0)$	q_1
$(1,0)$	q_2
$(0,1)$	q_3
$(1,1)$	$1 - q_1 - q_2 - q_3$

Algorithm 1 InfoMax: maximizing mutual information with respect to BAs

Require: Binding affinity (BA) range $S \leftarrow [-8.0, 2.0]$, perturbation $\delta\theta \leftarrow 10^{-4}$, termination threshold $T \leftarrow 10^6$, and function $I(\cdot)$ that computes the input-output mutual information of a given signaling network G with BA $\{\theta\}$.

procedure INFOMAX(G)
  $t \leftarrow 0$
  Draw BA uniformly from S: $\{\theta^{(t)}\} \leftarrow U(S)$
  while $t \leq T$ AND $\{\theta^{(t)}\} \in S$ **do**
    Draw a uniformly from $\{1, 0, -1\}$ and update BA $\{\theta^{(t)\star}\} \leftarrow \{\theta^{(t)} + a \cdot \delta\theta\}$
    Compute $p \leftarrow \min\{1, \exp\{\alpha[I(\{\theta^{(t)^\star}\}) - I(\{\theta^{(t)}\})]\}\}$, where $\alpha = \log t$
    Draw b uniformly from $[0, 1]$
    if $b \leq p$ **then**
      Accept $\{\theta^{(t+1)}\} \leftarrow \{\theta^{(t)\star}\}$
    else
      $\{\theta^{(t+1)}\} \leftarrow \{\theta^{(t)}\}$
    $t + 1 \leftarrow t$
  return $\{\theta^{(t)}\}$

From this, the connected correlation function reads

$$\begin{aligned} c &= q_1(1 - q_1 - q_2 - q_3) - q_2 q_3 \\ &= \mathbb{P}(0,0)\mathbb{P}(1,1) - \mathbb{P}(1,0)\mathbb{P}(0,1) \end{aligned} \tag{16}$$

Acknowledgments

We thank Amir Bitran and Henry Mattingly for helpful discussions. This work was also supported by NIH NIGMS grant 1R35GM119461, and by Simons Investigator in the Mathematical Modeling of Living Systems (MMLS) awards to PM. Part of the computations were carried out on the Boston University Shared Computing Cluster (SCC).

[1] A.-L. Barabasi and Z. N. Oltvai, Nature reviews genetics **5**, 101 (2004).
[2] A. Blais and B. D. Dynlacht, Genes & development **19**, 1499 (2005).
[3] B. D. MacArthur, A. Ma'ayan, and I. R. Lemischka, Nature Reviews Molecular Cell Biology **10**, 672 (2009).
[4] G. Martello and A. Smith, Annual review of cell and developmental biology **30**, 647 (2014).
[5] B. T. Seet, I. Dikic, M.-M. Zhou, and T. Pawson, Nature reviews Molecular cell biology **7**, 473 (2006).
[6] J. D. Scott and T. Pawson, Science **326**, 1220 (2009).
[7] W. Lim, B. Mayer, and T. Pawson, *Cell signaling: principles and mechanisms* (Taylor & Francis, 2014).
[8] S. Angers and R. T. Moon, Nature reviews Molecular cell biology **10**, 468 (2009).
[9] B. T. MacDonald, K. Tamai, and X. He, Developmental cell **17**, 9 (2009).
[10] J. Massagué, Nature reviews Molecular cell biology **13**, 616 (2012).
[11] J. N. Anastas and R. T. Moon, Nature Reviews Cancer **13**, 11 (2013).
[12] A. Moustakas and P. Heldin, Biochimica et Biophysica Acta (BBA)-General Subjects **1840**, 2621 (2014).
[13] Y. L. Deribe, T. Pawson, and I. Dikic, Nature structural & molecular biology **17**, 666 (2010).
[14] G. K. Ackers, A. D. Johnson, and M. A. Shea, Proceedings of the National Academy of Sciences **79**, 1129 (1982).
[15] T. L. Hill, *Cooperativity theory in biochemistry: steady-state and equilibrium systems* (Springer Science & Business Media, 2013).
[16] F. M. Weinert, R. C. Brewster, M. Rydenfelt, R. Phillips, and W. K. Kegel, Physical review letters **113**, 258101 (2014).
[17] L. Bintu, N. E. Buchler, H. G. Garcia, U. Gerland, T. Hwa, J. Kondev, and R. Phillips, Current opinion in genetics & development **15**, 116 (2005).
[18] J. B. Kinney, A. Murugan, C. G. Callan, and E. C. Cox, Proceedings of the National Academy of Sciences (2010).
[19] H. G. Garcia, A. Sanchez, T. Kuhlman, J. Kondev, and R. Phillips, Trends in cell biology **20**, 723 (2010).
[20] C. E. Shannon, ACM SIGMOBILE mobile computing and communications review **5**, 3 (2001).
[21] T. M. Cover and J. A. Thomas, *Elements of information theory* (John Wiley & Sons, 2012).
[22] H. A. Johnson, Science **168**, 1545 (1970).
[23] G. Tkačik, A. M. Walczak, and W. Bialek, Physical Review E **80**, 031920 (2009).
[24] A. M. Walczak, G. Tkačik, and W. Bialek, Physical Review E **81**, 041905 (2010).
[25] G. Tkačik and A. M. Walczak, Journal of Physics: Condensed Matter **23**, 153102 (2011).
[26] A. A. Granados, J. M. Pietsch, S. A. Cepeda-Humerez, I. L. Farquhar, G. Tkačik, and P. S. Swain, Proceedings of the National Academy of Sciences **115**, 6088 (2018).
[27] P. B. Detwiler, S. Ramanathan, A. Sengupta, and B. I. Shraiman, Biophysical Journal **79**, 2801 (2000).
[28] P. Mehta, S. Goyal, T. Long, B. L. Bassler, and N. S. Wingreen, Molecular systems biology **5**, 325 (2009).
[29] F. Tostevin and P. R. Ten Wolde, Physical review letters **102**, 218101 (2009).
[30] R. Cheong, A. Rhee, C. J. Wang, I. Nemenman, and A. Levchenko, science , 1204553 (2011).
[31] M. D. Brennan, R. Cheong, and A. Levchenko, Science **338**, 334 (2012).
[32] J. E. Ladbury and S. T. Arold, Trends in biochemical sciences **37**, 173 (2012).
[33] M. Voliotis, R. M. Perrett, C. McWilliams, C. A. McArdle, and C. G. Bowsher, Proceedings of the National Academy of Sciences **111**, E326 (2014).
[34] I. Chung, R. Akita, R. Vandlen, D. Toomre, J. Schlessinger, and I. Mellman, Nature **464**, 783 (2010).
[35] J. Schlessinger, Cell **103**, 211 (2000).

[36] S. M. Hill, The Anatomical Record: An Official Publication of the American Association of Anatomists **253**, 42 (1998).
[37] M. A. Schwartz and M. H. Ginsberg, Nature cell biology **4**, E65 (2002).
[38] T. Hunter, Molecular cell **28**, 730 (2007).
[39] S. Kontogeorgaki, R. J. Sánchez-García, R. M. Ewing, K. C. Zygalakis, and B. D. MacArthur, Scientific Reports **7**, 532 (2017).
[40] J. E. Ladbury and S. Arold, Chemistry & biology **7**, R3 (2000).
[41] J. E. Ladbury and S. T. Arold, in *Methods in enzymology*, Vol. 488 (Elsevier, 2011) pp. 147–183.
[42] E. D. Levy, C. R. Landry, and S. W. Michnick, Science **328**, 983 (2010).
[43] A. Breitkreutz, H. Choi, J. R. Sharom, L. Boucher, V. Neduva, B. Larsen, Z.-Y. Lin, B.-J. Breitkreutz, C. Stark, G. Liu, *et al.*, Science **328**, 1043 (2010).
[44] J. B. Kinney and G. S. Atwal, Proceedings of the National Academy of Sciences , 201309933 (2014).
[45] A. Tareen, N. S. Wingreen, and R. Mukhopadhyay, Physical Review E **97**, 020402 (2018).
[46] J. Mok, P. M. Kim, H. Y. Lam, S. Piccirillo, X. Zhou, G. R. Jeschke, D. L. Sheridan, S. A. Parker, V. Desai, M. Jwa, *et al.*, Sci. Signal. **3**, ra12 (2010).
[47] M. A. Fischbach, J. A. Bluestone, and W. A. Lim, Science translational medicine **5**, 179ps7 (2013).
[48] C. J. Bashor and J. J. Collins, Annual review of biophysics **47**, 399 (2018).

Supplemental Information for "The strength of protein-protein interactions controls the information capacity and dynamical response of signaling network"

I. SIGNAL-TO-NOISE RATIO (SNR) OF SIGNALING CIRCUIT

Consider a biochemical pathway that involves the relay of a (possibly continuous) signal c to the intracellular kinase X which intern activates an output Y. We assume that these proteins are catalytically active only when they undergo a post-translational modification (PTM). We represent the PTM-state of the proteins as binary random variables that take value 1 when catalytically active and 0 otherwise. Pictorially, this pathway can be summarized as the following channel: $c \to X \to Y$. Note that in the appendix, we first set $\beta = 1$ to simplify notation in the calculation and put it back in at the end using dimensional analysis.

To calculate the signal-to-noise ratio (SNR),we need the probability of the output given the input, $Q(c) \equiv \mathbb{P}(Y = 1|c)$. Specifically,

$$\begin{aligned}\mathbb{P}(Y=1|c) &= \sum_{X\in\{0,1\}} \mathbb{P}(Y=1|X)\mathbb{P}(X|c) &\text{(S1)}\\ &= \underbrace{\left(\frac{e^{-\theta_x}}{1+e^{-\theta_x}}\right)}_{\eta_b} q(c) + \underbrace{\left(\frac{e^{-W}}{1+e^{-W}}\right)}_{\eta_W}(1-q(c)), &\text{(S2)}\end{aligned}$$

where $q(c) \equiv \mathbb{P}(X = 1|c)$ is the probability that kinase X is phosphorylated in the presence of a ligand at concentration c. The functional form of $q(c)$ is not relevant so we simply assume that it is a monotonically increasing function of c and attains 1(0) when $c = 1(0)$. Note that when $q(c) = 1$ (i.e. full input signal), $Q(c)$ is purely dictated by bindings of phosphorylated kinase to its substrate (i.e. θ_x, see Fig. 1D), whereas when $q(c) = 0$ (i.e. no input signal), the contribution is solely from those involving unphosphorylated kinase (i.e. W, see Fig. 1D). Therefore, we define the signal-to-noise ratio (SNR) formally as:

$$\text{SNR} \equiv \frac{\langle Q(c=1)\rangle}{\langle Q(c=0)\rangle} = \frac{\langle \eta_b\rangle}{\langle \eta_W\rangle}, \tag{S3}$$

where $\langle\cdot\rangle$ denotes the average with respect the distribution of BAs. To simplify, we assume that W is a constant that sets the time scale of non-specific bindings and that the specific BA $\theta_x \sim \mathcal{N}(\mu, \sigma^2)$ is drawn from a Gaussian distribution with mean $\mu \ll -1$ (i.e. tight-binding) and variance σ^2. Since θ_x is normally distributed, e^{θ_x} follows log-normal distribution. In this tight-binding approximation,

$$\eta_b = \frac{1}{1+e^{\theta_x}} \approx 1 - e^{\theta_x} \equiv 1 - Z, \tag{S4}$$

where $Z \equiv e^{\theta_x} \sim \log\mathcal{N}(\mu, \sigma^2)$. From this, one can calculate its first two moments:

$$\begin{aligned}\langle\eta_b\rangle = \mathbb{E}[1-Z] &= \int_0^\infty dz\,(1-z)\frac{1}{z\sigma\sqrt{2\pi}}\exp\left[-\frac{(\log z-\mu)^2}{2\sigma^2}\right]\\ &= 1-\exp\left(\mu+\frac{\sigma^2}{2}\right) &\text{(S5)}\end{aligned}$$

and

$$\begin{aligned}\langle\eta_b^2\rangle = \mathbb{E}[(1-Z)^2] &= \int_0^\infty dz\,\frac{1-2z+z^2}{\sigma\sqrt{2\pi}}\exp\left[-\frac{(\log z-\mu)^2}{2\sigma^2}\right]\\ &= 1-2\exp\left(\mu+\frac{\sigma^2}{2}\right)+\exp[2\left(\mu+\sigma^2\right)]. &\text{(S6)}\end{aligned}$$

From this one can also derive its variance

$$\begin{aligned}\text{Var}(\eta_b) &= \langle\eta_b^2\rangle - \langle\eta_b\rangle^2\\ &= e^{2(\mu+\sigma^2)}(1-e^{-\sigma^2}). &\text{(S7)}\end{aligned}$$

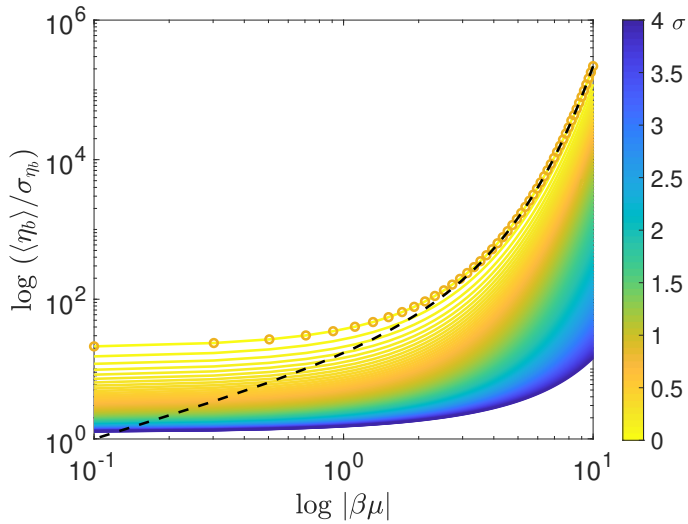

FIG. S1: Inverse coefficient of variation (CV^{-1}) of $Q(c)$ assuming negligible η_W is shown for different value of $\mu \equiv \langle\theta_{i,j}\rangle$ and $\sigma^2 \equiv \mathrm{Var}(\theta_{i,j})$, where $\theta_{i,j}$ is the binding affinity of K_i to K_j and the average $\langle\cdot\rangle$ is taken over different parameter realizations $\theta_{i,j}$. Note that $\mu < 0$ by definition. Different colors indicate different σ whose values are encoded in the color bar. Colored curves are theoretical predictions Eq.(S9)(S10), open circles are calculated via sampling network parameters as described in *Materials and Methods*. Black dashed curve is the analytical result in the tight-binding limit whose expression is given by Eq.(S5)(S7) with σ corresponding to that of the yellow open circles ($\sigma = 0.1$). Note the logarithmic scale.

These quantities can be used to analyze the effect of heterogeneity in θ_X on $Q(c)$ which we summarized in Fig. S1. Finally, after putting the energy unit $\beta^{-1} = k_B T$ back in and noting that $e^{-\beta W} \ll 1$ so that $\eta_W \approx e^{-\beta W}$, the signal-to-noise ratio is simply

$$\mathrm{SNR} = \frac{\langle\eta_b\rangle}{\langle\eta_W\rangle} = e^{\beta W}\left[1 - e^{\beta\left(\mu+\frac{\sigma^2}{2}\right)}\right]. \tag{S8}$$

Note that one can still calculate the SNR without assuming tight-binding, except in this case there's no closed form solution. Follow the same procedure while retaining $\eta_b = e^{-\theta_x}/(1+e^{-\theta_x})$ and performing change-of-variable, one ended up with the following integrals:

$$\langle\eta_b\rangle = \frac{1}{\sigma\sqrt{2\pi}}\int_0^\infty d\lambda\,\frac{1}{1+\lambda}\exp\left[-\frac{(\log\lambda+\mu)^2}{2\sigma^2}\right] \tag{S9}$$

$$\langle\eta_b^2\rangle = \frac{1}{\sigma\sqrt{2\pi}}\int_0^\infty d\lambda\,\frac{\lambda}{(1+\lambda)^2}\exp\left[-\frac{(\log\lambda+\mu)^2}{2\sigma^2}\right]. \tag{S10}$$

One can further apply Laplace method by assuming $M \equiv 1/(2\sigma^2) \gg 0$ (i.e. $\sigma^2 \to 0$, zero temperature limit) to get

$$\langle\eta_b\rangle_{M\gg1} \approx \frac{e^{-\mu}}{1+e^{-\mu}} \tag{S11}$$

$$\langle\eta_b^2\rangle_{M\gg1} \approx \left(\frac{e^{-\mu}}{1+e^{-\mu}}\right)^2, \tag{S12}$$

implying that $\mathrm{Var}(\eta_b) \sim 0$ and $\langle\eta_b\rangle$ is simply the mean-field value. In this case, the SNR with the energy unit in place reads

$$\mathrm{SNR}_{M\gg1} \approx \frac{e^{-\beta(\mu-W)}}{1+e^{-\beta\mu}}. \tag{S13}$$

II. DERIVING THE INFORMATION CAPACITY

In this section, we derive the mutual information transduced across a linear signaling network based on phosphorylation cascade. Note that linearity here refers to the network topology not that of the transfer function relating phosphorylation reaction downstream. Concretely, we consider a n-phosphorylation kinase cascade represented by a Bayesian network of the form: $x_{\text{in}} \to x_1 \to \cdots \to x_n \to x_{\text{out}}$, where $x_i, x_{\text{in}}, x_{\text{out}} \in \{0, 1\}$ as defined in the main text. For brevity, we denote $x_{\text{out}} \equiv x_{n+1}$ and $x_{\text{in}} \equiv x_0$. Due to the Markovian nature of this network, the joint distribution of kinase states can be factorized as

$$\mathbb{P}(\mathbf{x}) := \mathbb{P}(x_0, \cdots, x_{n+1}) = \left(\prod_{i=0}^{n} \mathbb{P}(x_{i+1}|x_i)\right) \mathbb{P}(x_0), \tag{S14}$$

where the conditionals are given by

$$\mathbb{P}(x_{i+1} = 1|x_i) = \frac{x_i e^{-\theta_{i,i+1}}}{1 + x_i e^{-\theta_{i,i+1}}}. \tag{S15}$$

Note that we denote the *relative binding affinity* of i to $i+1$ as $\theta_{i,i+1}$ (c.f. Eq. (S35)) . From now on, every energetic parameters are measured in units of $k_B T$. To simplify notation, we represent the conditional probability by a transfer matrix defined as:

$$\mathcal{M}_{i+1,i} = \begin{pmatrix} \mathbb{P}(x_{i+1} = 1|x_i = 1) & \mathbb{P}(x_{i+1} = 1|x_i = 0) \\ \mathbb{P}(x_{i+1} = 0|x_i = 1) & \mathbb{P}(x_{i+1} = 0|x_i = 0) \end{pmatrix} \tag{S16}$$

To calculate the mutual information between x_1 and x_n,

$$I(x_0; x_{n+1}) = \sum_{x_0} \sum_{x_{n+1}} \mathbb{P}(x_0)\mathbb{P}(x_{n+1}|x_0) \log_2 \left[\frac{\mathbb{P}(x_{n+1}|x_0)}{\mathbb{P}(x_{n+1})}\right], \tag{S17}$$

we need to get $\mathbb{P}(x_{n+1}|x_0)$ first. Using the matrix notation, we have

$$\begin{aligned} \mathbb{P}(x_{n+1}|x_0) &= \sum_{x_1} \cdots \sum_{x_n} \prod_{i=0}^{n} \mathbb{P}(x_{i+1}|x_i) \\ &= \sum_{x_n} \mathbb{P}(x_{n+1}|x_n)\mathbb{P}(x_n|x_{n-1}) \cdots \sum_{x_2} \mathbb{P}(x_3|x_2) \sum_{x_1} \mathbb{P}(x_2|x_1)\mathbb{P}(x_1|x_0) \\ &= \prod_{i=0}^{n} \mathcal{M}_{i+1,i} \equiv \mathbf{P}_{n+1,0}, \end{aligned} \tag{S18}$$

from which we can derive the marginal probability of x_{n+1}:

$$\mathbf{p}_{n+1} = \sum_{x_0} \mathbb{P}(x_{n+1}|x_0)\mathbb{P}(x_0) = \left(\prod_{i=0}^{n} \mathcal{M}_{i+1,i}\right) \mathbf{p}_0 = \mathbf{P}_{n+1,0}\mathbf{p}_0, \tag{S19}$$

where $\mathbf{p}_{n+1} \equiv (\mathbb{P}(x_{n+1} = 1), \mathbb{P}(x_{n+1} = 0))^T$, and similarly for $\mathbf{p}_0$. With this defined and for a given set of $\theta_{i,j}$, we can calculate mutual information Eq. (S17) by a series of matrix multiplications.

Now consider several realizations of such signaling circuits with $\theta_{i,j}$ drawn from some distribution, say, Gaussian with mean μ and variance σ^2. In the tight-binding limit, $\mu \ll -1$ and the transfer matrix Eq. (S16) approximates the identity matrix $\mathbf{I}$ due to Eq. (S15). From this, one can easily show that mutual information averaged over different realizations is given by:

$$\langle I(x_0; x_{n+1})\rangle = -q \log_2 q - (1-q) \log_2 (1-q) \equiv H_2(q), \tag{S20}$$

where $q \equiv \mathbb{P}(x_1 = 1)$, and $H_2(q)$ is the entropy function of Bernoulli process with probability q of one of the two values. Note that this calculation does not depend on the depth of the network (i.e. n), which implies as long as this approximation holds (i.e. tight-binding), mutual information is always peaked when input is least certain (i.e. $q = 0.5$), see Figure S2.

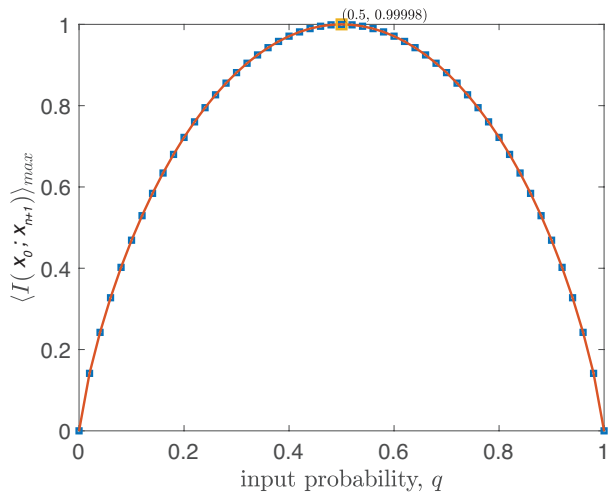

FIG. S2: At tight-binding mutual information for linear network reduces to binary entropy function. Linear network of depth $n = 8$ is used. Red squares are obtained by averaging the result over 100 different realizations of binding affinities using the methods detailed in this appendix. Dashed black curve is plotted using Eq. (S20). Parameters used are: $\beta\mu = -5, \sigma = 0.1$. The value and location of maximum mutual information obtained by averaging is indicated as $(q, I_{\max}) = (0.5, 0.99998)$.

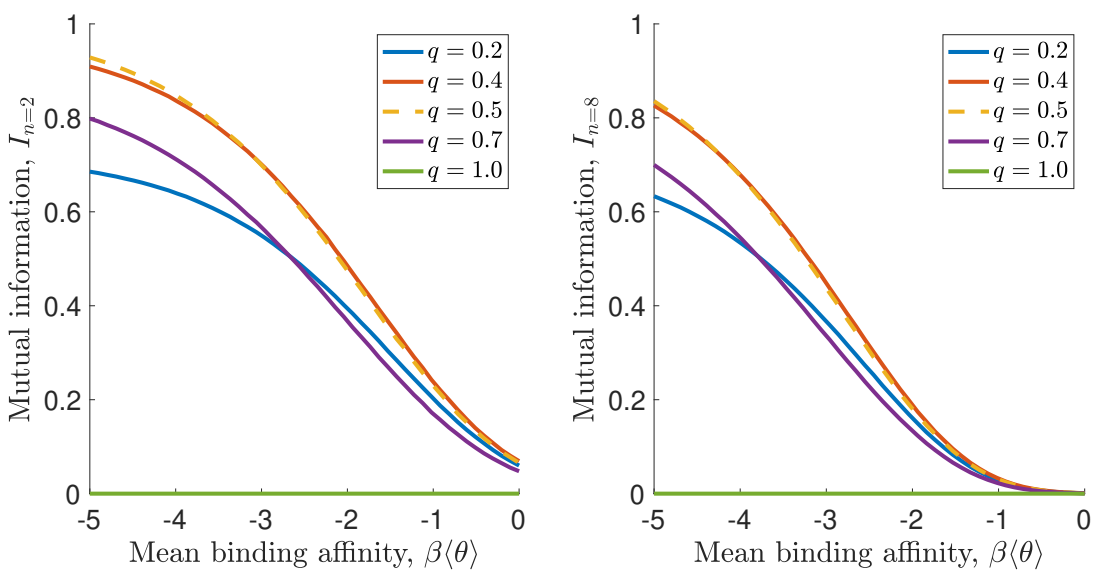

FIG. S3: Mutual information as a function of mean binding affinity $\mu \equiv \langle\theta\rangle$ at different inputs. Note that maximum mutual information may not occur at $q = 0.5$ (dashed orange curve) away from tight-binding (i.e. less negative $\beta\langle\theta\rangle$). In all panels, $\sigma = 0.1$.

III. OPTIMAL INPUT TO REACH MAXIMUM MUTUAL INFORMATION

In this section, we derive the optimal input that gives maximum mutual information. To simplify notation, let's define

$$b_n = \prod_{i=1}^{n-1} f_i = \prod_{i=1} \left(\frac{e^{-\theta_{i,i+1}}}{1 + e^{-\theta_{i,i+1}}} \right). \tag{S21}$$

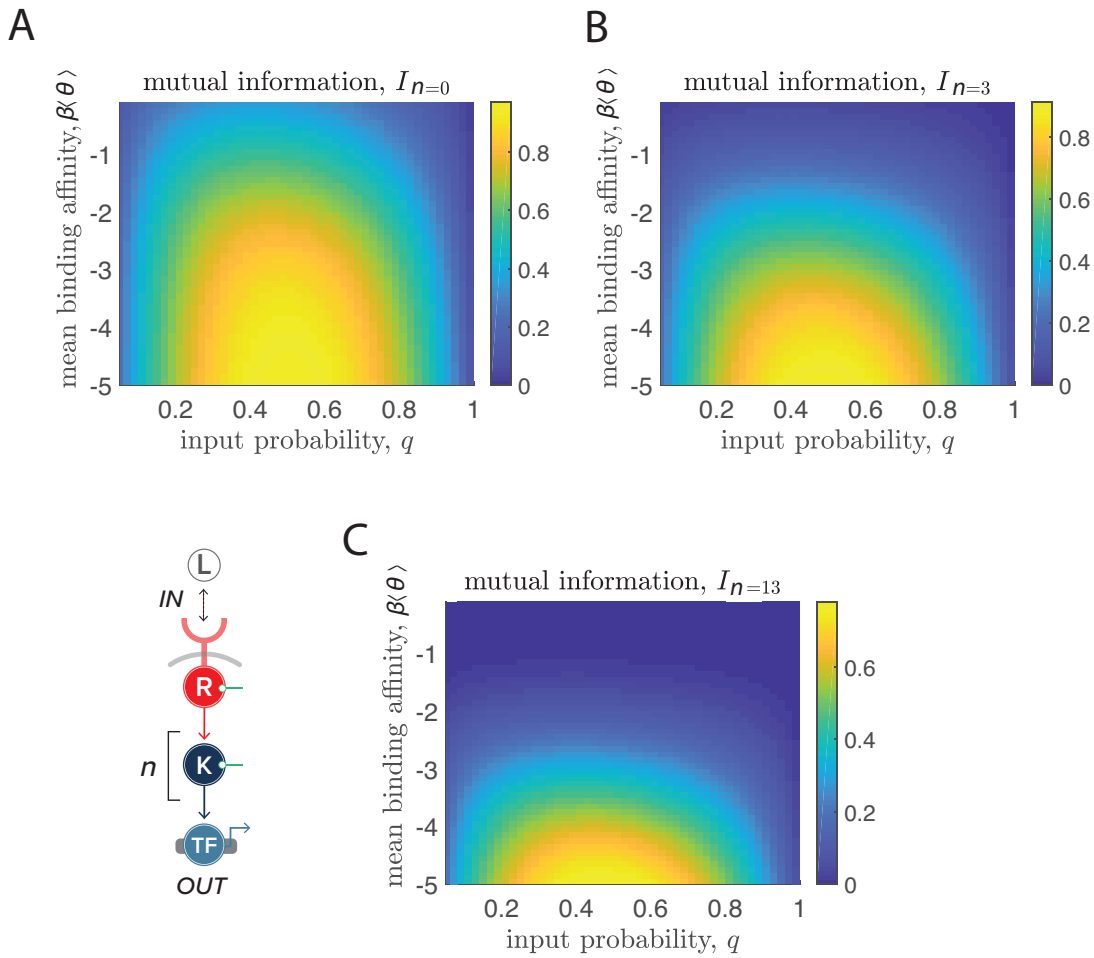

FIG. S4: **Tighter binding always increases information transmission for cascades of depths** $n = 0, 3, 13$. Here $\mathrm{Var}(\theta_{i,j}) \equiv \sigma^2 = 0.01$. All panels are generated by averaging 100 realizations of binding affinities using the scheme detailed in SI Section 2.

From this, one can re-write Eq. (S18) and Eq. (S19) as

$$\mathbf{P}_{n,1} = \begin{pmatrix} b_n & 0 \\ 1 - b_n & 1 \end{pmatrix}, \qquad \mathbf{p}_n = \begin{pmatrix} q\, b_n \\ 1 - q\, b_n \end{pmatrix}. \tag{S22}$$

Plugging this back in to the definition of mutual information,Eq. (S17), one gets,

$$I(x_1; x_n) = -q b_n \log q + q(1 - b_n)\log(1 - b_n) - (1 - q\, b_n)\log(1 - q\, b_n) \tag{S23}$$

After taking the derivate of Eq. (S23) with respect to q and setting it to zero, one finds that the optimal input $q^\star$ that gives the maximum mutual information $I(x_1; x_n)$ is the solution to the following transcendental equation:

$$b_n \log\left(\frac{q^\star}{1 - q^\star b_n}\right) = (1 - b_n)\log(1 - b_n) \tag{S24}$$

IV. RELATING THERMODYNAMICS TO A KINETIC MODEL OF PHOSPHORYLATION CASCADE

Here we derive the Eq.(3) in the main text (re-written as Eq. (S34) here) from chemical kinetics. Following Fig. S6, let X_i be the concentration of kinase i in its active (i.e. phosphorylated) form and $\tilde{X}_i$ be that of its inactive (i.e. unphosphorylated) form. For each step i of cascade except for $i = 1$, the rate of phosphorylation is dependent on the

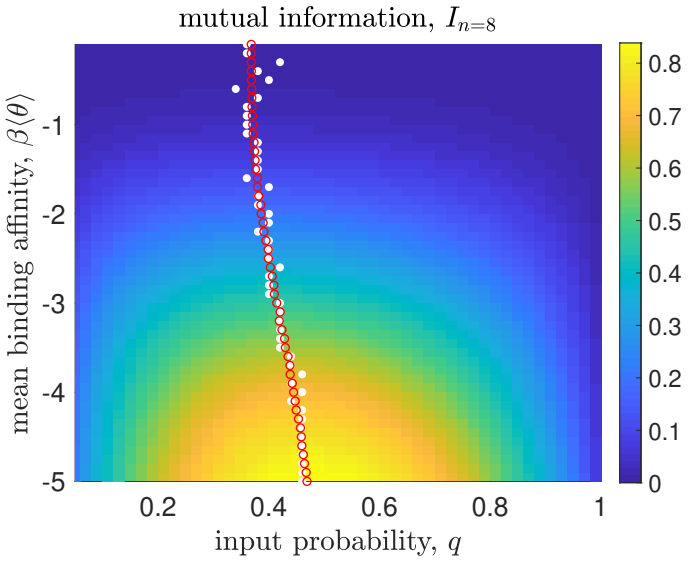

FIG. S5: **Location of maximum mutual information is not necessarily at input with most uncertainty ($q = 0.5$).** White filled circles are calculated by numerically searching for the input $q^\star$ on this color map that gives maximum mutual information. Red open circles are obtained through solving Eq. (S24). Here $\mathrm{Var}(\theta_{i,j}) \equiv \sigma^2 = 0.01$. The color map is generated by averaging 100 realizations of binding affinities using the scheme detailed in SI Section 2.

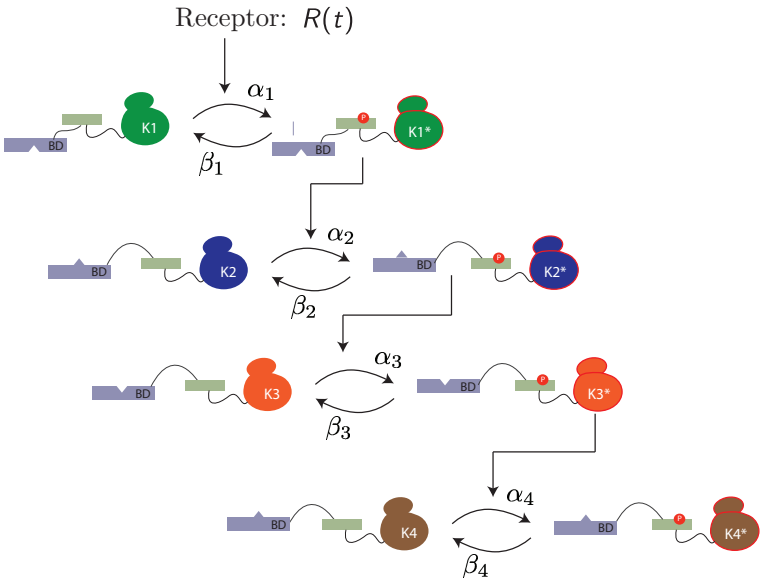

FIG. S6: **Signal network based on PK cascade.** The input of this signaling pathway is conceptualized as receptor kinase activation R which could potentially be time-varying. The phosphorylation cascade depicted here is as described in the main text. Here we denote α_i (β_i) as the phosphorylation (dephosphorylation) rate of the cascade stage $K_i \to K_{i+1}$.

concentration of active kinase X_{i-1} and that of the inactive downstream $\tilde{X}_i$. We describe the phosphorylation rate of kinase i by $\Phi_i^+ = \tilde{\alpha}_i X_{i-1} \tilde{X}_i$. Assuming the phosphatase concentration is constant, we can write down the dephosphorylation rate as $\Phi_i^- = \beta_i X_i$. Here $\tilde{\alpha}_i$, β_i are the kinetics rate constants of phosphorylation and dephosphorylation reactions, respectively. With this defined we can write down the kinetics equations for all kinases in the pathway

(except for the first one) as: ($\forall i > 1$)

$$\frac{dX_i}{dt} = \Phi_i^+ - \Phi_i^- \tag{S25}$$
$$= \tilde{\alpha}_i X_{i-1} \tilde{X}_i - \beta X_i \tag{S26}$$
$$= \alpha_i X_{i-1} \left(1 - \frac{X_i}{C_i}\right) - \beta_i X_i, \tag{S27}$$

where $\alpha_i = \tilde{\alpha}_i C_i$ is the pseudo-first order rate constant and $C_i = X_i + \tilde{X}_i$ is the total concentration of kinase i. For the first kinase, its phosphorylation is stimulated by active receptors whose concentration is denoted as $R(t)$. In addition, it is dephosphorylated by phosphatase at rate β_1. Combining this we have

$$\frac{dX_1}{dt} = \alpha_1 R(t) \left(1 - \frac{X_1}{C_1}\right) - \beta_1 X_1. \tag{S28}$$

At steady-state, we have for $i \neq 1$

$$X_i^{SS} = \frac{C_i X_{i-1}^{SS}}{\gamma_i C_i + X_{i-1}^{SS}}, \tag{S29}$$

where $\gamma_i = \beta_i / \alpha_i$. Divide both sides by the total concentration of kinase i, C_i, one gets the steady-state activation probability of i:

$$p_i^{SS} = \frac{C_i X_{i-1}^{SS} (\gamma_i C_i)^{-1}}{1 + X_{i-1}^{SS} (\gamma_i C_i)^{-1}}. \tag{S30}$$

Eq.(S29) is related to the Michaelis-Menton equation $V_{\max} S / (K_m + S)$ by recognizing

$$X_{i-1}^{SS} \rightarrow S \tag{S31}$$
$$C_i \rightarrow V_{\max} \tag{S32}$$
$$K_m \rightarrow C_i \gamma_i \tag{S33}$$

Finally, the steady-state probability model presented in the main text,

$$\mathbb{P}(x_i = 1 | x_{i-1}) = \frac{x_{i-1} e^{-\theta_{i-1,i}/(k_B T)}}{1 + x_{i-1} e^{-\beta \theta_{i-1,i}/(k_B T)}}, \tag{S34}$$

can be interpreted under this kinetic framework by relating

$$\theta_{i-1,i} = k_B T \ln \left(\frac{K_m}{X_{i-1}^{SS}}\right) = \Delta F - \tilde{\mu}_{i-1}, \tag{S35}$$

where $\Delta F = k_B T \ln K_m$ is the free energy difference between the bound and unbound state and $\tilde{\mu}_{i-1} = k_B T \ln X_{i-1}^{SS}$ is the chemical potential of active kinase $i-1$.

V. EFFECTS OF NETWORK DEPTH

Here we examine how the depth of network affects information transduction capacity. According to data processing inequality (DPI)[21, 44], information is never gained when transmitted through some noisy channel (or observation process). Formally, DPI states that suppose we have a Markov chain: $X_1 \rightarrow X_2 \rightarrow X_3$, where $X_1 \perp X_3 \| X_2$ (i.e. X_1 and X_3 are independent conditionally on X_2), then it must be that $I(X_1; X_3) \leq I(X_1; X_2)$. The pertinent question is therefore how much information degradation across signaling circuit is controlled by biochemical noise due to non-specific PPIs. In Fig.S7, we calculated mutual information for networks described in SI Sec.II of varying depth at two binding scenarios. At tight-binding, the noise due to promiscuity of PPIs is small and we observe that DPI is almost saturated (i.e. equality in DPI holds). In the other limit, information is always degraded when as it is relayed downstream.

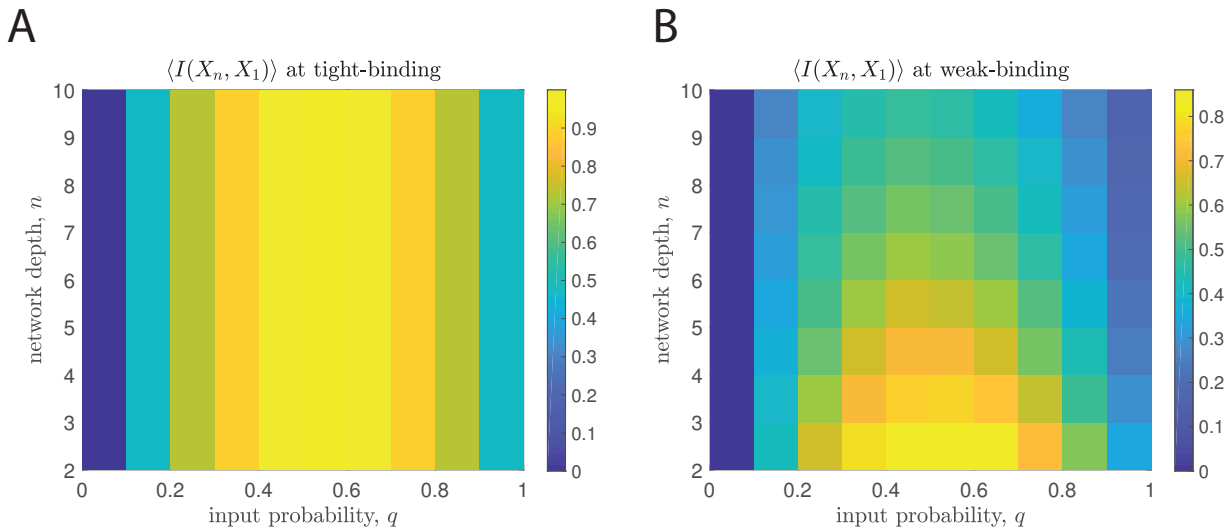

FIG. S7: Data processing inequality and biochemical noise: Mutual information across a linear network ($X_1 \to X_2 \to \cdots \to X_n$) is shown as the color map. Models are described in details in SI Sec.II. For tight/weak-binding (left/right panel), mean binding affinity $\beta\mu = -5/-0.1$. In all panels, $q = \mathbb{P}(x_1 = 1)$ and standard deviation of binding affinities $\sigma = 0.1$.

VI. IMPLEMENTATION OF COMPLEX NETWORKS

Here we consider all-to-all connected 2-n_w-2 networks, where $n_w \in \mathbb{N}$ is the number of intermediate nodes (see Figure 5B for an illustration). The goal is to compare the maximum mutual information of circuits with different n_w by using InfoMax (see Algorithm 1). To do so, we first construct such networks with varying n_w, subject them to inputs with different correlations, then maximize mutual information with respect to binding affinities. From a practical point of view, it is useful to define the followings. In the sequel, we use Roman letters $i, j, k, \cdots$ to denote the identity of nodes (i.e. protein species label) while reserving Greek letters $\mu, \nu, \sigma, \cdots$ for configurations (i.e. joint protein phosphorylation states).

Let $\mathbf{x} = (x_1, x_2)$ be the input (phosphorylation) state vector, $\mathbf{y} = (y_1, y_2, \cdots, y_{n_w})$ be the intermediates, and $\mathbf{z} = (z_1, z_2)$ be the outputs. Denote the binding affinity between y_i and x_j as $\theta_{i,j}$ and that between z_k and y_i as $\eta_{k,i}$ (all measured in units of $k_B T$). In other words, these energetic parameters can be summarized by the binding matrix $\boldsymbol{\theta} \in \mathbb{R}^{n_w \times 2}$ and $\boldsymbol{\eta} \in \mathbb{R}^{2 \times n_w}$. To distinguish the variable space (indexed by $i, j, k \cdots$) from the configuration space (indexed by $\mu, \nu, \sigma, \cdots$), let $\mathcal{X} = \{\mathbf{x}^{(\mu)}, \mu = 1, \cdots, 2^2\}$, $\mathcal{Y} = \{\mathbf{y}^{(\mu)}, \mu = 1, \cdots, 2^{n_w}\}$, and $\mathcal{Z} = \{\mathbf{z}^{(\mu)}, \mu = 1, \cdots, 2^2\}$ be the set of input, intermediate, output configurations, respectively. Define $\mathcal{P}_{\mu\nu} : \mathcal{X} \to \mathcal{Y}$ as a matrix that relates the joint states of two inputs (of dimensionality four) to that of the intermediates and $\mathcal{Q}_{\sigma\mu} : \mathcal{Y} \to \mathcal{Z}$ for that between the intermediates and the outputs. In matrix form,

$$\mathcal{P}_{\mu\nu} = \begin{pmatrix} 1 & \vdots & \vdots & \vdots \\ 0 & \mathcal{P}_{\mu,2} & \mathcal{P}_{\mu,2} & \mathcal{P}_{\mu,2} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \vdots & \vdots & \vdots \end{pmatrix}, \quad \text{and} \quad \mathcal{Q}_{\sigma\mu} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ \dots & \mathcal{Q}_{2,\mu} & \dots & \dots \\ \dots & \mathcal{Q}_{4,\mu} & \dots & \dots \\ \dots & \mathcal{Q}_{3,\mu} & \dots & \dots \end{pmatrix}, \tag{S36}$$

from which one can calculate the input-output and output marginal probability as

$$\mathbb{P}_{\sigma\nu}(\mathcal{Z}|\mathcal{X}) = \mathcal{Q}_{\sigma\mu}\mathcal{P}_{\mu\nu} \tag{S37}$$

$$\mathbb{P}_{\sigma}(\mathcal{Z}) = \mathcal{Q}_{\sigma\mu}\mathcal{P}_{\mu\nu}q_{\nu}, \tag{S38}$$

where q_ν is the joint probability of the inputs, namely, $\mathbb{P}(x_1, x_2)$. Note that we use the Einstein notation where repeated indices are implicitly summed over. To simplify notation, denote $y_i^{(\mu)}$ as the i-th component of the phosphorylation state vector of intermediate nodes $\mathbf{y}^{(\mu)} = (y_1^{(\mu)}, \cdots, y_{n_w}^{(\mu)})$. Let $u_\mu = \{i | i \in \mathcal{Y}, y_i^{(\mu)} = 1\}$ be the set of

intermediate nodes that are phosphorylated and $v_\mu = \mathcal{Y} \setminus u_\mu$ be those of that are not. The matrix element of $\mathcal{P}_{\mu\nu}$ is therefore:

$$\mathcal{P}_{\mu,2} = \prod_{i\in u_\mu} f(\theta_{i,1}) \prod_{j\in v_\mu} [1 - f(\theta_{j,1})] \tag{S39}$$

$$\mathcal{P}_{\mu,3} = \prod_{i\in u_\mu} f(\theta_{i,2}) \prod_{j\in v_\mu} [1 - f(\theta_{j,2})] \tag{S40}$$

$$\mathcal{P}_{\mu,4} = \prod_{i\in u_\mu} g(\theta_{i,1}, \theta_{i,2}) \prod_{j\in v_\mu} [1 - g(\theta_{i,1}, \theta_{i,2})], \tag{S41}$$

where

$$f(\zeta) = \frac{e^{-\zeta}}{1 + e^{-\zeta}} \tag{S42}$$

$$g(\zeta, \xi) = \frac{e^{-\zeta} + e^{-\xi}}{1 + e^{-\zeta} + e^{-\xi}}. \tag{S43}$$

Note that in writing down g, we ignored higher-order interactions such as those due to cooperativity $\sim e^{-\zeta-\xi}$ etc. We also choose $\mathbf{x}^{(\nu)} = \{(0,0), (1,0), (0,1), (1,1)\}$ for $\nu = 1, 2, 3, 4$ ordering. Similarly,

$$\mathcal{Q}_{2,\mu} = h(\eta_{1,u_\mu}) \, [1 - h(\eta_{2,u_\mu})] \tag{S44}$$

$$\mathcal{Q}_{3,\mu} = [1 - h(\eta_{1,u_\mu})] \, h(\eta_{2,u_\mu}) \tag{S45}$$

$$\mathcal{Q}_{4,\mu} = h(\eta_{1,u_\mu}) \, h(\eta_{2,u_\mu}), \tag{S46}$$

where the function h, with higher order interactions ignored, reads

$$h(\eta_{i,\mathcal{S}_i}) = \frac{\sum_{j\in\mathcal{S}_i} e^{-\eta_{i,j}}}{1 + \sum_{j\in\mathcal{S}_i} e^{-\eta_{i,j}}} \tag{S47}$$

With all these matrices defined, we can compute the mutual information Eq. (S17) by a series of matrix multiplications as we did in SI Appendix II.

VII. EFFECTS OF INPUT CORRELATIONS AND PATHWAY CROSS-TALKS ON INFORMATION CAPACITY

In this appendix, we show the results of a full analysis on pathway cross-talks to complement Figure 4. The network studied are depicted as labeled according Figure 4A.

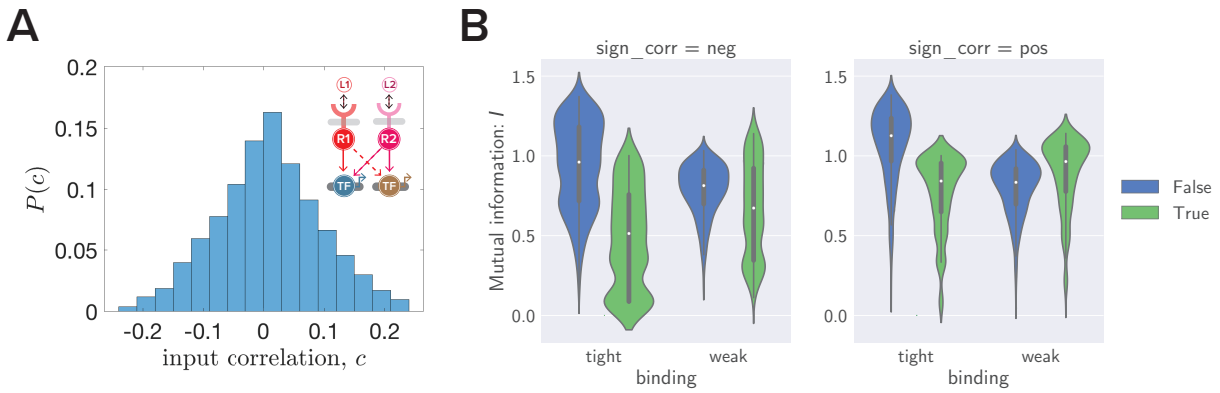

FIG. S8: (A) Network studied and the distribution of input correlation $P(c)$ (B) Violin plot for mutual information for network and input correlation shown in A. Violins are categorized according to their cross-talk levels and binding affinities. Blue for 'cross-talk' with $\beta\eta = -5.0$ and green for 'no cross-talk' with $\beta\eta = 0$ (see SI Section 4 for details). Tight-binding refers to $\beta\langle\theta\rangle = -5.0$ while weak-binding to $\beta\langle\theta\rangle = -1.0$. Note that $P(c)$ is constructed based on Eq.(16). See *Materials and Methods* for details.

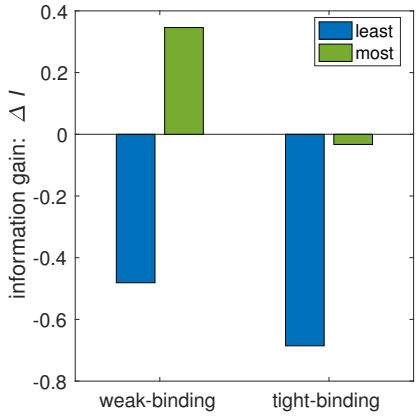

FIG. S9: Gain of mutual information through cross-talks for the network shown in Figure S8. Information gain ΔI is defined as the different between the mutual information with and that without cross-talk. The label 'most' and 'least' are annotated based on maximizing and minimizing ΔI with respect to input correlations c, respectively.