Boston University

OpenBU **http://open.bu.edu**

Theses & Dissertations Boston University Theses & Dissertations

2019

Statistical physics of information processing by cells

https://hdl.handle.net/2144/37062

Boston University

BOSTON UNIVERSITY

GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

STATISTICAL PHYSICS OF INFORMATION PROCESSING BY CELLS

by

CHING-HAO WANG

B.S., National Tsing-Hua University, 2009
M.S., National Tsing-Hua University, 2011

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2019

© 2019 by
CHING-HAO WANG
All rights reserved

Approved by

First Reader

Pankaj Mehta, PhD
Associate Professor of Physics

Second Reader

Kirill S. Korolev, PhD
Assistant Professor of Physics

There are two ways in which a science develops; in response to problems which is itself creates, and in response to problems that are forced on it from the outside. Ian Hacking, 1975

Sometimes it is the people no one imagines anything of who do the things that no one can imagine; Alan Turing in *The Imitation Game*

*Scholarship that is indifferent to human suffering is immoral.*Richard Levins

Acknowledgments

The acknowledgment is perhaps one of the few parts of this thesis that people will read. And this thesis is in some ways the culmination of decades of education, so there are many people without whom this thesis wouldn't have been possible. As such, I try my best to give credit when credit's due, but this list is unfortunately necessarily incomplete.

First and foremost, I'd like to thank my family for their unyielding support. They set me off this path long ago, perhaps by tolerating my never-ending questions and requests for more books, and, according to them, a right dose of stubbornness. I also thank my brother for always being a far more caring kid than I am, in particular when I'm far away from home to accompany my parents on holidays. Their encouragement helped me through many difficult times in both my life and education. Although they would never admit, they are indeed what makes everything possible.

Getting a Ph.D. is hard, and I really can't imagine without having Pankaj as my advisor. His limitless passion for science and uncompromising moral courage to fight for an emancipatory vision of knowledge have influenced me in myriad and profound ways. As a scientific mentor, Pankaj taught me to identify a problem, tinker with it, come up with rigorous analysis, and then finally condense all the efforts into a paper with an emphasis on clarity. He encourages, challenges, and refocuses even my silliest ideas with tireless enthusiasm and unprecedented insight, which always motivates me to think harder. As a friend and mentor in life, he showed me that there's more just work in a student's life. He espouses the rarely-mentioned truth that the key to success, in the long run, is to lead a rich and healthy lifestyle. As such, he views my wide range of non-scientific hobbies as strengths, not weaknesses or distractions, which significantly contributed to my work-life balance. Of course, his priceless laughter and ruthless prattle about random stuff is always the antidote to any mishaps a physics Ph.D. student can possibly encounter.

Of course, without the company of friends, I would definitely have an entirely different outcome in my scientific career. A special thank you to Alex D., Pranay, Chris B., Jonathan, Phil W., Phil. C, and Marin for reaching out and inviting me for food, drinks, hikes, and adventures in NOLA and LA. I will definitely remember the few nights we went carousing and got so wasted but somehow managed to wake up early the next morning for work. And yes, I should express my sincere gratitude to Alex D. for saving my life, literally, when we absconded to and got lost in the mountains in Santa Fe, for the hospitality during my visit to Quebec, for tolerating my rants and raves, and for providing me solid support when I went through dark moments in my life. I also thank Maddie for always making me laugh, even though inadvertently, and introducing me the "hotsy-totsy" — a hidden gem in Berkeley that has become my go-to whenever I find myself in the SF area. I also relish hanging out with folks outside the physics family. David B., Meghan T., Josh, and other members in the Segre lab are always my source of enjoyment. Particularly I thank David B. for those HEADY-TOPPERs, free Vermont commercials, and hosting my VT fall getaway. Well, too bad we didn't get to hike the Camel's Hump last time. I'll greatly miss those late-night drinks and chants with Josh. If I ever become a gutter punk, I think I know how to survive, thanks to him. Members of the Khalil lab are also much appreciated. I particularly thank Nikit, Meghan B., Szilvia, Divya, Brandon, Chris M., and Mo for enduring my silly questions in their lab meeting, and for inviting me to their fun lab outing. I'll greatly miss Nikit and Meghan B.'s banter for sure. I also thank my roomies, Amir and Ben C., for making 25 Cameron such a lovely living environment, and I can't wait to share with you guys another exciting year under the same roof.

I also benefit a great deal from many knowledgeable faculties and scholars at BU. Anatoli's class in quantum ergodicity provided me a new perspective on classical and quantum chaos, random matrix theory, (pre-)thermalization, and stat mech in general. Claudio's special-topic course got me into a zoo of research fields regarding spin glass, satisfiability, and statistical inference, which motivated in some ways many of the things presented in this thesis. Liam's class on particle physics is also a precious one. As he puts it, "every physicist should at least calculate the amplitude of gluon-gluon scattering once in their life." Although I've never taken Chris L.' class, I absolutely enjoy his impromptu lecture in the CMT lounge explaining quantum computation (or just theory of computation in general). I also thank Kirill for taking me under his wing as I was navigating my research direction when I first arrived at BU. Beyond BU, I thank Caleb for patiently teaching me biology and giving me useful advice one scientific presentation and career choice (and of course showing me around during my Houston visit!) I also thank Michael for a fruitful collaboration on the "thermodynamics of nuclear transport" project.

Finally, I thank my boyfriend, Ben K., for loving me back, giving me support, and bringing joy to my life. I thank him for showing me the profound ideal that love can embody, and how two people can become something greater than they once were. I'm thrilled to have him around to witness this hallmark of my life.

Ching-Hao Wang
Cambridge, Massachusettes

STATISTICAL PHYSICS OF INFORMATION PROCESSING BY CELLS

CHING-HAO WANG

Boston University, College of Arts and Sciences, 2019

Major Professor: Pankaj Mehta, PhD
Associate Professor of Physics

ABSTRACT

This thesis provides a physics account of the ability of cells to integrate environmental information to make complex decisions, a process commonly known as signaling. It strives to address the following questions: (i) How do cells relate the state of the environment (e.g. presence/absence of specific molecules) to a desired response such as gene expression? (ii) How can cells robustly transfer information? (iii) Is there a biophysical limit to a cells' ability to process information? (iv) Can we use the answers to the above questions to formulate biophysical principles that inform us about the evolution of signaling? Throughout, I borrow techniques from non-equilibrium statistical physics, statistical learning theory, information theory and information geometry to construct biophysical models capable of making quantitative experimental predictions. Finally, I address the connection of energy expenditure and biological efficiency by zeroing in on a process unique to eukaryotic cells– nuclear transport. The thesis concludes with a discussion of our theory and its implications for synthetic biology.

Contents

1 Introduction 1

1.1 All living cells are able to sense and response 2

1.2 How do cells process information? 3

1.2.1 Cellular information processing must operate at multiple scales in space and time 4

1.2.2 Information is encoded as the state of proteins and processed via the interactions between them 6

1.2.3 Proteins can change their states through various mechanisms 9

1.2.4 Protein-protein interaction mediates the currencies of signaling 10

1.3 What might be a plausible physical theory? 11

1.4 Organization of thesis 12

2 What controls the flow of information across signaling network? 17

2.1 Introduction 18

2.2 Including post-translational modifications in thermodynamic models 22

2.3 Mutual information and PPIs 25

2.4 Results 28

2.4.1 Weak binding affinities result in noise that limit the signal-to-noise ratio and information capacity 28

2.4.2 Information loss in signaling 'can' be mitigated by cross-talks when inputs are correlated 34

2.4.3 Information maximization for complex multi-input, multi-output circuits 35
2.5 Discussions 39

3 Can we design signaling network to implement desired functions? 43
3.1 Introduction 44
3.2 Results 48
3.2.1 Modular allostery and the "reader-writer-eraser" paradigm . 48
3.2.2 A thermodynamic model combining protein interactions with post translational modifications 49
3.2.3 Finding biophysical parameters that implement a given decision surface 51
3.2.4 Learning decision surfaces for multi-input signaling circuits . 53
3.2.5 Circuits with identical steady-state properties can display diverse dynamical behaviors 61
3.3 Discussion 63
3.4 Methods and Materials 68
3.4.1 Using gradient descent to design an input-output relation . . 68
3.4.2 Implementing decisions using CFFL and ICFFL 69
3.4.3 Robustness of decision from Fisher's information 72

4 Is there a limit to cell's ability to perform computation? 74
4.1 Introduction 75
4.2 Results 78
4.2.1 Generalized thermodynamic model of a synthetic protein-protein interaction network 78
4.2.2 Defining computational capacity 82

4.2.3 Computation is limited by the number of biophysically realizable configurations: notion of capacity 85

4.3 Computational capacity of a signaling network in the large n limit . 86

4.3.1 Writers only 86

4.3.2 Systems with both writers and erasers 86

4.4 Discussion 89

5 Example study: Thermodynamic paradigm of nuclear transport by living cells 94

5.1 Introduction 95

5.2 Thermodynamic formulation 98

5.3 Entropy production of futile cycle 101

5.4 Discussions 103

A Appendix of Chapter 2 107

A.1 Details of InfoMax 107

A.2 Signal-to-noise ratio (SNR) of signaling circuit 107

A.3 Deriving the information capacity 111

A.4 Optimal input to reach maximum mutual information 113

A.5 Relating thermodynamics to a kinetic model of phosphorylation cascade 115

A.6 Effects of network depth 119

A.7 Implementation of complex networks 120

A.8 Effects of input correlations and pathway cross-talks on information capacity 123

B Appendix of Chapter 3 125

B.1 Kinetics model on phosphorylation cascade 125

B.2 CFFL 127

B.3 ICCFL 129

B.3.1 Summary of parameters 130

B.4 Formulae used in the optimization 130

B.4.1 CFFL 130

B.4.2 ICFFL 132

B.4.3 Loss function used 133

C Appendix of Chapter 4 135

C.1 Statistical mechanical model for generic systems 135

C.2 Problem setup in our synthetic system 136

C.3 Bounds on admissible design matrix 138

C.4 Size of minimal configuration space 140

C.5 Capacity of computation: synthetic systems with writers only 145

C.6 Capacity of computation: synthetic systems with writers and erasers 148

C.7 Calculating capacity with Monte Carlo Sampling 153

D Appendix of Chapter 5 155

D.1 Biophysical models 155

D.1.1 Kinetics equations 157

D.2 Estimating the rate constants 159

D.2.1 Reaction 5: Ran exchange mediated by RanGEF 159

D.2.2 Reaction 2: Ran exchange mediated by RanGAP 162

D.3 Standard estimate of diffusion-limited reaction rate 163

D.4 Simulation codes 164

D.5 Entropy Production 164

D.6 Weak sensitivity to the GTP:GDP ratio θ 167

References 170

Curriculum Vitae **182**

List of Tables

4.1 Cover′s perceptron vs. Cell signaling 92

D.1 Reaction rate constants used in the thermodynamic model of nuclear transport 160

D.2 Molecular species involved in nuclear transport 160

List of Figures

1·1 **Cells respond to a wide range of inputs by invoking different responses.** Some commons inputs (green arrows) and outputs (red arrows) are shown. In many cases, the outputs will change the response to future inputs (i.e. feedback). The ability to monitor and self-regulate its internal state (i.e. homeostasis) is also indicated. This figure is adapted from(Lim et al., 2014) 3

1·2 **Cells integrate signaling and gene expression to achieve regulation.** An exemplary signaling system that combines signaling and gene regulatory unit's is shown as red and green boxes, respectively. Signaling input is assumed to be ligand (L) binding to the receptor that straddles the cell surface depicted as a purple box. This input is transmitted across the cytosol into the nucleus to induce gene expression. Contrast between common features of cell signaling and gene expression is summarized. This figure is adapted from(Lim et al., 2014) 6

1·3 **Information is encoded in the state proteins and further processed via protein-protein interactions.** A generic signaling pathway is depicted on the left where nodes represent proteins while directed edges illustrate their interactions (e.g. binding/dissociations, enzymatic reactions). Shown on the right is the biophysical representation of a node (in this case a protein). A node in the pathway can be thought of as a switch that can change between ON (usually catalytically active) and OFF (catalytically inactive) states, leading to the change of output. Such state switching is dictated by the protein-protein interactions. Conceptually, information is encoded as the ON and OFF states of these nodes. Note that more than two states are possible for each node. Since the switching between states typically involves protein-protein interactions, one can imagine it as a mechanism by which information is processed. 14

1·4 **Information currencies in cell signaling.** Unlike electronic systems, cell signaling employs different currencies to process and transmit information. The purple species indicates the protein that can switch between ON (orange-red haloed circle) and OFF states, as depicted in Figure 1·3. Such switching (or writing and erasing) can be manifested a variety of ways such as binding and dissociation, post-translational modifications, conformational changes, and localization to different part of the cell. In (B), the modification is shown as X with $X = \mathrm{P, Me, Ub}$ indicating phosphorylation, methylation, and ubiquitination, respectively. 15

1·5 **Protein-protein interactions mediate various mechanisms of information processing in cell signaling**. Interactions between proteins can change various aspects of cell signaling (indicated by solid arrows) such as protein's conformation, post-translational modifications, and catalytic activities. They may also localize proteins to specific site within the cell such as nucleus to allow for DNA-protein interactions. These aspects (colored boxes), regarded as the currencies of cell signaling, can directly or indirectly affect one another (dashed arrows). 16

2·1 (*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 epidermal growth factor receptor (EGFR) signaling, binding of EGF to the extracellular domain of EGFR leads ot 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. 27

2·2 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 **??** *C*, green circles indicates noise. (*B*) Color map shows the numerically simulated log-signal-to-noise ratio (log-SNR), defined by Eq.(2.10), 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 0.01 (see main text for simulation details). This quantity can also be obtained by solving Eq.(A.9) (see Appendix). In this panel, we show the result for $n = 2, 5$. (*C*) Input-output relation of the n-layer kinase cascade ($n = 2, 5$) at tight- and weak-binding is shown. 32

2·3 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. 36

2·4 Effect of cross-talk between pathways on information capacity. (*A*) Schematic of insulated and cross-talked networks are shown. Dashed connections represents 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 Appendix A 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 the binding affinities for protein-protein interaction (weak/tight-binding corresponds to $\beta\langle\theta\rangle = -5/-1$), and colors indicate the presence or absence of cross-talks. In the 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 Appendix A 4 for more details.) 38

2·5 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 $R \rightarrow K_1$ 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 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 $[-2, 8]$. 42

3·1 (A) Shown here are three representative network motifs in signaling based on cascade of two or more (phosphorylation based) enzymes, the activity of which is determined by opposing regulators (i.e. kinase and phosphatase). Coherent feedforward loop: epidermal growth factor receptor (EGFR) activation (i.e. input, not shown) activates the ERK-RSK-SRF pathway. Both phosphorylated RSK and SRF function to promote the transcription of FOS gene. This motif can filter out noise and distinguish transient from sustained inputs. Incoherent feedforward loop: EGFR signaling wherein the activation of EGFR induces both the expression of FOS and its inhibitor zinc-finger protein 36 (ZFP36), which promotes FOS mRNA degradation. This motif demonstrates the transient memory of FOS due to the initial activation of ERK before it turns on ZFP36 whose level accumulation serves to shut off FOS. The perceptron-like motif can orchestrate multi-input-multi-output (MIMO) decisions(Jordan et al., 2000). For example, cell division control protein 42 (Cdc42), a member of the Rho family GTPases, can be stimulated both by the receptor tyrosine kinase (RTK) as well as the G-protein coupled receptor (GPCR). Depending on the combination of these inputs, Cdc42 can activate different downstream kinases: p21 activation kinase (Pak), S6-kinase (S6K), and SRF. (B): Thermodynamics model of signaling based on kinase phosphorylation. The biophysical parameters shown here can be mapped to chemical kinetics rate constants measurable in experiments (see Appendix B for more details). (C): Our design flow. 55

3·2 Learning to engineer computations and decisions. (A): Learning to implement desired decision surface (decision surface) with 2-1-1 perceptron motif (c.f. Figure 3·1B). This network consists of three types of protein kinases (K1, K2, and K3) that can undergo phosphorylation cascade. The phosphorylation of K1 and K2 are controlled by binding of ligands L1 and L2, respectively. K1 and K2 can both interact with and phosphorylate K3 which in term regulates the expression of some target gene. The target and learned decision surface (i.e. input-output relation, I-O relation) are shown in blue and green, respectively. The parameters (effective binding energies in units of $k_B T$ and the kinase concentrations in units of 1M) learned that implement this specific design are shown on the right. For more details on learning, see and main text and Appendix B. (B,C): Similar to (A) but with more complex perceptron-like motifs. MIMO: multi-input-multi-output. 56

3·3 Learned parameters of panel B and C.Here we show the learned parameters of DS implemented in Figure 3·2B,C. Note that concentrations are held fixed through learning. In other words, optimization is taken over the effective binding affinities $\Delta\epsilon_{ji}$. 57

3·4 Robustness of learned DS. Here we use the 2-1-1 network (Figure 3·2) for demonstration. Based on the model in the main text we calculate the Fisher's information matrix (FIM), $g_{\mu\nu}$, defined in Eq.(3.26). To illustrate FIM in 2-dimensional space, we chose $\Delta\epsilon_{21} = \Delta\epsilon_{31} := \theta_1$ and $\Delta\epsilon_{43} := \theta_2$ while fixing the concentrations of all proteins constant so that we have effectively 2 parameters(i.e. $\mu, \nu \in \{\theta_1, \theta_2\}$, which we abbreviate as $\mu, \nu \in \{1, 2\}$). In panel (A) we showed the solution on the θ_1-θ_2 plane. White dot indicates the point where the solution found by our design flow (c.f. Eq.(3.26)). $\hat{e}_1, \hat{e}_2$ are the eigenvectors of FIM. (B-D) Components of FIM. The eigenvalues of FIM are: $\lambda_1 = 2.78 \times 10^{-4}$ and $\lambda_2 = 1.49 \times 10^{-7}$. 60

3·5 Learning to implement desired decision with naturally occurring motifs. (A) Motifs used are CFFL and ICFFL. Red curve indicates the target decision while blue circles and green squares represent that implemented by CFFL and ICFFL, respectively. Learned parameters are then used to simulate the kinetics (i.e. nodes in the network diagram) of kinase activity (B-G). (B): Constant temporal input stimulation. (C): Same as (B) but with exponentially decaying input profile $e^{-\lambda t}, \lambda = 1.0$. (D): Kinase activity of the CFFL subject to constant stimulation (i.e. panel (B)) using the learned parameters that implements the I-O relation in (A): $\alpha_i = 1.0, \beta_1 = e^{-8.2}, \beta_2 = e^{-7.5}, \beta_3 = \beta_4 = e^{-6.3}, K = e^{-2}, C_i = 2.5$ (See Appendix B for interpretations of parameters). (F): Kinase activity of ICFFL subject to constant stimulation with learned parameters: $\alpha_i = 1.0, \beta_1 = \beta_2 = e^{1.0}, \beta_3 = e^{-0.05}, \beta_4 = e^{-3.0}, K = e^{-2}, C_i = 2.5$. (E): Same as (D) on CFFL but with exponentially decaying stimulation (i.e. $\lambda = 1.0$). (G): Same as (F) on ICFFL but with exponentially decaying stimulation (i.e. $\lambda = 1.0$). In (D)(E)(F)(G), activities are normalized to the maximum value attainable by all kinases. Colors of curves are matched to that of the nodes in the networks shown in (A). 64

4·1 Number of biophysically realizable collective protein internal states (i.e. configurations) is limited by the diversity of domain interactions. (*A*) Shown here are the domain diagrams of n distinct protein species, each of which bears specific internal states. In particular, species i has n_i internal (e.g. allostery, phosphorylation) states. Species shown here indicate $n_1 = 3, n_2 = 2, \cdots n_n = 1$. Total number of collective (species) internal states N_{total} is given by $\Pi_{i=1}^{n} n_i$, which, in the case where $n_i = 2$ for all i, becomes 2^n. (*B*) Varieties of protein domain interactions can be summarized by a table of binding domain (BD) pairs with internal state-dependent binding affinity. In this example, there are M distinct domain interactions so M binding pairs. (*C*) Number of realizable configurations $N_{\text{realizable}}$ is far less than N_{config} due to the number of ways proteins can interact with one another. 79

4·2 Biophysical constraints associated with protein activity configurations. (*A*) Shown here are three configurations of a signaling systems consisting of three writer species (W), along with their configuration vector (left column) and the associated constraints in physical parameter space (right column): $\Delta\epsilon_{ij}$ is the *effective* binding energy between species i and j. The two terms shadowed in yellow in the right column indicates that the configurations they correspond to are not compatible. (*B*) Exemplary species diagram. This construction is similar to the synthetic example given in Fig.4·1. 83

4·3 **Incompatibility probability and capacity in synthetic systems with both writers and erasers**: $P_m(n)$ is given by Eq.(C.46) 89

4·4 Effect of network sparsity on capacity measured by the fraction of biophysically realizable configurations: $C/2^n$, where n is the number of protein species. Shown here are Monte Carlo simulation of different set of network ensembles with different sparsity p (connectivity probability) and fraction of writers α. (A) $\alpha = 0.2$ (B) $\alpha = 0.5$ (C) $\alpha = 0.9$. In all panels, dots are results of Monte Carlo simulations. Dashed lines are shown only for clarity. 90

5·1 (Color online) Demixing of cargo across the nuclear membrane is driven by Ran coupled to NTF2 and importin system. (A) With such coupling (upper panel), nuclear cargo accumulation is favored and Ran GTP/GDP exchange cycle proceeds faster than without coupling (lower panel). The thickness of arrowed curves in Ran cycle indicates strength of reaction flux; length of arrowed lines in cargo transport represents the rate at which the underlying processes occur. (B) Details of molecular demixing machine in the context of nuclear transport. Species labels as above. Reactions corresponding to Ran cycle and cargo transport are highlighted by red and green boxes, respectively. The orange dashed box includes all reactions coupled by the importin-NTF2 system. See also Fig S1 in the Supplementary Material. 96

5·2 (Color online) Phase diagram of nuclear localization. (A) The cargo nuclear localization NL:= $[C]_{nu}/[C]_{cyto}$ (color shadings) is obtained by varying the total importin and total NTF2 concentrations while keeping overall cargo level fixed at $[C]_{tot} = 100$ nM. (B) A family of curves shows NL for several cargo concentrations as a function of importin concentration with $[NTF2]_{tot} = 100$ nM. The 1D curve for $[C]_{tot} = 100$ nM is a cut across the plot of panel A. Locations of NL maximum are marked by diamonds (see FIG.5·4C as well). Kinetic rate constants used are given in the Supplementary Material. Total Ran concentration $[Ran]_{tot}$=75 nM. 100

5·3 (Color online) Competition between RanGTP and cargo to bind importin. (A) Schematic of the two competing reactions. (B) Reaction flux for importin-RanGTP formation $\tilde{\Phi}_4^+ \sim k_4^+[RanGTP]$ and (C) flux for importin-cargo formation $\tilde{\Phi}_7^- \sim k_7^-[C]_{nu}$. Fluxes are scaled by $[Im]_{nu}$ (see text). Parameters as in FIG.5·2 102

5·4 (Color online) Phase diagram of entropy production. (A) Entropy production is plotted as a 2D function of NTF2 and importin, at fixed cargo concentration $[C]_{tot} = 100$ nM. Compare with FIG. 2A. (Axes extend to 5 instead of 0 nM to avoid numerical divergence.) (B) A family of curves shows the entropy production for several cargo concentrations as a function of importin concentration; NTF2 concentration fixed at $[NTF2]_{tot} = 100$ nM. The 1D curve for $[C]_{tot} = 100$ nM is a cut across the plot of panel A. Compare with FIG. 2B. Peaks and troughs are marked by squares and circles, respectively. (C) Locations of entropy production max/min (square/circle) and that of nuclear localization maximum (diamond). Colors match curves in panel B. The importin concentration at which EP is minimum is close to but always less than $[Im^*]$, where NL is maximum. Thus, one strategy for maximizing the efficiency of demixing is to have the futile cycle operate in regime where its entropy production is minimized. (D) EP decreases at very high importin concentration. This reflects a loop around the energetic reaction Φ_4 via the reversible reaction Φ_8. Here $[NTF2]_{tot}$=100 nM, as in the panel B. In all panels, $[Ran]_{tot} = 75$ nM. Kinetics constants as in FIG.5·2 and 5·3 (see SM Section II). 104

A·1 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.(A.9)(A.10), open circles are calculated via sampling network parameters as described in Methods. Black dashed curve is the analytical result in the tight-binding limit whose expression is given by Eq.(A.5)(A.7) with σ corresponding to that of the yellow open circles ($\sigma = 0.1$). Note the logarithmic scale. 111

A·2 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. (A.20). Parameters used are: $\beta\mu = -5, \sigma = 0.1$. The value and location of maximum mutual information obtained by averaging is indicated as $(q, I_{\max}) = (0.5, 0.99998)$. 114

A·3 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, $\sigma = 0.1$. 114

A·4 **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 Appendix A 2. 116

A·5 **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. (A.24). 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 Appendix A 2. 117

A·6 **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 $K_i \to K_{i+1}$. 118

A·7 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.A.3. For tight/weak-binding (left/right panel), mean binding affinity $\beta\mu = -5/-0.1$. In all panels, $q = P(x_1 = 1)$ and standard deviation of binding affinities $\sigma = 0.1$. 120

A·8 (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 Appendix A 4 for details). Tight-binding refers to $\beta\langle\theta\rangle = -5.0$ while weak-binding to $\beta\langle\theta\rangle = -1.0$. 124

A·9 Gain of mutual information through cross-talks for the network shown in Figure A·8. 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. 124

B·1 **Signal transduction cascade that implements CFFL** 133

B·2 **Signal transduction cascade that implements ICFFL** 134

D·1 **Molecular reactions involved in nuclear transport**. (A) Molecular species in our nuclear transport model. (B) Schematics of the reactions involved in nuclear transport. Note that + and - signs represent self-consistently the start and end points of the reactions, rather than forward or reverse cycle orientations. Thus for example $k_{"+"}$ and $k_{"-"}$ for reaction 9 in Eq.(S15) take signs for loss and gain, respectively. (C) The subset of reactions in (B) that forms the futile cycle (i.e. the energy source for nuclear transport). 156

D·2 **Illustration of Ran GDP to GTP exchange reaction mediated by RanGEF** 159

D·3 **Illustration of RanGTP to RanGDP exchange reaction mediated by RanGAP** 162

D·4 **Effect of free GTP to free GDP ratio θ on nuclear localization**. (A) $[\mathrm{NTF2}]_{tot}$=100 nM (B) $[\mathrm{NTF2}]_{tot}$=10 nM. Other parameters are the same for both panels: $[\mathrm{C}]_{tot}$=10 nM, $[\mathrm{Ran}]_{tot}$=75 nM and $[\mathrm{Im}]_{tot}$=100 nM. Kinetics rate constants used are given in SM Section II. 168

D·5 **Phase diagram of nuclear localization and entropy production**. (A) total cargo concentration $[\mathrm{C}]_{\mathrm{tot}} = 20$ nM (B) $[\mathrm{C}]_{\mathrm{tot}} = 50$ nM and (C) $[\mathrm{C}]_{\mathrm{tot}} = 100$ nM. Other parameters used are the same as in FIG.3 and 4: $[\mathrm{Ran}]_{\mathrm{tot}} = 75$ nM. 169

Math Symbols

Scalars, Vectors, and Matrices

a	A scalar
a	A scalar random variable
$\boldsymbol{\mathsf{a}}$	A vector
$\mathbf{a}$	A random vector
$\boldsymbol{\mathsf{A}}$	A matrix
$\mathbf{A}$	A random matrix
$\boldsymbol{\mathsf{A}}^\mathrm{T}$	Transpose of matrix $\boldsymbol{\mathsf{A}}$
a_i	The i-th component of vector $\boldsymbol{\mathsf{a}}$
a_{-i}	All but the i-th component of vector $\boldsymbol{\mathsf{a}}$
a_i	The i-th component of random vector $\mathbf{a}$
a_{-i}	All but the i-th component of random vector $\mathbf{a}$
$A_{i,j}$	The i, j-th component of matrix $\boldsymbol{\mathsf{A}}$
$\boldsymbol{\mathsf{A}}_{i,:}$	Row i of matrix $\boldsymbol{\mathsf{A}}$
$\boldsymbol{\mathsf{A}}_{:,i}$	Column i of matrix $\boldsymbol{\mathsf{A}}$
$\mathsf{A}_{i,j,k}$	Element (i, j, k) of tensor A
$\mathsf{A}_{:,:,i}$	2-D slice of a 3-D tensor A

Probability and Information-theoretic Quantities

$\mathrm{a} \perp \mathrm{b}$	The random variables a and b are independent
$\mathrm{a} \perp \mathrm{b} \parallel \mathrm{c}$	a and b are independent conditioned on c
$P(\mathrm{x})$	Probability distribution over a random variable x
$\mathrm{x} \sim P$	Random variable x a has distribution P
$\mathbb{E}_{\mathrm{x}\sim P}(f(x))$ or $\mathbb{E}f(x)$	Expectation value of $f(x)$ with respect to $P(\mathrm{x})$
$\mathrm{Var}(f(x))$	Variance of $f(x)$ under $\mathrm{x}\sim P$
$\mathrm{Cov}(f(x), g(x))$	Covariance of $f(x)$ and $g(x)$ under $P(\mathrm{x})$
$H(\mathrm{x})$	Shannon entropy of random variable x
$D_{\mathrm{KL}}(P \parallel Q)$	Kullback-Leibler divergence of P and Q
$\mathcal{I}(\theta) \equiv \mathcal{I}_{\mathrm{x}\sim P(\cdot \mid \theta)}$	Fisher information of x about parameter θ under P
$MI(\mathrm{x};\mathrm{y})$	Mutual information between x and y

Functions

$\nabla_{\mathbf{x}} f$	Gradient of function f with respect to $\mathbf{x}$
$\nabla_{\mathbf{X}} f$	Matrix derivative of function f with respect to $\mathbf{X}$
$\frac{\partial f}{\partial \mathbf{x}}$	Jacobian matrix $\mathbf{J} \in \mathbb{R}^{m\times n}$ of function $f : \mathbb{R}^n \to \mathbb{R}^m$.
$\nabla^2_{\mathbf{x}} f(\mathbf{y})$	Hessian of function f evaluated at $\mathbf{y}$
$\partial f(\mathbf{x})$	Subdifferential of convex function f evaluated at $\mathbf{x}$

Abbreviations

GAP	GTPase-activating proteins
GEF	Guanine nucleotide exchange factors
GDP	Guanosine diphosphate
GTP	Guanosine triphosphate
PK	Protein kinase
PPI	Protein-protein interaction
PTM	Post-translational modifications
TF	Transcriptional factor

Chapter 1

Introduction

> *'Among substances are by general consent reckoned bodies and especially natural bodies; for they are the principles of all other bodies. Of natural bodies some have life in them, others not; by life we mean self-nutrition and growth (with its correlative decay). It follows that every natural body which has life in it is a substance in the sense of a composite. '* (Aristotle, 350 BC)

The cell is the basic unit of life. It defines the structural and functional unit of living organisms that display remarkable properties that preserves and furthers their existence. Irrespective of the differences in their identities, cells of all living organisms are equipped with the ability to sense and react to the environments. This is particularly important for the development of cells under evolutionary pressure. For example, the ability to move towards nutrients and to avoid environmental stress and toxins could confer a tremendous competitive advantage for early unicellular organisms. For multicellular organisms, such ability is also critical since their functioning relies substantially on the constant monitoring and exchange of information with the environment to coordinate the activities of individual cells. Furthermore, disruptions of communication between cells and the environments they live in are usually associated with disease such as cancer. To reduce the occurrence of such outcome and to maximize their chances of survival, it is conceivable that cells must have evolved some kind of mechanisms to process information.

In this chapter, I explore this idea that cells, being the smallest unit of life, are processing environmental information to decide on their behaviors. I begin with a discussion on the fundamental role of *cell signaling*, a process defined as the relay of environmental information to intracellular response, its importance, ubiquity, and molecular basis. I then contrast cell signaling with information processing done by human-made electronic devices such as computers and cells phone to argue that they are conceptually similar even though having different physical implementations. Based on these discussions, I motivate and outline the questions to be addressed in this thesis and highlight the ingredients of a plausible physical theory to tackle these questions. Throughout, I emphasize the general properties of cell signaling in terms of its core molecular biology, and showcase how these pieces of qualitative understandings culminate in a hint on the formulation of a plausible physical theory.

1.1 All living cells are able to sense and response

While there isn't a unanimous definition of life, it is commonly agreed that the ability to response to stimuli is an essential characteristic of a living being– an entity that preserves, furthers, and reinforces its existence(Koshland, 2002). This is perhaps not hard to imagine since we typically consider something life-like if it is responsive to some type of external probe (e.g. a gentle smudge). For single-celled organisms, such ability would allow them to evade environmental toxins and stresses, therefore conferring them an selective advantage in the course of evolution. But for that to have happened, a tiny being like cell must be able to discern a wide variety of environmental conditions such as the existence of molecular species, temperature, pH values etc., and to adjust aspects of their, say, gene expression, metabolism, and structure etc., in the wake of changes in these conditions (see Figure. 1·1). With the emergence of multicellular organisms, individual cells have evolved specialized biochemical machinery (to be discussed later) to communicate with other cells within the same organism, thereby implementing a high-level of coordination among these cells to function as an integrated entity. Therefore, by studying the stimulus-response behaviors observed in cells, one can learn a great deal about how living organisms function and came to being.

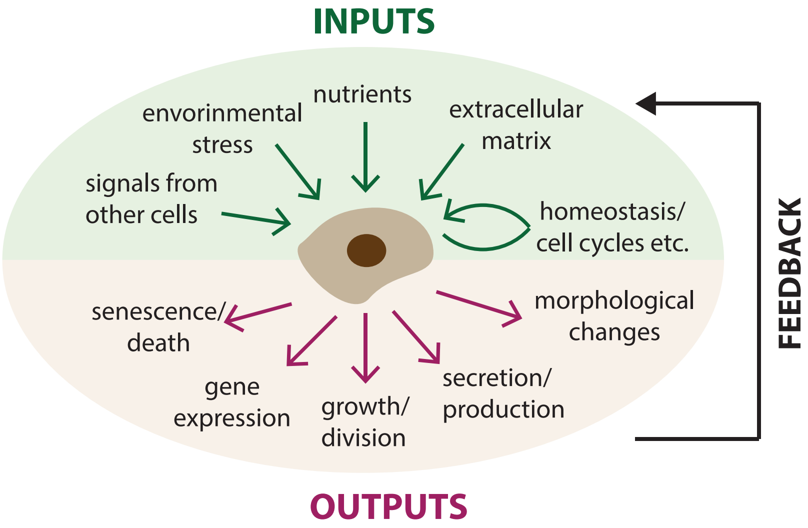

INPUTS
envorinmental stress
nutrients
extracellular matrix
signals from other cells
homeostasis/ cell cycles etc.
senescence/ death
morphological changes
gene expression
growth/ division
secretion/ production
FEEDBACK
OUTPUTS

Figure 1·1: Cells respond to a wide range of inputs by invoking different responses. Some commons inputs (green arrows) and outputs (red arrows) are shown. In many cases, the outputs will change the response to future inputs (i.e. feedback). The ability to monitor and self-regulate its internal state (i.e. homeostasis) is also indicated. This figure is adapted from(Lim et al., 2014)

1.2 How do cells process information?

Now having a picture of cells relating external information to some sort of internal response, which is commonly called cell signaling, it is perhaps not hard to imagine these teeny-tiny cells as some sort of information processing or decision-making systems. But unlike our human-made devices such as computers and cell phones which consists of gazillions of transistors that are tidily wired, cell signaling systems are made up of densely packed proteins, lipids, and many other biomolecules diffusing around, all surrounded by a water-impermeable membrane. So how is that a genetically-encoded information processing system that operates at such ‘messy’ environment like this is able to generate the diverse behavior that we observe at the cellular and even the organismic level? This is the question that is central to all biology.

In the following, I discuss a few features fundamental to cellular information processing. Before that, let me clarify a few jargons. A *signaling pathway* is a collection of molecules in a cell, most of which are proteins and lipids, that work in tandem to regulate and control the functions and behaviors of one or more cells. *Pathway activation* refers to the event that the first molecular species in the pathway, labeled sequentially in time, receives a signal (e.g. ligand binding, see Figure 1·1) and induce subsequent events.

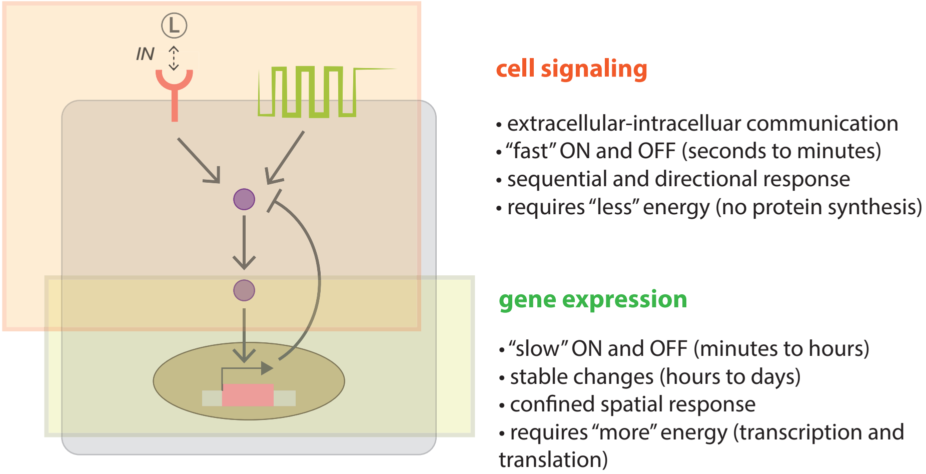
1.2.1 Cellular information processing must operate at multiple scales in space and time

A typical signaling pathway can be roughly divided into two parts, based on the location where events occurred – those at the plasma membrane and the cytosol (termed signaling for convenience), and those “around” the nucleus (termed gene expression[1]), see Figure 1·2. For most eukaryotic signaling pathways, upon activation, an extracellular signal has to “propagate” into the cell and across the cytosol, before reaching the nucleus to promote or repress the transcription of specific genes. To get a sense of scale, a mammalian cell is roughly $d = 10 - 100\mu$m in diameter, with a nucleus roughly 10% of its size. Proteins in the cytoplasm has diffusion constant of the order $D \approx 10\,\mu\text{m}^2/\text{sec}$ (Milo and Phillips, 2015), which implies that it takes roughly $\tau = d^2/6D \approx$ 1 sec - 3 mins for a protein to diffuse across a mammalian cell, and $\tau \sim$ 10 ms for a prokaryotic cell such as *E. coli* with $d \sim 1\,\mu$m. If the terminal stage of a pathway involves gene expression, it would take additional minutes to hours to complete. In this case, from signal activation to the terminal gene expression, a pathway would require a physical process that spans roughly 90% the body size of a mammalian cell at the sub-minute to minute time scale, plus a sequence of events leading to gene expression that happen within the nucleus but requires hours to finish. And remember we haven't taken the time scale for signal activation and the many enzymatic reactions that occur within the pathway into account! In contrast to human-made device like computers and cell phones, the sheer thought of the scales that a cellular signaling system operates on should amaze all!

---

[1] Note that not all signaling pathway leads to gene expression (e.g. those related to apoptosis). The terminology used here is just for convenience

Apart from the separation of space and time scales, signaling and gene expression have other important distinctions. For example, most signaling steps do not expend energy anywhere close that required to synthesize a new protein, usually requiring a few ATPs (whose hydrolysis to ADP leads to Gibbs free energy change at standard condition of roughly $\Delta G \sim -30$kJ/mol) to transmit information from one molecule to the other (Garrett and Grisham, 2010; Lim et al., 2014). Protein synthesis, on the other hand, requires tons of energy (roughly a few cellular energy coins such as ATP, GTP, etc., per amino acid), which, for a typical naturally-occurring protein of roughly 100 or more amino acid in length, is a few hundreds times of that expended at a single 'signaling step. However, this energy is not wasted since gene expression is rather stable and the effects are usually cell-wide, not limited spatially to specific part of a cell. This is probably not surprising since a wholescale change in the cellular state is likely to require the production of new proteins to help the cell adapt to its new environment. Figure 1·2 presents a typical cellular information processing system as a composition of signaling and gene expression subunits. Contrasts between both units are summarized in this figure.



cell signaling

- extracellular-intracelluar communication
- "fast" ON and OFF (seconds to minutes)
- sequential and directional response
- requires "less" energy (no protein synthesis)

gene expression

- "slow" ON and OFF (minutes to hours)
- stable changes (hours to days)
- confined spatial response
- requires "more" energy (transcription and translation)

Figure 1·2: Cells integrate signaling and gene expression to achieve regulation. An exemplary signaling system that combines signaling and gene regulatory unit's is shown as red and green boxes, respectively. Signaling input is assumed to be ligand (L) binding to the receptor that straddles the cell surface depicted as a purple box. This input is transmitted across the cytosol into the nucleus to induce gene expression. Contrast between common features of cell signaling and gene expression is summarized. This figure is adapted from(Lim et al., 2014)

1.2.2 Information is encoded as the state of proteins and processed via the interactions between them

Now let's discuss how information is defined in the cellular context. First let's go back to our favorite example to gain some intuition. For modern digital information processing systems such as computers, inputs are transformed into binary codes interpretable by the machine before central processing unit (CPU), a particular kind of electronic circuitry, carries out specific instructions (e.g. programs) based on these inputs. We can think of the information contained in the input as being relayed across a series of *nodes* or *switches* that *record* these binary states. Through a sequence of changes in the states of these node, input is "processed" into the output that defines state of the last node. If one has a system with many nodes wired together, information can be embedded, process and transmitted in a more complicated manner. Indeed, this is the basis of modern electronic information processing system whose architecture relies heavily on the sophisticated linkage between many transistors that collectively encode a binary representation (i.e. Boolean) of input signal.

On a similar vein, one can imagine that a cell signaling system is encoding the environmental information as the states of some sort. But unlike electronic systems where electrons flowing through the circuit is treated as a fundamental information currency, cell signaling employs multiple information units, depending on the pathway and molecular interactions it underlies (c.f. Figure 1·4). Indeed, there's a wider variety of molecular devices that cell can adopt to read and process different types of signals, see Figure. 1·1. In comparison, most electronic systems utilize one type of architecture that digitize signal before converting it to some output format at the end of processing. This is perhaps not so surprising since the task we subject these electronic systems to is standardized to achieve efficiency, consistency, and uniformity. For tiny beings like cells, the environments they live in can be noisy and unpredictable, therefore, it might be advantageous for them to develop diverse ways to process different information. This brings to the core questions– what is the workhorse of information currencies in cell signaling and what's the molecular basis of that?

As it turns out, proteins plays a dominant role in cell signaling. Proteins are diverse class of biomolecules that constitute more than 50% of the dry weight of cell body. They are essential in virtually every aspects of cell function and behavior, including metabolism, DNA replication, signaling, transportation of molecules, and the support of structure of cells and organisms. This is made possible due to the versatility inherent in proteins, each of which is tailored to its biological role. Of particular interest to signaling is the class of proteins called *enzymes* which act as catalysts that increase the rate at which certain biochemical reaction occurs. As we shall explore more in this thesis, these are the reactions that physically "records" and "modifies", therefore, process information. Simply put, *the information cells collects about the environment is encoded in the states of proteins that are written and erased (or switched back and forth) by specific enzymatic reactions.*

In Figure 1·3, I summarize the idea we discussed in this part by showing that cell signaling is like electronic systems where information is transferred as a series of "switches" or "nodes" (realized by transistors) that are wired to form complex circuitry. The difference is, however, that now these nodes represent proteins bearing some sort of states that can switch between, say, ON (e.g. catalytically active) and OFF (inactive) states through enzymatic reactions and other physical processes. One may naturally ask, what are the mechanism by which state transition happens? Before moving on, I'd like to point out that in addition to proteins, signaling also involves lipids, ions, and other biomolecules. However, these molecules plays mostly a supporting role as they facilitate and assist the main biochemical reactions that change the states of the proteins (i.e. switching the states that encode information). We'll come back to this point in later chapters.

1.2.3 Proteins can change their states through various mechanisms

In Figure 1·4, I illustrate a few common ways that the state of proteins can be changed. For example, they can switch to the ON state (activation) through binding/dissociation with other proteins, a change of their conformations, catalytic reactions catalyzed by their enzymes, or via localization to other parts within the cell.

Of particular importance is the binding and dissociation between proteins or other molecules that facilitate the so-called *post-translational modifications* (PTMs). PTMs are chemical modifications that include the addition or removal of small chemical groups such as phosphate, or big structure such as small regulatory protein ubiquitin (of mass 8.5 kDa), thereby changing the *catalytic properties* of the modified proteins (e.g. enzymatically active or inactive). These modifications are generally performed by specific signaling enzymes such as kinase that catalyzes phosphorylation and phosphatase that *undo* phosphorylations. In a loose sense, we refer the state of proteins as their PTM state (e.g. phosphorylation or methylation state)[2].

Since most signaling proteins are enzymes that catalyze reactions such as chemical modifications of other proteins (e.g. PTMs) and other biomolecules, their activation or inactivation by their upstream signals can lead to widespread downstream effects. For example, the phosphorylation of certain proteins can promote gene expression and protein production, thus driving a massive change in the cellular state. This is in part due to the specificity and efficiency of these proteins as enzymes. Moreover, a change in the enzymatic activity of proteins is usually associated with a change in their conformations. For this reason, these two signaling currencies (i.e. enzymatic activity and conformations) are usually considered together. One important example is the class of proteins that act as molecular switches inside cells, G proteins. These proteins can switch between active and inactive conformations, depending on whether they are bound to GTP or GDP[3] that are needed in the synthesis of RNA during transcription process. Again, since these are associated with changes in catalytic activities, the switching is facilitated by other enzymes such as Guanine nucleotide exchange factors (GEFs) and GTPase-activating proteins (GAPs)[4].

---

[2] Broadly speaking, state can include the conformational or the allosteric state of proteins. But since these different states are usually coupled, here we use this term loosely for convenience.

1.2.4 Protein-protein interaction mediates the currencies of signaling

Previously we discussed the many ways by which proteins can change their states and that such changes define a range of signal currencies (c.f. Figure 1·4). Moreover, in as much as the switching between logic states in electronic circuit determines the outcome of a digital computing system, changes in the protein state link signal inputs to outputs and guides cellular information processing.

By and large, most signaling currencies involve the interactions between proteins and other molecules. For example, binding of a kinase to its substrate can change the substrate's phosphorylation state (therefore the catalytic activity); interactions between protein and peptide or small chemical group can drastically alter protein's conformational (thus the catalytic) state; localization can bring proteins to specific site of the cell where their substrates are easily found, etc.. Figure 1·5 illustrates that protein-protein interactions (PPIs) serves as a mediator to the signaling curries that cells utilize to process information.

One important insight that came out of this observation is that *through PPIs, changes in one aspect of signaling often leads to changes of others*, as we enumerated in the previous paragraph and shown in Figure 1·5. Therefore, one can imagine

---

[3] Usually GTP-bound state is the active state while GDP-bound is inactive

[4] GEFs catalyze the switching from GDP-bound to GTP-bound state while GAPs catalyze the reverse.

PPIs as a knob that coordinates different aspects of cellular information processing whose concerted regulation gives rise to the diverse behaviors that we observe at both the cellular and organismic level. Indeed, this approach is in line with the field of synthetic biology that strives to understand cellular control programs through harnessing the interactions within the molecular circuits.

1.3 What might be a plausible physical theory?

In this chapter, we discussed that the ability to respond to stimuli is one of the defining features of life. Such ability allows living organisms to monitor the conditions of environment and to adjust aspects of their life in the wake of changes in these conditions. For that to happen, their cells have evolved specialized signaling pathway with proteins, lipids, and other biomolecules. We then argued that although cell signaling has to operate on multiple scales in space and time, it is operationally equivalent to information processing carried out by human-made devices such as computers and cell phones. However, unlike these devices that rely on one information currency, namely, the flow of electrons through the circuit, cells utilize multiple currencies that are coupled to each other by protein-protein interaction (c.f. Figure 1·4 and Figure 1·5). By doing so, cells develop a wide range of "molecular devices" that can relate different types of signals to different outputs, thereby implementing a higher level of regulation and coordination to function as a integrated organism.

From the physics point-of-view, it is natural to ask what's the level of abstraction based on the biology of signaling we've discussed so far would allow us to build a physical model to further our understanding? Particularly, what are the ingredients such as degrees of freedom, time scale, interactions etc relevant to cell signaling that would culminate in theory verifiable through quantitative experimentations? Here I argue that since PPIs mediate various mechanisms of cell signaling, a plausible physical theory about cellular information processing should incorporate some aspects of these interactions. To do so and to account for the fact that both the intracellular and extracellular environments are thermal in nature, I adopt a statistical physics approach, namely, starting from a statistical description of PPIs that lands at predictions on macroscopic and measurable quantities. However, to explore notions like information processing requires tools from disciplines apart from physics such as information theory and complexity theory. Therefore, a plausible physical theory should be able to bridge these tools with the statistical physics that aims at characterizing the interactions relevant in signaling. This goal epitomizes the model and methodology I developed in this thesis.

1.4 Organization of thesis

Finally, based on the biology and the provision for a physical model we discussed in this chapter, I summarize the questions I attempt to address in this thesis:

- What's the minimal model based on protein-protein interactions that would allow us to relate the inputs to the outputs of a given signaling pathway?
- How do these interactions give rise to a quantification of information transmission across a signaling network?
- What's the limit to cell's ability to transmit information?
- Can we harness these interactions to design signaling pathways that implement desired input-output relations or computations?
- What's the theoretical underpinning of cellular computations and how similar they are to those done by human-made electronic devices?

- Is is possible to define the capacity to cell's ability to perform computations?

At the end of this thesis, I discuss how our answers to these questions culminate in a deeper understanding of cellular information processing through signaling, and discuss their implications to synthetic biology whose ultimate goal is to engineer molecular building blocks to generate novel cellular control programs.

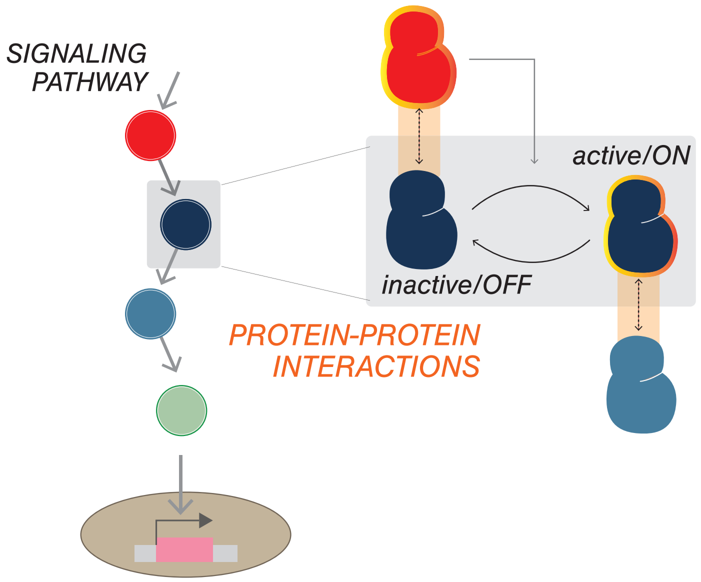


Figure 1·3: Information is encoded in the state proteins and further processed via protein-protein interactions. A generic signaling pathway is depicted on the left where nodes represent proteins while directed edges illustrate their interactions (e.g. binding/dissociations, enzymatic reactions). Shown on the right is the biophysical representation of a node (in this case a protein). A node in the pathway can be thought of as a switch that can change between ON (usually catalytically active) and OFF (catalytically inactive) states, leading to the change of output. Such state switching is dictated by the protein-protein interactions. Conceptually, information is encoded as the ON and OFF states of these nodes. Note that more than two states are possible for each node. Since the switching between states typically involves protein-protein interactions, one can imagine it as a mechanism by which information is processed.

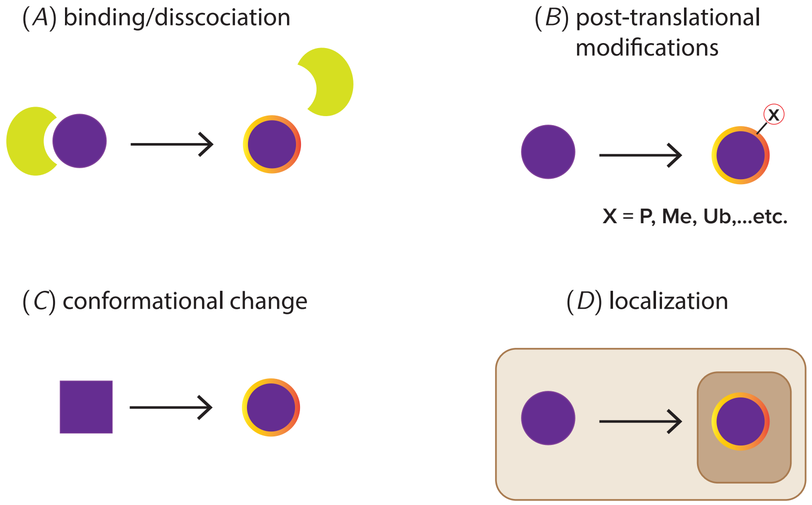

(A) binding/disscociation
(B) post-translational modifications
X
X = P, Me, Ub,...etc.
(C) conformational change
(D) localization

Figure 1·4: Information currencies in cell signaling. Unlike electronic systems, cell signaling employs different currencies to process and transmit information. The purple species indicates the protein that can switch between ON (orange-red haloed circle) and OFF states, as depicted in Figure 1·3. Such switching (or writing and erasing) can be manifested a variety of ways such as binding and dissociation, post-translational modifications, conformational changes, and localization to different part of the cell. In (B), the modification is shown as X with $X = \text{P, Me, Ub}$ indicating phosphorylation, methylation, and ubiquitination, respectively.

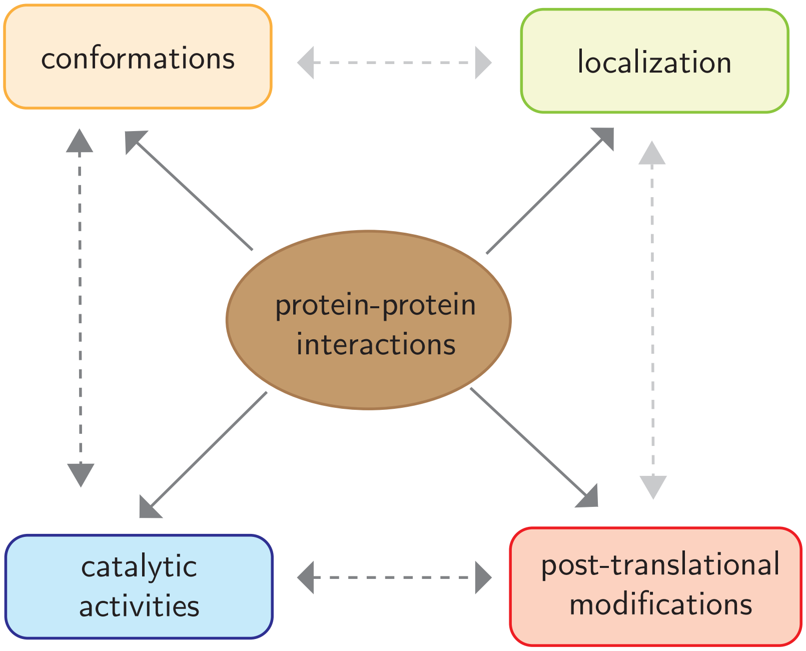

conformations
localization
protein-protein interactions
catalytic activities
post-translational modifications

Figure 1·5: Protein-protein interactions mediate various mechanisms of information processing in cell signaling. Interactions between proteins can change various aspects of cell signaling (indicated by solid arrows) such as protein's conformation, post-translational modifications, and catalytic activities. They may also localize proteins to specific site within the cell such as nucleus to allow for DNA-protein interactions. These aspects (colored boxes), regarded as the currencies of cell signaling, can directly or indirectly affect one another (dashed arrows).

Chapter 2

What controls the flow of information across signaling network?

> *'I formerly defined the possible as that which in a given state of information (real or feigned) we do not know not to be true. But this definition today seems to me only a twisted phrase which, by means of two negatives, conceals an anacoluthon. We know in advance of experience that certain things are not true, because we see they are impossible. '* (Charles Sanders Peirce, 1897)

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 suggest 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.

2.1 Introduction

Cells have evolved complex protein signaling networks to process information about their living environments (Barabasi and Oltvai, 2004; Blais and Dynlacht, 2005; MacArthur et al., 2009; Martello and Smith, 2014). These networks play a central role in cellular decision-making, development, growth, and migration (Seet et al., 2006; Scott and Pawson, 2009; Lim et al., 2014). In eukaryotic cells, signaling pathways such as Wnt/β-Catenin(Angers and Moon, 2009; MacDonald et al., 2009) and TGF-β pathways(Massagué, 2012) have important homeostatic functions (e.g., cell proliferation, differentiation, and fate determination), with disruptions in their signaling leading to tumorigenesis and drive metastasis (Anastas and Moon, 2013; Moustakas and Heldin, 2014).

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(Deribe et al., 2010; Scott and Pawson, 2009). By linking an activity to a substrate through binding, PTM-mediated PPI interactions serve as signaling network links by interconnecting writer/eraser cycles (Fig. 2·1A). 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. 2·1A,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 (Ackers et al., 1982; Hill, 2013; Weinert et al., 2014). Models of this class have been successfully used to understand the biophysics of promoter regulation in transcriptional networks(Bintu et al., 2005b; Kinney et al., 2010; **?**; Weinert et al., 2014). Here, we extend this approach to signaling networks by introducing variables representing PTM-dependent PPIs, thereby accounting for the non-equilibrium nature of reversible, enzyme-catalyzed phosphorylation. We combine this statistical physics approach with information theory(Shannon, 2001;

Cover and Thomas, 2012), which has seen widespread recent application in biology(Johnson, 1970). Examples include the modeling of information processing in gene networks(Tkačik et al., 2009; Walczak et al., 2010; Tkačik and Walczak, 2011; Granados et al., 2018), enzyme cascades (Detwiler et al., 2000), and bacterial signaling networks(Mehta et al., 2009; Tostevin and Ten Wolde, 2009), as well as calculating information capacity in canonical eukaryotic signaling networks from single cell measurement of input-output relationships(Cheong et al., 2011; Brennan et al., 2012).

Our joint a 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(Ladbury and Arold, 2012; Voliotis et al., 2014), often in combination with low levels of background kinase and phosphatase activity(Chung et al., 2010; Schlessinger, 2000).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(Hill, 1998; Schwartz and Ginsberg, 2002; Hunter, 2007; Voliotis et al., 2014; Kontogeorgaki et al., 2017)? 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.

The paper is organized as follows. We begin by discussing how to generalize thermodynamic models for binding to include PTMs. We then discuss how the basic elements of these models can serve as an input into information theoretic calculation. 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 such noise can be significantly decreased by increasing the strength of PPIs at the expense of a slower dynamical response, a biophysical manifestation of what in engineering is often called the “gain-bandwidth” tradeoff. We then show that cross-talks between pathways in complex signaling networks does not significantly alter information capacity. We conclude by discussing a new procedure ”InfoMax” for designing synthetic biochemical networks.

2.2 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 2·1A). 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(Bintu et al., 2005b; Kinney et al., 2010; Garcia et al., 2010; Weinert et al., 2014) 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 2·1C); 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 2·1C into three parts and enumerating the possible states within each. As depicted in Figure 2·1A,C, the receptor kinase only has two possible PTM states (phosphorylated or not). We label the probability of phosphorylated receptor kinase as $P(R = 1) = q$, where $q \in [0,1]$ is the parameter that encapsulates ligand activation. The probability of the complementary configuration is therefore given by $P(R = 0) = 1 - q$. (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 2·1D. 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.*

$$P(K_1 = 1 | R = 1) = \frac{e^{-\beta\theta_{R,K_1}}}{1 + e^{-\beta\theta_{R,K_1}}}, \tag{2.1}$$

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 Appendix A for its expression in terms of kinetic parameters.) By conservation, the complementary configuration has probability $P(K_1 = 0|R = 1) = 1 - P(K_1 = 1|R = 1) = 1/(1 + \exp(-\beta\theta_{R,K_1}))$. The cases where receptor is not phosphorylated (i.e. $R = 0$) are similar except that the binding affinity is now denoted as W, which is assumed to be 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.*

$$P(K_1 = 1 | R = 0) = \frac{e^{-\beta W}}{1 + e^{-\beta W}} \approx 0, \tag{2.2}$$

which implies $P(K_1 = 0|R = 0) = 1 - P(K_1 = 1|R = 0) \approx 1$. 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. 2·1D). (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 $P(TF|K_1)$ in a similar fashion by relating θ_{R,K_1} to $\theta_{K_1,TF}$ (see Fig. 2·1D.)

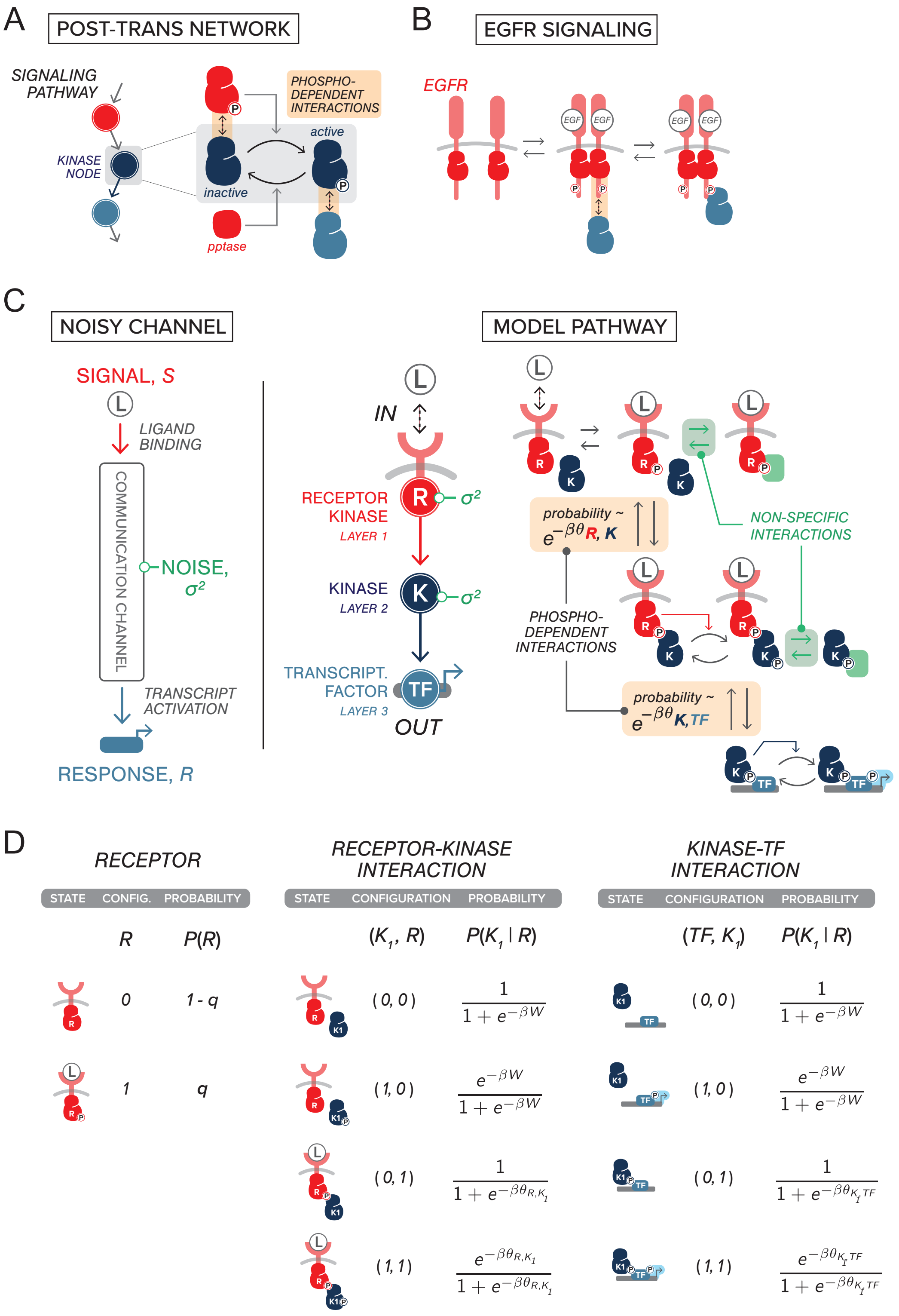
2.3 Mutual information and PPIs

Mutual information between two random variables measures how much knowing one tells us about the other, usually measured in units of bits(Shannon, 2001; Cover and Thomas, 2012). In biology, it has been widely used to characterize the information transfer by biochemical systems (Johnson, 1970; Detwiler et al., 2000; Tkačik et al., 2009; Mehta et al., 2009; Walczak et al., 2010; Tkačik and Walczak, 2011; Cheong et al., 2011; Brennan et al., 2012). 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,

$$I(R;TF)=\sum_{R}\sum_{TF}P(R)P(TF|R)\log_2\left[\frac{P(TF|R)}{P(TF)}\right]. \tag{2.3}$$

Note that since the summations in Eq.(2.3) are over $\{0,1\}$, this signaling network represents a discrete (binary) channel(Cover and Thomas, 2012). Physically speaking, $P(TF)$ quantifies the transcriptional readout, $P(TF|R)$ defines the input-output relation (i.e., channel transfer function), and $P(R)$ 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. 2·1C and detailed in Fig. 2·1D, all quantities in Eq.(2.3) can be explicitly calculated: signal input $P(R)$ is given in Fig. 2·1D while the channel input-output relation (i.e. transfer function), $P(TF|R)$, is obtained by first invoking the conditional independence of TF and R on K_1, then marginalizing contributions from K_1, *viz.* $P(TF|R)=\sum_{K_1}P(TF|K_1)P(K_1|R)$. Finally, the output is simply


A
POST-TRANS NETWORK
SIGNALING PATHWAY
KINASE NODE
P
PHOSPHO-DEPENDENT INTERACTIONS
active
inactive
pptase
B
EGFR SIGNALING
EGFR
EGF
C
NOISY CHANNEL
SIGNAL, S
L
LIGAND BINDING
COMMUNICATION CHANNEL
NOISE, σ^2
TRANSCRIPT ACTIVATION
RESPONSE, R
IN
RECEPTOR KINASE
LAYER 1
R
σ^2
KINASE
LAYER 2
K
TRANSCRIPT. FACTOR
LAYER 3
TF
OUT
MODEL PATHWAY
NON-SPECIFIC INTERACTIONS
probability ~ $e^{-\beta\theta_{R,K}}$
PHOSPHO-DEPENDENT INTERACTIONS
probability ~ $e^{-\beta\theta_{K,TF}}$
D
RECEPTOR
RECEPTOR-KINASE INTERACTION
KINASE-TF INTERACTION
STATE
CONFIG.
CONFIGURATION
PROBABILITY
K1

STATE	CONFIG. R	PROBABILITY $P(R)$
	0	1 - q
	1	q

STATE	CONFIGURATION (K_1, R)	PROBABILITY $P(K_1 \mid R)$
	(0, 0)	$\frac{1}{1+e^{-\beta W}}$
	(1, 0)	$\frac{e^{-\beta W}}{1+e^{-\beta W}}$
	(0, 1)	$\frac{1}{1+e^{-\beta\theta_{R,K_1}}}$
	(1, 1)	$\frac{e^{-\beta\theta_{R,K_1}}}{1+e^{-\beta\theta_{R,K_1}}}$

STATE	CONFIGURATION (TF, K_1)	PROBABILITY $P(K_1 \mid R)$
	(0, 0)	$\frac{1}{1+e^{-\beta W}}$
	(1, 0)	$\frac{e^{-\beta W}}{1+e^{-\beta W}}$
	(0, 1)	$\frac{1}{1+e^{-\beta\theta_{K_1 TF}}}$
	(1, 1)	$\frac{e^{-\beta\theta_{K_1 TF}}}{1+e^{-\beta\theta_{K_1 TF}}}$

Figure 2·1: (*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 epidermal growth factor receptor (EGFR) signaling, binding of EGF to the extracellular domain of EGFR leads ot 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.

given by $P(TF) = \sum_R P(TF|R)P(R)$. Explicitly, the transfer function is given by:

$$P(TF=1|R=1) \approx \left(\frac{e^{-\beta\theta_{K_1,TF}}}{1+e^{-\beta\theta_{K_1,TF}}}\right)\left(\frac{e^{-\beta\theta_{R,K_1}}}{1+e^{-\beta\theta_{R,K_1}}}\right) \tag{2.4}$$

$$P(TF=1|R=0) \approx 0 \tag{2.5}$$

$$P(TF=0|R=1) \approx \left(\frac{1}{1+e^{-\beta\theta_{K_1,TF}}}\right)\left(\frac{e^{-\beta\theta_{R,K_1}}}{1+e^{-\beta\theta_{R,K_1}}}\right) + \left(\frac{1}{1+e^{-\beta\theta_{R,K_1}}}\right) \tag{2.6}$$

$$P(TF=0|R=0) \approx 1, \tag{2.7}$$

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

$$P(TF=1) \approx \left(\frac{e^{-\beta\theta_{K_1,TF}}}{1+e^{-\beta\theta_{K_1,TF}}}\right)\left(\frac{e^{-\beta\theta_{R,K_1}}}{1+e^{-\beta\theta_{R,K_1}}}\right) q \tag{2.8}$$

$$P(TF=0) \approx \left[\left(\frac{1}{1+e^{-\beta\theta_{K_1,TF}}}\right)\left(\frac{e^{-\beta\theta_{R,K_1}}}{1+e^{-\beta\theta_{R,K_1}}}\right) + \left(\frac{1}{1+e^{-\beta\theta_{R,K_1}}}\right)\right] q + (1-q) \tag{2.9}$$

With all these at hand, we can express Eq. (2.3) as a function of BAs $\theta_{i,j}$. In Appendix A, we provide the analytic expression of mutual information Eq.(2.3) in terms of BAs. We have thus established an explicit functional relation between mutual information and PPIs.

2.4 Results

2.4.1 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. 2·2A). 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) (Detwiler et al., 2000; Cover and Thomas, 2012). 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 inherent 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 Appendix A 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 Appendix A for full derivations). When proteins bind tightly (i.e. large negative binding energies $\beta\mu \lesssim -1$), the SNR for the simplest signaling network $L \rightarrow R \rightarrow TF$ reduces to the following simple expression:

$$\mathrm{SNR} \equiv \frac{\langle Q(L=1)\rangle}{\langle Q(L=0)\rangle} = e^{\beta W}\left[1 - e^{\beta\left(\mu + \frac{\sigma^2}{2}\right)}\right] \tag{2.10}$$

For networks with n-layers of kinase between input R and output TF, as depicted in Figure 2·2A, we plot the color map of their log-SNR at different level of specific and non-specific PPIs in Figure 2·2B. 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. $P(TF = 1)$ as a function of $P(R = 1) \equiv q$) in Figure 2·2C both at tight- and weak-binding. As shown, networks with strong BAs always have a larger gain, implying a higher information capacity(Cover and Thomas, 2012; Detwiler et al., 2000). 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 Appendix A 1, 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 Appendix A 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 Appendix A 2 Figure A·2). Surprisingly, for weak binding we find numerically and analytically that $q^\star \leq 0.5$ (see Appendix A 2 Figure A·5).

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 (Ladbury and Arold, 2000; Ladbury and Arold, 2011). 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 2·3A into a kinetic model. Note that the thermodynamic model presented in Fig. 2·1D can be explicitly derived from the kinetic formulation. Here we invoke this duality to investigate both the signaling dynamics through kinetic

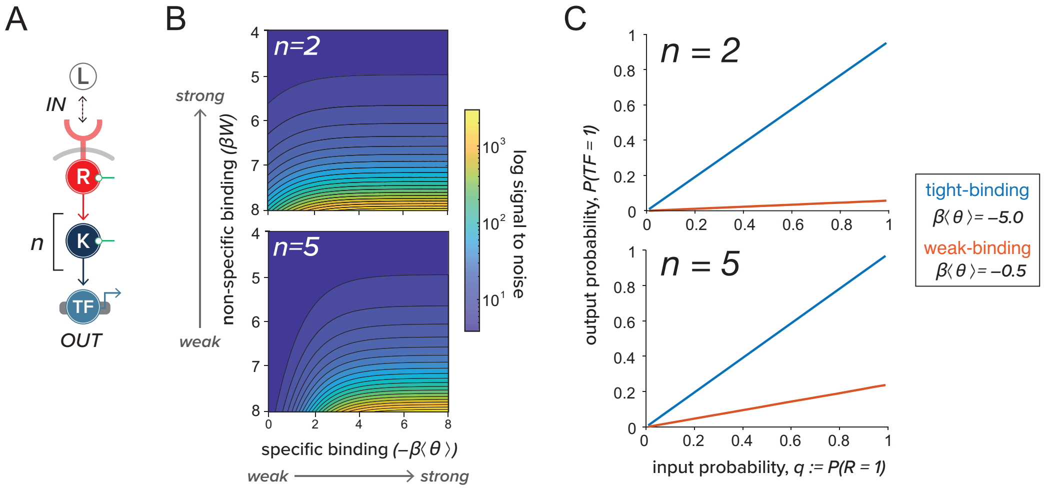


Figure 2·2: Noise due to non-specific protein-protein interactions (PPIs) limits the quality of information transmission. (*A*) A simple linear network that mediates information of input (L) through a *n*-layer kinase cascade (K_i, $i = 1, 2, \cdots, n$), to an output transcription factor (TF) which is active when phosphorylated. As in Figure **??** *C*, green circles indicates noise. (*B*) Color map shows the numerically simulated log-signal-to-noise ratio (log-SNR), defined by Eq.(2.10), 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 0.01 (see main text for simulation details). This quantity can also be obtained by solving Eq.(A.9) (see Appendix). In this panel, we show the result for $n = 2, 5$. (C) Input-output relation of the *n*-layer kinase cascade ($n = 2, 5$) at tight- and weak-binding is shown.

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. 2·1C, 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 Appendix A, 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 $P(K_i = 1)$ 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 2·3A). We characterized the response times by measuring the time τ it took the output to reach a new steady-state. We repeated this procedure for binding affinities drawn from distributions with different means $\langle\theta\rangle$. In Figure 2·3B, 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 2·3C).

This "speed-information" trade-off can be viewed as a biophysical manifestation of the gain-bandwidth tradeoff (Detwiler et al., 2000). 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.

2.4.2 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 2·4A). In this scheme, cross-talk refers to interactions where proteins in one pathway activate those in the other (i.e., dashed lines in Figure 2·4A). 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 $P(R_1, R_2)$ – and calculated the mutual information, $I(\{R_1, R_2\}; \{TF_1, TF_2\})$ between all the inputs outputs, (see Figure 2·4B 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 2·4C and Appendix A 6 Figure A·8 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.

2.4.3 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 (Levy et al., 2010; Breitkreutz et al., 2010). 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 affnity is known (c.f. Eq.(2.3) and Figure 2·1D), this procedure can be done

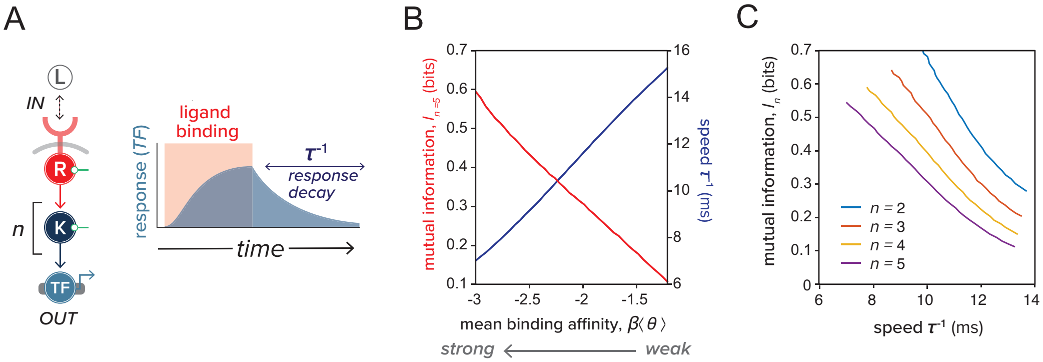


Figure 2·3: Specificity in PPIs mediates the information-speed trade-off. (*A*) A non-zero constant input (ligand binding of duration indicated in red) is administered to the network shown on the left. This signal turns on the output (TF) to its steady-state before switching off. The speed of response 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.

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 Methods for detailed description). 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 two-input-two-output networks where we systematically varied network depth (total number of nodes per layer, n in Figure 2·5A) and width (number of nodes per layer, n_w in Figure 2·5B). 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. 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 (Cover and Thomas, 2012; Kinney and Atwal, 2014)(see Figure 2·5A and Appendix A 4 Figure A·7). For the optimal solution found, mutual information almost saturates at the 1 bit limit, and deviation from this optimal, say, with weak-binding interactions for any given layer in the kinase cascade, will substantially diminish information transmission (see Figure 2·5A). In contrast, we found that increasing the width of the intermediate network can increase information transmission modestly for small widths. As seen in Figure 2·5B, 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 a PPI-design strategy crucial for networks with complex topology. For one input, one output networks, we have shown previously that tight-binding between proteins (i.e. specific interactions) helps transmit information. However, for MIMO networks with wide intermediate layers (e.g. those shown in Figure 2·5B with large n_w), such intuition is not straightforward, especially when inputs are correlated. This is illustrated in the solution found through InfoMax in Figure 2·5B. For example, a subset of cross-layer PPIs are substantially stronger than others, and such combination of BAs collectively achieves maximum mutual information.

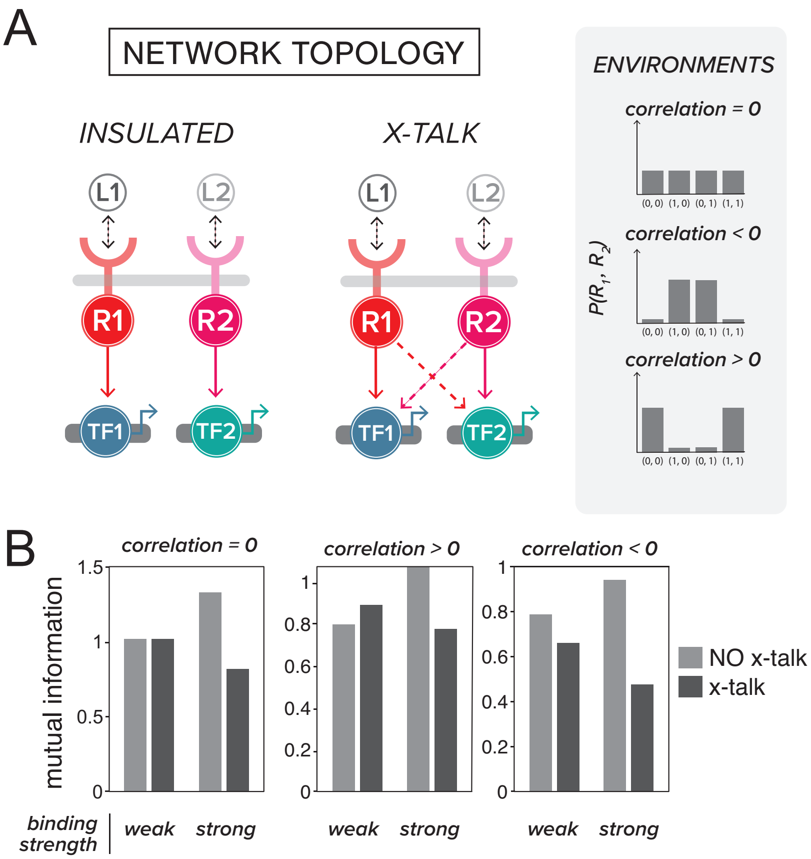


Figure 2·4: Effect of cross-talk between pathways on information capacity. (*A*) Schematic of insulated and cross-talked networks are shown. Dashed connections represents cross-talks between pathways. Three environmental conditions that differ in terms of the correlation between the two signals they provide to the network (left) are illustrated: inputs with zero, negative, and positive correlation (see Appendix A 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 the binding affinities for protein-protein interaction (weak/tight-binding corresponds to $\beta\langle\theta\rangle = -5/-1$), and colors indicate the presence or absence of cross-talks. In the 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 Appendix A 4 for more details.)

2.5 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 (Detwiler et al., 2000; Tkačik et al., 2009; Mehta et al., 2009; Walczak et al., 2010; Tostevin and Ten Wolde, 2009; Cheong et al., 2011; Brennan et al., 2012). 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 obtained in (Tareen et al., 2018) 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 Appendix A 4 Figure A·7). This is consistent with previous work on MAP kinase cascade (Detwiler et al., 2000), 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(Breitkreutz et al., 2010). Binding affinity clustering was particularly pronounced for the kinase/phosphatase catalytic domains that mediate phosphorylation-dependent binding(Mok et al., 2010). 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 (Ladbury and Arold, 2012).

Another intriguing observation from the yeast kinase-phosphatase interactions is the existence of extensive cross-talk between signaling pathways (Breitkreutz et al., 2010) that the authors describe a 'collaborative network of interactions' – a topology that suggests a distributed cellular decision-making strategy (Levy et al., 2010). 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(Levy et al., 2010). 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(Fischbach et al., 2013). 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 (Bashor and Collins, 2018). 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.

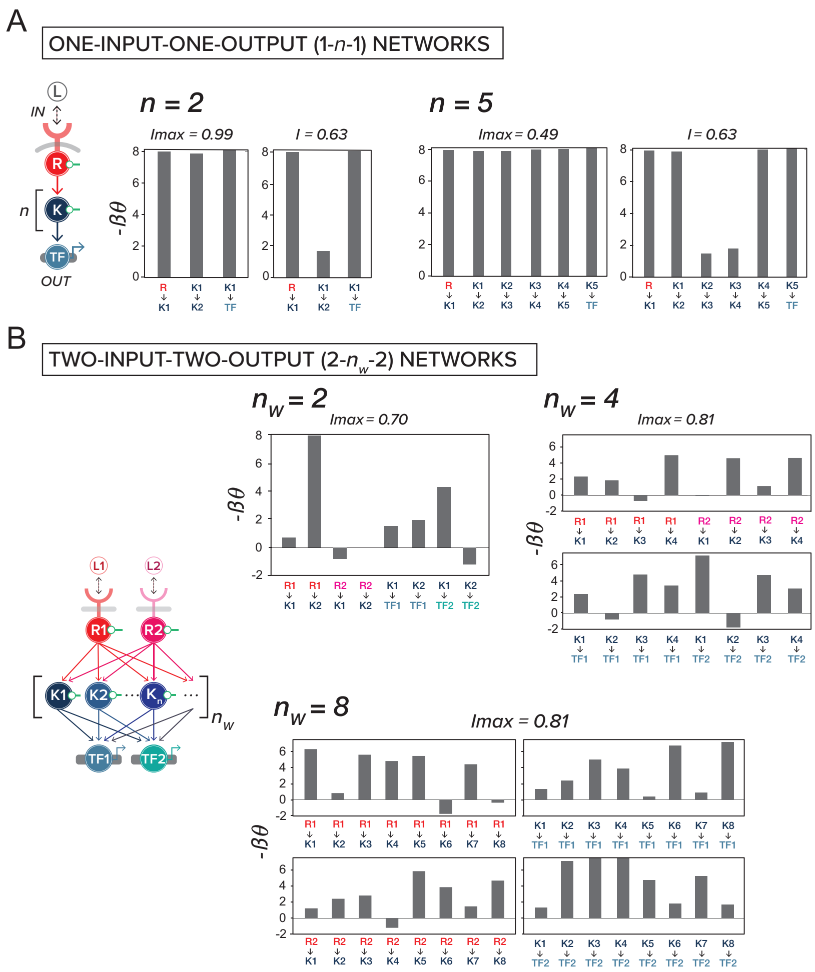


Figure 2·5: InfoMax design finds the PPIs that maximize information transmission. (*A*) InfoMax is applied to networks with one input, n layers of kinase, and one output. For $n = 2$ and 5, the bar graph on the left shows the binding affinities that give the maximum mutual information (indicated as I_{max} 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 $R \rightarrow K_1$ 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 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 $[-2, 8]$.

Chapter 3

Can we design signaling network to implement desired functions?

> *"If one wants to make a machine mimic the behaviour of the human computer in some complex operation one has to ask him how it is done, and then translate the answer into the form of an instruction table. Constructing instruction tables is usually described as 'programming.'"* (Alan Turing, 1950)

It is known that living cells use complex biochemical networks to perform sophisticated computational tasks. Yet, a major question in synthetic and systems biology remains: How are network level computational properties encoded in the biophysics of protein-protein interactions (PPIs)? We address this question by developing a new computational framework relating the thermodynamics of PPIs to network signaling properties in a simplified synthetic biochemical system inspired by the "reader-writer-eraser" signaling paradigm (RWE paradigm). We show with our framework that networks engineered to have identical steady-state decision surfaces can exhibit very different dynamical behaviors. More generally, our work shows that complex computational and information processing tasks can be programmed in cellular signaling circuits by manipulating biophysical parameters.

3.1 Introduction

One of the defining features of living organisms is the ability to sense and respond to environmental signals by invoking internal responses (i.e. decisions), which often leads to changes in gene expression or phenotypic properties. At the cellular level, this ability, often termed cell signaling, forms the basis of development(Kuchroo et al., 1995; Gerhart and Kirschner, 1997), immunity(Taganov et al., 2006), and tissue homeostasis(Pasparakis, 2009). In the past few decades, there has been a large effort to characterize and understand the basic principles governing cellular signaling (Koch et al., 1991; Pawson and Nash, 2003; Lim and Pawson, 2010; Lim et al., 2014; Hunter, 2000; Selimkhanov et al., 2014; Detwiler et al., 2000), and to use this knowledge to engineer novel signaling networks to implement complex decisions(Kiel et al., 2010; Lim, 2010; Khalil and Collins, 2010; Morsut et al., 2016; Roybal et al., 2016; Bashor et al., 2010; Lim et al., 2014). Despite these considerable successes, our understanding of how biophysical properties – such as the strength and types of protein-protein interactions (PPIs) – shape and constrain the cellular decisions via signaling remains limited(Dayarian et al., 2009; Letsou and Cai, 2016).

Cellular signaling networks are composed of multiple specialized proteins interacting with one another. These proteins often exist in multiple internal states (e.g. phosphorylated or unphosphorylated; allosteric conformations etc.) that function as the biochemical correlates of information. From the point of view of information processing and computation, the internal states of proteins (e.g. the phosphorylation state) play an analogous role in biochemical circuits to the voltage state (high or low) of a memory cell in modern computers. For example, a protein that bears two internal states can in principle store one bit of information. Just as computers perform computations by manipulating memory states through physically wired circuitry, biochemical signaling networks perform computations by modifying the internal states of proteins through PPIs. Thus, to be able to perform complex computations biochemical signaling networks must be able to store many bits of information and selectively and reliably change the internal states of proteins in the network.

These observations suggest that cellular decisions through signaling is intimately related to the structural organization of signaling proteins. Over the last few decades, it has become clear that many signaling proteins in eukaryotes have a modular structure. Proteins often possess multiple structural domains with distinct functional roles. Domains can be divided into two broad categories: interaction domains, which mediate interactions with other molecules within the cell (Koch et al., 1991; Pawson and Nash, 2003) and catalytic domains, which catalyze reversible enzymatic reactions through post-translational modifications (PTMs) such as phosphorylation or methylation (Jin and Pawson, 2012; Lim et al., 2014).

Even with the functional distinction between domains from different categories, they often work in tandem to regulate the catalytic activity of proteins through allosteric control. In contrast with conventional allostery where the catalytic activity and binding interactions of a protein are tightly coupled into a single structural unit(Cui and Karplus, 2008; Monod et al., 1965), the emerging ”modular allostery” paradigm suggests that in many signal proteins there exists a structural separation between the output catalytic and regulatory domains(Lim, 2002; Taylor and Radzio-Andzelm, 1997). Those that implement modular allostery usually contain catalytic domains that, when isolated, display constitutive activity but are otherwise auto-inhibited by other regions of the protein through steric blocking or conformational contortion. Indeed, this modular domain structure confers a wide variety of PPIs and allosteric control, which essentially provide the underlying

”biophysical hardware” cells can utilize to carry out computation and decision making.

Synthetic biologists have started to exploit these insights to start engineering sophisticated synthetic signaling networks. One of the prominent principles exploited in recent synthetic systems is the RWE paradigm(Bashor et al., 2010; Lim et al., 2014; Khalil and Collins, 2010). Writer proteins such as kinase or Guanine nucleotide exchange factor (GEF) enzymatically catalyze the transfer of chemical marks onto target molecules, whereas eraser proteins (e.g. phosphatase or GTPase-activating proteins, GAPs) catalyze the removal of the chemical mark. The presence of a mark is then read out by a reader module, usually a binding motif that recognizes and binds the modified chemical mark.

Inspired by these ideas, in this chapter we present a new theoretical framework to explicitly relate the thermodynamics of PPIs to cellular decisions carried by signaling networks. In this chapter, we consider synthetic-biology inspired circuits based on the reader-writer-eraser signaling paradigm. For concreteness, most of our examples focus on circuits that utilize phosphorylation cascades where the phosphorylation of an enzyme can modify its catalytic activity.

A novel feature of our framework is that we can systematically “program” network level computational and information processing properties by an appropriate choice of microscopic biophysical parameters (e.g. protein concentrations, binding energies, Michaelis-Menten constants). We also develop a method for examining the sensitivity of network level computational properties to changes in these biophysical parameters. Importantly, these biophysical properties can be easily manipulated in experiments allowing for the rationally-design synthetic circuits with a desired computational process. We concentrate on two types of design problems: designing synthetic circuits to implement a given multi-input multi-output (MIMO) steady-state response function (i.e. MIMO decision surfaces) and finding networks that optimize information transmission (InfoMax). The first of these is closely related to programming computation and the latter to information transmission. Though obviously related, the two objectives are different and optimize different objective functions since making decisions necessarily involves the loss of information (Mehta and Schwab, 2012; Mehta et al., 2016; ten Wolde et al., 2016).

The chapter is organized as follows. We begin by introducing a thermodynamics model for the steady-state properties of PPI networks. Note that although similar type of analysis has been applied to the context of transcriptional network and enzymatic reactions(Bintu et al., 2005c; Bintu et al., 2005a; Einav et al., 2016; Razo-Mejia et al., 2017), here we focus on how to properly frame it in the context of signaling and to incorporate the non-equilibrium nature of catalytic reactions into the network level (Mehta et al., 2016). After that we introduce a methodology to systematically learn how to implement decisions through experimentally accessible "parameters". We then expatiate the notion of decisions in terms of computation and information, with the former focusing on how to faithfully realize a target input-output relation (I-O relation) while latter emphasizing how to reliably transmit information. We then apply our framework to examine how commonly occurring signaling motifs such as coherent and incoherent feedforward loops and two-input, two-output decision surfaces and show that it is possible to tune biophysical parameters to have the same steady-state decision surface, but very different dynamical behaviors. We then discuss the implications of our work through the lens of synthetic biology, and discuss experimental realizations. We then conclude with a discussion of the implications of our analysis for understanding the more, complex, naturally occurring in eukaryotic cell signaling.

3.2 Results

3.2.1 Modular allostery and the "reader-writer-eraser" paradigm

Cell signaling relates extracellular cues (i.e. inputs) into intracellular responses. To model this process, we need an extended computational framework that combines both the thermodynamics of protein-protein interactions and a model of how the catalytic activity of proteins can change the post-translational state of a protein. Inspired by naturally occurring networks (see Figure 3·1A), we focus on phosphorylation cascades with two or more different protein kinases. For example, in the ERK pathway shown in the left panel of Figure 3·1A epidermal growth factor receptor (EGFR) activation (i.e. input) facilitates the phosphorylation of extracellular signal-regulated kinase (ERK), which in term drives the phosphorylation of its downstream effector ribosomal protein S6 kinase (RSK). Phosphorylated RSK then catalyzes the phosphorylation of its downstream serum response factor (SRF), which then promotes the transcription of FOS mRNA (i.e. output).

Natural circuits, such as the one described above, often posses complex proteins that can undergo complex allosteric changes in response to post-translational modifications. This significantly complicates efforts to both model and build synthetic systems. For this reason, we focus on simple synthetic systems based on the "reader-writer-eraser" signal relay paradigm(Bashor et al., 2010; Lim et al., 2014; Khalil and Collins, 2010). In our system, proteins are engineered to have modular allostery.

As in another other signaling systems, the catalytic activity of signaling proteins is not only affected by its own activator (e.g. the kinase that phosphorylates the proteins) , but also by an opposing regulator called a phosphatase that desphosphorylates the protein. Through out, we assume that phosphatases are constitutively expressed, resulting in a constant background dephosphorylation rate.

In the absence of a kinase, proteins are quickly desphosphorylated and a protein must be bound to an active kinase in order to be phosphorylated.

3.2.2 A thermodynamic model combining protein interactions with post translational modifications

We now discuss how we mathematically model such a systems. Concretely, let n be the number of distinct kinases in the protein kinase interaction network that can undergo phosphorylation and dephosphorylation, and let x_i, $i = 1, \cdots, n$ be the binary variable indicating the phosphorylation state of protein kinase i with $x_i = 1$ and 0 being phosphorylated and unphosphorylated, respectively. Due to the modular domain composition, these kinases are assumed to interact in a phosphorylation dependent manner. When a protein kinase is phosphorylated, it can bind its cognate binding partners (i.e. other kinase proteins) and change their phosphorylation state. Conversely, the phosphorylation state of a protein kinase can be modified if it is bound by another phosphorylated kinase. Since all molecules can only bind and modify the phosphorylation state of another kinase only if they are themselves phosphorylated and that their interactions are cognate, we will use the shorthand of calling phosphorylated kinase active ($x_i = 1$) and unphosphorylated kinase inactive ($x_i = 0$).

Next we can combine the catalytic activity and binding interactions of kinases into a single generalized thermodynamic model. It's worth noting that similar models have been proposed to help understand transcriptional regulation (Bintu et al., 2005c; Bintu et al., 2005a; Phillips et al., 2012). In particular, this type of analysis has been shown that for a wide class of regulatory architectures one can derive a set of "governing equations" that relates fold change in gene expression to physically tunable regulatory parameters(Garcia and Phillips, 2011). However, cell signaling is at many levels more complex than transcriptional regulation, in part due to the intricate PPIs that are coupled to enzymatic reactions. From thermodynamics standpoint, this implies that underlying process is often driven from the equilibrium, therefore, enzymatic steady state need not follow Boltzmann distribution(Einav et al., 2016). Here we show (see Appendix B) that in our framework the "probability" of enzymatic activity at steady state shares similar functional form as that derived from the Boltzmann formalism. Indeed, this was also discovered in simple repression models built on specific transcriptional circuits(Phillips, 2015; Razo-Mejia et al., 2017). Note that simple systems with kinase cascade was considered in (Heinrich et al., 2002) but via the conventional chemical kinetics approach.

We begin by introducing a *effective binding energy matrix* $\Delta\epsilon$ whose ji-th element encodes the change of "effective binding energy" when active kinase j binds i, thereby rendering i catalytically active (i.e. phosphorylated). Concretely, $\Delta\epsilon_{ji}$ is related to the Michaelis constant of the enzymatic reaction of i phosphorylation by j, $K_M^{(ji)}$, through $\Delta\epsilon_{ji} = k_B T \log K_M^{(ji)}$, where k_B is the Boltzmann constant and T is temperature (see Figure 3·1B for a table summary and Appendix B for details). In addition, we denote $[c_i]$ ($[\tilde{c}_i]$) as the concentration of protein kinase i at its phosphorylated (unphosphorylated) form. Let $P(x_1 = 1|\mathbf{x}_{\setminus i}, \boldsymbol{\theta})$ be the *steady-state* probability of kinase i being in phosphorylated form, given the phosphorylation states of other kinases in the system $\mathbf{x}_{\setminus i}$, and the physical parameters $\boldsymbol{\theta} := \{\Delta\boldsymbol{\epsilon}, [\mathbf{c}], [\tilde{\mathbf{c}}]\}$. Heuristically, one can express it in a suggestive Boltzmann form:

$$P(x_i = 1|\mathbf{x}_{\setminus i}, \boldsymbol{\theta}) = \frac{\sum_{j \neq i} [c_j] x_j e^{-\Delta\epsilon_{ji}/(k_B T)}}{1 + \sum_{j \neq i} [c_j] x_j e^{-\Delta\epsilon_{ji}/(k_B T)}}.. \tag{3.1}$$

In writing this we have set the weight of the unbound state to 1, and have made use of the assumption that j can bind its cognate partner only if it is phosphorylated (i.e. $x_j = 1$). We also assume that the probability for i to be phosphorylated is directly proportional to the probability to be bound by a kinase, which is simply the weighted Boltzmann factor. This, together with the condition that $[c_i] = P(x_i = 1|\mathbf{x}_{\setminus i}, \boldsymbol{\theta})C_i$, where $C_i = [c_i] + [\tilde{c}_i]$ is the total concentration of i, defines a generalized thermodynamic model for the state of our protein kinase interaction network. Note that this can actually be derived by the conventional approach based on the chemical kinetics of kinases and phosphatases(See Appendix B). Finally, we note that our formalism can be generalized to more complex signaling circuit by replacing Eq. 3.1 with an appropriate partition function.

3.2.3 Finding biophysical parameters that implement a given decision surface

To see how these interactions can be used to relate environmental cues (i.e. input) to cellular decisions (i.e. outputs), it is helpful to view the signaling network as a directed graphical model. First one represents different kinases by different nodes whose activities are labeled by x_i, and use directed links to indicate the phosphorylation reactions with arrow pointing from kinase to its substrate. For instance, the directed arrow in the left panel of Figure 3·1, $x_2 \rightarrow x_3$, encodes the fact that 2 (RSK) is the kinase of 3 (SRF). Indeed, one can also replace the summation over $j \neq i$ in Eq. (4.3) by its specific kinase denoted as $j \in \mathrm{pa}(j)$. Following the same example, we can write $\mathrm{pa}(3) = \{2\}$ since the kinase of 3 (SRF) is 2 (RSK). Since every such phosphorylation event is directional and can be characterized by a conditional probability Eq. (4.3), we can simply apply chain rule to arrive at the joint probability of kinase activity:

$$P(\mathbf{x}) := P(x_1, \cdots, x_n) = \prod_{i=1}^{n} P(x_i|\mathrm{pa}(i), \boldsymbol{\theta}), \tag{3.2}$$

where $P(x_i|\mathrm{pa}(i), \boldsymbol{\theta})$ is given by Eq. (4.3). Note that $\mathbf{x}_{\setminus i}$ in Eq. (4.3) is now replaced by $\mathrm{pa}(i)$ due to the specificity and directionality of interactions.

In the examples considered below, we assume that the ultimate output of our signaling network is transcriptional (e.g. a promoter for *gfp* controlled by an output kinase). We model this by assuming that kinases in the last layer can bind bind (multiple) transcription factor binding site(s), with the binding probability again governed by a simple thermodynamic function of the form given in Eq. (3.1) [CITE]. In assuming this form, we have assumed a hill coefficient of one for transcription factor binding. However, cooperatively can be easily incorporated by modifying the Michelis-Mentin form in Eq. (3.1) to a Hill equation. This highlights another advantage of the framework discussed here: post-translational and transcriptional information processing can be treated on equal footing using a single thermodynamic model.

The equations above depend on the phosphorylation states of all proteins in our network. To calculate the input-output relation for a signaling network, we simply marginalize out the phosphorylation states of all kinases except for the input and output (see 3.4 for details). For a signaling networks with m intermediate proteins, in general this involves summing over 2^m states and can become computationally intractable when trying to design input-output surfaces for large networks $(m \gg 1)$. To circumvent this computational bottleneck, we have written code that allows us to perform this step quickly using a sophisticated algorithm based on Belief Propagation (BP) (Yedidia et al., 2003; Yedidia, 2001; Mezard and Montanari, 2009).

Importantly, even after we marginalize, the input-output relation depends on *all* the biophysical parameters $\boldsymbol{\theta}$ in the signaling network: the concentration of all proteins as well as the binding energy between all protein pairs. Our goal is to find $\boldsymbol{\theta}$ that reproduce a desired input-output relation. From a computational perspective, the ability to reproduce complex, multi-input, multi-output relations allows for the programming of complex decision surfaces in response to environmental inputs.

Algorithmically speaking, for a given graphical representation of signaling pathway associated with some parameters $\boldsymbol{\theta}$ and a target decision surface, we first use Eq. (4.3) and Eq. (3.2) to build a probabilistic description of kinase activity. We then marginalize out the phosphorylation states of all kinases except for the input and output to arrive at the decision surface defined through $\boldsymbol{\theta}$. Based on this, we construct a objective function that measures the difference between the target decision surface and the decision surface implemented through the current $\boldsymbol{\theta}$. This function is then minimized by performing gradient descent on this objective function with respect to $\boldsymbol{\theta}$. Such gradient descent algorithms form the backbone of many machine learning and optimization algorithms (Bishop, 2006). The output of this procedure is a "learned" decision surface which is a local minimum of the objective function (see Figure 3·1C for the algorithm flowchart).

3.2.4 Learning decision surfaces for multi-input signaling circuits

We demonstrate our framework on a variety of two-input circuits shown in In Figure 3·2. The inputs to these circuits are the concentrations of two ligands L_1 and L_2 which can bind the receptors and induce a conformational change between an active and inactive state. We model this process using a simple thermodynamically-inspired two state model where the probability that the receptor is active depends on the external ligand concentration. Such a simple two-state model of receptors have been successfully used to make quantitative predictions in the context bacterial chemotaxis (Keymer et al., 2006) and quorum sensing (Mehta et al., 2009; Teng et al., 2011). Within the two-state receptor model, there is a one-to-one mapping between ligand concentration L_i and the probability that the receptor is active q_i. For this reason, we represent the inputs to the circuits directly in probability space rather than working in concentration space (see (Mehta et al., 2009; Teng et al., 2011) for detailed discussion).

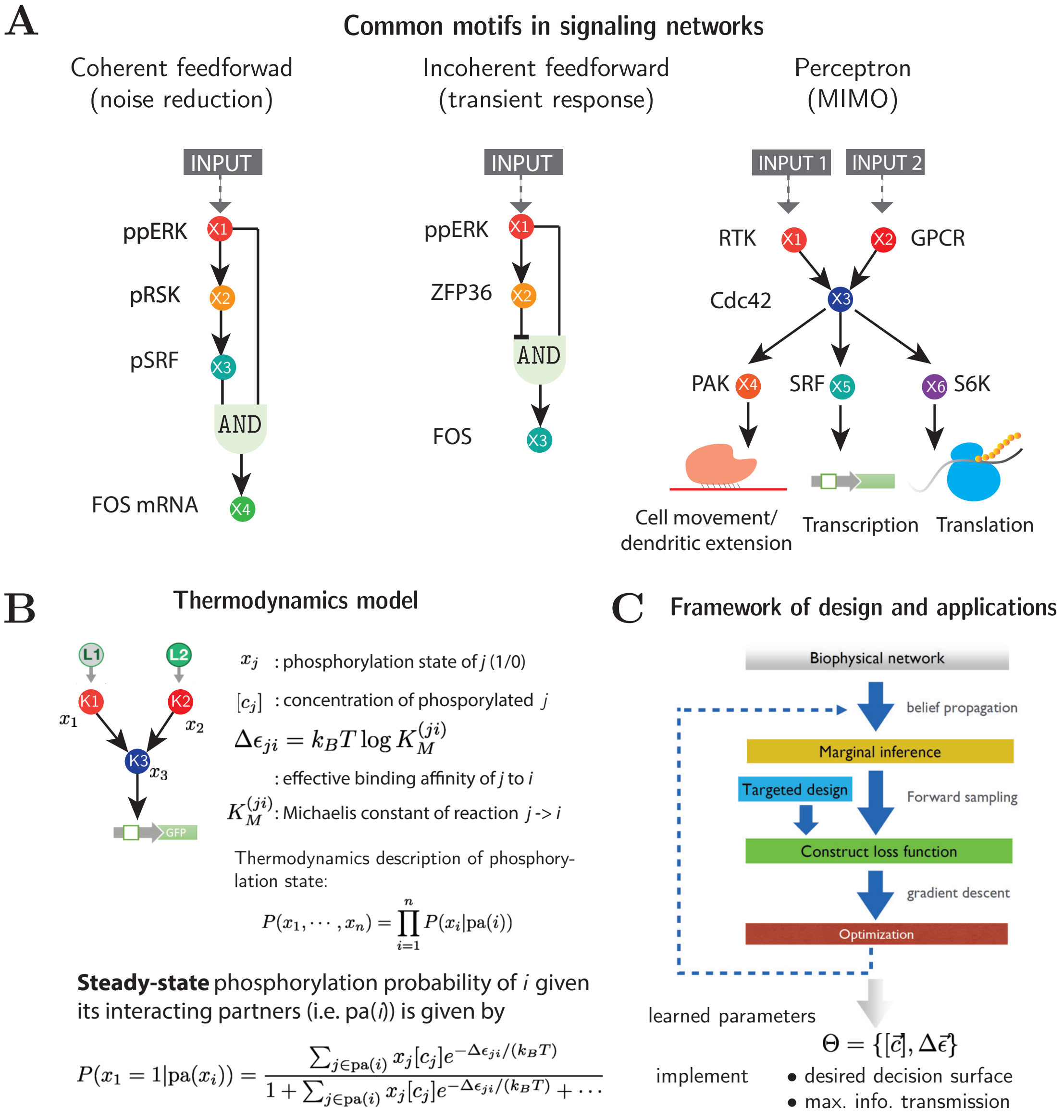

A
Common motifs in signaling networks
Coherent feedforwad
(noise reduction)
INPUT
ppERK
X1
pRSK
X2
pSRF
X3
AND
FOS mRNA
X4
Incoherent feedforward
(transient response)
INPUT
ppERK
X1
ZFP36
X2
AND
FOS
X3
Perceptron
(MIMO)
INPUT 1
INPUT 2
RTK
X1
X2
GPCR
Cdc42
X3
PAK
X4
SRF
X5
X6
S6K
Cell movement/
dendritic extension
Transcription
Translation
B
Thermodynamics model
L1
L2
K1
K2
x_1
x_2
K3
x_3
GFP
x_j : phosphorylation state of j (1/0)
$[c_j]$: concentration of phosporylated j
$\Delta\epsilon_{ji} = k_B T \log K_M^{(ji)}$
: effective binding affinity of j to i
$K_M^{(ji)}$: Michaelis constant of reaction j -> i
Thermodynamics description of phosphorylation state:
$P(x_1, \cdots, x_n) = \prod_{i=1}^{n} P(x_i|\mathrm{pa}(i))$
Steady-state phosphorylation probability of i given its interacting partners (i.e. pa(i)) is given by
$P(x_1 = 1|\mathrm{pa}(x_i)) = \frac{\sum_{j\in\mathrm{pa}(i)} x_j [c_j] e^{-\Delta\epsilon_{ji}/(k_B T)}}{1 + \sum_{j\in\mathrm{pa}(i)} x_j [c_j] e^{-\Delta\epsilon_{ji}/(k_B T)} + \cdots}$
C
Framework of design and applications
Biophysical network
belief propagation
Marginal inference
Targeted design
Forward sampling
Construct loss function
gradient descent
Optimization
learned parameters
$\Theta = \{[\vec{c}], \Delta\vec{\epsilon}\}$
implement
• desired decision surface
• max. info. transmission

Figure 3·1: (A) Shown here are three representative network motifs in signaling based on cascade of two or more (phosphorylation based) enzymes, the activity of which is determined by opposing regulators (i.e. kinase and phosphatase). Coherent feedforward loop: epidermal growth factor receptor (EGFR) activation (i.e. input, not shown) activates the ERK-RSK-SRF pathway. Both phosphorylated RSK and SRF function to promote the transcription of FOS gene. This motif can filter out noise and distinguish transient from sustained inputs. Incoherent feedforward loop: EGFR signaling wherein the activation of EGFR induces both the expression of FOS and its inhibitor zinc-finger protein 36 (ZFP36), which promotes FOS mRNA degradation. This motif demonstrates the transient memory of FOS due to the initial activation of ERK before it turns on ZFP36 whose level accumulation serves to shut off FOS. The perceptron-like motif can orchestrate multi-input-multi-output (MIMO) decisions(Jordan et al., 2000). For example, cell division control protein 42 (Cdc42), a member of the Rho family GTPases, can be stimulated both by the receptor tyrosine kinase (RTK) as well as the G-protein coupled receptor (GPCR). Depending on the combination of these inputs, Cdc42 can activate different downstream kinases: p21 activation kinase (Pak), S6-kinase (S6K), and SRF. (B): Thermodynamics model of signaling based on kinase phosphorylation. The biophysical parameters shown here can be mapped to chemical kinetics rate constants measurable in experiments (see Appendix B for more details). (C): Our design flow.

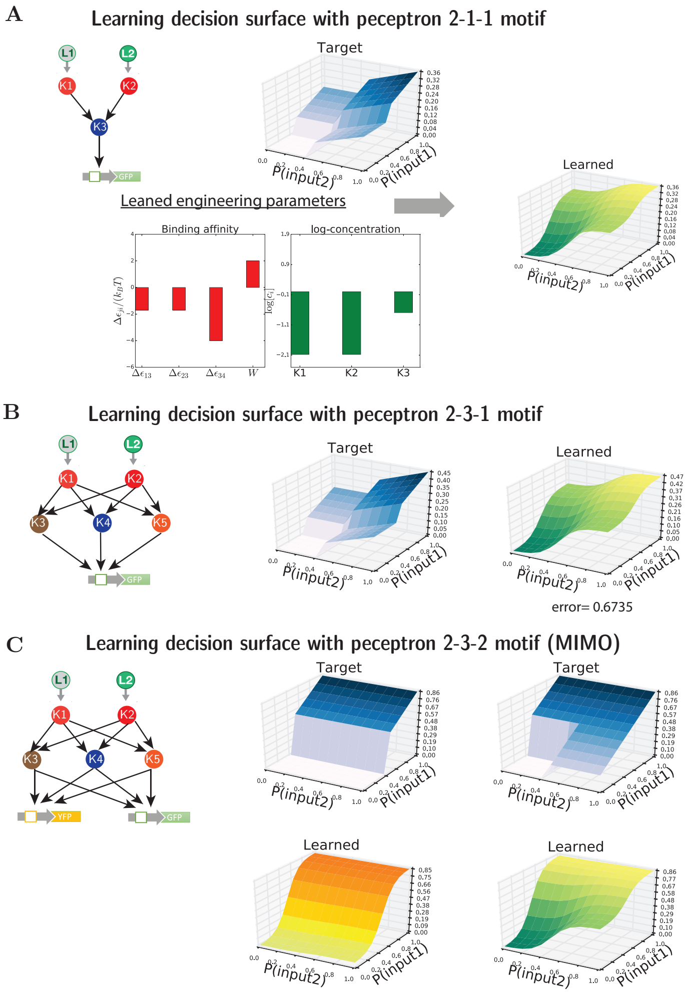


Figure 3·2: Learning to engineer computations and decisions. (A): Learning to implement desired decision surface (decision surface) with 2-1-1 perceptron motif (c.f. Figure 3·1B). This network consists of three types of protein kinases (K1, K2, and K3) that can undergo phosphorylation cascade. The phosphorylation of K1 and K2 are controlled by binding of ligands L1 and L2, respectively. K1 and K2 can both interact with and phosphorylate K3 which in term regulates the expression of some target gene. The target and learned decision surface (i.e. input-output relation, I-O relation) are shown in blue and green, respectively. The parameters (effective binding energies in units of $k_B T$ and the kinase concentrations in units of 1M) learned that implement this specific design are shown on the right. For more details on learning, see and main text and Appendix B. (B,C): Similar to (A) but with more complex perceptron-like motifs. MIMO: multi-input-multi-output.

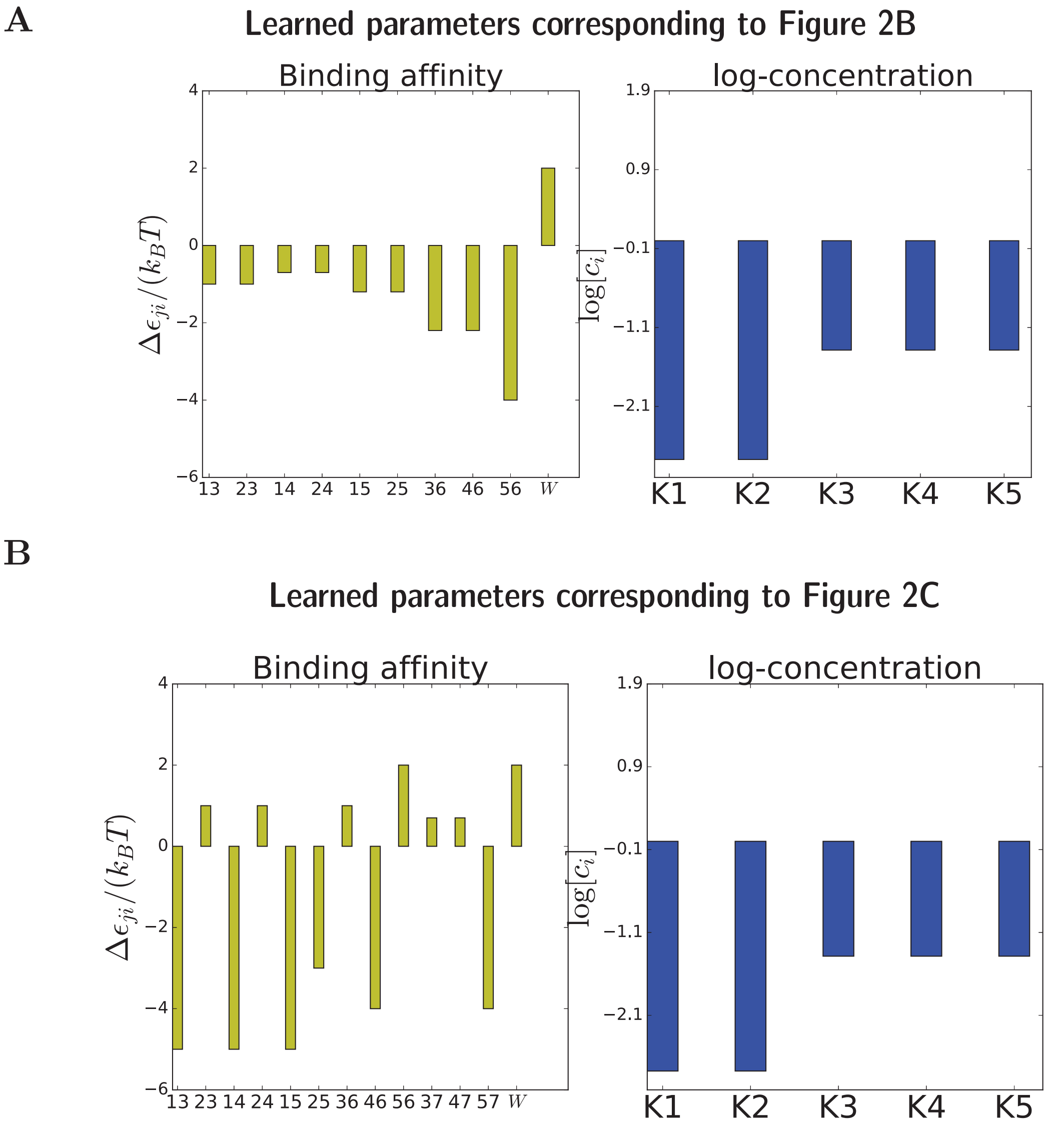


Figure 3·3: Learned parameters of panel B and C.Here we show the learned parameters of DS implemented in Figure 3·2B,C. Note that concentrations are held fixed through learning. In other words, optimization is taken over the effective binding affinities $\Delta\epsilon_{ji}$.

In Figure 3·2A , we considered a simple 2-1-1 cascade where the two receptors both phosphorylate a single kinase which regulates an output gene (e.g GFP expression). The goal is to train this simple circuit to learn the discontinuous input-output relation shown in blue. In the target decision surface, the output is negligible if the both ligand concentrations are below some thresholds (denoted by q_1^* and q_2^* and both chose to be 0.5), proportional to q_1 if $q_1 \geq q_1^*$ and $q_2 < q_2^*$, L_2 if $q_2 > q_2^*$ and $q_1 < q_1^*$, and $q_1 + q_2$ if $q_1 \geq q_1^*$ and $q_2 \geq q_2^*$. The learned decision surface (green) and the corresponding biophysical parameters that generate this surface are also shown. Notice that the learned surface captures the essential features of the target decision surface (blue). For example, it showcases the transition of expression around q1, q2$\sim$ 0.5, between the three regimes discussed above.

It is also easy to rationalize the learned biophysical parameters. Notice that the desired output is always linear in the receptor probabilities. For this reason, it is essential that the concentration of kinase three (K3) be in excess of the concentration of the two receptor (K1 and K2) to prevent saturation effect. This is exactly what is found by our algorithm. Furthermore, since for our example $q_1^* = q_2^* = 0.5$, there is a symmetry between the two ligands. This is reflected in the fact that the binding energies of the two receptors to the intermediate kinase K3 are nearly identical ($\Delta\epsilon_{13} \approx \Delta\epsilon_{23}$).

In Figure 3·2 B, we show that the same surface can also be learned by a 2-3-1 network where two inputs can regulate three distinct protein kinases, all of which down regulate the expression of a common gene of interest. Not surprisingly, this architecture also can learn the desired surface. Examining the underlying biophysical parameters, one again we see that the binding energies of the two receptor for all three intermediate kinases are symmetric. Moreover, even though the decision circuit can be in principle implemented using just of one of the kinases in the middle layer (K3, K4, K5) as in the 2-1-1 circuit in Figure 3·2A, our algorithm chooses to utilize all three kinases to form the decision surface. The reason for this is that the fit to the target surface can be slightly improved by using all three kinases, since increasing the number of parameters always improves fits. This suggests that in the presence of a strong evolutionary pressure to make high fidelity decisions and in the absence of an opposing evolutionary pressure favoring small circuits, implementation of decision surfaces in cellular signaling circuits maybe distributed over many proteins. We return to this point in the discussion.

Our algorithm can also be used to design mulit-input-multi-output (MIMO) decision surfaces (Jordan et al., 2000). Figure 3·2 C shows the target and learned input-output relation for a 2-3-2 circuit with two inputs and two outputs. One of the outputs (YFP) depends only on the concentration L_1 of ligand 1, where as the target surface for the other output (GFP) is the same as for Figure 3·2A,B. Once again our algorithm can identify biophysical parameters that reproduce many of the essential features of the target MIMO circuit. Interestingly, an examination of the binding energies indicates that the L_2-receptor binds primarily to the kinase K5 whereas the L_1 receptor binds all three intermediate kinases (K3,K4,K5). Furthermore, kinase K5 regulates the YFP output and does not bind to GFP. Thus, we see that we can think of the YFP as being essentially governed by the linear pathway $K1 \rightarrow K5 \rightarrow YFP$, even though there is cross-talk between the L_1 receptor and $K5$ and $K5$ is also used to produce the GFP-output. If one naively examined the circuit, the binding of K5 to GFP may seem like superfluous crosstalk, however we know that this binding in fact is a crucial component of creating the desired decision surface for GFP. These examples highlight how implementing even a simple MIMO decision surface can give rise to extremely complex phenotypes at the level of PPIs.

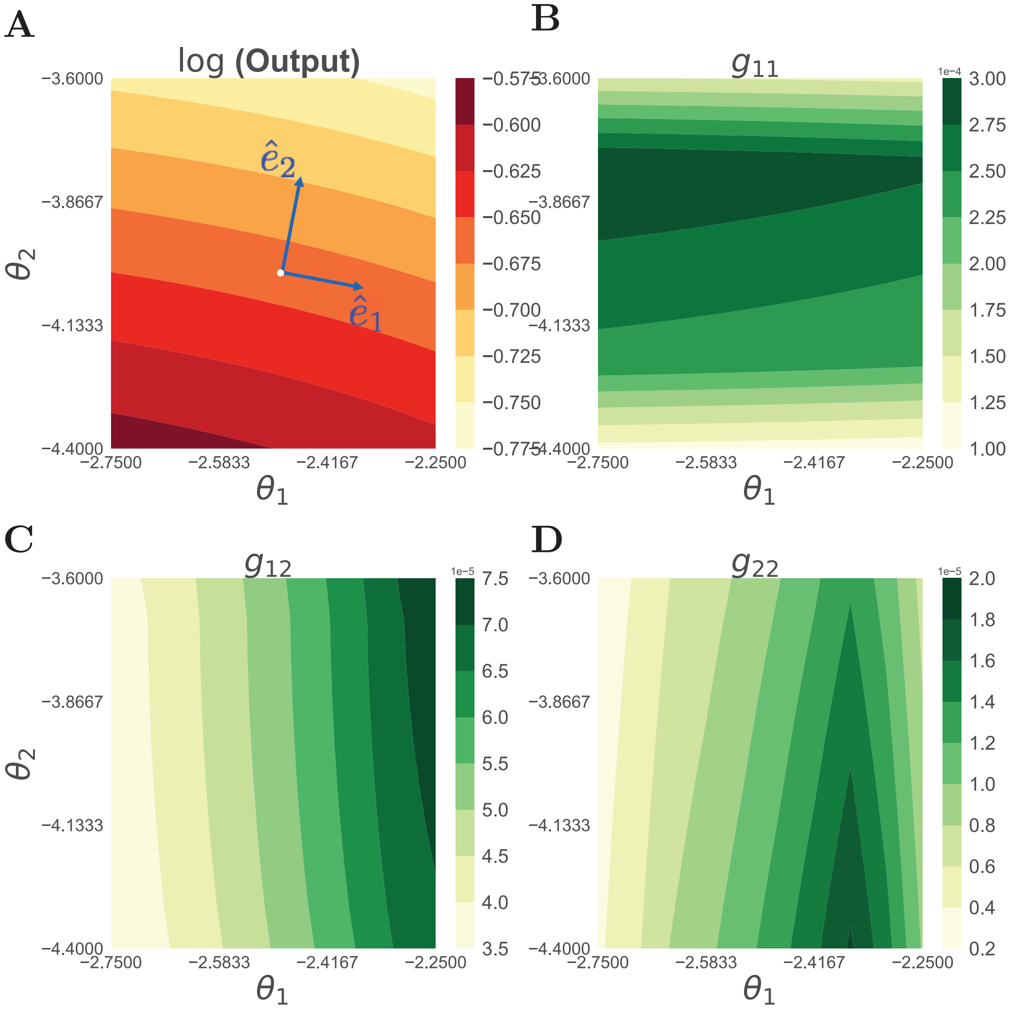


Figure 3·4: Robustness of learned DS. Here we use the 2-1-1 network (Figure 3·2) for demonstration. Based on the model in the main text we calculate the Fisher's information matrix (FIM), $g_{\mu\nu}$, defined in Eq.(3.26). To illustrate FIM in 2-dimensional space, we chose $\Delta\epsilon_{21} = \Delta\epsilon_{31} := \theta_1$ and $\Delta\epsilon_{43} := \theta_2$ while fixing the concentrations of all proteins constant so that we have effectively 2 parameters(i.e. $\mu, \nu \in \{\theta_1, \theta_2\}$, which we abbreviate as $\mu, \nu \in \{1, 2\}$). In panel (A) we showed the solution on the θ_1-θ_2 plane. White dot indicates the point where the solution found by our design flow (c.f. Eq.(3.26)). $\hat{e}_1, \hat{e}_2$ are the eigenvectors of FIM. (B-D) Components of FIM. The eigenvalues of FIM are: $\lambda_1 = 2.78 \times 10^{-4}$ and $\lambda_2 = 1.49 \times 10^{-7}$.

3.2.5 Circuits with identical steady-state properties can display diverse dynamical behaviors

In the previous sections, we have shown that it is possible to program complex computational and information processing tasks in cellular signaling circuits by designing the energetics of PPIs. Our focus has been almost entirely on the steady-state properties of these circuits. However, there is mounting experimental evidence that cellular signaling circuits also display complex dynamical behaviors. The advances in time-lapse microscopy techniques have allowed for the direct visualization of transcriptional circuit dynamics in living cells and it is now well-established that many key transcription and regulatory factors display pulsatile behaviors. Beyond transcription, recent work using quantitative mass spectroscopy has shown that external signals can induce multiple waves of distinct phosphorylation events and protein interactions in cellular signaling networks. It has also been suggested that cells can expand their signaling pathway capabilities through dynamical multiplexing, a process that encodes the stimulus information through the dynamics of regulatory molecules(Detwiler et al., 2000; Hao and O'shea, 2012; Tay et al., 2010; Purvis and Lahav, 2013). These observations inspired us to ask can signaling circuits that implement the same steady-state decisions surface display distinct dynamical behaviors?

To address this, we examined the steady-state and dynamical properties of two signaling motifs that have been widely studied in the systems and synthetic biology literature : the coherent feedforward loop (CFFL) and the incoherent feedforward loop (ICFFL) (see Figure 3·1A) (Alon, 2007; Shoval and Alon, 2010). These motifs appear commonly in genetic networks(Shen-Orr et al., 2002; Mangan et al., 2003; Milo et al., 2002; Lee et al., 2002) and are also common in many key signaling pathways such as EGFR pathway (Anjum and Blenis, 2008; Kolch et al., 2015). For both these motifs, we applied our framework to identify biophysical parameters that implement the sigmoid input-output relation shown in Figure 3·5, with target decision surface in depicted in red, the response of the CFFL and ICFFL in blue and green respectively. As can be seen from the figure, both network topologies can reproduce the desired input-output relation with high precision.

The learned biophysical parameters can be related directly to dynamical parameters for both of these motifs using a kinetic model of enzymatic activity. The kinetic model uses ordinary differential equations (ODE) to model the push-pull enzyme kinetics at each layer in the cascade. By comparing the steady-state solutions of the kinetic ODE model with those from the biophysics based thermodynamic model, it is possible to directly map biophysical parameters to kinetic parameters and vice versa. For example, Michaelis-Menten constants characterizing the enzyme kinetics can be calculated directly from PPI binding energies (see Sec. Kinetics for details).

We used the inferred kinetic models to examine the dynamic response of the learned CFFL and ICFFL to two different dynamic input profiles: a sustained, constant input (Figure 3·1 panel BDF) and an input that decays exponentially in time (Figure 3·1 panel CEG). As can be seen from the graph, the CFFL turns "on" more quickly in response to a signal than a ICFFL but turns "off" extremely slowly, on time scales of $\sim 100s$. In fact, the response of the CFFL is barely evident on the twenty second time scale shown in Panels E and G. This shows that the dynamic responses of this circuits are markedly different despite having the same steady-state behavior.

We can understand these differences by noting that the target steady-state decision surface requires an extremely high output signal when input signal is high. For CFFL, this can only be achieved by guaranteeing that the PPI interaction energies are extremely large. A direct consequence for this is that once bound, a kinase will remain bound and active for an extremely long time even after the signal has been removed. The ICFFL, on the other hand, displays another subtle feature in terms of learned parameters. Attaining a large output signal in the ICFFL requires combining a minimal activation of the repressive pathway (blue→ red → orange in panel A) with a strong activation (blue → purple). This implies that a scant stimulus on the input (blue) is sufficient to activate the output (purple) due to the insufficient repressive regulator activities. Furthermore, this leads to a time scale separation in the rates at which upstream proteins are inactivated when compared to the output signal(panel G).

The kinetics described above suggest that CFFL-based circuits can buffer temporal fluctuations in the inputs and hence are more robust to input noise than those utilizing an ICCFL motif. However, this robustness to input noise and temporal fluctuations comes at a steep price: an extremely sluggish dynamic response. This suggests that depending on the desired dynamic properties (fast dynamic responses or robustness to fluctuations), it maybe advantageous to use different motifs for implementing the same static steady-state response.

3.3 Discussion

We showed that in cell signaling the input stimulus is intimately related to the output decision through the thermodynamics of signal protein interactions. To reify this notion, we construct a theoretical framework based on thermodynamics of these interactions, and use it to further explore how cellular decisions can be implemented on par with the desired functions. We exemplify this idea through the computation perspective which emphasizes carrying out target decision surface or I-O relation. In response to the emerging experimental capability to diagnose

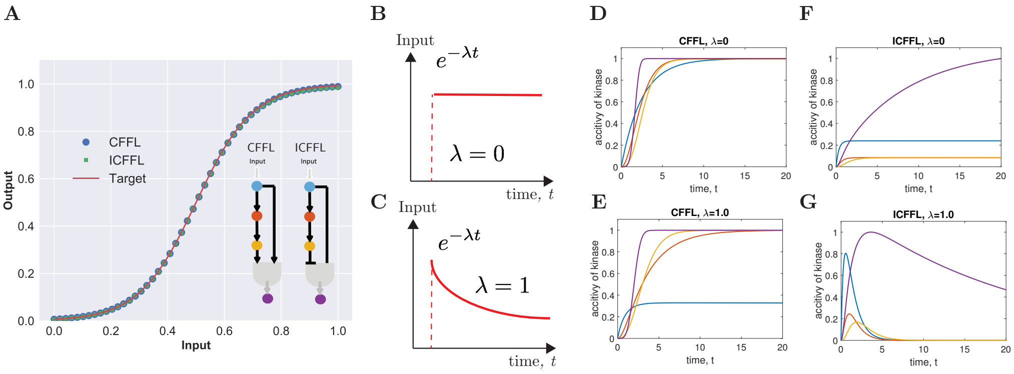


Figure 3·5: Learning to implement desired decision with naturally occurring motifs. (A) Motifs used are CFFL and ICFFL. Red curve indicates the target decision while blue circles and green squares represent that implemented by CFFL and ICFFL, respectively. Learned parameters are then used to simulate the kinetics (i.e. nodes in the network diagram) of kinase activity (B-G). (B): Constant temporal input stimulation. (C): Same as (B) but with exponentially decaying input profile $e^{-\lambda t}, \lambda = 1.0$. (D): Kinase activity of the CFFL subject to constant stimulation (i.e. panel (B)) using the learned parameters that implements the I-O relation in (A): $\alpha_i = 1.0, \beta_1 = e^{-8.2}, \beta_2 = e^{-7.5}, \beta_3 = \beta_4 = e^{-6.3}, K = e^{-2}, C_i = 2.5$ (See Appendix B for interpretations of parameters). (F): Kinase activity of ICFFL subject to constant stimulation with learned parameters: $\alpha_i = 1.0, \beta_1 = \beta_2 = e^{1.0}, \beta_3 = e^{-0.05}, \beta_4 = e^{-3.0}, K = e^{-2}, C_i = 2.5$. (E): Same as (D) on CFFL but with exponentially decaying stimulation (i.e. $\lambda = 1.0$). (G): Same as (F) on ICFFL but with exponentially decaying stimulation (i.e. $\lambda = 1.0$). In (D)(E)(F)(G), activities are normalized to the maximum value attainable by all kinases. Colors of curves are matched to that of the nodes in the networks shown in (A).

signaling through monitoring the dynamics of regulatory activities, we illustrate how our framework can be used to expand our understanding of signaling capacity beyond the steady-state decisions.

It's worth noting that the computations signaling networks implement is fundamentally different from the InfoMax design we presented in the previous chapter. For computations, signaling network can be thought of as a machine that incorporates a sensor that can detect environmental stimuli, and a circuit that can process them into some sort of response. For example, the perceptron-like pathway shown in Figure 3·1A illustrates that cell division control protein 42 (Cdc42) can be stimulated both by the receptor tyrosine kinase (RTK) as well as the G-protein coupled receptor (GPCR). Depending on the combination of these inputs, Cdc42 can activate different downstream kinases, say, p21 activation kinase (Pak) essential for cell movement. In this case, we have a *mapping* from certain combination of RTK and GPCR stimulation to the downstream Pak activation, a *well-defined I-O relation* (i.e. computation). Biologically speaking, the states of the signaling molecules and the network they intertwined into constitute the circuitry by which designated computation with specific physiological implications is done. On the information realm, however, the *context* is more obscure since the very notion of information could in general be defined in a manner that is irrespective of teleology of the physical design. Translating this to the previous example, it means that maximizing mutual information between the input and output does not necessarily guarantee to activate Pak even when both RTK and GPCR are at their right level. This is because generically speaking information is construed as the reduction of uncertainty of variables with the knowledge of their correlates, whose raison d'être is agnostic about the functionalities relevant to the biology (i.e. activate Pak when appropriate level of GPCR and Pak is provided). Therefore, to relate the InfoMax calculation presented here to the physiological meaning of signaling requires special scrutiny. For example, we showed that cells are able to transmit more information by using a more complex network. From the physics standpoint, deploying complex network demands more signal proteins and more reversible enzymatic reactions, which means more energy consumption(Mehta et al., 2016). One relevant question is why does this energy go to information and what is it good for? A plausible and intuitive answer lies in the information theoretic interpretation of mutual information: the higher it is, the more certain cells are about their output (input) when knowing their input (output). However, this answer is not an exact statement about the physiological functionality relevant to biology (e.g. express certain genes when specific inputs are present); rather, it's a modifier for statements about functionality. One can sensibly think of higher mutual information implies being able to more reliably discern input signals for the output desired, a notion of robustness. To further explore what this sort of information calculation entails, and to further relate that to the physical meaning of energy consumed and biological functionality, we suspect a deeper understanding to the relevant biology is required (e.g. undiscovered cross-talks between pathways might be a leaky channel for information).

On the theoretical side, since our approach relies only on the steady-state statistics of signaling activities, there's no need to solve the complex kinetics equations especially when dealing with large signaling networks to learn a specific decision. In addition, the thermodynamics parameters learned are those accessible in experiments: Michaelis constants and concentrations. This implies that cellular decision can in principle be *physically maneuvered* by tuning the interactions between signaling molecules and changing their concentrations. To accomplish this in synthetic networks, one would need to have control over protein modifications and interactions, which are often done via mutagenesis, insertion of unstructured recognition sequences, or alteration of domain composition. However, this is in general harder than engineering genetic circuit since mutation inside domains or on globular proteins can easily induce structural changes that lead to protein misfolding and aggregation. Beyond this structural constraint, the promiscuity of protein interactions and crosstalks between pathways also makes it difficult to engineering specific interactions. Furthermore, one also need to design modules that can be selectively activated by small molecules or external perturbations such as pH value or temperature for signal activation, as well as linear polypeptides that can be post-translationally modified and recognized by other modular domains. Therefore, to experimentally probe the applicability of our theoretical framework, one would require a better engineering protocol to address the issue mentioned above.

From the standpoint of systems biology, it is natural to ask how is signaling circuit more special than its genetic counterpart whose regulatory behaviors are simpler and easier to implement, if the goal of engineering is to utilize cellar components to carry out certain function? Component-wise speaking, signaling operates on specialized proteins and their interactions, which is far more modular and harder to manipulate than DNA and DNA binding proteins deployed in genetic circuit. However, since signaling pathways usually involve large number of such molecules, they are usually less stochastic than genetic circuits where bursting of mRNA transcription and protein synthesis often render the process noisy. On the systems scale, signaling spans across different length scales, ranging from the size of a cell (where signal activation occurs) down the that of the nucleus (where downstream decision through gene expression take place). Nevertheless, it usually has fast response (milliseconds to minutes), as opposed to genetic circuit where response ranges from minutes for prokaryotes and hours for eukaryotes. In addition, many signaling pathways are built on the signaling cascades (e.g. phosphorylation cascade), which can be used to achieve signal amplifications and other functions reminiscent of those in electronic circuits. Therefore, it is highly advantageous to employ signaling network which is modular in design to achieve functionalities with fast response and low stochasticity (i.e. robustness).

Finally, we stress that in this chapter we devise a theoretical framework to connect the systems level behavior of signaling to the underlying molecular interactions. It is our belief that our framework, along with solutions to the experimental hurdles mentioned above, should provide a practical protocol towards synthetically engineering networks to create new functionalities that at the broader scale might impact biotechnology and biomedicine.

3.4 Methods and Materials

3.4.1 Using gradient descent to design an input-output relation

Here we briefly describe how to construct the loss function in order to learn the target decision surface shown in Figure 3·2A. Other motifs can be done in a similar fashion. Let $q_i = P(x_i = 1|L_i)$, $i = 1, 2$ be the probability of input parameterized by ligand concentration L_i (see Figure 3·2 and (Keymer et al., 2006; Mehta et al., 2009; Teng et al., 2011)):

$$q_i = P(x_i = 1|L_i) = \frac{1}{1 + e^{-\beta(L_i - L_i^{(0.5)})}}, \tag{3.3}$$

where $L_i^{(0.5)}$ is the ligand concentration at which the receptor is on half the time. We also introduce $T(\mathbf{L})$ where $\mathbf{L} = (L_1, L_2)$ as the target decision surface. Learning then amounts to finding the parameters $\mathbf{\Delta}\boldsymbol{\epsilon}$ that produce an output activation probability which recapitulates $T(\mathbf{L})$. For the 2-1-1 motif shown in Figure 3·2A, the loss function used to get the leanred decision surface reads:

$$\begin{aligned}\mathcal{L}(\boldsymbol{\Delta\epsilon}, \mathbf{L}, x_4 = 1) &\equiv \frac{1}{M}\sum_{m=1}^{M}\Big[\log P(x_4 = 1|\mathbf{x}_{\backslash 4}^{(m)}, \mathbf{L}, \boldsymbol{\Delta\epsilon}) - \log T(\mathbf{L})\Big]^2 - C||\boldsymbol{\epsilon}||^2 \\ &= \frac{1}{M}\sum_{m=1}^{M}\left[\log \frac{P(x_4 = 1|\mathbf{x}_{\backslash 4}^{(m)}, \mathbf{L}, \boldsymbol{\Delta\epsilon})}{T(\mathbf{L})}\right]^2 - C||\boldsymbol{\epsilon}||^2, \qquad (3.4)\end{aligned}$$

where M is the number of samples, η is the learning rate, C is the regularization, and $P(x_4 = 1|\mathbf{x}_{\backslash 4}^{(m)}, \mathbf{L}, \boldsymbol{\Delta\epsilon})$ is the output probability based on the inputs and parameters given, which can be obtained by forward sampling $\mathbf{x}_{\backslash 4}^{(m)} := (x_1^{(m)}, x_2^{(m)}, x_3^{(m)})$ from the graphical model defined through Eq. (4.3) and Eq. (3.2). Concretely,

$$P(x_4 = 1|\mathbf{x}_{\backslash 4}^{(m)}, \mathbf{L}, \boldsymbol{\Delta\epsilon}) = P(x_4|x_3^{(m)})P(x_3^{(m)}|x_1^{(m)}, x_2^{(m)})P(x_1^{(m)}|L_1)P(x_2^{(m)}|L_2). \qquad (3.5)$$

Note that the target decision surface used in Figure 3·2 is given by a piece-wise discontinuous linear surface in the receptor probability space:

$$T(\mathbf{L}) = \begin{cases} q_1 + q_2 & \text{if } q_1 \geq q_1^* \text{ and } q_2 \geq q_2^* \\ q_1 & \text{if } q_1 \geq q_1^* \text{ and } q_2 < q_2^* \\ q_2 & \text{if } q_1 < q_1^* \text{ and } q_2 \geq q_2^* \\ \delta \ll 1 & \text{if } q_1 < q_1^* \text{ and } q_2 < q_2^*, \end{cases} \qquad (3.6)$$

where δ is set to be 10^{-3} and $q_1^* = q_2^* = 0.5$ in all panels of Figure 3·2.

3.4.2 Implementing decisions using CFFL and ICFFL

Here we only summarize how we implement desired decisions using CFFL and ICFFL illustrated in Figure 3·5 using Eq. (4.3) and Eq. (3.2). The detailed calculation based on kinetics equations is presented in the Appendix B.

Coherent feedforward loop (CFFL)

Let $T(q_1)$ be the input-output relation we want to implement (i.e. target red curve in Figure 3·5A). This input-output function is a sigmoid in the probability that the input receptor is in the on state: $T(q_1) = \frac{1}{1+e^{-10*(q_1-0.5)}}$. As such, this requires an extremely large amplification of the signal at high inputs. Fulfilling this basic feature dominates the kinetic properties of this circuits. All the probability functions used in the optimization procedure are given below:

$$\begin{aligned}
T(q_1) &= \frac{1}{1+e^{-10*(q_1-0.5)}} && (3.7)\\
P(x_2=1|x_1) &= \frac{1}{Z_2}\left(x_1[c_1]e^{-\Delta\epsilon_{13}/(k_BT)}\right) && (3.8)\\
P(x_3=1|x_2) &= \frac{1}{Z_3}\left(x_2[c_2]e^{-\Delta\epsilon_{23}/(k_BT)}\right) && (3.9)\\
P(x_4=1|x_1,x_3) &= \frac{1}{Z_4}\left(x_1x_3[c_1][c_3]e^{-(\Delta\epsilon_{14}+\Delta\epsilon_{34}+J)/(k_BT)}\right), && (3.10)
\end{aligned}$$

where the partition functions Z_i are given by

$$\begin{aligned}
Z_2 &= 1+x_1[c_1]e^{-\Delta\epsilon_{12}/(k_BT)}+(1-x_1)[\tilde{c}_1]e^{-W/(k_BT)} && (3.11)\\
Z_3 &= 1+x_2[c_2]e^{-\Delta\epsilon_{23}/(k_BT)}+(1-x_2)[\tilde{c}_2]e^{-W/(k_BT)} && (3.12)\\
Z_4 &= 1+x_1[c_1]e^{-\Delta\epsilon_{14}/(k_BT)}+x_3[c_3]e^{-\Delta\epsilon_{34}/(k_BT)}\\
&+ x_1x_3[c_1][c_3]e^{-(\Delta\epsilon_{14}+\Delta\epsilon_{34}+J)/(k_BT)}\\
&+ \cdots, && (3.13)
\end{aligned}$$

where in Z_4 we neglect terms proportional to $e^{-W/(k_BT)}$ since those events are probabilistically unlikely. We optimize the loss function with respect to $\Delta\epsilon = (\Delta\epsilon_{12}, \Delta\epsilon_{23}, \Delta\epsilon_{14}, \Delta\epsilon_{34})$. We also include a cooperativity energetic reward J to encourage kinase 4 phosphorylation due to the activation of both of its upstreams: kinase 1 and kinase 3.

Incoherent feedforward loop (ICFFL)

All the probability functions used in the optimization procedure are given below:

$$
\begin{aligned}
f(p_1) &= \frac{1}{1+e^{-10*(q_1-0.5)}} && (3.14)\\
P(x_1=1|L) &= \frac{1}{1+e^{-10*(L-0.5)}} && (3.15)\\
P(x_2=1|x_1) &= \frac{1}{Z_2}\left(x_1[c_1]e^{-\Delta\epsilon_{12}/(k_BT)}\right) && (3.16)\\
P(x_3=1|x_2) &= \frac{1}{Z_3}\left(x_2[c_2]e^{-\Delta\epsilon_{23}/(k_BT)}\right) && (3.17)\\
P(x_4=1|x_1,x_3) &= \frac{1}{Z_4}\left(x_1(1-x_3)[c_1][\tilde{c}_3]e^{-(\Delta\epsilon_{14}+\Delta\tilde{\epsilon}_{34}+J)/(k_BT)}\right), && (3.18)
\end{aligned}
$$

with the partition functions Z_i given by

$$
\begin{aligned}
Z_2 &= 1+x_1[c_1]e^{-\Delta\epsilon_{12}/(k_BT)}+(1-x_1)[\tilde{c}_1]e^{-W/(k_BT)} && (3.19)\\
Z_3 &= 1+x_2[c_2]e^{-\Delta\epsilon_{23}/(k_BT)}+(1-x_2)[\tilde{c}_2]e^{-W/(k_BT)} && (3.20)\\
Z_4 &= 1+x_1[c_1]e^{-\Delta\epsilon_{14}/(k_BT)}+x_3[c_3]e^{-\Delta\epsilon_{34}/(k_BT)}+(1-x_3)[\tilde{c}_3]e^{-\Delta\tilde{\epsilon}_{34}/(k_BT)}\\
&+ x_1(1-x_3)[c_1][\tilde{c}_3]e^{-(\Delta\epsilon_{14}+\Delta\tilde{\epsilon}_{34}+J)/(k_BT)}+x_1x_3[c_1][c_3]e^{-(\Delta\epsilon_{14}+\Delta\epsilon_{34})/(k_BT)}\\
&+ \cdots, && (3.21)
\end{aligned}
$$

where $\Delta\tilde{\epsilon}_{34}$ is the binding energy of inactive kinase 3 to kinase 4, as opposed to $\Delta\epsilon_{34}$ for that between active kinase 3 to kinase 4. We treated $\Delta\tilde{\epsilon}_{34}$ as a constant while optimizing the loss function with respect to $\Delta\boldsymbol{\epsilon} = (\Delta\epsilon_{12}, \Delta\epsilon_{23}, \Delta\epsilon_{14}, \Delta\epsilon_{34})$. We also include a cooperativity energetic reward J to encourage kinase 4 phosphorylation due to the inactivation of its upstream repressive kinase 1 and the activation of upstream kinase 1. In Z_4, we neglect a few terms pertaining to non-specific binding (i.e. $\sim e^{-W/(k_BT)}$) since these events are probabilistically unlikely.

Loss function used

$$\mathcal{L} \equiv \frac{1}{N} \sum_{\alpha=1}^{N} \Big(f(p_1) - P(x_4 = 1 | \mathbf{x}_{\setminus 4}^{(\alpha)}, \Delta\boldsymbol{\epsilon}, p_1) \Big)^2 + C||\Delta\boldsymbol{\epsilon}||_2^2, \tag{3.22}$$

where $\{\mathbf{x}_{\setminus 4}^{(\alpha)}\}_{\alpha=1}^{N} = \{(x_1^{(\alpha)}, x_2^{(\alpha)}, x_3^{(\alpha)})\}_{\alpha=1}^{N}$ is the set of samples generated from our model defined above and in the main text using the parameters $\Delta\boldsymbol{\epsilon}$ at each optimization step, and C is the regularization constant introduced to avoid trivial solutions (i.e. all energies are infinitely negative). The parameters are updated according to

$$\Delta\boldsymbol{\epsilon} \leftarrow \Delta\boldsymbol{\epsilon} - \eta \nabla_{\Delta\boldsymbol{\epsilon}} \mathcal{L}, \tag{3.23}$$

where η is the learning rate.

3.4.3 Robustness of decision from Fisher's information

One of the most intuitive ways to quantify the uncertainties or robustness against perturbation in parameters in a statistical model is to calculate the Fisher Information Matrix (FIM). More concretely, for a probability function associated with a parametric family of parameters Θ: $P(\mathbf{x}|\boldsymbol{\theta})$, the robustness of the model can be gauged by the difference between the probability generated by a fixed set of parameter $\boldsymbol{\theta}$ and that by a small deviation from that parameter $\boldsymbol{\theta} + \delta\boldsymbol{\theta}$. In the Bayesian's language, this difference measures how more likely one is to produce typical data from oneself than the other would be. A typical measure is the Kullback-Leibler divergence (KL divergence):

$$D_{KL}[P(\mathbf{x}|\boldsymbol{\theta})||P(\mathbf{x}|\boldsymbol{\theta}+\delta\boldsymbol{\theta})] = \int d\mathbf{x} P(\mathbf{x}|\boldsymbol{\theta}) \log \frac{P(\mathbf{x}|\boldsymbol{\theta})}{P(\mathbf{x}|\boldsymbol{\theta}+\delta\boldsymbol{\theta})} \tag{3.24}$$

By expanding $P(\mathbf{x}|\boldsymbol{\theta}+\delta\boldsymbol{\theta})$ around $\boldsymbol{\theta}$ and assuming $|\delta\boldsymbol{\theta}| << 1$, one gets,

$$D_{KL}[P(\mathbf{x}|\boldsymbol{\theta})||P(\mathbf{x}|\boldsymbol{\theta}+\delta\boldsymbol{\theta})] \simeq g_{\mu\nu}\delta\theta^{\mu}\delta\theta^{\nu} + O(\delta\theta^3), \tag{3.25}$$

where $g_{\mu\nu}$ is the Fisher information matrix (FIM)

$$g_{\mu\nu} = -\langle \partial_\mu \partial \nu \log P(\mathbf{x}|\boldsymbol{\theta}) \rangle_{P(\mathbf{x}|\boldsymbol{\theta})} = -\int d\mathbf{x} P(\mathbf{x}|\boldsymbol{\theta}) \frac{\partial^2}{\partial\theta^\mu \partial\theta^\nu} \log P(\mathbf{x}|\boldsymbol{\theta}) \tag{3.26}$$

One can show that FIM, $g_{\mu\nu}$, is indeed a legal metric and is symmetric, positive semi-definite. In addition, it is invariant under transformation on θ. The eigenvector corresponding to small eigenvalue represents the combination of parameters that is irrelevant to model behavior (i.e. sloppiness) while that for the large eigenvalue indicates the combination that is crucial (i.e. stiffness). In statistics, for unbiased estimator $\mathbf{T}(X)$ of parameter $\boldsymbol{\theta}$, the Cramér-Rao bound is just the inverse of FIM:

$$\mathrm{cov}_{\boldsymbol{\theta}}(\mathbf{T}(X)) \geq g(\boldsymbol{\theta})^{-1} \tag{3.27}$$

Chapter 4

Is there a limit to cell's ability to perform computation?

> *'Progress imposes not only new possibilities for the future but new restrictions. It seems almost as if progress itself and our fight against the increase of entropy intrinsically must end in the downhill path from which we are trying to escape.'*
>
> (Norbert Wiener, 1950)

Cells have evolved sophisticated signaling networks to perform computations and make decisions in response to environmental cues. The proteins in these signaling networks often possess a modular structure that combines interaction domains that mediate protein-protein interactions (PPIs), allosteric domains, and specialized domains that can read, write, or erase a post-translational modification such as phosphorylation. Here, we adapt ideas from the statistical physics to develop a novel theoretical framework for quantifying the computational capacity of such signaling networks. We show that the biophysics underlying PPIs fundamentally limits the ability of such networks to perform computations and derive a bound for the computational capacity of a generic signaling network. We relate our bound to statistical learning theory (i.e. Cover's Theorem for the computational capacity of perceptrons) and discuss the implications of our framework for identifying design bottlenecks in synthetic circuits.

4.1 Introduction

Living organisms must sense and respond to dynamic environments. One common strategy to accomplish this feat is signal transduction. At the cellular level, this ability to sense and respond, often called cell signaling, forms the basis of development(Kuchroo et al., 1995; Gerhart and Kirschner, 1997), immunity(Taganov et al., 2006), and tissue homeostasis(Pasparakis, 2009). In the past few decades, there has been a large effort to characterize and understand the basic principles governing cellular signaling networks (Koch et al., 1991; Pawson and Nash, 2003; Lim and Pawson, 2010; Lim et al., 2014; Hunter, 2000; Selimkhanov et al., 2014; Brennan et al., 2012; Cheong et al., 2011; Detwiler et al., 2000) and to use this knowledge to engineer novel signaling networks capable of carrying out complex computations(Lim, 2010; Khalil and Collins, 2010; Morsut et al., 2016; Roybal et al., 2016; Bashor et al., 2010; Lim et al., 2014). Despite these considerable successes, our understanding of how biophysical properties – such as the strength and types of protein-protein interactions – shape and constrain the computational capabilities of cellular signaling networks remains limited.

Cellular signaling networks are composed of multiple, interacting proteins. Proteins often exist in multiple internal states (e.g. phosphorylated or unphosphorylated; a particular conformation, etc.) that function as the biochemical correlates of information. From the point of view of information processing and computation, the internal states of proteins (e.g. the phosphorylation state) play an analogous role in biochemical circuits to the voltage state (high or low) of a memory cell in modern computers. For example, a protein that can be in two internal states can in principle store 1 bit of information. Just as computers perform computations by manipulating memory states, biochemical signaling networks perform computations by modifying the internal states of proteins through protein-protein interactions (PPIs). Thus, to be able to perform complex computations biochemical signaling networks must be able to store many bits of information and selectively and reliably change the internal states of proteins in the network. Whereas the former is directly related to the number of protein species, the latter is related to the number and specificity of realizable PPIs (see Figure 4·1).

These observations suggest that the computational capacity of signaling networks is intimately related to the structural organization of signaling proteins. Over the last few decades, it has become clear that many signaling proteins in eukaryotes have a modular structure. Proteins often possess multiple structural domains with distinct functional roles. Domains can be divided into two broad categories: interaction domains, which mediate interactions with other molecules within the cell (Koch et al., 1991; Pawson and Nash, 2003) and catalytic domains, which catalyze reversible enzymatic reactions through post-translational modifications (PTMs) such as phosphorylation or methylation (Jin and Pawson, 2012; Lim et al., 2014). Different proteins species possess different combinations of catalytic and interaction domains.

In contrast with conventional allostery where the catalytic activity and binding interactions of a protein are tightly coupled into a single structural unit(Cui and Karplus, 2008; Monod et al., 1965), the emerging "modular allostery" paradigm suggests that in many proteins there exists a structural separation between the output catalytic and regulatory domains(Lim, 2002; Taylor and Radzio-Andzelm, 1997). Proteins that implements modular allostery usually contain catalytic domains that, when isolated, display constitutive activity but are otherwise autoinhibited by other regions of the protein through steric blocking or conformational contortion. This modular domain structure provides the underlying "biophysical hardware" that allows cells to evolve signaling networks to carry out complex computations using proteins.

Synthetic biologists have started to exploit these insights to start engineering sophisticated synthetic signaling networks. One of the prominent principles exploited in recent synthetic systems is the RWE signal relay paradigm(Bashor et al., 2010; Lim et al., 2014; Khalil and Collins, 2010) (see Chapter 3). Writer proteins such as kinase or Guanine nucleotide exchange factor (GEF) enzymatically catalyze the transfer of chemical marks onto target molecules, whereas eraser proteins (e.g. phosphatase or GTPase-activating proteins, GAPs) catalyze the removal of the chemical mark. The presence of a mark is then read out by a reader module, usually a binding motif that recognizes and binds the modified chemical mark.

Inspired by these ideas, in this paper we present a new theoretical framework for quantifying the computational capacity of signaling networks based on modular allostery. We focus primarily on the reader-writer-eraser paradigm but our results and methods easily generalize to more complex signaling networks. We show that for large networks, the computational capacity of a signaling network is fundamentally limited by the number of interaction domains rather than the number of distinct protein species. The underlying reason for this is that cross-talk between proteins prevents the signaling network from realizing most potential configurations of the network (see Figure 4·1). In this paper we focus on the steady-state computational capacity of signaling networks. Real signaling network likely use important dynamical properties to transmit information and perform computation (Dayarian et al., 2009; Cheong et al., 2011; Selimkhanov et al., 2014; ten Wolde et al., 2016; Detwiler et al., 2000) and it will be interesting to understand how to generalize our framework to study dynamics.

The chapter is organized as follows. We begin by introducing a thermodynamic biophysical model for the steady-state properties of protein-protein interaction networks. We then define computational capacity and discuss how it can be calculated using ideas for statistical physics(Engel and Van den Broeck, 2001). We then use this framework to show that the computational capacity of signaling networks based on the reader-writer-eraser framework scales polynomially (rather than exponentially) in the number of protein species. We then frame our results in the context of statistical learning theory and discuss the implications of our results for understanding natural circuits and designing complex synthetic signaling circuits.

4.2 Results

4.2.1 Generalized thermodynamic model of a synthetic protein-protein interaction network

Inspired by synthetic biology, we focus on modeling signaling networks that implements a simple form of the reader-writer-eraser paradigm. The fundamental building blocks of our synthetic system consists of four basic types of domains: catalytic kinase domains that phosphorylate a particular residue (e.g. tyrosine kinase , serine/threonine kinases), phosphatase domains (tyrosine phosphatase, serine/theonine phosphatase), constitutive interaction domains and their cognate binding sites (e.g PDZ domains, Leudine zippers), and phosphorylation-state dependent interaction domains and their cognate binding sites (e.g. SH 2 domain, FHA domain, WW domain). In general, the binding sites of phosphorylation-state dependent interactions contain or are near the phosphorylation site that controls binding. We call the type of interaction domain or binding site (i.e SH2 or WW domain) its flavor.

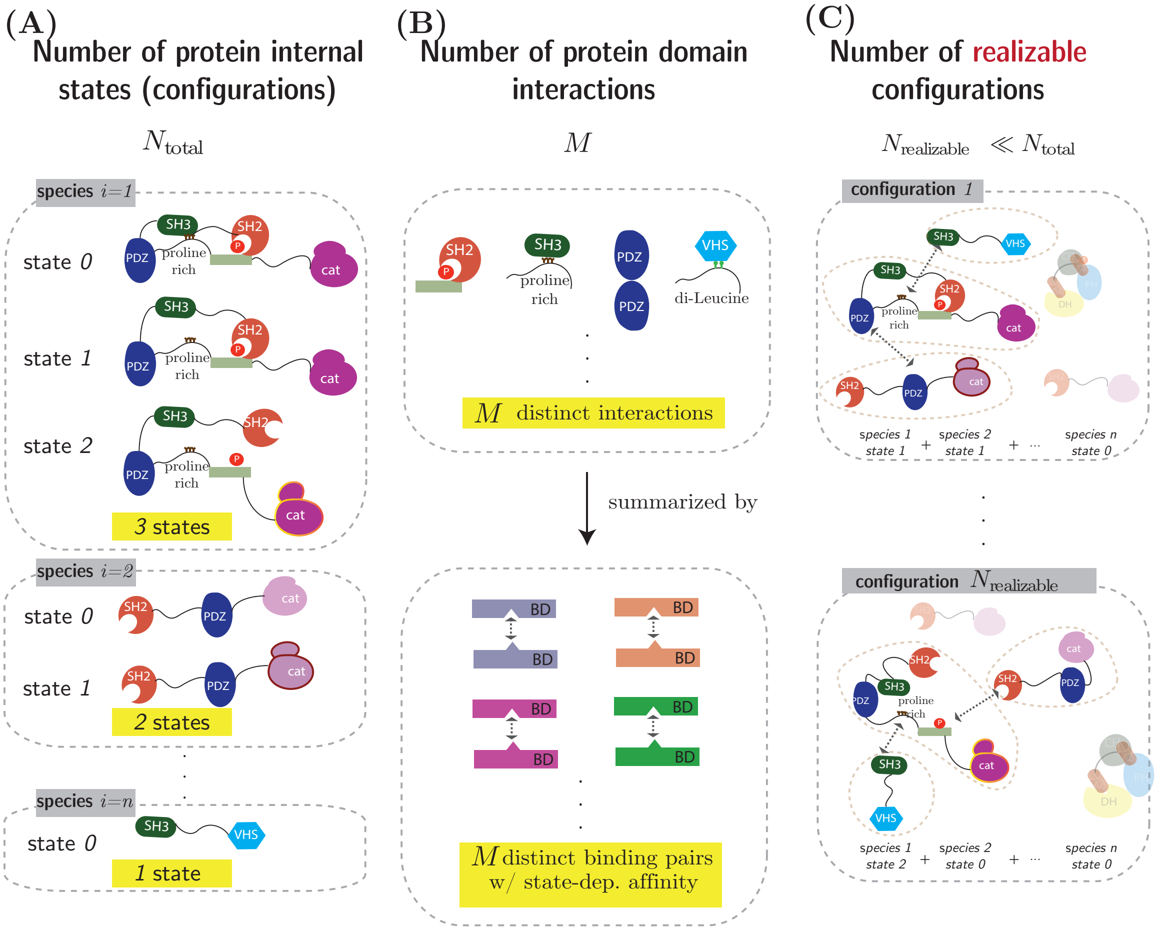

(A)
Number of protein internal states (configurations)
N_{total}
species i=1
state 0
state 1
state 2
SH3
SH2
PDZ
proline rich
P
cat
3 states
species i=2
state 0
state 1
2 states
species i=n
state 0
VHS
1 state
(B)
Number of protein domain interactions
M
di-Leucine
M distinct interactions
summarized by
BD
M distinct binding pairs w/ state-dep. affinity
(C)
Number of realizable configurations
$N_{\text{realizable}} \ll N_{\text{total}}$
configuration 1
DH
species 1 state 1 + species 2 state 1 + ··· species n state 0
configuration $N_{\text{realizable}}$
species 1 state 2 + species 2 state 0 + ··· species n state 0

Figure 4·1: Number of biophysically realizable collective protein internal states (i.e. configurations) is limited by the diversity of domain interactions. (*A*) Shown here are the domain diagrams of n distinct protein species, each of which bears specific internal states. In particular, species i has n_i internal (e.g. allostery, phosphorylation) states. Species shown here indicate $n_1 = 3$, $n_2 = 2$, $\cdots n_n = 1$. Total number of collective (species) internal states N_{total} is given by $\Pi_{i=1}^{n} n_i$, which, in the case where $n_i = 2$ for all i, becomes 2^n. (*B*) Varieties of protein domain interactions can be summarized by a table of binding domain (BD) pairs with internal state-dependent binding affinity. In this example, there are M distinct domain interactions so M binding pairs. (*C*) Number of realizable configurations $N_{\text{realizable}}$ is far less than N_{config} due to the number of ways proteins can interact with one another.

Our synthetic system contains two kinds of protein species, writer proteins which contain a kinase domain, and eraser proteins which contain a phosphatase domain. A protein consists of three parts: a catalytic domain which is a kinase domain ($\mathcal{C} = \mathcal{K}$) for writers and a phosphatase domain ($\mathcal{C} = \mathcal{P}$) for erasers, phosphorylation-dependent interaction domain with flavor α, and a binding site for a phosphorlyation dependent domain with flavor β (See Figure). In general, the flavors of the binding site and interaction domain will be different. For this simple synthetic system, each species of protein i can be represented by a triplet $\mathbf{v}_i = (\alpha_i, \beta_i, \mathcal{C}_j)$ and its phosphorylation state is represented by a single binary number $x_i = 0, 1$. If there are M flavors of interaction domains, there are generally $n = 2M(M-1)$ different potential species since $\alpha_i \neq \beta_i$. Note that in our simple system we do not make use of constitutive interaction domains all interactions are dependent on phosphorylation.

Proteins in our synthetic system interact in a phosphorylation dependent manner (see Figure 4·1). When a protein is phosphorylated, it can bind other proteins and change their phosphorylation state. Conversely, the phosphorylation state of a protein can be modified if it is bound by another phosphorylated protein. In all cases, the the strength binding depends on the flavor of the binding site and interaction domains. Since both writers and erasers can only bind and modify the phosphorylation-state of another protein if they are themselves phosphorylated, we will use the shorthand of calling phosphorylated proteins active ($x_i = 1$) and unphosphorylated proteins inactive ($x_i = 0$).

We can combine the catalytic activity and binding interactions of our protein interaction into a single generalized thermodynamic model. Similar models have played an important role in understanding transcriptional regulation(Phillips et al., 2012; Bintu et al., 2005b; Garcia et al., 2010). We begin by introducing a $M \times M$

binding energy matrix $\Delta\epsilon_{\alpha\beta}$ that encodes the change in binding energy when an interaction domain with flavor α binds a phosphorylated binding site of flavor β. We can also introduce $2(M^2 - M)$ concentrations for each protein species i, $[c_i]$. If we denote the probability for protein i to in state x_i by $P_i(x_i)$, then we can introduce the state-dependent concentrations

$$[c(x_i)] = P_i(x_i)[c_i]. \tag{4.1}$$

The probability that a protein i is bound by a another protein j can then simply be written using the thermodynamic model based on Boltzmann statistics as

$$p_{ij}^{bound} = \frac{[c_j(x_j)]x_j e^{-\beta\Delta\epsilon(\alpha_i,\beta_j)}}{1 + \sum_k [c_k(x_k)]x_k e^{-\beta\Delta\epsilon(\alpha_i,\beta_k)}}, \tag{4.2}$$

with $\beta = 1/k_B T$ is the inverse temperature). In writing this, we have set the weight of the unbound state to 1 and we have made use of the assumption that a protein k can bind only if it is phosphorylated (i.e. $x_k = 1$). This simple thermodynamic model captures the equilibrium energetics of binding.

We must extend this model to account for the catalytic activity of proteins. Once a protein is bound by a writer or eraser, its phospho-state can be changed. We assume that writes favor phosphorylation ($x_i = 1$) while erasers favor the unphosphorylated state $x_i = 1$. We assume that the probability for a protein i to be phosphorylated is directly proportional to the probability to be bound by a kinase. If we consider a system with n protein species, this gives us the following equations

$$P_i(x_i = 1|\mathcal{W}_{\setminus i}) \equiv P_i(x_i = 1) = \sum_{j s.t. \mathcal{C}_j = \mathcal{K}} \frac{[c_j(x_j)]x_j e^{-\beta\Delta\epsilon(\alpha_i,\beta_j)}}{1 + \sum_k [c_k(x_k)]x_k e^{-\beta\Delta\epsilon(\alpha_i,\beta_k)}}., \tag{4.3}$$

where $\mathcal{W}_{\setminus i} = \{j \neq i \text{ such that } \mathcal{C}_j = \mathcal{K}\}$ Note, these must be supplemented by the self-consistency equations (Eq.(4.1)). Together, this defines a generalized thermodynamic model for the state of our protein interaction network.

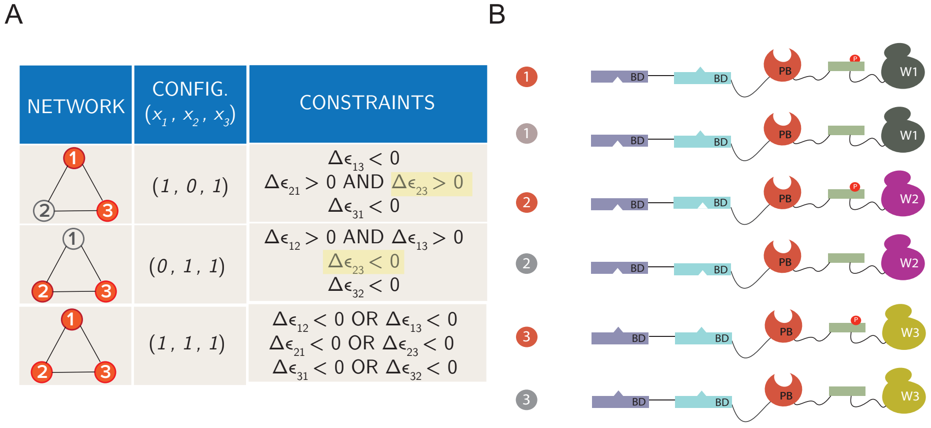
4.2.2 Defining computational capacity

The computational capacity of a network depends on its ability to manipulate information. In our synthetic reader-writer-eraser system, the biochemical correlate of information is contained in the phosphorylation states of the proteins that make up the network. Since each protein i can be in one of two states, phosphorylated or unphosphorylated state $x_i = 1, 0$, the state of the network can be described by a binary vector $\mathbf{x} = (x_1, x_2, \dots, x_n)$, where n is the number of protein species in the network.

Notice that the maximum number of possible configurations in a network with n species is 2^n. However, many of these configurations may in practice be inaccessible due to biophysical constraints on protein-protein interactions (PPIs). In particular, given a set of binding energies $\Delta\epsilon(\alpha, \beta)$ between M interaction-domains, we can ask how many of these configurations can actually be realized with high probability (we will define this more carefully below). When calculating this number, we allow the cell to choose the concentrations c_i of all protein species. We define the computational capacity, of the network as the maximum number of high-probability states that can be realized for any fixed choice of c_i.

We can make this more precise by taking the zero temperature limit $\beta \to \infty$ in the expression Eq.(4.3). In this case, the shape of the Boltzmann distribution reduces from sigmoid to a step function and the protein has either probability strictly 1 or 0 to be in a given phosphorylation state (i.e $P_i(x_i = 1) \in \{0, 1\}$ when $\beta \to \infty$). In other words, in this limit is assigned to its most probable configuration.

To see how this works, let us consider species activation probability in a system containing just writers first. Note that by taking the zero temperature limit $\beta \to \infty$, $P_i(x_i) \to x_i \in \{0, 1\}$, as can be inferred self-consistently from $[c_j(x_j)] =$



NETWORK	CONFIG. (x_1, x_2, x_3)	CONSTRAINTS
	$(1, 0, 1)$	$\Delta\epsilon_{13} < 0$ $\Delta\epsilon_{21} > 0$ AND $\Delta\epsilon_{23} > 0$ $\Delta\epsilon_{31} < 0$
	$(0, 1, 1)$	$\Delta\epsilon_{12} > 0$ AND $\Delta\epsilon_{13} > 0$ $\Delta\epsilon_{23} < 0$ $\Delta\epsilon_{32} < 0$
	$(1, 1, 1)$	$\Delta\epsilon_{12} < 0$ OR $\Delta\epsilon_{13} < 0$ $\Delta\epsilon_{21} < 0$ OR $\Delta\epsilon_{23} < 0$ $\Delta\epsilon_{31} < 0$ OR $\Delta\epsilon_{32} < 0$

Figure 4·2: Biophysical constraints associated with protein activity configurations. (*A*) Shown here are three configurations of a signaling systems consisting of three writer species (W), along with their configuration vector (left column) and the associated constraints in physical parameter space (right column): $\Delta\epsilon_{ij}$ is the *effective* binding energy between species i and j. The two terms shadowed in yellow in the right column indicates that the configurations they correspond to are not compatible. (*B*) Exemplary species diagram. This construction is similar to the synthetic example given in Fig.4·1.

$P_j(x_j)[c_j] \to x_j[c_j]$. Concretely,

$$P(x_i = 1)|_{\beta\to\infty} \;=\; \begin{cases} 1, \text{ if } \exists j \in \mathcal{W}_{\setminus i} \text{ s.t. } x_j = 1 \text{ AND } \Delta\epsilon_{ij} < 0 \\ 0, \text{ if } \forall j \in \mathcal{W}_{\setminus i},\ x_j = 0 \\ \quad \text{OR } \forall j \in \mathcal{W}_{\setminus i} \text{ s.t. } x_j = 1, \Delta\epsilon_{ij} > 0 \end{cases} \tag{4.4}$$

(4.5)

Note that since $x_j \in \{0,1\}$ and $[c_j(x_j)] \to x_j[c_j]$ in this limit, the binding energy matrix $\Delta\epsilon(\alpha_i, \beta_j)$ reduces to a single scalar $\Delta\epsilon_{ij}$ under the summation over $x_j \in \{0,1\}$ in Eq. (4.3). We term $\Delta\epsilon_{ij}$ the *effective* species binding energy between species i and species j.

This "zero-temperature limit" procedure smoothens out the irrelevant but principally rugged probability landscape in high dimensional parameter space (i.e. Eq. (4.3)), thus singling out the probability states that capture the underlying physics (i.e. Eq. (4.4)). In spirit, this transition between two functional resolutions (i.e. from smooth to binary) is the same as that in the Fermi-Dirac distribution function of fermions in physics or as that in Hill function in biochemistry and pharmacology. In Fig.4·2A, we illustrate this idea by considering a systems with 3 writer protein species whose activities are collectively denoted by the binary vector (x_1, x_2, x_3). By reading off Eq. (4.4), one immediately sees that, for example, the species state configuration $(1, 0, 1)$ implies that (i) $\Delta\epsilon_{13} < 0$ (ii) $\Delta\epsilon_{21} > 0$ AND $\Delta\epsilon_{23} > 0$ (iii) $\Delta\epsilon_{31} < 0$. Similarly, we can read off the binarized probability for all other configurations, two of which are shown in Fig.4·2A. Indeed, the construction of the constraint table in Fig.4·2A expatiates that the interactions mediating cell signaling, namely, binding affinities between species $\Delta\epsilon_{ij}$, and their concentrations, $[c_j]$, fundamentally determines the realizability of configurations. To see this, consider the same system in Fig.4·2A but now the interaction between species 1 and 2 are designed such that $\Delta\epsilon_{13} > 0$. Note that this is always possible by re-engineering, say, the constitutive binding domains, of the proteins involved (e.g. utilize different binding domain in Fig.4·1B). However, such construction explicitly violates condition (i), one of the existence criteria of configuration (1,0,1). This implies that one can never use this system with this particularly designed interaction (e.g. one set of reader binding domain flavor table in Fig.4·1B) to implement (1,0,1) configuration. In abstraction, this means the realizability of configurations is contingent on the underlying protein interactions.

4.2.3 Computation is limited by the number of biophysically realizable configurations: notion of capacity

Since the realizability of activity configurations is dependent on the underlying interactions, it is natural to speculate that for any given signaling systems with specially designed interactions, not all configurations are biophysically implementable. To illustrate this, we consider configurations (1,0,1) and (0,1,1) in the previous example (see Fig.4·2A). Reading off Eq. (4.4) one finds the first configuration implies (i) $\Delta\epsilon_{13} < 0$ (ii) $\Delta\epsilon_{21} > 0$ AND $\Delta\epsilon_{23} > 0$ (iii) $\Delta\epsilon_{31} < 0$ while the latter implies (iv) $\epsilon_{12} > 0$ AND $\Delta\epsilon_{13} > 0$ (v) $\Delta\epsilon_{23} < 0$ (iii) $\Delta\epsilon_{32} < 0$. However, condition (ii) and (v) are obviously incompatible, meaning there's no way to design the interactions of a three-writer-protein system that constitute an all-to-all network to simultaneously implement (1,0,1) and (0,1,1) configurations. It is, on other hand, possible to realize both (1,0,1) and (1,1,1) or both (0,1,1) and (1,1,1) by fine-tuning the interactions (e.g. changing the binding domains of proteins, see Fig.4·1C) such that there's no contradiction in the constraints of interactions, as we illustrate in Fig.4·2A. This observation imparts a critical message: for a given set of interactions (e.g. number of binding domain pairs, see Fig.4·1C) in a given signaling system with fixed number of protein species, be it writer or eraser, there is a inherent limit to the number of mutually compatible configurations. For instance, in the three-writer-species system we considered, this number is obviously smaller than $2^3 = 8$ since we already identified two pairs of mutually incompatible configurations. As we argued that in the digital logic how powerful a computational unit is, namely, it's capacity, depends crucially on the number of configurations it can implement and assign meaning (e.g. targeted Boolean function) to, one can define in parallel the cell signaling version:

Definition of capacity

The computational capacity C of a signaling network is the maximum number of collective species internal states (i.e. configurations) this network can realize, pursuant to the underlying biophysical laws.

It is obvious that if the activities of signal proteins that constitute the network are labeled by binary variables (e.g. 0 or 1), then $C \leq 2^n$.

4.3 Computational capacity of a signaling network in the large n limit

4.3.1 Writers only

In Appendix B, I show that

Result 1 (Computational capacity: writer only)**.** *Suppose that the biophysics governing the compatibility of any two proteins are independent. Then the capacity of computation C in the large n (i.e. number of proteins in the system) limit is given by*

$$C \succeq n^{-1/2} \left(\frac{4}{3}\right)^{\frac{n-1}{2}} \tag{4.6}$$

4.3.2 Systems with both writers and erasers

In the previous section we considered systems with n writer species (e.g. kinases). Here we incorporate eraser species (e.g. phosphatases) into our analysis. Focus on our synthetic system, now every species is either one of the two categories $\mathcal{C}_i \in \{\mathcal{W}, \mathcal{E}\}$, namely, writer or eraser. We order the species label such that the phosphorylation state vector $\mathbf{x} = (x_1, \cdots, x_m, x_{m+1}, \cdots, n)$ is partitioned into writer (kinase) sector $i \in \mathcal{W} = \{1, \cdots, m\}$ and the eraser (phosphatase) sector $i \in \mathcal{E} = \{m+1, \cdots, n\}$. In other words, such system contains m writer species and $n - m$ eraser species. Similar to the case without erasers, we can write down the probability of species i at state $x_i = 1$ (c.f. Eq.(4.3))

$$P_i(x_i = 1) = \sum_{j \text{ s.t. } \mathcal{C}_j = \mathcal{W}} \frac{x_j e^{-\beta[\Delta\epsilon(\alpha_i,\beta_j)]}}{1 + \sum_{k,\forall \mathcal{C}_k} x_k e^{-\beta[\Delta\epsilon(\alpha_i,\beta_k)(x_k)]}}. \tag{4.7}$$

Note that the summation over k in the denominator is over all erasers and writers $\mathcal{E}, \mathcal{W}$ whereas that over j in the numerator is over writers $\mathcal{W}$ only. In the zero temperature limit, this probability becomes binary so that we can use $\Delta\epsilon_{ij}$ to characterize the exponents in the Boltzmann factor. Similar to the argument made in the previous section, the interpretation of Eq.(C.40) reads:

$$P_i(x_i = 1)|_{\beta\to\infty} = \begin{cases} 1, \text{ if } \{\, \exists j \in \mathcal{W}_{\setminus i} \text{ s.t. } x_j = 1 \texttt{ AND } \Delta\epsilon_{ij} < 0 \,\} \\ \qquad \text{AND} \{\, \forall\ k \in \mathcal{E},\ x_k = 0 \texttt{ OR } \forall k \in \mathcal{E} \text{ s.t. } x_k = 1, \Delta\epsilon_{ik} > \Delta\epsilon_{ij} \,\} \\ 0, \text{ if } \{\, \forall j \in \mathcal{W}_{\setminus i},\ x_j = 0 \texttt{ OR } \forall j \in \mathcal{W}_{\setminus i} \text{ s.t. } x_j = 1, \Delta\epsilon_{ij} > 0 \,\} \\ \qquad \text{OR} \{\, \exists\ k \in \mathcal{E}, \text{ s.t. } x_k = 1 \texttt{ AND } \Delta\epsilon_{ik} < 0 \} \end{cases} \tag{4.8}$$

The biophysical interpretation of this equation line by line is summarized as follows.

Species i is phosphorylated if it is

1. At least bound to one active writer AND
2. All active erasers bind i loosely than active writers do

Whereas species i is not phosphorylated $(x_i = 0)$ if

3. There's no active writers present OR active writers exit but none of them binds i OR
4. It is bound to active erasers

Note that since in the zero temperature limit proteins are either in strongly bound or strongly unbound state and that case 1 dictates $\Delta\epsilon_{ij} < 0$, we can reasonably replace $\Delta\epsilon_{ik} > \Delta\epsilon_{ij}$ in the second line of Eq.(C.41) by $\Delta\epsilon_{ik} > 0$ (active erasers bind i extremely loose). With this we can rewrite Eq.(C.41) as

$$P_i(x_i = 1)|_{\beta\to\infty} = \begin{cases} 1, \text{ if } \{\, \exists j \in \mathcal{W}_{\backslash i} \text{ s.t. } x_j = 1 \texttt{ AND } \Delta\epsilon_{ij} < 0 \,\} \\ \qquad \text{AND} \{\, \forall\, k \in \mathcal{E},\ x_k = 0 \texttt{ OR } \forall k \in \mathcal{E} \text{ s.t. } x_k = 1, \Delta\epsilon_{ik} > 0 \,\} \\ 0, \text{ if } \{\, \forall j \in \mathcal{W}_{\backslash i},\ x_j = 0 \texttt{ OR } \forall j \in \mathcal{W}_{\backslash i} \text{ s.t. } x_j = 1, \Delta\epsilon_{ij} > 0 \,\} \\ \qquad \text{OR} \{\, \exists\, k \in \mathcal{E}, \text{ s.t. } x_k = 1 \texttt{ AND } \Delta\epsilon_{ik} < 0 \} \end{cases} \tag{4.9}$$

In Appendix B, I show the following:

Result 2 (Incompatibility probability: with eraser)**.** *Consider systems with n protein species among with m are writers and the remaining $n - m$ are erasers. Assume that the biophysics governing the compatibility of any two proteins are independent. Then under the writer-eraser symmetry, the probability of two configurations being incompatible $P_m(n)$, is given by*

$$P_m(n) = \left(\frac{2m}{3} \left[1 - \frac{m-1}{3} \left(\frac{2}{3} \right)^{m-2} \right] \left(\frac{3}{4} \right)^{n} + \frac{m(m-1)}{2^{n+2}} \right) + \begin{pmatrix} m \to n - m \\ n \to m \end{pmatrix} \tag{4.10}$$

Result 3 (Computational capacity: with eraser)**.** *Consider systems with n protein species among with m are writers and the remaining $n - m$ are erasers. Assume that the biophysics governing the compatibility of any two proteins are independent. Then the capacity of computation C in the large n (i.e. number of proteins in the system) limit is given by*

$$C \succeq P_m(n)^{-1/2} \tag{4.11}$$

One can easily check that in the limit where $m \to n$, Eq.(4.11) reduces to Eq.(4.6).

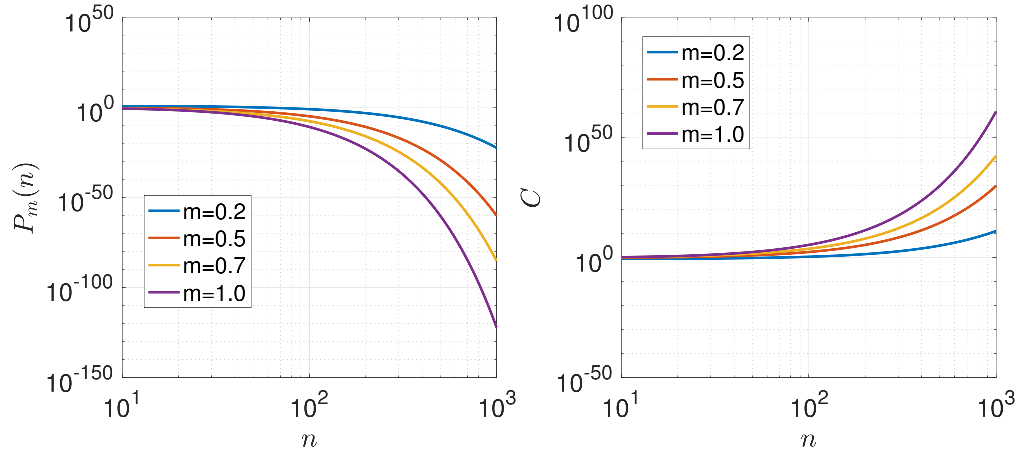


Figure 4·3: Incompatibility probability and capacity in synthetic systems with both writers and erasers: $P_m(n)$ is given by Eq.(C.46)

4.4 Discussion

We develop a framework to describe the computation process by which cells utilize to relay messages derived from extracellular cues to intracellular effectors, namely, signal transduction. We also demonstrate that such computation is fundamentally limited by the underlying interactions between as well as the modular nature of signal proteins, which are the molecules that carry out the relay. In particular, we derive the capacity of computation explicitly and show that it is far less than what a cell signaling system can accommodate following the combinatorial logic paradigm (i.e. exponential in the number of protein species). The computational capacity is also shown to be dependent on the fraction of writer proteins and the complexity of protein binding interactions, which reinforces the idea that the bottleneck stems from the cell signaling system itself. Since the origin of capacity is the thermodynamics of protein interactions, our result also underscores the importance of thermodynamics in computation(Bennett, 1982; Mehta et al., 2016), particularly pertaining to molecular processes within living cells.

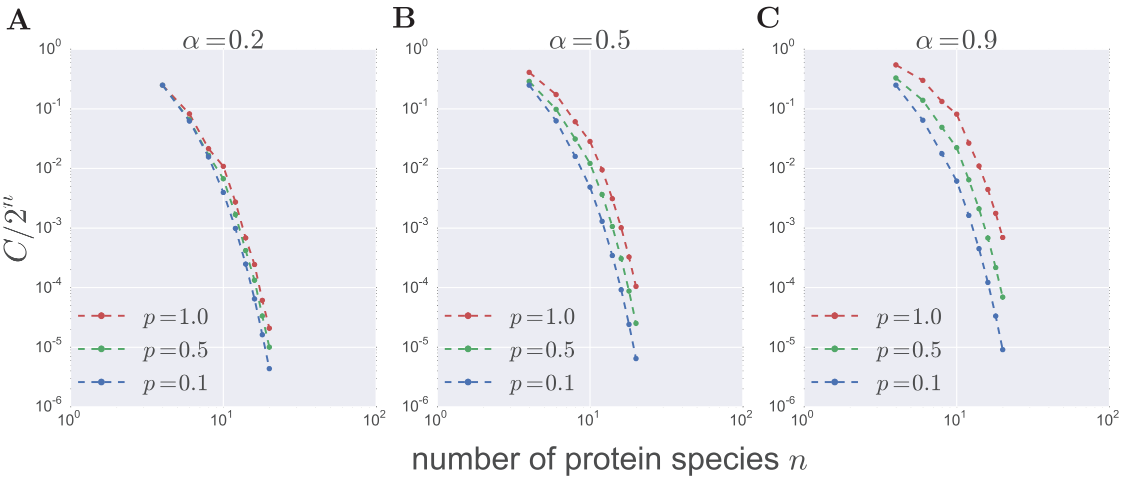


Figure 4·4: Effect of network sparsity on capacity measured by the fraction of biophysically realizable configurations: $C/2^n$, where n is the number of protein species. Shown here are Monte Carlo simulation of different set of network ensembles with different sparsity p (connectivity probability) and fraction of writers α. (*A*) $\alpha = 0.2$ (*B*) $\alpha = 0.5$ (*C*) $\alpha = 0.9$. In all panels, dots are results of Monte Carlo simulations. Dashed lines are shown only for clarity.

The theory we present presumes that the underlying signaling network is all-to-all connected, namely, every single protein is allowed to interact with one another, as long as the underlying energetics permits. This raises a question regarding the effect of network sparsity on capacity. To see how capacity scales with the number of protein species at signaling networks with different sparsity, we introduce in our Monte Carlo sampling procedure (*Methods*) a sparsity measure p (with $p = 1$ meaning all-to-all connected and $p \ll 1$ barely connected). In Fig.4·4, we show $P_C \equiv C/2^n$ for networks with different p at varying fraction of writers α. It is clear that higher connectivity always guarantees larger capacity. And this advantage is more pronounced when the network contains more writers. Concretely, P_C at sparsity $p = 1.0$ is two orders-of-magnitude of that at $p = 0.1$, when $\alpha = 0.9$ (see Fig. 4·4 C). In addition, the exponentially suppressed tail of P_C echos the result in Fig.4·3.

Generally speaking, the suppression of capacity can be understood by counting the number of parameters in a given signaling network. Consider a system of n species whose activities are labeled by binary variables (0 or 1), meaning each species can only be one of the two states. Combinatorially speaking, the maximum number of distinct configurations is 2^n. On the energetics side, however, the number of elements in the relative interaction matrix $F_{ij} = \epsilon_{ij} - k_B T \log[c_j]$ is n^2. Suppose elements of F_{ij} can only take value from a finite *flavor set* $\mathcal{M}$ with cardinality M, then the total number of energetics parameters is Mn^2. Therefore, the difference in the number of parameters in a system with n protein species between these two categories is $2^n - Mn^2 \gg 1$ for finite M. This implies that there will always be far less energetics parameters cells can utilize to carry out computation than the that enabled by the number of protein species (c.f. $2Mn \ll 2^n$). In other words, due to biophysical constraints cells always under exploit the combinatorial nature of digital computation. Therefore, the fraction of attainable computation, P_C, is strongly suppressed at large n.

This counting argument also highlights an important message pertaining to synthetic biology: in designing sophisticated synthetic circuit to perform complex computation (i.e. demanding high capacity C), it is always more advantageous to increase the domain interaction complexity M (e.g. having more binding domain pairs, see Fig.4·1B) rather than employing more proteins (i.e. increase merely n). This is because while utilizing more protein species certainly enhance the flexibility of design (e.g. more energetics parameters $\sim Mn^2$), our previous argument shows that this will introduce far too many configurations that even utilizing the increased energetics parameters can never implement, thus higher degree of under-exploitation.

On the methodology side, our result is reminiscent of Cover's theorem on func-

Table 4.1: Cover's perceptron vs. Cell signaling

Parameters	Perceptrons	Cell signaling
Capacity C	# of dichotomies	# of compatible config.
data $\mathbf{x}$	$\mathbf{x} \in \mathbb{R}^d$	binary protein act. vector
n	# of data points	# of protein species
d	data dimension	(q, q', m): writer & eraser
C_{max}	2^n (max. # of dichotomies)	2^n (max. # of config.)
Incomp.	geometry	biophysics

tion counting in statistical learning theory(Cover, 1965). This theorem places a bound on the number of dichotomies realizable by linear separators (i.e. perceptrons) for a generic set of n points in $\mathbb{R}^d$, provided that they satisfy general position (i.e. any subset of d or fewer points is linearly independent). Specifically, Cover showed that when the number of data points n is much greater than the data dimension d, the fraction of linearly separable points is exponentially suppressed in n, similar to P_C in cell signaling. Intuitively, linear separability depends on the number of data points n as well as the their dimension d, since they both affect the degree of freedom we can maneuver the separating hyperplane with to realize the desired dichotomies. In the context of cell signaling, this translates into the computational capacity is contingent on the variety of protein species n as well as how they interact through binding domain interactions (q, q', m), for they collectively dictates the physical implementation of computation (see Table 4.1). In as much the same way as in cell signaling that the suppression of capacity is attributed to the insufficient number of utilizable energetics parameters to exploit the combinatorial nature of computation, perceptron's separation capacity is drastically circumscribed due to the limited spatial degree of freedom d available to separate the colossally many data points n. Therefore, P_C in both cases bear similar degree of suppression at large n (see *Appendix B*). We include in Table 4.1 a comparison between the two realms of capacity.

Python class to generate signaling network and to infer capacity using this

Monte Carlo procedure is freely available at: `https://github.com/chinghao0703/MC_capacity`

Chapter 5

Example study: Thermodynamic paradigm of nuclear transport by living cells

> *'Scientific revolutions are inaugurated by a growing sense... that an existing paradigm has ceased to function adequately in the exploration of an aspect of nature to which that paradigm itself had previously led the way.'*
>
> (Thomas Kuhn, 1962)

The chapter is adapted from the following publication: *Thermodynamic paradigm for solution demixing inspired by nuclear transport in living cells*, Ching-Hao Wang, Pankaj Mehta, Michael Elbaum, *Phys. Rev. Lett* **118**, 158101 (2017)

Living cells display a remarkable capacity to compartmentalize their functional biochemistry. A particularly fascinating example is the cell nucleus. Exchange of macromolecules between the nucleus and the surrounding cytoplasm does not involve traversing a lipid bilayer membrane. Instead, large protein channels known as nuclear pores cross the nuclear envelope and regulate the passage of other proteins and RNA molecules. Beyond simply gating diffusion, the system of nuclear pores and associated transport receptors is able to generate substantial concentration gradients, at the energetic expense of guanosine triphosphate (GTP) hydrolysis. In contrast to conventional approaches to demixing such as reverse osmosis or dialysis, the biological system operates continuously, without application of cyclic changes in pressure or solvent exchange. Abstracting the biological paradigm, we examine this transport system as a thermodynamic machine of solution demixing. Building on the construct of free energy transduction and biochemical kinetics, we find conditions for stable operation and optimization of the concentration gradients as a function of dissipation in the form of entropy production.

5.1 Introduction

Demixing of solutions is a difficult thermodynamic problem with important practical consequences (Dijkstra and Frenkel, 1994). Examples include the desalination of seawater, medical dialysis, and chemical purification. In all of these processes, free energy is consumed in order to balance entropy of mixing. Typical engineering approaches to demixing involve application of hydrostatic pressure (reverse osmosis), solution exchange (dialysis), or phase change (crystallization or distillation) (Mistry et al., 2011; Glynn and Reardon, 1990). In this context living cells adopt a fundamentally different paradigm by establishing and maintaining concentration gradients at *steady-state* under a fixed set of intrinsic thermodynamic parameters. This recalls the similar capacity to operate mechanochemical motors isothermally (Parmeggiani et al., 1999; Parrondo et al., 2000).

A prominent example of molecular separation is the eukaryotic cell nucleus, wherein the concentrations of many proteins and RNA differ significantly from those in the cell body (cytoplasm). These gradients are maintained by a transport system that shuttles molecular cargo in and out via large protein channels known as nuclear pores (Maul and Deaven, 1977; Talcott and Moore, 1999). This system has been under intensive study in the biological (Görlich and Kutay, 1999; Stewart, 2007; D'Angelo and Hetzer, 2008; Wente and Rout, 2010; Kimura and

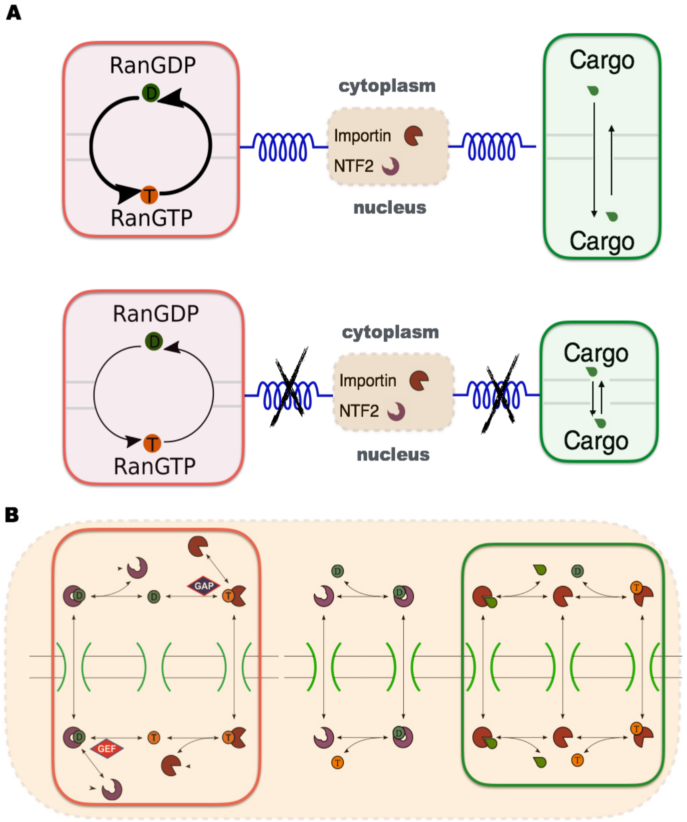

A
RanGDP
D
T
RanGTP
cytoplasm
Importin
NTF2
nucleus
Cargo
Cargo
RanGDP
D
T
RanGTP
cytoplasm
Importin
NTF2
nucleus
Cargo
Cargo
B
GAP
GEF
D
T

Figure 5·1: (Color online) Demixing of cargo across the nuclear membrane is driven by Ran coupled to NTF2 and importin system. (A) With such coupling (upper panel), nuclear cargo accumulation is favored and Ran GTP/GDP exchange cycle proceeds faster than without coupling (lower panel). The thickness of arrowed curves in Ran cycle indicates strength of reaction flux; length of arrowed lines in cargo transport represents the rate at which the underlying processes occur. (B) Details of molecular demixing machine in the context of nuclear transport. Species labels as above. Reactions corresponding to Ran cycle and cargo transport are highlighted by red and green boxes, respectively. The orange dashed box includes all reactions coupled by the importin-NTF2 system. See also Fig S1 in the Supplementary Material.

Imamoto, 2014) and biophysical (Peters, 2003; Kopito and Elbaum, 2007; Wagner et al., 2015; Zahn et al., 2016; Vovk et al., 2016) literatures, with particular emphasis on single-molecule interactions at the pore itself (Keminer and Peters, 1999; Yang et al., 2004; Kubitscheck et al., 2005; Grünwald et al., 2011). Simple thermodynamic considerations make clear that equilibrium pore-molecule interactions are insufficient to support concentration gradients in solution. Demixing between two compartments cannot occur spontaneously, but must be coupled to a free energy source (Hill, 2012). At the same time, demixing does not require rectified translocation (Görlich et al., 2003). Concentration gradients may be established in the presence of a balanced, bi-directional exchange (Kopito and Elbaum, 2007; Kopito and Elbaum, 2009; Lolodi et al., 2016).

Nuclear pores represent an unusual transporter in that there is no membrane to cross. Water, ions, and small molecules diffuse freely across the nuclear envelope to equilibrate between the two compartments. Generally, the permeability drops between molecular weight 20 kDa and 40 kDa (Peters, 1984; Samudram et al., 2016). Transport of larger macromolecules relies on a special class of proteins, called transport receptors (i.e. "importin"), that usher their cargoes across the nuclear pores by virtue of specific interactions with the channel components. Recognition between importins and their molecular cargo depends on the presence of particular amino acid sequences known as nuclear localization signals (NLS) (Grote et al., 1995; Görlich and Kutay, 1999; Rexach and Blobel, 1995). The affinity between importin and cargo is regulated by a small GTP-binding protein called Ran (Smith et al., 2002; Terry et al., 2007). When associated with GTP (RanGTP), Ran binds strongly to importin in a manner that is competitive to NLS binding. By contrast, Ran associated with GDP (RanGDP) binds importin very weakly. Ran interconverts between these two forms through GTP hydrolysis and GTP/GDP exchange, facilitated by the GTPase Activating Protein (RanGAP) and the Guanosine Exchange Factor (RanGEF), respectively (Bos et al., 2007). RanGAP is structurally bound to the cytoplasmic face of the nuclear pore and RanGEF is bound to chromatin. Their activities generate a high concentration of RanGTP in the nucleus and RanGDP in the cytoplasm (see FIG.5·1).

Demixing is powered by transducing free energy from GTP hydrolysis through the interactions of transport receptor with Ran. The transport machinery has been formulated in terms of coupled chemical kinetics (Görlich et al., 2003; Riddick and Macara, 2005; Kim and Elbaum, 2013b) but the energetics have not yet been addressed. In particular, we ask: How does the rate of dissipation (energy consumption) relate to the achieved concentration gradient? What is the proper definition of transport efficiency? Is there an optimal working point given the nonequilibrium nature of this cellular machine? To address these questions, it is helpful to reformulate the problem in a thermodynamic language. For consistency with the literature we retain the biological nomenclature, yet the aim is to understand the natural engineering in a more abstract sense that might ultimately be implemented synthetically.

5.2 Thermodynamic formulation

In the thermodynamic formulation, a central role is played by energy transduction in a "futile cycle" among the components (see FIG. 5·1). This is roughly analogous to heat flow in a Carnot cycle. The importin receptor binds RanGTP, and a second receptor known as nuclear transport factor 2 (NTF2) binds specifically RanGDP. The forward cycle takes RanGTP out to the cytoplasm with importin and RanGDP back to the nucleus with NTF2. Detailed balance is broken by the distribution of RanGAP and RanGEF as described above, so that the reverse cycle is scarcely populated.

Free energy from the Ran cycle is transduced by importin to bias the steady-state free cargo concentrations in the nuclear and cytoplasmic compartments. Details of the underlying biochemical reactions are shown in FIG. 5·1B and can be modeled on the basis of mass action. The corresponding kinetic parameters can be found from the literature or estimated from simple scaling arguments (see FIG. S1 and Supplementary Material for details of the kinetic model(sup,)). Numerical solutions are obtained by solving all the coupled rate equations using a standard Runge-Kutta method (The code used for simulation is available in the Supplementary Material). We emphasize that the present aim is not so much to model the biological implementation as to explore the generic operation of the thermodynamic machine. Relations between parameters are therefore more important than specific values.

Energetics enter the model via the charging of Ran with GTP and its subsequent hydrolysis to GDP (reactions 5 and 2 in FIG. S1, respectively). The flux through these two reactions must be equal in steady state. Energy is drawn from the non-equilibrium ratio of free GTP to GDP, θ, which is maintained by cellular metabolism and defines an effective "free energy" $F_\theta := k_B T \log(\theta)$. A typical value of θ is roughly a few tens to a hundred (Garrett and Grisham, 2008; **?**). Independent of the complex operational details of RanGEF and RanGAP with associated co-factors, we can look at the steady states and relate the reactions to θ. (See Supplementary Material for details.) On the nuclear side, the complex NTF2-RanGDP exchanges for NTF2 and RanGTP. The dissociation constant K_D (forward divided by reverse flux) can be shown to be proportional to θ. Conversely, on the cytoplasmic side the corresponding K_D is proportional to $1/\theta$. As a result, any enhancement of flux through the futile cycle in the forward reaction conferred by

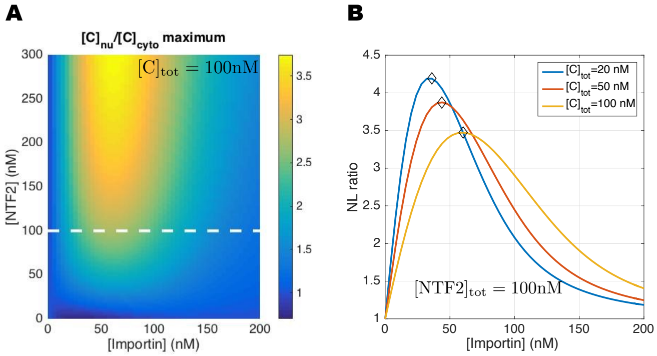


Figure 5·2: (Color online) Phase diagram of nuclear localization. (A) The cargo nuclear localization NL:= $[C]_{nu}/[C]_{cyto}$ (color shadings) is obtained by varying the total importin and total NTF2 concentrations while keeping overall cargo level fixed at $[C]_{tot} = 100$ nM. (B) A family of curves shows NL for several cargo concentrations as a function of importin concentration with $[NTF2]_{tot} = 100$ nM. The 1D curve for $[C]_{tot} = 100$ nM is a cut across the plot of panel A. Locations of NL maximum are marked by diamonds (see FIG.5·4C as well). Kinetic rate constants used are given in the Supplementary Material. Total Ran concentration $[Ran]_{tot}$=75 nM.

increasing θ (i.e. reaction 5 in FIG.S1C) is balanced by the contradicting counterpart in preventing RanGTP release (i.e. reaction 2 in FIG.S1C).

A useful measure of cargo demixing is the nuclear localization ratio, NL, defined as the ratio between nuclear and cytoplasmic cargo concentrations: $[C]_{nu}/[C]_{cyto}$. This ratio defines a chemical potential, $\Delta\mu = -k_B T \log [C]_{nu}/[C]_{cyto}$, that measures the excursion from equilibrium. FIG. 5·2A shows NL as a function of importin and NTF2 concentrations. The most striking feature is that NL is maximum for intermediate levels of importin. The importin concentration at which NL is maximized, $[Im^*]$, grows with the total cargo load, $[C]_{tot}$ (see FIG. 5·2B). Furthermore, $[Im^*]$ is largely independent of NTF2 concentration for different cargo concentration considered (see FIG. S5). This suggests an inherent optimization.

At first sight it is surprising that augmenting the importin concentration, which increases the number of molecules that can transport cargo to nucleus, may decrease the localization ratio. The optimal dependence of NL on importin reflects the dual role importin plays as the *inbound carrier of cargo protein* as well as the *outbound carrier of RanGTP*. Powering the futile cycle requires that importin bind RanGTP, whereas cargo transport requires importin to bind cargo. This establishes a binding competition in the nucleus that is a characteristic feature of protein import (FIG.5·3A). In spite of the higher affinity of RanGTP for importin, the cycle analysis shows that importin in the nucleus binds cargo more rapidly. As seen in FIG. 5·3BC, NL is maximized close to the point at which the difference between the reaction fluxes of importin-cargo formation ($\Phi_7^- := \tilde{\Phi}_7^-[Im]_{nu} = (k_7^-[C]_{nu})[Im]_{nu}$) and importin-RanGTP formation ($\Phi_4^+ := \tilde{\Phi}_4^+[Im]_{nu} = (k_4^+[RanGTP]_{nu})[Im]_{nu}$) is maximal. Intuitively, this is the realm where importin can bind cargo effectively while maintaining its coupling to the reaction cycle that transduces energy for cargo transport.

5.3 Entropy production of futile cycle

To understand the thermodynamics of nuclear transport, we formulate the transport system as a nonequilibrium Markov process. Since a nonequilibrium steady state (NESS) necessarily breaks detailed balance in the underlying Markov process, the system has a nonzero entropy production (Hill, 2012; Mehta and Schwab, 2012; Lang et al., 2014). This is the energy per unit time required to maintain the NESS, with units of power. Following the Schnakenberg description, the EP for a NESS is given by (Lebowitz and Spohn, 1999)

$$EP = k_B T \sum_{i,j} P_i^{SS} W(i,j) \log \frac{W(i,j)}{W(j,i)}, \tag{5.1}$$

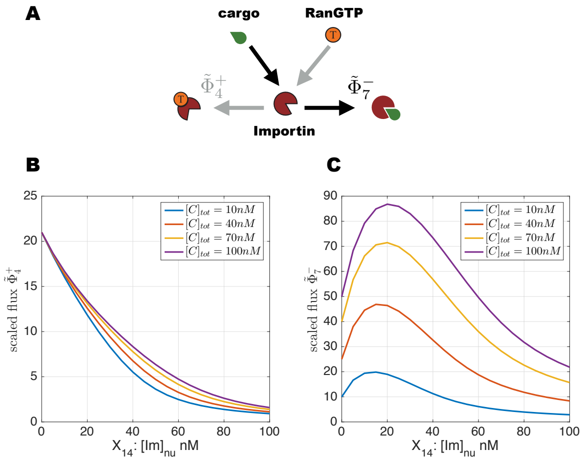


Figure 5·3: (Color online) Competition between RanGTP and cargo to bind importin. (A) Schematic of the two competing reactions. (B) Reaction flux for importin-RanGTP formation $\tilde{\Phi}_4^+ \sim k_4^+[RanGTP]$ and (C) flux for importin-cargo formation $\tilde{\Phi}_7^- \sim k_7^-[C]_{nu}$. Fluxes are scaled by $[Im]_{nu}$ (see text). Parameters as in FIG.5·2

where P_i^{SS} is the steady state probability distribution of state i while $W(i,j)$ denotes the transition probability from state i to state j. Concretely, P_i^{SS} is the fraction of reactants that participate in the transition reaction starting from state i while $W(i,j)$ can be calculated from the relevant reaction fluxes. Note that the sum in Eq.(5.1) is taken over all links of the reaction network. This is equivalent to summing over the links pertaining to the Ran futile cycle. (See Supporting Material for details).

This entropy production provides a direct measure of the power input to the underlying biochemical circuit. FIG. 5·4A shows EP for various importin and NTF2 concentrations. FIG. 5·4B adds various cargo concentrations for a fixed level of $[NTF2]$. In each case, as the importin concentration increases, EP first drops to a minimum and then peaks before slowly decaying. Note that the minimal dissipation (entropy production) tracks closely with the value at which the NL ratio peaks (see FIG.5·4C). These conditions define an optimal efficiency of the demixing machine. With further increasing importin concentration, the futile cycle decouples from cargo translocation and EP increases. At a still higher concentration, EP peaks and then decreases. This can be understood qualitatively as a short circuit via reaction 8, where importin moves between compartments carrying neither cargo nor RanGTP. As seen in FIG.5·4D, for such high importin levels the corresponding flux Φ_8 exceeds that of the RanGTP loading to importin, Φ_4.

5.4 Discussions

To the best of our knowledge the optimal steady-state has not been observed experimentally. The *kinetic* rate of nuclear protein uptake was found to be reduced by microinjection of importin receptor to live cells; rate equation simulations done in parallel also pointed to the dual role of importin (FIG. 5·3A) (Ribbeck and Görlich, 2001). Steady-states were not reported in that study, however. Other possible experimental tests include titration of importin protein to nuclei reconstituted in vitro in *Xenopus* egg extract and optical activation of importin receptors, similarly to induction of nuclear transport by NLS activation (Niopek et al., 2014). An important point in comparison with literature is that we have considered a single, collective "cargo" for transport. In reality, multiple cargoes compete for binding to relatively few but promiscuous transport receptors. This competition leads to a partitioning according to equilibrium binding affinities and may lead to vastly different kinetics. However the steady-state NL ratio (in solution) is independent of the affinity, reflecting thermodynamic control and equilibration of the chemical potentials (Kopito and Elbaum, 2009; Kim and Elbaum, 2013b; Lolodi et al., 2016). Consistent with this paradigm, in which a net accumulation occurs together with a balanced bidirectional flux, the simulations show that the nuclear and cytoplasmic concentrations of the importin-cargo complex (X_4 and X_{11}, respectively) equilibrate in

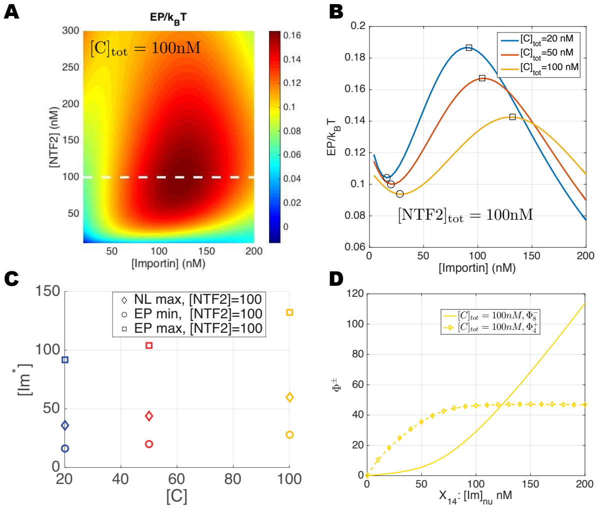


Figure 5·4: (Color online) Phase diagram of entropy production. (A) Entropy production is plotted as a 2D function of NTF2 and importin, at fixed cargo concentration $[C]_{tot} = 100$ nM. Compare with FIG. 2A. (Axes extend to 5 instead of 0 nM to avoid numerical divergence.) (B) A family of curves shows the entropy production for several cargo concentrations as a function of importin concentration; NTF2 concentration fixed at $[NTF2]_{tot} = 100$ nM. The 1D curve for $[C]_{tot} = 100$ nM is a cut across the plot of panel A. Compare with FIG. 2B. Peaks and troughs are marked by squares and circles, respectively. (C) Locations of entropy production max/min (square/circle) and that of nuclear localization maximum (diamond). Colors match curves in panel B. The importin concentration at which EP is minimum is close to but always less than $[Im^*]$, where NL is maximum. Thus, one strategy for maximizing the efficiency of demixing is to have the futile cycle operate in regime where its entropy production is minimized. (D) EP decreases at very high importin concentration. This reflects a loop around the energetic reaction Φ_4 via the reversible reaction Φ_8. Here $[NTF2]_{tot}$=100 nM, as in the panel B. In all panels, $[Ran]_{tot} = 75$ nM. Kinetics constants as in FIG.5·2 and 5·3 (see SM Section II).

steady-state. It is also interesting to note that RanGTP loading onto importin (reaction 4) was identified in the earlier analysis as the primary rate-limiting step in accumulation kinetics (Kim and Elbaum, 2013b).

In summary, we have analyzed the biological paradigm for nuclear transport from a thermodynamic point of view. Building upon prior understanding that protein cargo demixing is facilitated by hydrolysis of GTP, we draw the connection between consumption of chemical energy and maintenance of the cargo concentration gradient at non-equilibrium steady states. We show that the efficacy of nuclear localization ratio peaks at intermediate importin level, which is not far from the power consumption (entropy production) minimal. It is likely that the cell maintains an importin concentration at an advantageous level with respect to these operating points defined by the thermodynamic analysis. Interestingly, the system as configured is robust to the quality of the chemical energy source, in the sense that the NL ratio is almost independent of the GTP:GDP ratio θ when $\theta \gtrsim 20$, FIG. S4. A thermodynamic definition of the system efficiency remains elusive, however. Whereas conventional efficiency of an engine is a dimensionless ratio of mechanical to thermal power, in the NESS a constant free energetic gradient (chemical potential in the present case) is maintained by a constant power input. The ratio has units of time. This could be renormalized sensibly by a characteristic remixing time, e.g., the permeability of the nuclear pores to the cargo-importin complex. There is no guarantee of a bound at unity, however, so the definition remains ad hoc, a useful figure of merit. It is also interesting to contrast the competitive interactions between receptor and RanGTP in nuclear protein accumulation (import) with the cooperative interactions in nuclear protein depletion (export). While these are often considered as simple inverse processes, they differ in this essential aspect (Kim and Elbaum, 2013a).

This work is part of a larger literature that seeks to examine basic biophysical processes from a thermodynamic perspective. It is now clear that thermodynamics fundamentally constrains the ability of cells to perform various task ranging from detecting external signals (Berg and Purcell, 1977; Endres and Wingreen, 2009; Mora and Wingreen, 2010), to adaptation (Sartori et al., 2014), to making fidelity decisions (Lang et al., 2014), generating oscillatory behavior (Elowitz and Leibler, 2000), and of course generating forces and dynamic structures (Nédélec et al., 1997; Surrey et al., 2001; Karsenti, 2008). In all these examples, it is possible to map these tasks to Markov processes and compute the corresponding entropy production rate. This suggests that there may be general theorems about thermal efficiency in cells that are independent of the particular task under consideration (Mehta and Schwab, 2012; Mehta et al., 2016; Barato and Seifert, 2015; Gingrich et al., 2016). It will be interesting to explore if this is actually the case and to see if these principles can be applied to synthetic biology and ultimately biomimetic engineering (Mehta et al., 2016).

Appendix A

Appendix of Chapter 2

A.1 Details of InfoMax

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 2·5, 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 2.

A.2 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

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 $\{\theta^{(t)}\}$
```

binary random variables that take value 1 when catalytically active and 0 otherwise. Pictorially, this pathway can be summarized as the following channel: $c \rightarrow X \rightarrow Y$. 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, $Q(c) \equiv P(Y = 1|c)$. Specifically,

$$P(Y = 1|c) = \sum_{X \in \{0,1\}} P(Y = 1|X) P(X|c) \tag{A.1}$$

$$= \underbrace{\left(\frac{e^{-\theta_x}}{1 + e^{-\theta_x}}\right)}_{\eta_b} f(c) + \underbrace{\left(\frac{e^{-W}}{1 + e^{-W}}\right)}_{\eta_W} (1 - f(c)), \tag{A.2}$$

where $f(c) \equiv P(X = 1|c)$ is the probability that kinase X is phosphorylated in the presence of a ligand at concentration c. The functional form of $f(c)$ 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 $f(c) = 1$ (i.e. full input signal), $Q(c)$ is purely dictated by bindings of phosphorylated kinase to its substrate (i.e. θ_x, see Fig. 2·1D), whereas when $f(c) = 0$ (i.e. no input signal), the contribution is solely from the those involving unphosphorylated kinase (i.e. W, see Fig. 2·1D). Therefore, we define the signal-to-noise ratio (SNR) formally as:

$$SNR \equiv \frac{\langle Q(c=1)\rangle}{\langle Q(c=0)\rangle} = \frac{\langle \eta_b \rangle}{\langle \eta_W \rangle}, \tag{A.3}$$

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,

$$\eta_b = \frac{1}{1 + e^{\theta_x}} \approx 1 - e^{\theta_x} \equiv 1 - Z, \tag{A.4}$$

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

$$\begin{aligned} \langle \eta_b \rangle = \mathbb{E}[1 - Z] &= \int_0^\infty dz\, (1 - z) \frac{1}{z\sigma\sqrt{2\pi}} \exp\left[-\frac{(\log z - \mu)^2}{2\sigma^2} \right] \\ &= 1 - \exp\left(\mu + \frac{\sigma^2}{2} \right) \end{aligned} \tag{A.5}$$

and

$$\begin{aligned} \langle \eta_b^2 \rangle = \mathbb{E}[(1 - Z)^2] &= \int_0^\infty dz\, \frac{1 - 2z + z^2}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(\log z - \mu)^2}{2\sigma^2} \right] \\ &= 1 - 2\exp\left(\mu + \frac{\sigma^2}{2} \right) + \exp[2\left(\mu + \sigma^2 \right)]. \end{aligned} \tag{A.6}$$

From this one can also derive its variance

$$\begin{aligned}\mathrm{Var}(\eta_b) &= \langle \eta_b^2 \rangle - \langle \eta_b \rangle^2 \\ &= e^{2(\mu+\sigma^2)}(1 - e^{-\sigma^2}). \end{aligned} \tag{A.7}$$

These quantities can be used to analyze the effect of heterogeneity in θ_X on $Q(c)$ which we summarized in Fig. A·1. 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

$$\mathrm{SNR} = \frac{\langle \eta_b \rangle}{\langle \eta_W \rangle} = e^{\beta W}\left[1 - e^{\beta\left(\mu + \frac{\sigma^2}{2}\right)}\right]. \tag{A.8}$$

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:

$$\langle \eta_b \rangle = \frac{1}{\sigma\sqrt{2\pi}} \int_0^\infty d\lambda\, \frac{1}{1+\lambda} \exp\left[-\frac{(\log\lambda + \mu)^2}{2\sigma^2}\right] \tag{A.9}$$

$$\langle \eta_b^2 \rangle = \frac{1}{\sigma\sqrt{2\pi}} \int_0^\infty d\lambda\, \frac{\lambda}{(1+\lambda)^2} \exp\left[-\frac{(\log\lambda + \mu)^2}{2\sigma^2}\right]. \tag{A.10}$$

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

$$\langle \eta_b \rangle_{M \gg 1} \approx \frac{e^{-\mu}}{1 + e^{-\mu}} \tag{A.11}$$

$$\langle \eta_b^2 \rangle_{M \gg 1} \approx \left(\frac{e^{-\mu}}{1 + e^{-\mu}}\right)^2, \tag{A.12}$$

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

$$\text{SNR}_{M \gg 1} \approx \frac{e^{-\beta(\mu - W)}}{1 + e^{-\beta\mu}}. \tag{A.13}$$

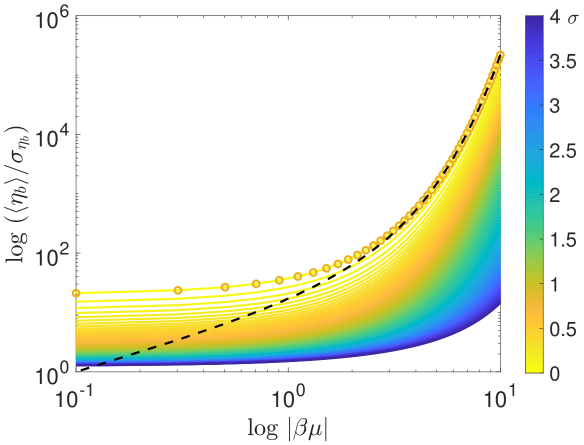


Figure A·1: Inverse coefficient of variation (CV^{-1}) of $Q(c)$ assuming negligible η_W is shown for different value of $\mu \equiv \langle\theta_{i,j}\rangle$ and $\sigma^2 \equiv \text{Var}(\theta_{i,j})$, where $\theta_{i,j}$ is the binding affinity of K_i to K_j and the average $\langle\cdot\rangle$ 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.(A.9)(A.10), open circles are calculated via sampling network parameters as described in Methods. Black dashed curve is the analytical result in the tight-binding limit whose expression is given by Eq.(A.5)(A.7) with σ corresponding to that of the yellow open circles ($\sigma = 0.1$). Note the logarithmic scale.

A.3 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

$$P(\mathbf{x}) := P(x_0, \cdots, x_{n+1}) = \left(\prod_{i=0}^{n} P(x_{i+1}|x_i) \right) P(x_0), \tag{A.14}$$

where the conditionals are given by

$$P(x_{i+1} = 1|x_i) = \frac{x_i e^{-\theta_{i,i+1}}}{1 + x_i e^{-\theta_{i,i+1}}}. \tag{A.15}$$

Note that we denote the *relative binding affinity* of i to $i+1$ as $\theta_{i,i+1}$ (c.f. Eq. (A.35)) . 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:

$$\mathcal{M}_{i+1,i} = \begin{pmatrix} P(x_{i+1} = 1|x_i = 1) & P(x_{i+1} = 1|x_i = 0) \\ P(x_{i+1} = 0|x_i = 1) & P(x_{i+1} = 0|x_i = 0) \end{pmatrix} \tag{A.16}$$

To calculate the mutual information between x_1 and x_n,

$$I(x_0; x_{n+1}) = \sum_{x_0} \sum_{x_{n+1}} P(x_0) P(x_{n+1}|x_0) \log_2 \left[\frac{P(x_{n+1}|x_0)}{P(x_{n+1})} \right], \tag{A.17}$$

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

$$\begin{aligned} P(x_{n+1}|x_0) &= \sum_{x_1} \cdots \sum_{x_n} \prod_{i=0}^{n} P(x_{i+1}|x_i) \\ &= \sum_{x_n} P(x_{n+1}|x_n) P(x_n|x_{n-1}) \cdots \sum_{x_2} P(x_3|x_2) \sum_{x_1} P(x_2|x_1) P(x_1|x_0) \\ &= \prod_{i=0}^{n} \mathcal{M}_{i+1,i} \equiv \mathbf{P}_{n+1,0}, \end{aligned} \tag{A.18}$$

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

$$\mathbf{p}_{n+1} = \sum_{x_0} P(x_{n+1}|x_0)P(x_0) = \left(\prod_{i=0}^{n} \mathcal{M}_{i+1,i}\right)\mathbf{p}_0 = \mathbf{P}_{n+1,0}\mathbf{p}_0, \tag{A.19}$$

where $\mathbf{p}_{n+1} \equiv (P(x_{n+1}=1), P(x_{n+1}=0))^T$, and similarly for $\mathbf{p}_0$. With this defined and for a given set of $\theta_{i,j}$, we can calculate mutual information Eq. (A.17) 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. (A.16) approximates the identity matrix $\mathbf{I}$ due to Eq. (A.15). From this, one can easily show that mutual information averaged over different realizations is given by:

$$\langle I(x_0; x_{n+1})\rangle = -q\log_2 q - (1-q)\log_2(1-q) \equiv H_2(q), \tag{A.20}$$

where $q \equiv P(x_1 = 1)$, 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 A·2.

A.4 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

$$b_n = \prod_{i=1}^{n-1} f_i = \prod_{i=1} \left(\frac{e^{-\theta_{i,i+1}}}{1+e^{-\theta_{i,i+1}}}\right). \tag{A.21}$$

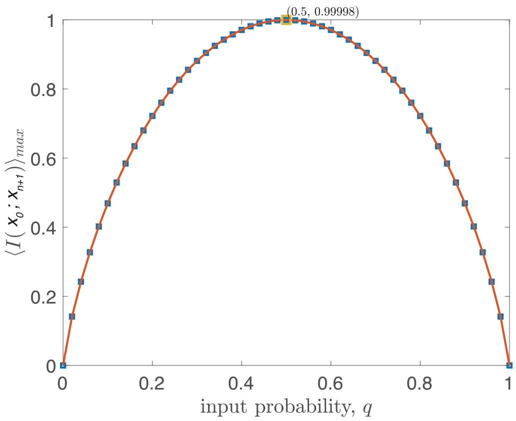


Figure A·2: At tight-binding mutual information for linear network reduces to binary entropy function. Linear network of depth $n = 8$ is used. Red squares are obtained by averaging the result over 100 different realizations of binding affinities using the methods detailed in this appendix. Dashed black curve is plotted using Eq. (A.20). Parameters used are: $\beta\mu = -5, \sigma = 0.1$. The value and location of maximum mutual information obtained by averaging is indicated as $(q, I_{\max}) = (0.5, 0.99998)$.

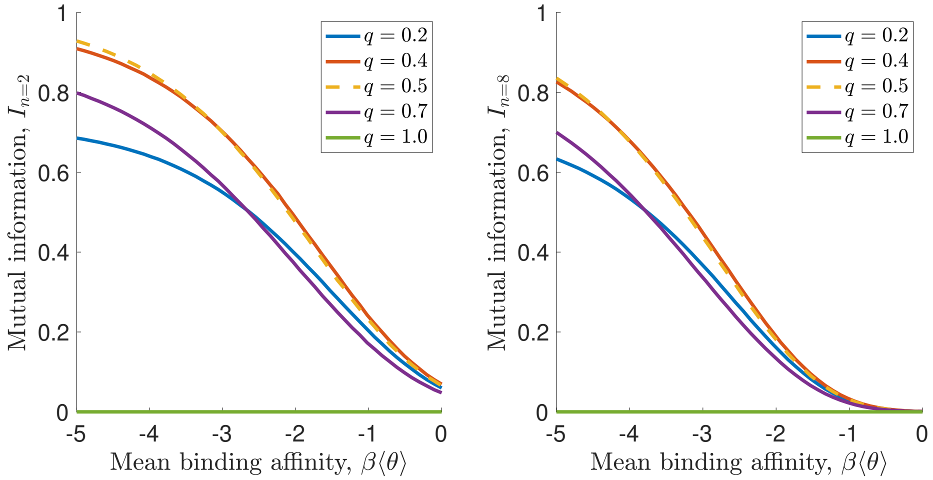


Figure A·3: Mutual information as a function of mean binding affinity $\mu \equiv \langle\theta\rangle$ at different inputs. Note that maximum mutual information may not occur at $q = 0.5$ (dashed orange curve) away from tight-binding (i.e. less negative $\beta\langle\theta\rangle$). In all panels, $\sigma = 0.1$.

From this, one can re-write Eq. (A.18) and Eq. (A.19) as

$$\mathbf{P}_{n,1} = \begin{pmatrix} b_n & 0 \\ 1-b_n & 1 \end{pmatrix}, \qquad \mathbf{p}_n = \begin{pmatrix} q\,b_n \\ 1-q\,b_n \end{pmatrix}. \tag{A.22}$$

Plugging this back in to the definition of mutual information,Eq. (A.17), one gets,

$$I(x_1; x_n) = -qb_n \log q + q(1-b_n)\log(1-b_n) - (1-q\,b_n)\log(1-q\,b_n) \tag{A.23}$$

After taking the derivate of Eq. (A.23) 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:

$$b_n \log\left(\frac{q^\star}{1-q^\star b_n}\right) = (1-b_n)\log(1-b_n) \tag{A.24}$$

A.5 Relating thermodynamics to a kinetic model of phosphorylation cascade

Here we derive the Eq.(2.3) in the main text (re-written as Eq. (A.34) here) from chemical kinetics. Following Fig. A·6, 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 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 path-

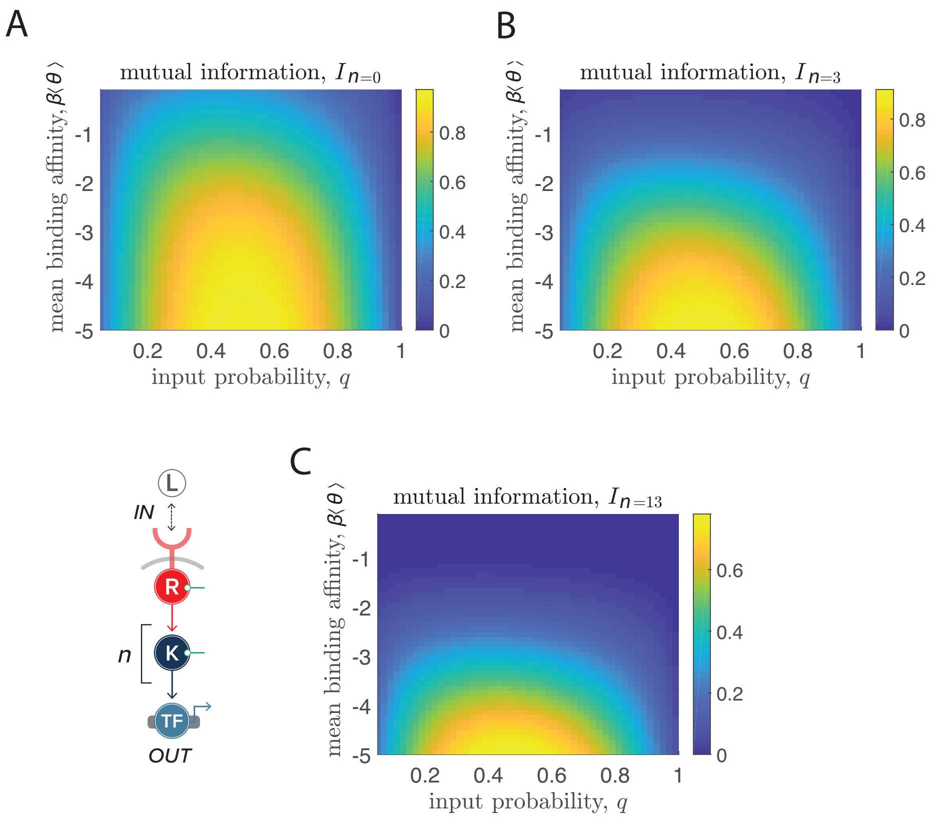


Figure A·4: Tighter binding always increases information transmission for cascades of depths $n = 0, 3, 13$. Here $\mathrm{Var}(\theta_{i,j}) \equiv \sigma^2 = 0.01$. All panels are generated by averaging 100 realizations of binding affinities using the scheme detailed in Appendix A 2.

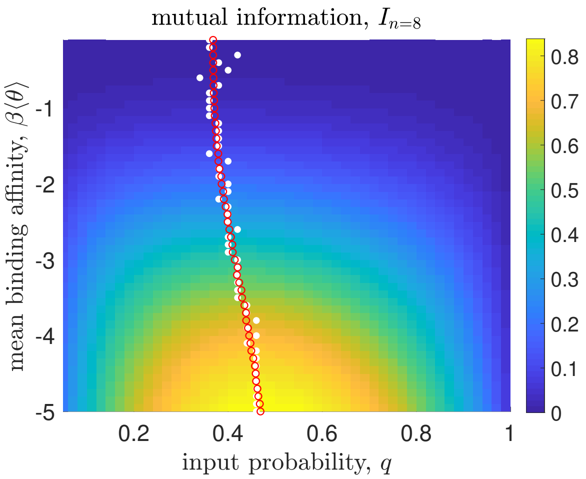


Figure A·5: Location of maximum mutual information is not necessarily at input with most uncertainty ($q = 0.5$). White filled circles are calculated by numerically searching for the input $q^\star$ on this color map that gives maximum mutual information. Red open circles are obtained through solving Eq. (A.24). 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 Appendix A 2.

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

$$\begin{aligned} \frac{dX_i}{dt} &= \Phi_i^+ - \Phi_i^- && \text{(A.25)} \\ &= \tilde{\alpha}_i X_{i-1} \tilde{X}_i - \beta X_i && \text{(A.26)} \\ &= \alpha_i X_{i-1}\left(1 - \frac{X_i}{C_i}\right) - \beta_i X_i, && \text{(A.27)} \end{aligned}$$

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

$$\frac{dX_1}{dt} = \alpha_1 R(t)\left(1 - \frac{X_1}{C_1}\right) - \beta_1 X_1. \quad \text{(A.28)}$$

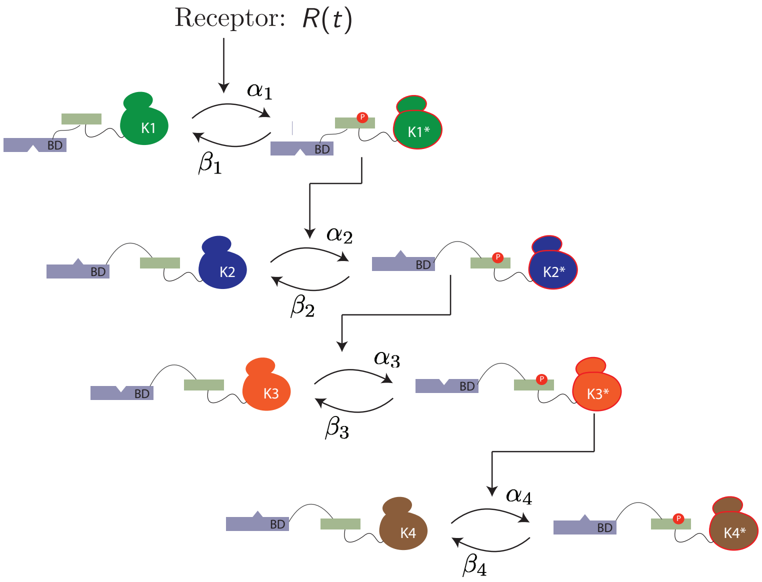

Receptor: $R(t)$
α_1
β_1
BD
K1
K1*
P
α_2
β_2
K2
K2*
α_3
β_3
K3
K3*
α_4
β_4
K4
K4*

Figure A·6: 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 $K_i \rightarrow K_{i+1}$.

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

$$X_i^{SS} = \frac{C_i X_{i-1}^{SS}}{\gamma_i C_i + X_{i-1}^{SS}}, \tag{A.29}$$

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:

$$p_i^{SS} = \frac{X_{i-1}^{SS} (\gamma_i C_i)^{-1}}{1 + X_{i-1}^{SS} (\gamma_i C_i)^{-1}}. \tag{A.30}$$

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

$$X_{i-1}^{SS} \rightarrow S \tag{A.31}$$

$$C_i \rightarrow V_{\max} \tag{A.32}$$

$$K_m \rightarrow C_i \gamma_i \tag{A.33}$$

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

$$P(x_i = 1 | x_{i-1}) = \frac{x_{i-1} e^{-\theta_{i-1,i}/(k_B T)}}{1 + x_{i-1} e^{-\beta\theta_{i-1,i}/(k_B T)}}, \tag{A.34}$$

can be interpreted under this kinetics frameworks by relating

$$\theta_{i-1,i} = k_B T \ln \left(\frac{K_m}{X_{i-1}^{SS}} \right) = \Delta F - \tilde{\mu}_{i-1}, \tag{A.35}$$

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 $i-1$.

A.6 Effects of network depth

Here we examine how the depth of network affects information transduction capacity. According to data processing inequality (DPI)(Cover and Thomas, 2012; **?**), information is never gained when transmitted through some noisy channel (or

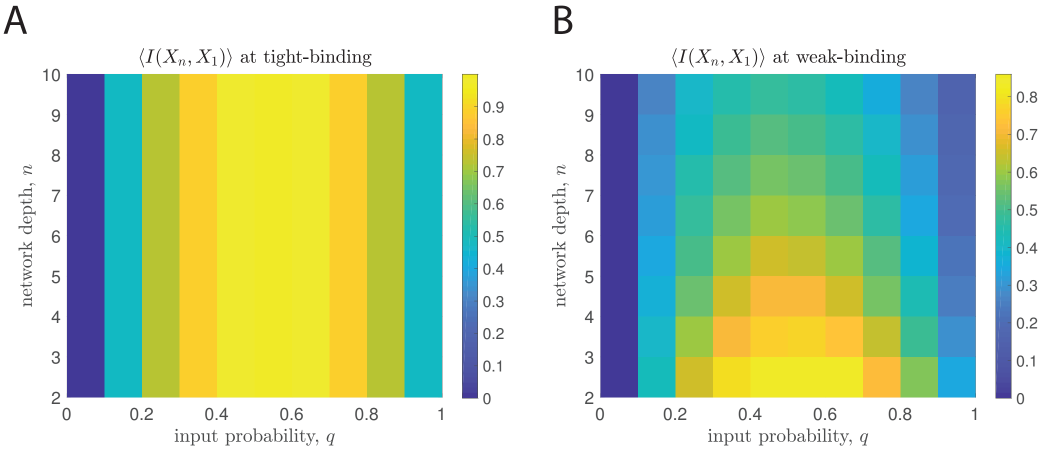


Figure A·7: Data processing inequality and biochemical noise: Mutual information across a linear network ($X_1 \to X_2 \to \cdots \to X_n$) is shown as the color map. Models are described in details in SI Sec.A.3. For tight/weak-binding (left/right panel), mean binding affinity $\beta\mu = -5/-0.1$. In all panels, $q = P(x_1 = 1)$ and standard deviation of binding affinities $\sigma = 0.1$.

observation process). Formally, DPI states that suppose we have a Markov chain: $X_1 \to X_2 \to X_3$, 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.A·7, we calculated mutual information for networks described in SI Sec.A.3 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.

A.7 Implementation of complex networks

Here we consider all-to-all connected 2-n-2 networks, where $n \in \mathbb{N}$ is the number of intermediate nodes. The goal is to compare the maximum mutual information conferred by optimal binding affinities (i.e. InfoMax) to that of the simplistic 2-2 network with and without cross-talks. To do so, we first construct such networks with varying n, 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, $\mathbf{y} = (y_1, y_2, \cdots, y_n)$ 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 energetics parameters can be summarized by the binding matrix ` $\in \mathbb{R}^{n\times 2}$ and $\mathbf{J} \in \mathbb{R}^{2\times n}$. To distinguish the variable space (indexed by $i, j, k \cdots$) from the configuration space (indexed by $\mu, \nu, \sigma, \cdots$), let $\mathcal{X} = \{\mathbf{x}^{(\mu)}, \mu = 1, \cdots, 2^2\}$, $\mathcal{Y} = \{\mathbf{y}^{(\mu)}, \mu = 1, \cdots, 2^n\}$, 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,

$$\mathcal{P}_{\mu\nu} = \begin{pmatrix} 1 & \vdots & \vdots & \vdots \\ 0 & \mathcal{P}_{\mu,2} & \mathcal{P}_{\mu,2} & \mathcal{P}_{\mu,2} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \vdots & \vdots & \vdots \end{pmatrix}, \quad \text{and} \quad \mathcal{Q}_{\sigma\mu} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ \dots & \mathcal{Q}_{2,\mu} & \dots & \dots \\ \dots & \mathcal{Q}_{4,\mu} & \dots & \dots \\ \dots & \mathcal{Q}_{3,\mu} & \dots & \dots \end{pmatrix}, \qquad \text{(A.36)}$$

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

$$\mathbb{P}_{\sigma\nu}(\mathcal{Z}|\mathcal{X}) = \mathcal{Q}_{\sigma\mu}\mathcal{P}_{\mu\nu} \tag{A.37}$$

$$\mathbb{P}_{\sigma}(\mathcal{Z}) = \mathcal{Q}_{\sigma\mu}\mathcal{P}_{\mu\nu}q_{\nu}, \tag{A.38}$$

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 $\mathbf{y}^{(\mu)} = (y_1^{(\mu)}, \cdots, y_n^{(\mu)})$. 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:

$$\mathcal{P}_{\mu,2} = \prod_{i\in u_\mu} f(\theta_{i,1}) \prod_{j\in v_\mu}[1 - f(\theta_{j,1})] \tag{A.39}$$

$$\mathcal{P}_{\mu,3} = \prod_{i\in u_\mu} f(\theta_{i,2}) \prod_{j\in v_\mu}[1 - f(\theta_{j,2})] \tag{A.40}$$

$$\mathcal{P}_{\mu,4} = \prod_{i\in u_\mu} g(\theta_{i,1}, \theta_{i,2}) \prod_{j\in v_\mu}[1 - g(\theta_{i,1}, \theta_{i,2})], \tag{A.41}$$

where

$$f(\zeta) = \frac{e^{-\zeta}}{1 + e^{-\zeta}} \tag{A.42}$$

$$g(\zeta, \xi) = \frac{e^{-\zeta} + e^{-\xi}}{1 + e^{-\zeta} + e^{-\xi}}. \tag{A.43}$$

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,

$$\begin{aligned} \mathcal{Q}_{2,\mu} &= h(\eta_{1,u_\mu})\,[1 - h(\eta_{2,u_\mu})] && \text{(A.44)} \\ \mathcal{Q}_{3,\mu} &= [1 - h(\eta_{1,u_\mu})]\,h(\eta_{2,u_\mu}) && \text{(A.45)} \\ \mathcal{Q}_{4,\mu} &= h(\eta_{1,u_\mu})\,h(\eta_{2,u_\mu}), && \text{(A.46)} \end{aligned}$$

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

$$h(\eta_{i,\mathcal{S}_i}) = \frac{\sum_{j\in\mathcal{S}_i} e^{-\eta_{i,j}}}{1 + \sum_{j\in\mathcal{S}_i} e^{-\eta_{i,j}}} \tag{A.47}$$

With all these matrices defined, we can compute the mutual information Eq. (A.17) by a series of matrix multiplications as we did in Sec. A.3.

A.8 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 2·4. The network studied are depicted as labeled according Figure 2·4A.

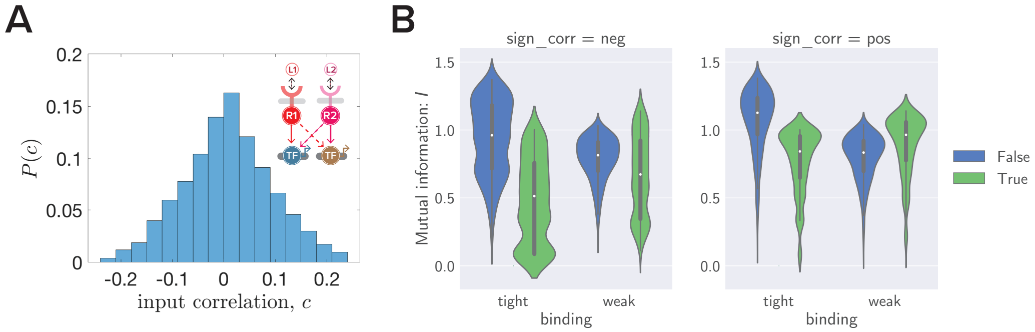


Figure A·8: (A) Network studied and the distribution of input correlation $P(c)$ (B) Violin plot for mutual information for network and input correlation shown in A. Violins are categorized according to their cross-talk levels and binding affinities. Blue for 'cross-talk' with $\beta\eta = -5.0$ and green for 'no cross-talk' with $\beta\eta = 0$ (see Appendix A 4 for details). Tight-binding refers to $\beta\langle\theta\rangle = -5.0$ while weak-binding to $\beta\langle\theta\rangle = -1.0$.

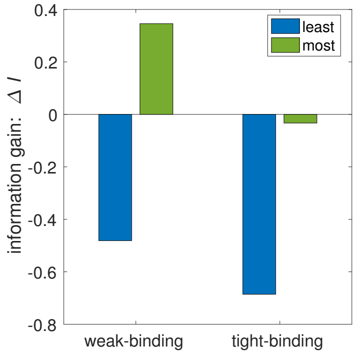


Figure A·9: Gain of mutual information through cross-talks for the network shown in Figure A·8. Information gain ΔI is defined as the different between the mutual information with and that without cross-talk. The label 'most' and 'least' are annotated based on maximizing and minimizing ΔI with respect to input correlations c, respectively.

Appendix B

Appendix of Chapter 3

B.1 Kinetics model on phosphorylation cascade

Here we consider kinase phosphorylation cascade without the ADD-gate logic. Following FIG. B·1, 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 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: $(\forall i > 1)$

$$\begin{aligned} \frac{dX_i}{dt} &= \Phi_i^+ - \Phi_i^- && \text{(B.1)} \\ &= \tilde{\alpha}_i X_{i-1} \tilde{X}_i - \beta X_i && \text{(B.2)} \\ &= \alpha_i X_{i-1} \left(1 - \frac{X_i}{C_i}\right) - \beta_i X_i, && \text{(B.3)} \end{aligned}$$

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

$$\frac{dX_1}{dt} = \alpha_1 R(t)\left(1 - \frac{X_1}{C_1}\right) - \beta_1 X_1. \tag{B.4}$$

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

$$X_i^{SS} = \frac{C_i X_{i-1}^{SS}}{\gamma_i C_i + X_{i-1}^{SS}}, \tag{B.5}$$

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:

$$p_i^{SS} = \frac{X_{i-1}^{SS}(\gamma_i C_i)^{-1}}{1 + X_{i-1}^{SS}(\gamma_i C_i)^{-1}}. \tag{B.6}$$

Note Eq.(B.7) can be mapped to our steady-state probability model presented in the main text, namely,

$$\begin{aligned} P(x_i = 1|\mathrm{pa}(x_i)) &= \frac{\sum_{j\in \mathrm{pa}(i)} x_j[c_j]e^{-\Delta\epsilon_{ji}/(k_BT)}}{1 + \sum_{j\in \mathrm{pa}(i)} x_j[c_j]e^{-\Delta\epsilon_{ji}/(k_BT)} + (1-x_j)[\tilde{c}_j]e^{-W/(k_BT)}} && \text{(B.7)} \\ &\approx \frac{\sum_{j\in \mathrm{pa}(i)} x_j[c_j]e^{-\Delta\epsilon_{ji}/(k_BT)}}{1 + \sum_{j\in \mathrm{pa}(i)} x_j[c_j]e^{-\Delta\epsilon_{ji}/(k_BT)}}, && \text{(B.8)} \end{aligned}$$

through

$$\begin{aligned} X_{i-1}^{SS} &\rightarrow x_{i-1}[c_{i-1}] && \text{(B.9)} \\ \gamma_i C_i &\rightarrow e^{\Delta\epsilon_{i,i-1}/(k_BT)}. && \text{(B.10)} \end{aligned}$$

Note that in Eq.(B.7) $[c_j]$ and $[\tilde{c}_j]$ are the concentration of kinase j in active and inactive form, respectively, at steady state. Also, Eq.(B.5) is related to the Michaelis-

Menton equation $V_{\max} S/(K_m + S)$ via

$$X_{i-1}^{SS} \rightarrow S \tag{B.11}$$

$$C_i \rightarrow V_{\max} \tag{B.12}$$

$$K_m \rightarrow C_i \gamma_i \tag{B.13}$$

Henrich et′ al showed that signal amplification at each stage i is possible only if

$$X_{i-1} < C_i \left(1 - \gamma_i\right). \tag{B.14}$$

If the concentration of activated kinase i is not much smaller than the total, then amplification is possible only if γ_i is almost but still smaller than 1 (i.e. $\alpha_i \approx \beta_i$ but $\alpha_i > \beta_i$). Whereas in the weakly activated regime (i.e. $X_i \ll C_i$), amplification only requires $\beta_i < \alpha_i$. In addition, they showed that longer signaling pathways do not always have a longer signal duration since they can distribute amplification over more steps.

B.2 CFFL

Following FIG. B·1, we simulate the following kinetics equations

$$\frac{dX_1}{dt} = \alpha_1 R(t) \left(1 - \frac{X_1}{C_1}\right) - \beta_1 X_1, \tag{B.15}$$

$$\frac{dX_2}{dt} = \alpha_2 X_1 \left(1 - \frac{X_2}{C_2}\right) - \beta_2 X_2, \tag{B.16}$$

$$\frac{dX_3}{dt} = \alpha_3 X_2 \left(1 - \frac{X_3}{C_3}\right) - \beta_3 X_3, \tag{B.17}$$

$$\frac{dX_4}{dt} = \alpha_4 \left(\frac{X_1}{K}\right) X_3 \left(1 - \frac{X_4}{C_4}\right) - \beta_4 X_4, \tag{B.18}$$

where one can interpret $\hat{\alpha}_4 = \alpha_4 \left(\frac{X_1}{K}\right)$ as the perturbed excitatory phosphorylation rate of kinase 4 by kinase 1. Note that the receptor activation function is chosen to be $R(t) = e^{-\lambda t}$. To see how this related to the model presented in the main text, lets solve for the steady-state fraction of phosphorylated kinase 4:

$$p_4^{SS} \equiv \frac{X_4}{C_4} = \frac{X_3(C_4\gamma_4)^{-1}\left(\frac{X_1}{K}\right)}{1 + X_3(C_4\gamma_4)^{-1}\left(\frac{X_1}{K}\right)}. \tag{B.19}$$

Next recall our mapping: $X_1 \to x_1[c_1]$, $X_3 \to x_3[c_3]$, $(C_4\gamma_4)^{-1} \to e^{-\Delta\epsilon_{34}/(k_B T)}$, and

$$\frac{X_1}{K} \to x_1[c_1]e^{-\Delta\epsilon_{14}/(k_B T)},$$

we can cast Eq.(B.19) into

$$p_4^{SS} \equiv \frac{X_4}{C_4} = \frac{x_3 x_1[c_1][c_3]e^{-(\Delta\epsilon_{14}+\Delta\epsilon_{34})/(k_B T)}}{1 + x_3 x_1[c_1][c_3]e^{-(\Delta\epsilon_{14}+\Delta\epsilon_{34})/(k_B T)}}, \tag{B.20}$$

which is approximately $P(x_4 = 1|x_1, x_3)$ except that a few terms pertaining to $x_j[c_j]e^{-\Delta\epsilon_{j4}}$ etc. are neglected in the partition function to be consistent with the AND nature of this CFFL. At steady state, one can solve these equations and get the input-output relation:

$$\begin{aligned} \frac{1}{X_4} &= \frac{1}{C_4} + \frac{K}{C_1}\left(\frac{\gamma_4}{C_3} + \frac{\gamma_4\gamma_3}{C_2}\right) + \frac{K\gamma_4\gamma_3\gamma_2}{C_1^2} \\ &+ \frac{K}{R}\left(\frac{\gamma_4\gamma_1}{C_3} + \frac{\gamma_4\gamma_3\gamma_1}{C_2} + \frac{2\gamma_4\gamma_3\gamma_2\gamma_1}{C_1}\right) + \frac{K\gamma_4\gamma_3\gamma_2\gamma_1^2}{R^2}. \end{aligned} \tag{B.21}$$

Note that due to the AND gate nature, the input-output relation is no longer Michaelis-Menton like. To see this, one can solve Eq.(B.1) and (B.4) to get:

$$\frac{1}{X_i} = \sum_{j=1}^{i} \frac{1}{C_j} \prod_{k=j+1}^{i} \gamma_k + \frac{1}{R}\prod_{k=1}^{i} \gamma_k, \tag{B.22}$$

which is linear in $1/C_j$ and $1/R$ but not in CFFL implemented here.

B.3 ICCFL

Here the kinetics equations are slightly modified (c.f. FIG. B·2)

$$\begin{aligned}
\frac{dX_1}{dt} &= \alpha_1 R(t)\left(1-\frac{X_1}{C_1}\right)-\beta_1 X_1, && \text{(B.23)}\\
\frac{dX_2}{dt} &= \alpha_2 X_1\left(1-\frac{X_2}{C_2}\right)-\beta_2 X_2, && \text{(B.24)}\\
\frac{dX_3}{dt} &= \alpha_3 X_2\left(1-\frac{X_3}{C_3}\right)-\beta_3 X_3, && \text{(B.25)}\\
\frac{dX_4}{dt} &= \left(\frac{\alpha_4}{1+\frac{X_3}{K}}\right) X_1\left(1-\frac{X_4}{C_4}\right)-\beta_4 X_4, && \text{(B.26)}
\end{aligned}$$

where one can intepret $\hat{\alpha}_4 = \alpha_4/(1+X_3/K)$ as the phosphoryation rate of kinase 4 perturbed by the ''inhibitory'' kinase 1 with strength K. Note that this choice of model allows us to relate K to our probabilistic model in the main text. To see this, note that at steady-state Eq.(B.26) implies:

$$p_4^{SS} \equiv \frac{X_4}{C_4} = \frac{X_1(C_4\gamma_4)^{-1}\left(1+\frac{X_3}{K}\right)^{-1}}{1+X_1(C_4\gamma_4)^{-1}\left(1+\frac{X_3}{K}\right)^{-1}}. \tag{B.27}$$

Recall that $X_1 \to x_1[c_1]$, $(C_4\gamma_4)^{-1} \to e^{-\Delta\epsilon_{14}/(k_B T)}$ and note that when $(X_3/K) \ll 1$ (c.f. the repressive nature of kinase 3 to 4):

$$\left(1+\frac{X_3}{K}\right)^{-1} \sim 1-\frac{X_3}{K} \to (1-x_3)[\tilde{c}_3]e^{-\Delta\epsilon_{34}/(k_B T)},$$

we can cast Eq.(B.27) into

$$p_4^{SS} \equiv \frac{X_4}{C_4} = \frac{x_1(1-x_3)[c_1][\tilde{c}_3]e^{-(\Delta\epsilon_{14}+\Delta\epsilon_{34})/(k_B T)}}{1+x_1(1-x_3)[c_1][\tilde{c}_3]e^{-(\Delta\epsilon_{14}+\Delta\epsilon_{34})/(k_B T)}}, \tag{B.28}$$

which is approximately $P(x_4 = 1|x_1, x_3)$ except that a few terms pertaining to $x_j[c_j]e^{-\Delta\epsilon_{j4}}, j = 1, 3$ and non-specific bindings are neglected in the partition function. Similarly, one can solve for the input-output. To simplify notation, we only write down the solution for X_3/K is small so that $\hat{\alpha}_4 = \alpha_4(1 - X_3/K)$:

$$\begin{aligned}\frac{1}{X_4} &= \frac{1}{C_4} + \left(\frac{\gamma_4}{C_1} + \frac{\gamma_4\gamma_1}{R}\right)\left(\frac{1}{C_3} + \frac{\gamma_3}{C_2} + \frac{\gamma_3\gamma_2}{C_1} + \frac{\gamma_3\gamma_2\gamma_1}{R}\right) \times \\ &\left(\frac{1}{C_3} + \frac{\gamma_3}{C_2} + \frac{\gamma_3\gamma_2}{C_1} + \frac{\gamma_3\gamma_2\gamma_1}{R} - \frac{1}{K}\right)\end{aligned} \tag{B.29}$$

B.3.1 Summary of parameters

Parameters	Meaning or definition
$X_i(\tilde{X}_i)$	conc. of active (inactive) kinase i
α_i	PPlation. rate of kinase i
β_i	De-PPlation rate of kinase i
γ_i	Propensity of de-phosphorylation
C_i	Total conc. of kinase i
$\gamma_i C_i$	Michaelis constant of reaction $j \to i$: $K_M^{(ji)}$
K	Regulatory strength of AND gate

Here is the summary of parameters extracted from the sigmoid decision shown in Figure 4 of main text.

Parameters	Kinase $i = 1$	Kinase $i = 2$	Kinase $i = 3$	Kinase $i = 4$
α_i	1.0 (1.0)	1.0 (1.0)	1.0 (1.0)	1.0 (1.0)
β_i	$e^{-8.2}$ $(e^{1.0})$	$e^{-7.5}$ $(e^{1.0})$	$e^{-6.3}$ $(e^{-0.05})$	$e^{-6.3}$ $(e^{-3.0})$
C_i	2.5 (2.5)	2.5 (2.5)	2.5 (2.5)	2.5 (2.5)

B.4 Formulae used in the optimization

B.4.1 CFFL

Let $f(p_1)$, where $p_1 = P(x_1 = 1)$, be the input-output relation we want to implement. All the probability functions used in the optimization procedure are given below:

$$
\begin{aligned}
f(p_1) &= \frac{1}{1+e^{-10*(p_1-0.5)}} && \text{(B.30)}\\
P(x_2=1|x_1) &= \frac{1}{Z_2}\left(x_1[c_1]e^{-\Delta\epsilon_{13}/(k_BT)}\right) && \text{(B.31)}\\
P(x_3=1|x_2) &= \frac{1}{Z_3}\left(x_2[c_2]e^{-\Delta\epsilon_{23}/(k_BT)}\right) && \text{(B.32)}\\
P(x_4=1|x_1,x_3) &= \frac{1}{Z_4}\left(x_1x_3[c_1][c_3]e^{-(\Delta\epsilon_{14}+\Delta\epsilon_{34}+J)/(k_BT)}\right), && \text{(B.33)}
\end{aligned}
$$

where the partition functions Z_i are given by

$$
\begin{aligned}
Z_2 &= 1+x_1[c_1]e^{-\Delta\epsilon_{12}/(k_BT)}+(1-x_1)[\tilde{c}_1]e^{-W/(k_BT)} && \text{(B.34)}\\
Z_3 &= 1+x_2[c_2]e^{-\Delta\epsilon_{23}/(k_BT)}+(1-x_2)[\tilde{c}_2]e^{-W/(k_BT)} && \text{(B.35)}\\
Z_4 &= 1+x_1[c_1]e^{-\Delta\epsilon_{14}/(k_BT)}+x_3[c_3]e^{-\Delta\epsilon_{34}/(k_BT)}\\
&+ x_1x_3[c_1][c_3]e^{-(\Delta\epsilon_{14}+\Delta\epsilon_{34}+J)/(k_BT)}+\cdots, && \text{(B.36)}
\end{aligned}
$$

where in Z_4 we neglect terms proportional to $e^{-W/(k_BT)}$ since those events are probabilistically unlikely. We optimize the loss function with respect to $\Delta\epsilon = (\Delta\epsilon_{12}, \Delta\epsilon_{23}, \Delta\epsilon_{14}, \Delta\epsilon_{34})$. We also include a cooperativity energetic reward J to encourage kinase 4 phosphorylation due to the activation of both of its upstreams: kinase 1 and kinase 3. This factor is reported as $\Delta\epsilon_{14}$ in the table shown above since it can be absorbed into it through K.

B.4.2 ICFFL

All the probability functions used in the optimization procedure are given below:

$$
\begin{aligned}
f(p_1) &= \frac{1}{1+e^{-10*(p_1-0.5)}} && \text{(B.37)}\\
P(x_1=1|L) &= \frac{1}{1+e^{-10*(L-0.5)}} && \text{(B.38)}\\
P(x_2=1|x_1) &= \frac{1}{Z_2}\left(x_1[c_1]e^{-\Delta\epsilon_{12}/(k_BT)}\right) && \text{(B.39)}\\
P(x_3=1|x_2) &= \frac{1}{Z_3}\left(x_2[c_2]e^{-\Delta\epsilon_{23}/(k_BT)}\right) && \text{(B.40)}\\
P(x_4=1|x_1,x_3) &= \frac{1}{Z_4}\left(x_1(1-x_3)[c_1][\tilde{c}_3]e^{-(\Delta\epsilon_{14}+\Delta\tilde{\epsilon}_{34}+J)/(k_BT)}\right), && \text{(B.41)}
\end{aligned}
$$

with the partition functions Z_i given by

$$
\begin{aligned}
Z_2 &= 1+x_1[c_1]e^{-\Delta\epsilon_{12}/(k_BT)}+(1-x_1)[\tilde{c}_1]e^{-W/(k_BT)} && \text{(B.42)}\\
Z_3 &= 1+x_2[c_2]e^{-\Delta\epsilon_{23}/(k_BT)}+(1-x_2)[\tilde{c}_2]e^{-W/(k_BT)} && \text{(B.43)}\\
Z_4 &= 1+x_1[c_1]e^{-\Delta\epsilon_{14}/(k_BT)}+x_3[c_3]e^{-\Delta\epsilon_{34}/(k_BT)}+(1-x_3)[\tilde{c}_3]e^{-\Delta\tilde{\epsilon}_{34}/(k_BT)}\\
&+ x_1(1-x_3)[c_1][\tilde{c}_3]e^{-(\Delta\epsilon_{14}+\Delta\tilde{\epsilon}_{34}+J)/(k_BT)}+x_1x_3[c_1][c_3]e^{-(\Delta\epsilon_{14}+\Delta\epsilon_{34})/(k_BT)}\\
&+ \cdots, && \text{(B.44)}
\end{aligned}
$$

where $\Delta\tilde{\epsilon}_{34}$ is the binding energy of inactive kinase 3 to kinase 4, as opposed to $\Delta\epsilon_{34}$ for that between active kinase 3 to kinase 4. We treated $\Delta\tilde{\epsilon}_{34}$ as a constant while optimizing the loss function with respect to $\Delta\epsilon = (\Delta\epsilon_{12}, \Delta\epsilon_{23}, \Delta\epsilon_{14}, \Delta\epsilon_{34})$. We also include a cooperativity energetic reward J to encourage kinase 4 phosphorylation due to the inactivation of its upstream repressive kinase 1 and the activation of upstream kinase 1. This factor is reported as $\Delta\epsilon_{34}$ in the table shown above since it can be absorbed into it through K. In Z_4, we neglect a few terms pertaining to non-specific binding (i.e. $\sim e^{-W/(k_BT)}$) since these events are probabilistically unlikely.

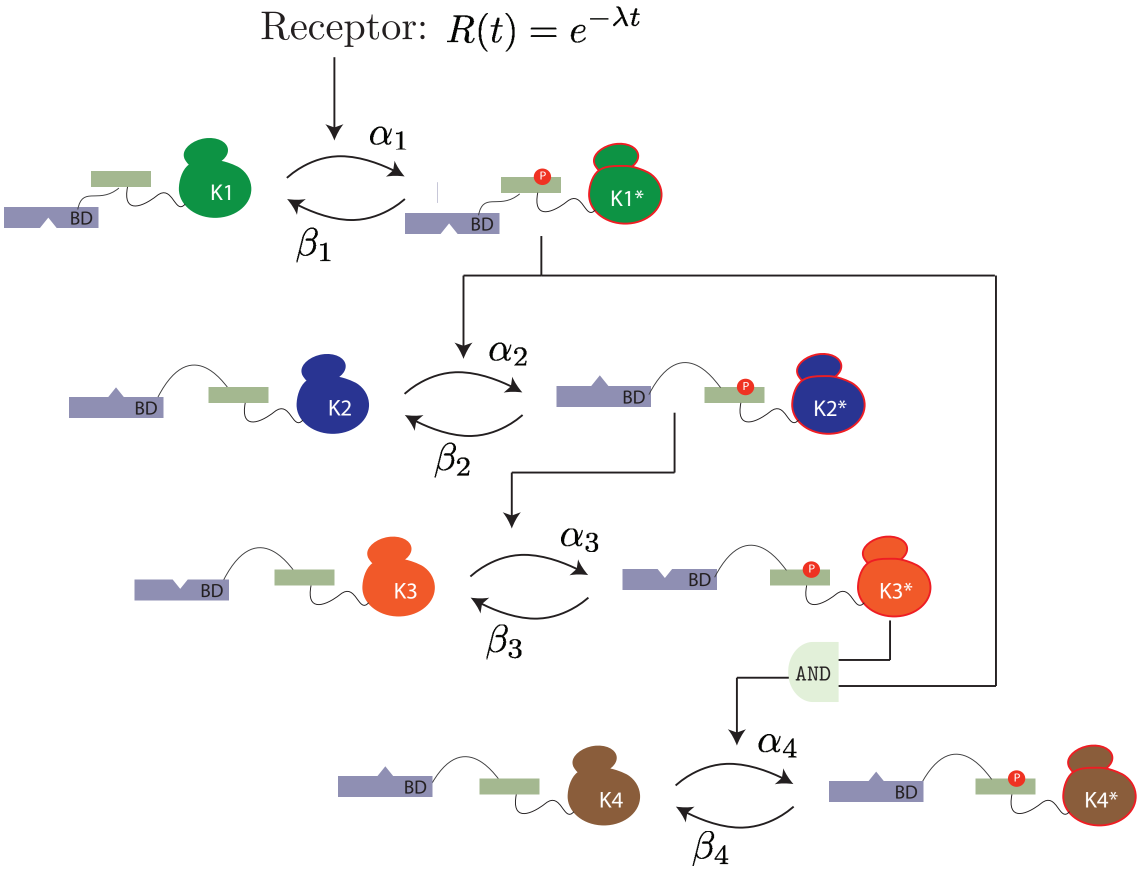

Receptor: $R(t) = e^{-\lambda t}$
α_1
β_1
BD
K1
P
BD
K1*
α_2
β_2
BD
K2
BD
P
K2*
α_3
β_3
BD
K3
BD
P
K3*
AND
α_4
β_4
BD
K4
BD
P
K4*

Figure B·1: Signal transduction cascade that implements CFFL

B.4.3 Loss function used

$$\mathcal{L} \equiv \frac{1}{N} \sum_{\alpha=1}^{N} \left(f(p_1) - P(x_4 = 1|\Delta\epsilon, p_1)\right)^2 + C||\Delta\epsilon||_2^2, \tag{B.45}$$

where C is the regularization constant introduced to avoid trivial solutions (i.e. all energies are infinitely negative). The parameters are updated according to

$$\Delta\epsilon \leftarrow \Delta\epsilon - \eta \nabla_{\Delta\epsilon} \mathcal{L}, \tag{B.46}$$

where η is the learning rate.

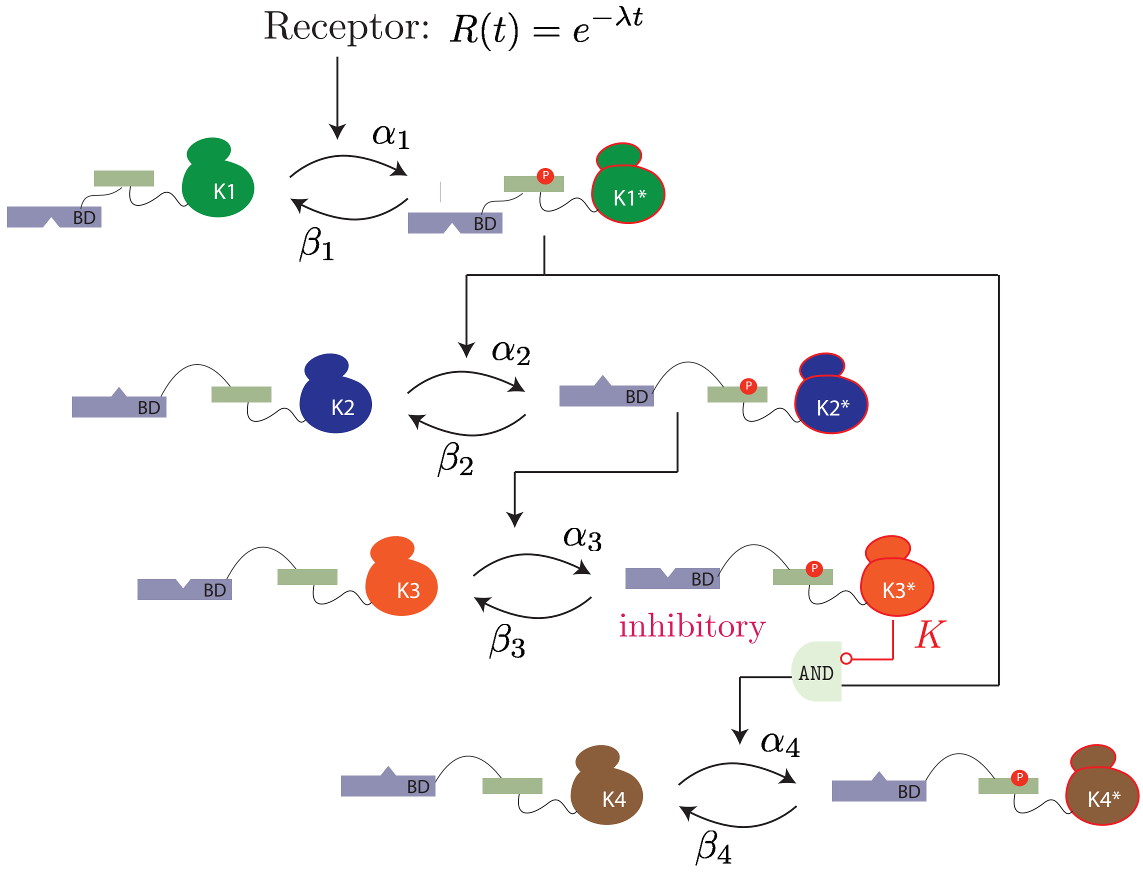

Receptor: $R(t) = e^{-\lambda t}$
α_1
β_1
K1
K1*
BD
P
α_2
β_2
K2
K2*
α_3
β_3
K3
K3*
inhibitory
K
AND
α_4
β_4
K4
K4*

Figure B·2: Signal transduction cascade that implements ICFFL

Appendix C

Appendix of Chapter 4

C.1 Statistical mechanical model for generic systems

Consider a system of n protein species indexed by $i = 1, \cdots, n$. Suppose for each species i, there are n_i internal states (e.g. phosphorylation, methylation, or allosteric states) labelled by $x_i \in \{0, \cdots, n_i - 1\}$. Denote $E^{(ij)}(x_i, x_j)$ as the energy of species i at internal state x_i and when bound to species j at state x_j. Let $E_0^{(i)}(x_i)$ be the energy of species i at state x_i in solution. Note that for any pair of i, j, $E^{(ij)}$ is a $n_i \times n_j$ matrix and for any i, $E_0^{(i)}$ is a n_i-dimensional column vector. We can write down for each protein species i at each of its internal state $x_i \in \mathcal{X}_i = \{0, 1, \cdots, n_i - 1\}$, the following conditional probability

$$P(x_i|\mathcal{X}_{\setminus i}) = \frac{\sum_{j \neq i}^{n} \sum_{x_j=0}^{n_j-1} [c_j(x_j)] e^{-\beta[E^{(ij)}(x_i,x_j) - E_0^{(i)}(x_i)]}}{\sum_{x_i=0}^{n_i-1} \sum_{j \neq i}^{n} \sum_{x_j=0}^{n_j-1} [c_j(x_j)] e^{-\beta[E^{(ij)}(x_i,x_j) - E_0^{(i)}(x_i)]}}, \tag{C.1}$$

where $\mathcal{X} = \bigcup_{i=1}^{n} \mathcal{X}_i$ is the collection of all internal state labels and $[c_j(x_j)]$ is the concentration of protein j at internal state x_j. The symbol $\mathcal{X}_{\setminus i}$ indicates $\mathcal{X}_{\setminus i} = \bigcup_{j \neq i}^{n} \mathcal{X}_j$. Note that $[c_i(x_i)]$ is related to the concentration of protein species i, $[c_i]$ via

$$[c_i(x_i)] = [c_i] P(x_i|\mathcal{X}_{\setminus i}) \tag{C.2}$$

Note that definition of concentration retains proper normalization of probability since after summing over x_i in Eq.(C.2), one gets

$$\sum_{x_i}[c_i(x_i)] = [c_i]\sum_{x_i} P(x_i|\mathcal{X}_{\setminus i}) = [c_i] \tag{C.3}$$

Consider for a simple two-species system $i = 1, 2$ with $n_1 = 3$ and $n_2 = 2$. Suppose both the species concentrations $[c_1], [c_2]$ and the energy matrices $E^{(12)}$, $E^{(21)}$, $E_0^{(1)}$, $E_0^{(2)}$ are known, one can utilize 5 equations exemplified by Eq.(C.1)(C.2) to solve for the 5 variables: $[c_1(0)], [c_1(1)], [c_1(2)], [c_2(0)], [c_2(1)]$ of interest. Namely,

$$\begin{aligned}
[c_1(0)] &= [c_2(0)]e^{-\beta[E^{(12)}(0,0)-E_0^{(1)}(0)]} + [c_2(1)]e^{-\beta[E^{(12)}(0,1)-E_0^{(1)}(0)]} && \text{(C.4)}\\
[c_1(1)] &= [c_2(0)]e^{-\beta[E^{(12)}(1,0)-E_0^{(1)}(1)]} + [c_2(1)]e^{-\beta[E^{(12)}(1,1)-E_0^{(1)}(1)]} && \text{(C.5)}\\
[c_1(2)] &= [c_2(0)]e^{-\beta[E^{(12)}(2,0)-E_0^{(1)}(2)]} + [c_2(1)]e^{-\beta[E^{(12)}(2,1)-E_0^{(1)}(2)]} && \text{(C.6)}\\
[c_2(0)] &= [c_1(0)]e^{-\beta[E^{(21)}(0,0)-E_0^{(2)}(0)]} + [c_1(1)]e^{-\beta[E^{(21)}(0,1)-E_0^{(2)}(0)]} \\
&+ [c_1(2)]e^{-\beta[E^{(21)}(0,2)-E_0^{(2)}(0)]} && \text{(C.7)}\\
[c_2(1)] &= [c_1(0)]e^{-\beta[E^{(21)}(1,0)-E_0^{(2)}(1)]} + [c_1(1)]e^{-\beta[E^{(21)}(1,1)-E_0^{(2)}(1)]} \\
&+ [c_1(2)]e^{-\beta[E^{(21)}(1,2)-E_0^{(2)}(1)]} && \text{(C.8)}
\end{aligned}$$

Note that to relate this to $P(x_i|\mathcal{X}_{\setminus i})$, one can invoke Eq.(C.2).

C.2 Problem setup in our synthetic system

In this section, we recast the generic model in the previous section into the one presented in the main text (i.e. synthetic systems). We first consider a system with writer proteins only that interact with on another in phosphorylation-dependent (P-dependent) manner. When a protein is phosphorylated, it can bind other proteins and change their phosphorylation state. Each species i is characterized by the triplet $\mathbf{v}_i = (\alpha_i, \beta_i, \mathcal{C}_i)$, where α_i, β_i are the flavor of constitutive interaction domain and P-dependent binding domain, respectively, and $\mathcal{C}_i = \mathcal{W}$ is the label for the catalytic domain, which in the case is simply writers $\mathcal{W}$(see main text). We use $x_i = 1$ to label that phosphorylated state and $x_i = 0$ for unphosphorylated state. In this simplest setting, the probability of writer species i bound to another writer species j is given by

$$p_{ij}^{bound} = \frac{x_j e^{-\beta[\Delta\epsilon(\alpha_i,\beta_j)-\mu_j(x_j)]}}{1+\sum_k x_k e^{-\beta[\Delta\epsilon(\alpha_i,\beta_k)-\mu_k(x_k)]}}, \tag{C.9}$$

where $\beta = 1/(k_B T)$ is the inverse temperature, $\Delta\epsilon(\alpha_i, \beta_j)$ is the energy difference between bound and unbound state of species i and j, $\mu_j(x_j)$ is the chemical potential of j which is related to concentration through $\mu_j(x_j) = -\beta^{-1}\log[c_j(x_j)]$. Note that $[c_j(x_j)] = [c_j]P_j(x_j)$ where $P_i(x_i)$ is the probability that species j is at state x_j. Concretely,

$$P_i(x_i = 1) = \sum_{j \text{ s.t. } \mathcal{C}_j=\mathcal{W}} \frac{x_j e^{-\beta[\Delta\epsilon(\alpha_i,\beta_j)-\mu_j(x_j)]}}{1+\sum_k x_j e^{-\beta[\Delta\epsilon(\alpha_i,\beta_k)-\mu_k(x_k)]}}. \tag{C.10}$$

In the low temperature limit, $P_j(x_j = 1) \to \{0,1\}$. Thus, $\mu_j(x_j = 1) = \beta^{-1}\log[c_j] \equiv \mu_j$ if $P_j(x_j = 1) = 1$ while $\mu_j(x_j = 1) \to -\infty$ if $P_j(x_j = 1) = 0$ (or in other words, $P_j(x_j = 0) = 1$). For the former, the quantity $\Delta\epsilon(\alpha_i, \beta_j)|_{x_i=1,x_j=1} - \mu_j(x_j = 1) \equiv \Delta\epsilon_{ij} - \mu_j$ only depends on i, j. Its Boltzmann factor reads,

$$x_j e^{-\beta[\Delta\epsilon(\alpha_i,\beta_j)|_{x_i=1,x_j=1}-\mu_j(x_j=1)]} \to 1 \cdot e^{-\beta[\Delta\epsilon(\alpha_i,\beta_j)|_{x_i=1,x_j=1}-\mu_j(x_j=1))]} \equiv e^{-\beta[\Delta\epsilon_{ij}-\mu_j]}, \tag{C.11}$$

The latter is always killed by the $x_j = 0$ factor in the Boltzmann weight. Specifically, it reads (note $\mu_j(x_j = 0) \to -\infty$)

$$x_j e^{-\beta[\Delta\epsilon(\alpha_i,\beta_j)|_{x_i=1,x_j=0}-\mu_j(x_j=0)]} \to 0 \cdot e^{-\beta\Delta\epsilon(\alpha_i,\beta_j)|_{x_i=1,x_j=0}} e^{+\beta(-\infty)} = 0, \tag{C.12}$$

In sum, we can always use $\Delta\epsilon_{ij} - \mu_j$ to summarize these two cases if x_j is binary. That dependence of μ_j on x_j is already taken care of by the x_j factor in the Boltzmann weight. Therefore, under this limit one has,

$$P_i(x_i = 1)|_{\beta\to\infty} = \begin{cases} 1, \text{ if } \exists\, j \neq i \text{ s.t. } x_j = 1 \text{ AND } \Delta\epsilon_{ij} < 0 \\ 0, \text{ if } \forall\, j \neq i,\ x_j = 0 \text{ OR } \forall j \neq i \text{ s.t. } x_j = 1, \Delta\epsilon_{ij} > 0 \end{cases} \tag{C.13}$$

Throughout the rest of this note, we consider only in this zero temperature limit.

Definition 1 (Design matrix). *The design matrix Θ is a $n \times n$ matrix whose elements are defined through $\theta_{ij} = sgn(\Delta\epsilon_{ij} - \mu_j)$, $\forall\, 1 \leq i, j \leq n$, where n is the number of distinct proteins in the system and $\Delta\epsilon_{ij}$ is the binding affinities between protein i and j and μ_j is the chemical potential of protein j. It is symmetric for homogenous systems in which chemical potential of all proteins are the same.*

Note that due the structure of our model, the diagonal elements don't matter (since we don't consider self-interaction at this moment).

C.3 Bounds on admissible design matrix

To interpret the design matrix one needs to look at the biophysics that dictates the resresponding configurations. For example, we would like to incorporate the fact that protein i is active whenever the rest of others are active. This naturally constrain the existence of any design matrices (again, we ignore the diagonal entries) having any row consisting of purely 1's (i.e. $\exists\, i$ s.t. $\Delta\epsilon_{ij} > 0, \forall j \neq i, j = 1, \cdots, n$). This motivates the notion of admissible design matrix.

Definition 2 (Admissible design matrix). *An $n \times n$ design matrix Θ is called admissible at level n if without regard to its diagonal elements each row of Θ, namely, $\boldsymbol{\theta}_i$, $i = 1, \cdots n$, satisfies $\|\boldsymbol{\theta}_i\|_0 \leq n - 2$. In other words, setting aside the diagonal elements, each row cannot consists of purely 1's.*

The following gives the bound on the allowed admissible $n \times n$ design matrix. As a reminder, the goal is to first see how simple biophysics limits the combinatorial complexity in the design matrix space. In the next section, we'll see a further reduction when translating these feasible matrices to *compatible configurations*, where the notion of capacity of computation becomes clear.

Result 4. *Let R_n be the number of admissible design at level n, then*

$$L_n < R_n < U_n, \tag{C.14}$$

where

$$U_n = 2^{\sum_{k=1}^{n-1} k} - \sum_{l=1}^{n-2} \binom{n-1}{l} \cdot 2^{l+\sum_{k=1}^{n-2-l} k} \tag{C.15}$$

$$L_n = 2^{\sum_{k=1}^{n-1} k} - \sum_{l=1}^{n-2} \binom{n-1}{l} \cdot 2^{l+\sum_{k=1}^{n-3} k} \tag{C.16}$$

Derivation. By definition, R_n is the number of distinct symmetric $n \times n$ binary matrices that have at most $n-2$ 1's in each row, without regard to the diagonal elements. For a given $n \times n$ symmetric matrix, the degree-of-freedom is $2^{\mathrm{DOF}(n)} = 2^{\sum_{k=1}^{n-1} k}$ without considering any constraints . Therefore, to find R_n, it is sufficient to count the possible ways of violation. Denote $f_n(l,k)$ as the number of configurations having exactly l zeros in the first row among which k fails, namely, there are k rows consists of $n-1$ zeros, apart from the diagonal elements. Since the constraint is applied to each row where there are only $n-1$ elements that matters and that one can choose to violate $k = 1, \cdots, l$ rows for a fixed l, R_n simply reads

$$R_n = 2^{\sum_{k=1}^{n-1} k} - \sum_{l=1}^{n-2} \sum_{k=1}^{l} \binom{n-1}{l} \binom{l}{k} f_n(l,k). \tag{C.17}$$

Note that the summation of l extends up to $n-2$ only due to the symmetric property of the matrix. Since for all $n \in \mathbb{Z}^+,\ n \geq 3, \forall\, 1 \leq l \leq n-1$ and $\forall\, 1 \leq k \leq l$,

$$f_n(l,l) \leq f_n(l,k) \leq f_n(l,1) \leq f_n(1,1), \tag{C.18}$$

and that $f_n(1,1) = 2^{\mathrm{DOF}(n-2)} = 2^{\sum_{k=1}^{n-3} k}$, $f_n(l,l) = 2^{\mathrm{DOF}(n-1-l)} = \sum_{k=1}^{n-2-l} k$, U_n and

L_n are obtained simply by replacing $f_n(l,k)$ in Eq.(C.17) with $f_n(l,l)$ and $f_n(1,1)$, respectively, and invoking binomial res. □

C.4 Size of minimal configuration space

Definition 3 (Activity vector)). *An activity vector $u \in d^n$ is a n-dimensional vector whose elements take value only from the alphabet set $\chi = \{0, 1, \cdots, d-1\}$. Concretely, $u = (u_1, \cdots, u_n)$, where $u_i \in \chi$, $\forall i = 1, \cdots, n$. A protein configuration vector is an activity vector with $\chi = \{0, 1\}$, namely, $d = 2$.*

Definition 4 (Writer-eraser partition). *Let $u \in 2^n$ be a binary protein configuration vector indexed by the alphabet set $\chi = \{0, 1\}$. Let $\alpha_m(u)$ be a partition of u such that it divides u up into a group of m and a group of $n-m$ elements, for all $m \in \mathbb{N}$, $m \in [0, n]$, in the following manner:*

$$\begin{aligned} \alpha_m(u) &= u_1, u_2 \cdots, u_m | u_{m+1}, \cdots, u_n \\ &\equiv \{\{u_1, u_2 \cdots, u_m\}, \{u_{m+1}, \cdots, u_n\}\} \end{aligned}$$

The element of α_m with cardinality m is termed the writer index set thus denoted as $\mathcal{W}$ while the remaining is that of the erasers and is labeled as $\mathcal{E}$. In other words, $\alpha_m = \mathcal{W} \cup \mathcal{E}$. We call $\alpha_m(u)$ a writer-eraser partition of u.

Definition 5 (Subset of activity vectors). *Let $u \in 2^n$ be a binary protein configuration vector and let $\alpha_m(\cdot) = (\mathcal{W} \cup \mathcal{E})(\cdot)$ be a writer-reader partition. Then $u_{\mathcal{W}}$ and $u_{\mathcal{E}}$ refers to the activity vector of writers and erasers, respectively, of this particular configuration u. For example, let $n = 4$, $m = 3$ and let $u = (1, 0, 1, 1)$. Then $u_{\mathcal{W}} = (1, 0, 1)$ and $u_{\mathcal{E}} = (1)$*

Definition 6. *The mapping $\mathbf{1}(\cdot) : 2^n \to \mathbb{N}$ takes in a binary vector of length n and returns the bit positions where that vector has value 1. For example, let $u = (1, 0, 1, 0) \in 2^4$. Then $\mathbf{1}(u) = \{1, 3\}$.*

Definition 7 (Configuration space). *Let $\mathcal{B}_n$ be the space of all unique n-bit binary vectors. The configuration space (CS) is the subset $\mathcal{C}_n \subset \mathcal{B}_n$ so that all elements in this set are consistent with physical constraints.*

Definition 8 (CS of size k at level n). *CS of size k at level n, denoted as $\mathcal{D}_k^{(n)}$, is the collection of cardinality-k sets whose elements are all in $\mathcal{B}_n$. Formally speaking, $\mathcal{D}_k^{(n)} = \{\mathcal{C}_n \subset \mathcal{B}_n | \, |\mathcal{C}_n| = k\}$. For convenience, the cardinality of $\mathcal{D}_k^{(n)}$ is denoted as $D_k^{(n)}$*

Definition 9 (Compatibility)**.** *Two n-bit binary vectors are compatible if they can coexist in the configuration space.*

Definition 10 (Subset of $\mathcal{B}_n$)**.** *Let $\mathcal{B}_n$ be the set of all unique binary vectors of length n. The subset $W_m^{(n)} \subseteq \mathcal{B}_n$ is the collection of binary vectors in $\mathcal{B}_n$ having exactly m 1's. Namely, for all* $0 \leq m \leq n$

$$W_m^{(n)} = \{I_n \in \mathcal{B}_n | \, \|I_n\|_0 = m\} \tag{C.19}$$

The following facts can be proven by scrutinizing the constraints imposed by the resresponding configurations (see Section VI).

Fact 1. $W_1^{(n)}$ *is never physical. That is, elements in* $W_1^{(n)}$ *are never in* $\mathcal{C}_n$.

Fact 2. $W_{n-1}^{(n)}$ *is not compatible with* $W_n^{(n)}$.

Fact 3. $W_0^{(n)}$ *does not violate any constraints and should always be included in any CSs.*

From the biophysics point of view, the catalytic nature of proteins implies one should always include $W_n^{(n)}$. This, along with Fact 1, 2, and 3, motivates the definition of *non-trivial CS*.

Definition 11 (Non-trivial CS)**.** *A CS is called non-trivial as long as it contains more than the following two elements:* $W_0^{(n)}$ *and* $W_n^{(n)}$*. It is called trivial if otherwise.*

Result 5 (Compatibility check)**.** *Let* $u^{(n)} \in W_m^{(n)}$ *and* $v^{(n)} \in W_{m'}^{(n)}$, $2 \leq m, m' \leq n-2$. $u^{(n)}$ *and* $v^{(n)}$ *are compatible if and only if for every bit i that u and v don't agree, say,* $(u_i, v_i) = (1, 0)$, *there exists at least one element in* $\mathbf{1}(u^{(n)}) \setminus i$ *but not in* $\mathbf{1}(v^{(n)})$, *where* $\mathbf{1}(u^{(n)})$ *is the set of bit positions in which* $u^{(n)}$ *is one; and similarly for* $(u_i, v_i) = (0, 1)$. *In other words,* $u^{(n)}$ *and* $v^{(n)}$ *are incompatible if and only if there is exactly one pair of* $(u_i, v_i) = (1, 0)$ *or* $(u_i, v_i) = (0, 1)$ *or both,* $\forall i \in [1, n]$.

Derivation. For simplicity in notation, we remove the superscript (n) from both u and v but keep in mind that they're of length n. Also since $(u_i, v_i) = (1, 0)$ and $(0, 1)$ are symmetric, we prove only the former. More specifically, we prove the statement that [*compatibility* $\Leftrightarrow \forall i \in [1, n], \exists \, w \in \mathbf{1}(u^{(n)}) \setminus i, \; w \notin \mathbf{1}(v^{(n)})$]. For the sufficient part, it is equivalent to prove that the non-existence of such element implies incompatibility. For any i, $(u_i, v_i) = (1, 0)$ translates into $\theta_{ik} > 0$, $\forall k \in \mathbf{1}(v)$

as well as that there exists $k \in \mathbf{1}(u) \setminus i$ such that $\theta_{ik} < 0$ in the parameter space. If such element does not exist, then the set $\{\mathbf{1}(u) \setminus i\} \subseteq \{\mathbf{1}(v)\}$. However, this implies that $\theta_{ij} > 0, \forall j \in \{\mathbf{1}(v)\}$ as well as the existence of $k \in \{\mathbf{1}(u) \setminus i\} \subseteq \{\mathbf{1}(v)\}$ such that $\theta_{ik} < 0$, a contradiction thus incompatibility. As for the necessary part, the result follows simply by definition. ☐

Result 6 (bitwise OR check: sufficiency)**.** *Let* $u^{(n)} \in W_m^{(n)}$ *and* $v^{(n)} \in W_{m'}^{(n)}$, $2 \leq m, m' \leq n-2$. *If* $u^{(n)} \oplus_2 v^{(n)} = \mathbf{1}_n$, *namely, their bit-wise OR is exactly an* $n-$*dimensional vector consists of purely 1's, then* $u^{(n)}$ *and* $v^{(n)}$ *are compatible.*

Derivation. Let x, y, w, z be the number of $(u_i, v_i) = (1,0), (0,1), (1,1), (0,0)$ pairs in $u^{(n)}$ and $v^{(n)}$, respectively. Clearly $u^{(n)} \oplus_2 v^{(n)} = \mathbf{1}_n$ implies that $z = 0$. Since $u \in W_m^{(n)}, v \in W_{m'}^{(n)}$ and $2 \leq m, m' \leq n-2$, we should have more than one configurations of $(1,0)$ and $(0,1)$, namely, $x, y \geq 2$. Thus by invoking res 5, u and v are compatible. ☐

Remark 1. *Bit-wise OR check is not necessary for compatibility. That is, two bit string could be compatible even though bitwise check is not satisfied, as illustrated by the following example.*

Example 1. *Let* $u = (1,0,0,1,1,1)$ *and* $v = (0,0,0,1,0,1)$. *Clearly* $u \in W_4^{(6)}$, $v \in W_2^{(6)}$ *and* $u \oplus_2 v \neq \mathbf{1}_n$. *However, there's no violating constraints in the parameter space. In particular, the only places where this could possibly happen are:*

$$\left[(\theta_{14} \vee \theta_{15} \vee \theta_{16}) < 0 \quad \bigwedge \quad (\theta_{14} \wedge \theta_{16}) > 0\right] \Longrightarrow \theta_{14} > 0,\ \theta_{15} < 0,\ \theta_{16} > 0 \tag{C.20}$$

$$\left[(\theta_{51} \vee \theta_{54} \vee \theta_{56}) < 0 \quad \bigwedge \quad (\theta_{54} \wedge \theta_{56}) > 0\right] \Longrightarrow \theta_{54} > 0,\ \theta_{51} < 0,\ \theta_{56} > 0 \tag{C.21}$$

Example 2. *Let* $|\mathcal{W}| = 5$, $|\mathcal{E}| = 2$ *and*

$$\begin{pmatrix} u \\ v \\ w \\ p \\ q \end{pmatrix} = \begin{pmatrix} 10011 \\ 00011 \\ 01011 \\ 11001 \\ 00111 \end{pmatrix}$$

Then (u, v), (v, w) *and* (u, w) *are incompatible while* (p, q) *are compatible because the number of bit positions of* $(1,0)$ *and* $(0,1)$ *are both 2.*

Since any non-trivial CS contains elements other than $W_n^{(n)}$ and $W_0^{(n)}$, and that any elements in $W_m^{(n)}$, $2 \leq m \leq n-2$ are compatible with $W_n^{(n)}$ and $W_0^{(n)}$, an intuitive definition of *minimal non-trivial CS* is the set of configurations that contains, in addition to $W_n^{(n)}$ and $W_0^{(n)}$, two additional compatible elements in $W_m^{(n)}$, $2 \leq m \leq n-2$. Formally speaking, we have the following definition:

Definition 12 (minimal non-trivial CS). *A CS $\mathcal{D}^{(n)}$ is called minimally non-trivial if $\mathcal{D}^{(n)} = \mathcal{D}_4^{(n)}$. Specifically, let $u^{(n)} \in W_m^{(n)}$ and $v^{(n)} \in W_{m'}^{(n)}, 2 \leq m, m' \leq n-2$ be two compatible configurations, the minimal non-trivial CS $\mathcal{D}^{(n)}$ consists of exactly 4 elements:*

$$\mathcal{D}^{(n)} = \mathcal{D}_4^{(n)} = \{W_n^{(n)}, W_0^{(n)}, u^{(n)}, v^{(n)}\} \tag{C.22}$$

The following Result establishes the number of minimal non-trivial CSs.

Result 7. *Let $D_4^{(n)}$ be the number of CS of size 4 at level n, namely, the cardinality of minimal non-trivial CS $\mathcal{D}_4^{(n)}$. Then for all $n \geq 4$,*

$$\begin{aligned} D_4^{(n)} &= 2^{n-1} - (n+1) + 2\sum_{k=4}^{n-4}(k-3)\binom{n}{k}\mathbb{1}_{n\geq 8} \\ &+ \left\{\sum_{k=3}^{n-3}(n-k-2)(k-2) - \binom{n}{k}\sum_{k=3}^{\lfloor \frac{n}{2} \rfloor}(n-k-3)(k-2)\binom{n}{k}\right\}\mathbb{1}_{n\geq 6} \\ &+ 2 \times \left\{\sum_{k=2}^{\lfloor \frac{n-1}{2} \rfloor}\binom{n}{k} \times \left(n-k-\frac{5}{2}\right) + \sum_{k=\lceil \frac{n-1}{2} \rceil}^{n-3}(n-k-2)\binom{n}{k}\right\}\mathbb{1}_{n\geq 5}, \end{aligned} \tag{C.23}$$

where $\mathbb{1}_{n\geq m}$ is an indicator function that is 1 whenever $n \geq m$ and 0 otherwise, $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$ are the floor and ceiling function, respectively.

Derivation. Since we're considering only $D_4^{(n)}$, it is sufficient to identify all possible compatible pairs of configurations (u, v), where $u \in W_m^{(n)}$ and $v \in W_{m'}^{(n)}$, $2 \leq m, m' \leq n-2$. Let x, y, w, z be the number of $(u_i, v_i) = (1,0), (0,1), (1,1), (0,0)$ pairs in u and v, respectively. By Result 5, $x, y \in \{0\} \cup \{n \geq 2 | n \in \mathbb{Z}_+\}$. We can parse it into different cases.

1. $z = w = 0$

   In this case $x, y \geq 2$ since $2 \leq m, m' \leq n-2$. Then it amounts to count the number of configurations that solves $x + y = n$, which is

$$\frac{1}{2}\sum_{k=2}^{n-2}\binom{n}{k} = 2^{n-1} - (n+1) \tag{C.24}$$

where the $1/2$ factor is to mitigate over-counting since x and y are symmetric. It is inherent that this applies only when $n - 2 \geq 2 \Rightarrow n \geq 4$.

2. $w, z \geq 1$ and $x, y \geq 2$
   The problem translates into the nonnegative integer solution to $x + y + z + w = n$, where $x, y \geq 2$ and $w, z \geq 0$. Define $x' = x + w$. Suppose $x' = k$, then $y + z = n - k$. The number of solution to this equation is $H^2_{n-k-3} = \binom{n-k-2}{1} = n - k - 2$. In addition, since $x' := x + w = k$ there are $H^2_{k-3} = \binom{k-2}{1} = k - 2$ solutions to this equation. Finally, there are $\binom{n}{k}$ permutations of such configuration pairs. In sum, we have

$$\sum_{k=3}^{n-3}(n-k-2)(k-2)\binom{n}{k}.$$

   However, we double-count whenever $x = y$. This can be mitigated by subtracting these cases:

$$\sum_{k=3}^{n-3}(n-k-2)(k-2)\binom{n}{k} - \sum_{k=3}^{\lfloor \frac{n}{2} \rfloor}(n-k-3)(k-2)\binom{n}{k}, \tag{C.25}$$

   where by writing this way it is inherent that $n - 3 \geq 3 \Rightarrow n \geq 6$. In other words, we need to take this case into account only when $n \geq 6$.

3. $w, z \geq 1$ and $x, y = 0$
   This is the null case since we're not pairing two distinct pairs at all.

4. $w, z \geq 1$ and $x = 0$ and $y \geq 2$ (or $y = 0$ and $x \geq 2$)
   In this case $y + z + w = n$ and $y \geq 2$ since if otherwise we would include $W^{(n)}_n$ and $W^{(n)}_{n-1}$. This means $y' \equiv y + z \in [4, n - 4]$ and the number of configurations that solves this equation is

$$2 \times \sum_{k=4}^{n-4} H^2_{k-4}\binom{n}{k} = 2\sum_{k=4}^{n-4}(k-3)\binom{n}{k}, \tag{C.26}$$

where the factor 2 is to take into account to two symmetric cases ($x = 0$ or $y = 0$). It is inherent that this applies only when $n - 4 \geq 4 \Rightarrow n \geq 8$.

5. $w = 0,\ z \geq 1$ and $x, y \geq 2$ (also the symmetric case $z = 0,\ w \geq 1$ and $x, y \geq 2$) Let $x = k \geq 2$, then $y + z = n - k$. Since $y \geq 2$ and $z \geq 1$, $k \in [2, n-3]$. Intuitively the number of configurations that solve this equation is

   $$2 \times \sum_{k=2}^{n-3} H^2_{n-k-3} \binom{n}{k} = 2 \sum_{k=2}^{n-3} (n-k-2) \binom{n}{k},$$

   where factor 2 is to take into account the symmetric case: ($z = 0, w \geq 1$) OR ($w = 0, z \geq 1$). However, we double-count whenever $x = y$, which occurs once for all $z = n - x - y = n - 2k \geq 1 \Rightarrow k \leq \lceil \frac{n-1}{2} \rceil$. To single this out, we have

   $$2 \times \left\{ \sum_{k=2}^{\lfloor \frac{n-1}{2} \rfloor} \binom{n}{k} \times \left[(n-k-3) + \frac{1}{2} \right] + \sum_{k=\lceil \frac{n-1}{2} \rceil}^{n-3} (n-k-2) \binom{n}{k} \right\} \tag{C.27}$$

   It is inherent that this applies only when $n - 3 \geq 2 \Rightarrow n \geq 5$.

6. Cases in which at least one of x, y is zero is not feasible due to biophysical constraints.

By summing over all cases the result follows. □

C.5 Capacity of computation: synthetic systems with writers only

In the previous section, we started from the biophysics model and proved the necessary and sufficient condition for two configurations to be compatible (i.e. res 5). In addition, we derived the size of the minimal non-trivial CS in res 7. In this section we use res 5 to build up the scaling for capacity. Note that since we're interested in asymptotic behavior of capacity, the restriction on $W_1^{(n)}$ and $W_{n-1}^{(n)}$ can be relieved since they have nearly zero contribution to capacity ($2 \ll 2^n$).

Result 8 (Probability of incompatible configurations)**.** *The probability that two configurations are incompatible in an n-writer-protein system, $P(n)$, is given by*

$$
\begin{aligned}
P(n) &= \frac{1}{2^{n+2}}\left\{4n\cdot\left(\frac{3}{2}\right)^{n-1} - n(n-1)\right\} && \text{(C.28)}\\
&\xrightarrow{n\gg 1} \frac{1}{2^{n+2}}\left[4n\left(\frac{3}{2}\right)^{n-1} - n^2\right] \\
&\approx \frac{n}{2}\left(\frac{3}{4}\right)^{n-1} && \text{(C.29)}
\end{aligned}
$$

Derivation. Let's consider two random configurations (binary bit-stings of length n). From res 5, the only chance for them to be incompatible is either (i) there exists *exactly* one bit position of $(1,0)$ or $(0,1)$ (ii) or both. Thus, the probability of such *incompatible* events reads:

$$
\begin{aligned}
P(n) &= \frac{2\cdot n\cdot[3^{n-1}-(n-1)\cdot 2^{n-2}] + \binom{n}{2}\cdot 2!\cdot 2^{n-2}}{4^n} && \text{(C.30)}\\
&= \frac{2^{n-2}}{4^n}\left\{2n\left[2\cdot\left(\frac{3}{2}\right)^{n-1} - (n-1)\right] + n(n-1)\right\}\\
&= \frac{1}{2^{n+2}}\left\{4n\cdot\left(\frac{3}{2}\right)^{n-1} - n(n-1)\right\}\\
&\xrightarrow{n\gg 1} \frac{1}{2^{n+2}}\left[4n\left(\frac{3}{2}\right)^{n-1} - n^2\right]\\
&\approx \frac{n}{2}\left(\frac{3}{4}\right)^{n-1}, && \text{(C.31)}
\end{aligned}
$$

where the first term and the second term in Eq.(C.30) represents the two cases (i) and (ii), respectively. Note that in case (i) we need to first select a bit position that has $(1,0)$ or $(0,1)$. Then for the remaining $n-1$ bit positions we need to subtract the situation in which the complementary $(0,1)$ or $(1,0)$ exists since that will be taken care of in case (ii). □

Result 9 (Capacity of computation)**.** *Suppose that the biophysics governing the compatibility of any two proteins are independent. Then the capacity of computation C, which is defined as the size of CS, in the large n (i.e. number of proteins in the system) limit is* *given by*

$$C = \frac{1}{2} + 2\sqrt{\delta}n^{-1/2}\left(\frac{4}{3}\right)^{\frac{n-1}{2}} \tag{C.32}$$

Moreover, the large deviation rate function of $P_C^{(\delta)}(n) := C/2^n$*, which measure the degree to which the capacity is constrained by biophysics scales as* $\log n/n$ *and converges asymptotically to* $-\frac{1}{2}\log\frac{4}{3} + \log 2 = 0.5493$*. Concretely,*

$$\begin{aligned} \tilde{I}_C &\cong \frac{1}{2n}\left(\log n - 2\log(2\sqrt{\delta}) + \log\frac{4}{3}\right) - \frac{1}{2}\log\frac{4}{3} + \log 2 \\ &\xrightarrow{n \gg 1} -\frac{1}{2}\log\frac{4}{3} + \log 2 = 0.5493 \end{aligned} \tag{C.33}$$

Derivation. For a CS with capacity C, every pair of configurations within it should be compatible. Since there are $\binom{C}{2}$ pairs, the probability such a CS, which we denote as P_C, reads

$$P_C = [1 - P(n)]^{\frac{C(C-1)}{2}} \tag{C.34}$$

By taking the natural logarithm and solve for the quadratic equation one has:

$$C = \frac{1}{2}\left[1 \pm \sqrt{1 + \frac{8\log P_C}{\log(1 - P(n))}}\right] \tag{C.35}$$

Alternatively, we can define $\delta = -\log P_C > 0$ and reexpress Eq.(C.34) using Eq.(C.31) as

$$\begin{aligned} \delta &= -\log P_C \\ &= -\frac{C(C-1)}{2}\log[1 - P(n)] \\ &\approx \frac{C(C-1)}{2} \cdot \frac{n}{2}\left(\frac{3}{4}\right)^{n-1}, \end{aligned} \tag{C.36}$$

where in the last step we used res 8 and the fact that $P(n) \ll 1$ when $n \gg 1$. By rearranging it we have

$$C = \frac{1}{2} \pm 2\sqrt{\delta}n^{-1/2}\left(\frac{4}{3}\right)^{\frac{n-1}{2}} \tag{C.37}$$

One can extract the large deviation rate function $I_C(n)$, which is related to $C/2^n$, namely, an estimate of how much capacity is suppressed by biophysics, through $C/2^n \simeq e^{-nI_C(n)}$. The result in the large C limits reads

$$\begin{aligned}
\tilde{I}_C &:= \lim_{C\to\infty} I_C(n) \\
&= \lim_{C\to\infty}\left(\lim_{n\to\infty} -\frac{1}{n}\log\frac{C}{2^n}\right) \\
&\cong \frac{1}{2n}\left(\log n - 2\log(2\sqrt{\delta}) + \log\frac{4}{3}\right) - \frac{1}{2}\log\frac{4}{3} + \log 2 \qquad \text{(C.38)} \\
&\xrightarrow{n\gg 1} -\frac{1}{2}\log\frac{4}{3} + \log 2 = 0.5493 \qquad \text{(C.39)}
\end{aligned}$$

□

In Figure. **??** we plotted P_C as a function of C and n as well as the $P(n)$, the probability that two randomly chosen configurations of length n are incompatible. The scaling of capacity and the large deviation rate functions are illustrated there as well.

C.6 Capacity of computation: synthetic systems with writers and erasers

In the previous section we considered systems with n writer species (e.g. kinases). Here we incorporate eraser species (e.g. phosphatases) into our analysis. Focus on our synthetic system, now every species is either one of the two categories $\mathcal{C}_i \in \{\mathcal{W}, \mathcal{E}\}$, namely, writer or eraser. We order the species label such that the phosphorylation state vector $\mathbf{x} = (x_1, \cdots, x_m, x_{m+1}, \cdots, n)$ is partitioned into writer (kinase) sector $i \in \mathcal{W} = \{1, \cdots, m\}$ and the eraser (phosphatase) sector $i \in \mathcal{E} = \{m+1, \cdots, n\}$. In other words, such system contains m writer species and $n-m$ eraser species. Similar to the case without erasers, we can write down the probability of species i at state $x_i = 1$ (c.f. Eq.(C.10))

$$P_i(x_i = 1) = \sum_{j \text{ s.t. } \mathcal{C}_j = \mathcal{W}} \frac{x_j e^{-\beta[\Delta\epsilon(\alpha_i,\beta_j) - \mu_j(x_j)]}}{1 + \sum_{k,\forall \mathcal{C}_k} x_k e^{-\beta[\Delta\epsilon(\alpha_i,\beta_k) - \mu_k(x_k)]}}. \tag{C.40}$$

Note that the summation over k in the denominator is over all erasers and writers $\mathcal{E}, \mathcal{W}$ whereas that over j in the numerator is over writers $\mathcal{W}$ only. In the zero temperature limit, this probability becomes binary so that we can use $\Delta\epsilon_{ij} - \mu_j$ to characterize the exponents in the Boltzmann factor (see the discussion that led to Eq.(C.11) and Eq.(C.12)). Similar to Eq.(C.13), the interpretation of Eq.(C.40) reads:

$$P_i(x_i = 1)|_{\beta\to\infty} = \begin{cases} 1, \text{ if } \{\, \exists j \in \mathcal{W}_{\setminus i} \text{ s.t. } x_j = 1 \texttt{ AND } \Delta\epsilon_{ij} < 0 \,\} \\ \qquad \text{AND} \{\, \forall\ k \in \mathcal{E},\ x_k = 0 \texttt{ OR } \forall k \in \mathcal{E} \text{ s.t. } x_k = 1, \Delta\epsilon_{ik} > \Delta\epsilon_{ij} \,\} \\ 0, \text{ if } \{\, \forall j \in \mathcal{W}_{\setminus i},\ x_j = 0 \texttt{ OR } \forall j \in \mathcal{W}_{\setminus i} \text{ s.t. } x_j = 1, \Delta\epsilon_{ij} > 0 \,\} \\ \qquad \text{OR} \{\, \exists\ k \in \mathcal{E},\ \text{s.t. } x_k = 1 \texttt{ AND } \Delta\epsilon_{ik} < 0 \} \end{cases} \tag{C.41}$$

The biophysical interpretation of this equation line by line is summarized as follows.

Species i is phosphorylated if it is

1. At least bound to one active writer AND
2. All active erasers bind i loosely than active writers do

Whereas species i is not phosphorylated $(x_i = 0)$ if

3. There's no active writers present OR active writers exit but none of them binds i OR
4. It is bound to active erasers

Note that since in the zero temperature limit proteins are either in strongly bound or strongly unbound state and that case 1 dictates $\Delta\epsilon_{ij} < 0$, we can reasonably replace $\Delta\epsilon_{ik} - \mu_k > \Delta\epsilon_{ij} - \mu_j$ in the second line of Eq.(C.41) by $\Delta\epsilon_{ik} > 0$ (active erasers bind i extremely loose). With this we can rewrite Eq.(C.41) as

$$P_i(x_i = 1)|_{\beta\to\infty} = \begin{cases} 1,\ \text{if}\ \{\,\exists j \in \mathcal{W}_{\setminus i}\ \text{s.t.}\ x_j = 1\ \texttt{AND}\ \ \Delta\epsilon_{ij} < 0\,\} \\ \qquad \text{AND}\,\{\,\forall\ k \in \mathcal{E},\ x_k = 0\ \texttt{OR}\ \ \forall k \in \mathcal{E}\ \text{s.t.}\ x_k = 1, \Delta\epsilon_{ik} > 0\,\} \\ 0,\ \text{if}\ \{\,\forall j \in \mathcal{W}_{\setminus i},\ x_j = 0\ \texttt{OR}\ \ \forall j \in \mathcal{W}_{\setminus i}\ \ \text{s.t.}\ x_j = 1, \Delta\epsilon_{ij} > 0\,\} \\ \qquad \text{OR}\,\{\,\exists\, k \in \mathcal{E},\ \text{s.t.}\ \ x_k = 1\ \texttt{AND}\ \ \Delta\epsilon_{ik} < 0\} \end{cases} \tag{C.42}$$

Fact 4. *By contrasting Eq.*(C.13) *with Eq.*(C.42)*, it is clear that the only modification is the additional constraints imposed on the eraser sector $\mathcal{E}$. If the activity of i is ON (phosphorylated), these constraints on $\mathcal{E}$ are* hard*(i.e. $\forall$) whereas they are* soft *(i.e. $\exists$)when writer is OFF.*

Fact 5. *To check for compatibility for two configurations u and v, one should still focus on the bit positions i such that there is a disagreement. Let u, v be the two configurations of interest and suppose $(u_i, v_i) = (1, 0)$. The constraint imposed by $u_i = 1$ alone is on both $\theta_{ij}, \forall j \neq i, j \in \mathbf{1}(u_{\mathcal{W}})$ and $\theta_{ik}, \forall k \in \mathcal{E}, k \in \mathbf{1}(u_{\mathcal{E}})$:*

$$\begin{aligned} &\{\,\exists j \in \mathcal{W}_{\setminus i}\ \textit{s.t.}\ x_j = 1\ \textit{AND}\ \Delta\epsilon_{ij} < 0\,\} \\ \texttt{AND}\,&\{\,\forall\ k \in \mathcal{E},\ x_k = 0\ \textit{OR}\ x_k = 1, \Delta\epsilon_{ik} > 0\,\} \end{aligned} \tag{C.43}$$

However, that by the complementary $v_i = 0$ is rather subtle. It could,

1. *as if without eraser sector, regularize only its writer sector $\theta_{ij}, j \in \mathbf{1}(v_{\mathcal{W}})$,*

$$\{\,\forall j \in \mathcal{W}_{\setminus i},\ x_j = 0 \quad \textit{OR} \quad \forall j \in \mathcal{W}_{\setminus i}\ \textit{s.t.}\ x_j = 1, \Delta\epsilon_{ij} > 0\,\} \tag{C.44}$$

2. *or impose no constraint on the writer sector and a set of (possibly contradicting) constraints on the eraser:*

$$\{\exists\, k \in \mathcal{E},\ \textit{s.t.}\ \ x_k = 1\ \textit{AND}\ \Delta\epsilon_{ik} < 0\} \tag{C.45}$$

Note that imposing Eq.(C.43) *and Eq.*(C.44) *is the same as the case without erasers since the eraser sector of Eq.*(C.43) *and Eq.*(C.44) *will always be satisfied. However, Eq.*(C.43) *and Eq.*(C.45) *implies that one can simply compare the constraints on the eraser part since they always agree on the writer sector. Thus, the only way for incompatibility to occur is when there's a disagreement in the writer sector and the eraser sector fail to mitigate that through Eq.*(C.45)*, or vice versa. Since this two situations are symmetric, we state the incompatibility criterion for the former.*

Result 10 (Compatibility check: systems with erasers)**.** *Let $u^{(n)}, v^{(n)}$ be two configurations in an n species system among which m of them are erasers. Then under the writer-eraser symmetry, $u^{(n)}$ and $v^{(n)}$ are incompatible if and only if they satisfy at least one of the following*

1. *there exists exactly one i such that $(u_i^{(n)}, v_i^{(n)}) = (1,0)$ and $\mathbf{1}(v_{\mathcal{E}}) \subseteq \mathbf{1}(u_{\mathcal{E}})$*
2. *there exists exactly one i such that $(u_i^{(n)}, v_i^{(n)}) = (0,1)$ and $\mathbf{1}(u_{\mathcal{E}}) \subseteq \mathbf{1}(v_{\mathcal{E}})$*
3. *there exists exactly one pair of i, j such that $(u_i^{(n)}, v_i^{(n)}) = (1,0)$ and $(u_j^{(n)}, v_j^{(n)}) = (0,1)$ and $\mathbf{1}(v_{\mathcal{E}}) = \mathbf{1}(u_{\mathcal{E}})$*

Derivation. First note that the first and the second cases are symmetric so we can just prove the former. Let i be the bit position of disagreement, say, $(u_i, v_i) = (1,0)$. By Eq.(C.42), Fact 4 and Fact 5, it suffices to check if the condition on eraser sector can mitigate the incompatibility implied by comparing *only* the *writer* part of u and v. Specifically, $(u_i, v_i) = (1,0)$ implies $\forall k \in \mathbf{1}(u_{\mathcal{E}}), \theta_{ik} > 0$ as well as $\exists l \in \mathbf{1}(v_{\mathcal{E}})$ such that $\theta_{il} < 0$. For u and v to be compatible, there should exist at least one element in $\mathbf{1}(v_{\mathcal{E}})$ but not in $\mathbf{1}(u_{\mathcal{E}})$. In other words, $\exists \omega \in \mathbf{1}(v_{\mathcal{E}})$ but $\omega \notin \mathbf{1}(u_{\mathcal{E}}) \Rightarrow u$ and v are compatible. This is equivalent to the statement that if u and v are incompatible, then $\mathbf{1}(v_{\mathcal{E}}) \subseteq \mathbf{1}(u_{\mathcal{E}})$. Similarly, by applying this argument to case 3 one gets $\mathbf{1}(v_{\mathcal{E}}) \subseteq \mathbf{1}(u_{\mathcal{E}})$ and $\mathbf{1}(u_{\mathcal{E}}) \subseteq \mathbf{1}(v_{\mathcal{E}})$, meaning $\mathbf{1}(u_{\mathcal{E}}) = \mathbf{1}(v_{\mathcal{E}})$. By Result 5, these are the only three cases where incompatibilities could emerge from the writer sector. Since we have demonstrated the situation where the eraser sector fail to mitigate these incompatibilities, the sufficient part is established. One can easily check that these statements are also necessary due to the definition of incompatibility. □

Example 3. *Let* $|\mathcal{W}| = 5$*,* $|\mathcal{E}| = m = 2$ *and*

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} \textcolor{blue}{\boldsymbol{1}}\boldsymbol{00}\textcolor{red}{\boldsymbol{1}}\boldsymbol{1}\underline{01} \\ \textcolor{blue}{\boldsymbol{0}}\boldsymbol{0011}\underline{01} \\ \textcolor{blue}{\boldsymbol{0}}\boldsymbol{00}\textcolor{red}{\boldsymbol{0}}\boldsymbol{1}\underline{10} \end{pmatrix}$$

Then (u, v) *and* (w, v) *are incompatible while* u, w *are compatible because they share more than one (1,0) pairs and* $\boldsymbol{1}(w_{\mathcal{E}}) \not\subseteq \boldsymbol{1}(u_{\mathcal{E}})$*.*

With this we can calculate the probability that two randomly chosen configurations are incompatible.

Result 11 (Incompatibility probability: with eraser)**.** *Consider systems with* n *protein species among with* m *are writers and the remaining* $n - m$ *are erasers. Assume that the biophysics governing the compatibility of any two proteins are independent. Then under the writer-eraser symmetry, the probability of two configurations being incompatible* $P_m(n)$*, is given by*

$$P_m(n) = \left(\frac{2m}{3} \left[1 - \frac{m-1}{3} \left(\frac{2}{3} \right)^{m-2} \right] \left(\frac{3}{4} \right)^n + \frac{m(m-1)}{2^{n+2}} \right) + \begin{pmatrix} m \to n-m \\ n \to m \end{pmatrix} \tag{C.46}$$

Derivation. By Result 10, $P_m(n)$ is simply the product of the probability without erasers (i.e. the first term in the numerator of Eq.(C.30) in Result 8 times another factor that characterizes the eraser part. Since the first two cases are exactly the same under $u \leftrightarrow v$ symmetry, we prove only the first. To get the eraser probability, first fix the size of $u_{\mathcal{E}}$ and then count all possible $v_{\mathcal{E}}$ such that $\boldsymbol{1}(v_{\mathcal{E}}) \subseteq \boldsymbol{1}(u_{\mathcal{E}})$. Suppose there are $n - m - k$, $k \in [0, n - m]$ elements in $\boldsymbol{1}(u_{\mathcal{E}})$, namely, there are only k zeros in $u_{\mathcal{E}}$, then all allowable $\boldsymbol{1}(v_{\mathcal{E}})$ can never contain the k indices that $\boldsymbol{1}(u_{\mathcal{E}})$ missed, otherwise $\boldsymbol{1}(v_{\mathcal{E}})$ won't be contained in $\boldsymbol{1}(u_{\mathcal{E}})$. The number of such $\boldsymbol{1}(v_{\mathcal{E}})$ is basically the cardinality of the power set of the remaining $n - m - k$ degree of freedom: 2^{n-m-k}. Since there are $\binom{n-m}{k}$ such $\boldsymbol{1}(u_{\mathcal{E}})$, the probability that describes the eraser condition reads:

$$P_m^{\mathcal{E}_{1,2}}(n) = \frac{1}{4^{n-m}} \sum_{k=0}^{n-m} \binom{n-m}{k} \cdot 2^{n-m-k} = \left(\frac{3}{4} \right)^{n-m}, \tag{C.47}$$

where the second equality is obtained by invoking Binomial theorem and the symbol $\mathcal{E}_{1,2}$ indicates the relevant eraser sector for case 1 and 2. By Result 8, the probability for case 1 and 2 in Result 10 reads

$$P_m^{(\mathcal{W},\mathcal{E})_{1,2}}(n) = P^{\mathcal{W}_{1,2}}(m) \cdot P_m^{\mathcal{E}_{1,2}}(n) = \frac{1}{4^m}\left\{2m[3^{m-1} - (m-1)2^{m-2}]\right\} \cdot \left(\frac{3}{4}\right)^{n-m} \tag{C.48}$$

For case 3, applying the same argument to the eraser sector gives

$$P_m^{\mathcal{E}_3}(n) = \frac{1}{4^{n-m}} \sum_{k=0}^{n-m} \binom{n-m}{k} = \frac{1}{2^{n-m}}, \tag{C.49}$$

implying

$$P_m^{(\mathcal{W},\mathcal{E})_3}(n) = P^{\mathcal{W}_3}(m) \cdot P_m^{\mathcal{E}_3}(n) = \frac{m(m-1)\cdot 2^{m-2}}{4^m} \cdot \frac{1}{2^{n-m}}. \tag{C.50}$$

After taking into account the writer-eraser symmetry and summing all cases, one has

$$\begin{aligned} P_m(n) &= \left(P_m^{(\mathcal{W},\mathcal{E})_{1,2}}(n) + P_m^{(\mathcal{W},\mathcal{E})_3}(n)\right) + \begin{pmatrix} m \to n-m \\ n \to m \end{pmatrix} \\ &= \left(\frac{2m}{3}\left[1 - \frac{m-1}{3}\left(\frac{2}{3}\right)^{m-2}\right]\left(\frac{3}{4}\right)^n + \frac{m(m-1)}{2^{n+2}}\right) + \begin{pmatrix} m \to n-m \\ n \to m \end{pmatrix} \end{aligned}$$

□

C.7 Calculating capacity with Monte Carlo Sampling

Note that the symbol $\succ$ indicates ”elementwise greater” since *prob* is a vector of size *capacity*. Since this procedure is applied to all ensembles, we take their average as the estimate:

$$C = \frac{1}{\mathcal{N}} \sum_{l=1}^{\mathcal{N}} C^{(l)} \tag{C.51}$$

Python class to generate signaling network and to infer capacity using this Monte Carlo procedure is freely available at: `https://github.com/chinghao0703/MC_capacity`

Algorithm 2 MC capacity sampling

```
Require: Cutoff probability $cutoff \leftarrow 0.5$ and algorithm termination threshold $threshold \leftarrow 10^3$.
procedure CapacityMC(configurations in ensemble $l$)
    $capacity \leftarrow 2$
    $noupdates \leftarrow 0$
    while $noupdates < threshold$ do
        draw a sample of size $capacity$ from ensemble $l$
        record their probability as $prob$
        if $prob \succ cutoff$ then
            $capacity \leftarrow capacity + 1$
        else
            $noupdates \leftarrow noupdates + 1$
    $C^{(l)} \leftarrow capacity$
    return $C^{(l)}$
```

Appendix D

Appendix of Chapter 5

D.1 Biophysical models

The 11 basic reactions constituting the whole transport process are depicted in Figure D·1. Our model incorporates the known (simplified) mechanism of nuclear transport of cargo through binding with importin and the active consumption of energy through hydrolysis of GTP. Such process is facilitated by Ran's intrinsic GTPase activity, which is activated via interaction with the Ran GTPase activating protein (RanGAP). In addition, we also include the reverse conversion of RanGDP to RanGTP through the action of guanine Exchange Factor RCC1 (known as RanGEF). In addition to the *standard model* of nuclear transport whose biochemistry is summarized below, we also incorporates the backward reactions to account for the reversibile nature of this transport process (Kim and Elbaum, 2013b).

- (Reaction 10, 9) In the cytoplasm, say, compartment A, the complex formed by importin protein (transport receptor) and the cargo C interacts with the nuclear pore complex and passes through the channel into the nucleus (compartment B).
- (Reaction 7, 4, 3) In the nucleus, RanGTP competes for binding with the receptor and causes the receptor to dissociate from the cargo. The new complex formed by RanGTP and receptor then translocates to the cytoplasm while the cargo is left inside the nucleus.

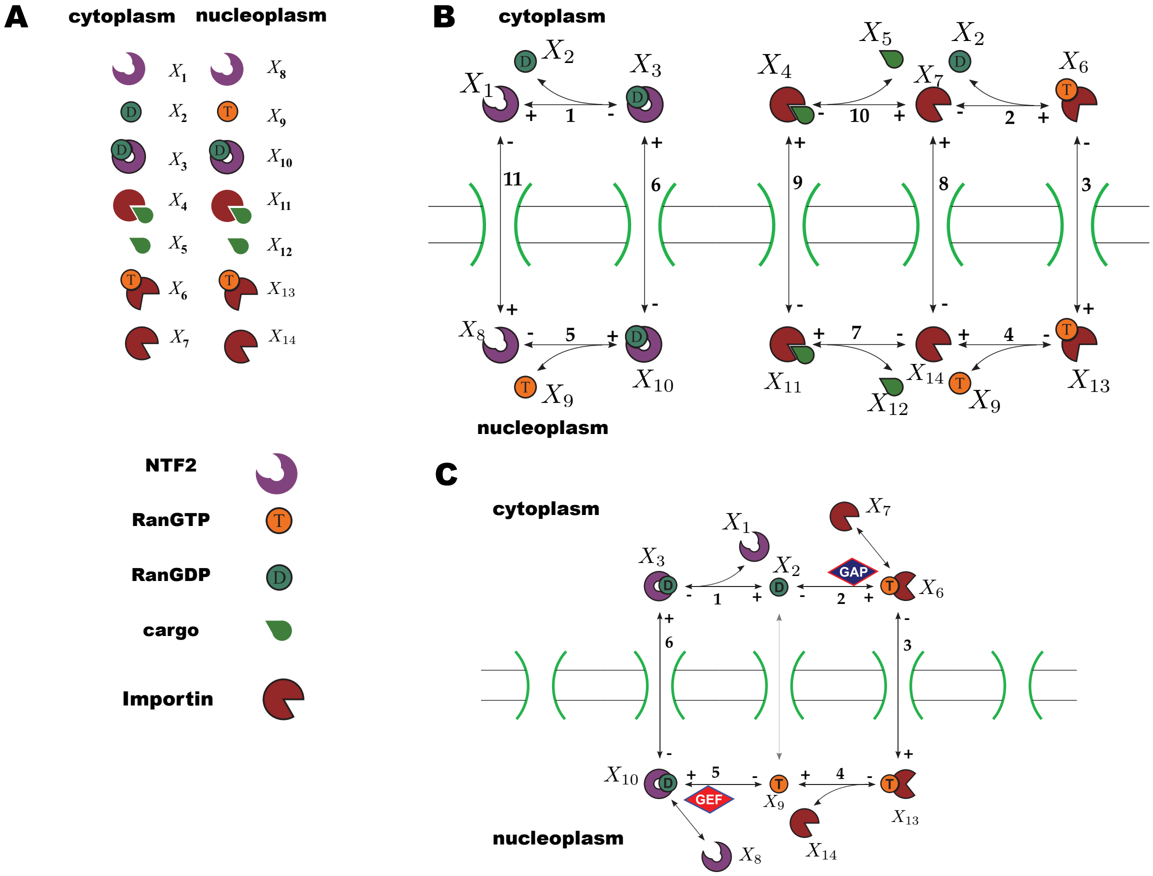

A
cytoplasm
nucleoplasm
X1
X8
X2
X9
X3
X10
X4
X11
X5
X12
X6
X13
X7
X14
D
T
NTF2
RanGTP
RanGDP
cargo
Importin
B
cytoplasm
nucleoplasm
X1
X2
X3
X4
X5
X2
X6
X7
X8
X9
X10
X11
X12
X14
X9
X13
1
10
2
11
6
9
8
3
5
7
4
+
-
C
cytoplasm
nucleoplasm
X3
X1
X2
X7
X6
GAP
X10
X9
X13
X14
X8
GEF
1
2
6
3
5
4
+
-

Figure D·1: Molecular reactions involved in nuclear transport. (A) Molecular species in our nuclear transport model. (B) Schematics of the reactions involved in nuclear transport. Note that + and - signs represent self-consistently the start and end points of the reactions, rather than forward or reverse cycle orientations. Thus for example $k_{``+"}$ and $k_{``-"}$ for reaction 9 in Eq.(S15) take signs for loss and gain, respectively. (C) The subset of reactions in (B) that forms the futile cycle (i.e. the energy source for nuclear transport).

- (Reaction 2) Once in the cytoplasm, the GTPase activating protein (RanGAP) then binds to RanGTP, causing the hydrolysis of GTP to GDP and release of energy.

- (Reaction 1, 6) The RanGDP produced in this process then binds the nuclear transport factor NTF2 which returns it to the nucleus.

- (Reaction 5) Now in the nucleus, RanGDP interacts with a guanine nucleotide exchange factor (GEF) which replaces GDP with GTP, resulting again a RanGTP from, and beginning a new cycle.

D.1.1 Kinetics equations

The whole process can be formulated by a set of kinetics equations involving both cargo protein translocation and Ran regulation. The molecular species in the kinetics equations are labelled according to Figure D·1.

$$[X_1] + [X_2] \underset{k_1^-}{\overset{k_1^+}{\rightleftharpoons}} [X_3] \tag{D.1}$$

$$[X_6] \underset{k_2^-}{\overset{k_2^+}{\rightleftharpoons}} [X_7] + [X_2] \tag{D.2}$$

$$[X_{13}] \underset{k_3^-}{\overset{k_3^+}{\rightleftharpoons}} [X_6] \tag{D.3}$$

$$[X_{14}] + [X_9] \underset{k_4^-}{\overset{k_4^+}{\rightleftharpoons}} [X_{13}] \tag{D.4}$$

$$[X_{10}] \underset{k_5^-}{\overset{k_5^+}{\rightleftharpoons}} [X_8] + [X_9] \tag{D.5}$$

$$[X_3] \underset{k_6^-}{\overset{k_6^+}{\rightleftharpoons}} [X_{10}] \tag{D.6}$$

$$[X_{11}] \underset{k_7^-}{\overset{k_7^+}{\rightleftharpoons}} [X_{14}] + [X_{12}] \tag{D.7}$$

$$[X_7] \underset{k_8^-}{\overset{k_8^+}{\rightleftharpoons}} [X_{14}] \tag{D.8}$$

$$[X_4] \underset{k_9^-}{\overset{k_9^+}{\rightleftharpoons}} [X_{11}] \tag{D.9}$$

$$[X_7] + [X_5] \underset{k_{10}^-}{\overset{k_{10}^+}{\rightleftharpoons}} [X_4] \tag{D.10}$$

$$[X_8] \underset{k_{11}^-}{\overset{k_{11}^+}{\rightleftharpoons}} [X_1] \tag{D.11}$$

From this we can write down the following kinetics:

$$\frac{d[X_1]}{dt} = -k_1^+[X_1][X_2] + k_1^-[X_3] + k_{11}^+[X_8] - k_{11}^-[X_1] \tag{D.12}$$

$$\frac{d[X_2]}{dt} = -k_1^+[X_1][X_2] + k_1^-[X_3] + k_2^+[X_6] - k_2^-[X_7][X_2] \tag{D.13}$$

$$\frac{d[X_3]}{dt} = k_1^+[X_1][X_2] - k_1^-[X_3] - k_6^+[X_3] + k_6^-[X_{10}] \tag{D.14}$$

$$\frac{d[X_4]}{dt} = -k_9^+[X_4] + k_9^-[X_{11}] + k_{10}^+[X_7][X_5] - k_{10}^-[X_4] \tag{D.15}$$

$$\frac{d[X_5]}{dt} = -k_{10}^+[X_5][X_7] + k_{10}^-[X_4] \tag{D.16}$$

$$\frac{d[X_6]}{dt} = -k_2^+[X_6] + k_2^-[X_2][X_7] + k_3^+[X_{13}] - k_3^-[X_6] \tag{D.17}$$

$$\frac{d[X_7]}{dt} = k_2^+[X_6] - k_2^-[X_2][X_7] - k_8^+[X_7] + k_8^-[X_{14}] - k_{10}^+[X_7][X_5] + k_{10}^-[X_4] \tag{D.18}$$

$$\frac{d[X_8]}{dt} = k_5^+[X_{10}] - k_5^-[X_8][X_9] - k_{11}^+[X_8] + k_{11}^-[X_1] \tag{D.19}$$

$$\frac{d[X_9]}{dt} = k_5^+[X_{10}] - k_5^-[X_8][X_9] - k_4^+[X_{14}][X_9] + k_4^-[X_{13}] \tag{D.20}$$

$$\frac{d[X_{10}]}{dt} = -k_5^+[X_{10}] + k_5^-[X_8][X_9] + k_6^+[X_3] - k_6^-[X_{10}] \tag{D.21}$$

$$\frac{d[X_{11}]}{dt} = -k_7^+[X_{11}] + k_7^-[X_{12}][X_{14}] + k_9^+[X_4] - k_9^-[X_{11}] \tag{D.22}$$

$$\frac{d[X_{12}]}{dt} = k_7^+[X_{11}] - k_7^-[X_{12}][X_{14}] \tag{D.23}$$

$$\frac{d[X_{13}]}{dt} = -k_3^+[X_{13}] + k_3^-[X_6] + k_4^+[X_{14}][X_9] - k_4^-[X_{13}] \tag{D.24}$$

$$\frac{d[X_{14}]}{dt} = -k_4^+[X_{14}][X_9] + k_4^-[X_{13}] + k_7^+[X_{11}] - k_7^-[X_{14}][X_{12}] + k_8^+[X_7] - k_8^-[X_{14}] \tag{D.25}$$

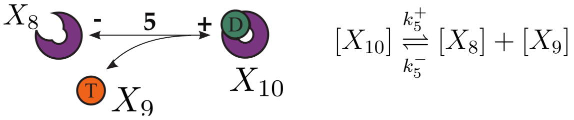
D.2 Estimating the rate constants

Here we list the kinetics rate constants used in the simulation. Some of them are directly available from literature while others are estimated as described below. In the following, $a = 100\,\mu\mathrm{m}^3\ \mathrm{s}^{-1}$ is the nuclear pore permeability and $v_N = 100\,\mu\mathrm{m}^3$ and $v_C = 500\,\mu\mathrm{m}^3$ are the nuclear and cytoplasm compartment volumes, respectively. The exponential free energy difference defined in Eq.(D.35)(D.43) are set to be: $e^{\Delta F} = e^{\Delta \tilde{F}} = 50$. Note that volume factors modulate the permeabilities in the usual manner (see Eq.(1) in (Kim and Elbaum, 2013b)): Namely, rate constants of cytosolic species (i.e. X_1, X_3, X_4, X_6, X_7) across the nuclear membrane is given by $k_\alpha^\pm = a/v_C$ with $\alpha = 3, 6, 8, 9, 11$ (i.e., reactions that involve crossing the nuclear pores). Rate constants for the nuclear counterparts (i.e. $X_8, X_{10}, X_{11}, X_{13}, X_{14}$) are, on the other hand, given by $k_\alpha^\pm = a/v_N$, with $\alpha = 3, 6, 8, 9, 11$. For example, kinetics equations for X_3 (cytosolic NTF2-RanGDP complex) and X_{10} (nuclear NTF2-RanGDP complex) should read (c.f. Eq.(D.14) and Eq.(D.21)):

$$\frac{d[X_3]}{dt} = k_1^+[X_1][X_2] - k_1^-[X_3] - \frac{a}{v_C}\left([X_3] - [X_{10}]\right) \tag{D.26}$$

$$\frac{d[X_{10}]}{dt} = -k_5^+[X_{10}] + k_5^-[X_8][X_9] + \frac{a}{v_N}\left([X_3] - [X_{10}]\right). \tag{D.27}$$

D.2.1 Reaction 5: Ran exchange mediated by RanGEF



$$[X_{10}] \underset{k_5^-}{\overset{k_5^+}{\rightleftharpoons}} [X_8] + [X_9]$$

Figure D·2: Illustration of Ran GDP to GTP exchange reaction mediated by RanGEF

Table D.1: Reaction rate constants used in the thermodynamic model of nuclear transport

reaction	K_D or k_{in}/k_{out}	k^+	k^-
1	25 nM	0.1 (nM^{-1} s^{-1})	2.5 (s^{-1})
2	$\sim$	$e^{\Delta\tilde{F}}/\theta$ (s^{-1})	1 (nM^{-1} s^{-1})
3	$\sim$	1 or 0.2(s^{-1})	1 or 0.2(s^{-1})
4	10 nM	0.1 (nM^{-1} s^{-1})	1 (s^{-1})
5	$\sim$	$e^{\Delta F} \times \theta$ (s^{-1})	1 (nM^{-1} s^{-1})
6	$\sim$	1 or 0.2 (s^{-1})	1 or 0.2 (s^{-1})
7	20 nM	20 (s^{-1})	1 (nM^{-1} s^{-1})
8	$\sim$	1 or 0.2 (s^{-1})	1 or 0.2(s^{-1})
9	$\sim$	1 or 0.2 (s^{-1})	1 or 0.2(s^{-1})
10	20 nM	1 (nM^{-1} s^{-1})	20 (s^{-1})
11	$\sim$	1 or 0.2(s^{-1})	1 or 0.2(s^{-1})

Table D.2: Molecular species involved in nuclear transport

Labels	Species
N	NTF2
Im	Importin (importin)
RD	RanGDP
RT	RanGTP
N·RD	NTF2+RanGDP complex
N·RT	NTF2+RanGTP complex
Im·RD	Importin+RanGDP complex
Im·RT	Importin+RanGTP complex
fD	(free) GDP
fT	(free) GTP

The goal is to estimate the K_D for the following reaction:

$$[N\cdot RD] \underset{k_5^-}{\overset{k_5^+}{\rightleftharpoons}} [N]+[RT], \tag{D.28}$$

namely,

$$\frac{k_5^+}{k_5^-} = \frac{[N][RT]}{[N\cdot RD]} \tag{D.29}$$

Consider the following two constituting reactions

$$[N\cdot RD]+[fT] \underset{k_\alpha^-}{\overset{k_\alpha^+}{\rightleftharpoons}} [N\cdot RT]+[fD] \tag{D.30}$$

$$[N\cdot RT] \underset{k_\beta^-}{\overset{k_\beta^+}{\rightleftharpoons}} [N]+[RT] \tag{D.31}$$

This implies (neglecting labels of steady states SS),

$$\frac{k_\alpha^+}{k_\alpha^-} = \frac{[N\cdot RT][fD]}{[N\cdot RD][fT]} \tag{D.32}$$

$$\frac{k_\beta^+}{k_\beta^-} = \frac{[N][RT]}{[N\cdot RT]} \tag{D.33}$$

Thus we can reexpress Eq.(D.32) using Eq.(D.33):

$$\frac{k_\alpha^+}{k_\alpha^-} = \frac{1}{[N\cdot RD]}\frac{[fD]}{[fT]}\cdot\left(\frac{k_\beta^-}{k_\beta^+}[N][RT]\right) = \left(\frac{[N][RT]}{[N\cdot RD]}\right)\cdot\frac{[fD]}{[fT]}\frac{k_\beta^-}{k_\beta^+} = \frac{k_5^+}{k_5^-}\cdot\frac{[fD]}{[fT]}\frac{k_\beta^-}{k_\beta^+} \tag{D.34}$$

Thus

$$\frac{k_5^+}{k_5^-} = \frac{k_\alpha^+}{k_\alpha^-}\cdot\frac{k_\beta^+}{k_\beta^-}\cdot\frac{[fT]}{[fD]} \sim O(1)\cdot k_0 e^{\Delta F}\cdot \exp\left(\log\frac{[fT]}{[fD]}\right) \tag{D.35}$$

The first term (i.e. k_α^+/k_α^-) comes from guanine nucleotide exchange reaction and is of order one while the second (i.e. k_β^+/k_β^-) is related to the free energy difference between binding and un-binding of NTF2+RanGTP complex which is much larger than 1: $\Delta F >> 1$. This can also be understood using Eq.(D.33) by noting that in the nucleus NTF2 seldom binds to RanGTP. Finally, since the free GTP to GDP ratio, $[fT]/[fD]$, is buffered by cellular metabolism, we simply treat the last term as a free parameter θ. Note that there is far more free GTP than Ran on a molar basis. After rescaling time by $\tau \leftarrow tc_0 k_{\text{diff}}$, with $k_{\text{diff}} = 10$ sec^{-1} nM^{-1} and c_0 represent the diffusion-limited reaction rate and the characteristic molar concentration (set to 1nM), respectively, and approximating $e^{\Delta F} \approx 10 \sim 100$, one can estimate $(k_5^+/k_5^-) \sim (10 \sim 100) \times \theta$, where $\theta := [fT]/[fD]$ is treated as a free parameter.

D.2.2 Reaction 2: Ran exchange mediated by RanGAP

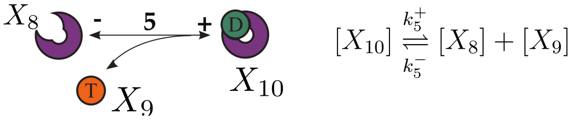

X_8
-
5
+
D
T
X_9
X_{10}
$[X_{10}] \underset{k_5^-}{\overset{k_5^+}{\rightleftharpoons}} [X_8] + [X_9]$

Figure D·3: Illustration of RanGTP to RanGDP exchange reaction mediated by RanGAP

We aim to approximate K_D for such reaction:

$$[Im \cdot RT] \underset{k_2^-}{\overset{k_2^+}{\rightleftharpoons}} [Im] + [RD], \tag{D.36}$$

$$\frac{k_2^+}{k_2^-} = \frac{[Im][RD]}{[Im \cdot RT]} \tag{D.37}$$

Similarly the estimation is based on the following two steps:

$$[Im \cdot RT] + [fD] \underset{k_\gamma^-}{\overset{k_\gamma^+}{\rightleftharpoons}} [Im \cdot RD] + [fT] \tag{D.38}$$

$$[Im \cdot RD] \underset{k_\delta^-}{\overset{k_\delta^+}{\rightleftharpoons}} [Im] + [RD] \tag{D.39}$$

This implies (neglecting labels of steady states SS),

$$\frac{k_\gamma^+}{k_\gamma^-} = \frac{[Im \cdot RD][fT]}{[Im \cdot RT][fD]} \tag{D.40}$$

$$\frac{k_\delta^+}{k_\delta^-} = \frac{[Im][RD]}{[Im \cdot RD]} \tag{D.41}$$

Thus we can reexpress Eq.(D.40) using Eq.(D.41):

$$\frac{k_\gamma^+}{k_\gamma^-} = \frac{1}{[Im \cdot RT]} \frac{[fT]}{[fD]} \cdot \left(\frac{k_\delta^-}{k_\delta^+} [Im][RD] \right) = \left(\frac{[Im][RD]}{[Im \cdot RT]} \right) \cdot \frac{[fT]}{[fD]} \frac{k_\delta^-}{k_\delta^+} = \frac{k_2^+}{k_2^-} \cdot \frac{[fT]}{[fD]} \frac{k_\delta^-}{k_\delta^+} \tag{D.42}$$

Thus

$$\frac{k_2^+}{k_2^-} = \frac{k_\gamma^+}{k_\gamma^-} \cdot \frac{k_\delta^+}{k_\delta^-} \cdot \frac{[fD]}{[fT]} \sim O(1) \cdot k_0 e^{\Delta \tilde{F}} \cdot \exp\left(\log \frac{[fD]}{[fT]} \right) \sim k_0 \times (10 \sim 100) \times \frac{1}{\theta} \tag{D.43}$$

D.3 Standard estimate of diffusion-limited reaction rate

Considering two type of molecules A and B diffusing in a viscous environment. According the Fick's law the diffusion flux of one type of molecule assuming the other is at stationary is given as

$$\vec{J}_\mu = -D_\mu \nabla [\mu], \tag{D.44}$$

where $\mu = A, B$ and D_μ is the diffusion constant of molecule μ. Assuming spherical symmetry one can integrate Fick's law to get the total number of molecules diffusing through a given surface area:

$$\phi_{tot} = 4\pi R (D_A + D_B)[A][B], \tag{D.45}$$

where R is the sum of molecular radii of A and B. The factor $k_a := 4\pi R(D_A + D_B)$ is exactly the reaction rate of the overall catalytic reaction under the assumption that the process is diffusion-limited (i.e. upon A and B are in contact, the intermediate complex AB immediately reacts to form the final product P):

$$A + B \xrightarrow{k_a} P \tag{D.46}$$

Finally, recall Stokes-Einstein relation: $D = k_B T/(6\pi\eta a)$ with molecule (spherical) particle radius a, we have

$$k_a = 4\pi(2a)\left(2 \times \frac{k_B T}{6\pi\eta a}\right)\left[\frac{\mathrm{m}^3}{\mathrm{sec}}\right] \to \left(\frac{8k_B T}{3\eta}\right) N_A 10^3 \left[\frac{1}{\mathrm{M}\cdot\mathrm{sec}}\right], \tag{D.47}$$

where N_A is the Avogadro's constant. The factor 10^3 appears because we convert the SI unit of volume m^3 to liter. Using $\eta = 10^{-1}$ (Pa·sec)= 10^{-3} kg/m/sec, we get

$$k_{\mathrm{diff}} := k_a \sim 10 \times 10^9\,[\mathrm{M}^{-1}\cdot\ \mathrm{sec}^{-1}] = 10\,[\mathrm{nM}^{-1}\cdot\ \mathrm{sec}^{-1}] \tag{D.48}$$

D.4 Simulation codes

MATLAB® simulation codes are available for download at `http://physics.bu.edu/~chinghao/thermo_transport/codes/`

D.5 Entropy Production

The distinct feature of systems out of thermodynamics equilibrium is the continuous production of entropy. The rate of entropy change (in time) consists of two parts: (i) the internal entropy change and (ii) the exchange of entropy with the environment

$$\frac{dS}{dt} = \Pi - \Phi, \tag{D.49}$$

where S is the entropy of the system and Π is the rate of entropy production and Φ denotes the rate of entropy flow from the system to the outside. Within this context, the 2nd law of thermodynamics dictates $\Pi \geq 0$ and the notion of steady states translates into $\Pi = \Phi$: entropy produced is continuously given away to the environment. One can further distinguishes the equilibrium from the nonequilibrium steady states by

- Equilibrium steady states (ESS): $\Pi = \Phi = 0$
- Nonequilibrium steady states (NESS, i.e. irreversible): $\Pi = \Phi > 0$

Consider systems that can be described by a continuous time Markov process such that the probability flow can be written as a master equation:

$$\frac{d}{dt}P_i(t) = \sum_j [P_j(t)W_{ji} - P_i(t)W_{ij}], \tag{D.50}$$

where W_{ij} is the transition rate from state j to state i and $P_i(t)$ is the probability of state i at time t. An appropriate microscopic description for the nonequilibrium system amounts to (i) having well-defined entropy for the irreversible systems and (ii) the entropy production rate Π should respects the non-negativity and should vanish when system equilibrates (i.e. when it exhibits reversibility). For systems described by the master equation, thermodynamics equilibrium is essentially the detailed-balanced condition: $P_i W_{ij} = P_j W_{ji}$. The solution for the first is the Boltzmann-Gibbs entropy:

$$S(t) = -k_B \sum_i P_i(t) \log P_i(t), \tag{D.51}$$

while the entropy production rate is advanced by the Schnakenberg description (Lebowitz and Spohn, 1999):

$$\Pi(t) = \frac{k_B}{2} \sum_{ij} [P_i(t)W_{ij} - P_j(t)W_{ji}] \log \frac{W_{ij}P_j(t)}{W_{ji}P_i(t)} \tag{D.52}$$

By imposing $dS(t)/dt = 0$ to Eq.(D.51) at steady state and using Eq.(D.50)(D.52) to simply, one gets the steady state entropy production rate:

$$\Pi = k_B \sum_{ij} W_{ij} P_j \log \frac{W_{ij}}{W_{ji}}, \tag{D.53}$$

where P_i is the stationary probability distribution. It's easy to check that $\Pi - \Pi(t) = dS/dt \to 0$ in the stationary state.

One can map the network in FIG. D·1B to a nonequilibrium Markov process. A non-equilibrium steady state (NESS) essentially necessitates breaking the detailed balance in the underlying Markov process and therefore, the system has a nonzero entropy production that is continuously given away to the environment. Such entropy production is exactly the power consumed by the circuit to maintain NESS. Now defining $EP := \Pi \times T$ using (D.53) (in the same spirit as $F = U - TS$, where F is the Helmholtz free energy), we have

$$EP = k_B T \sum_{i,j} P_i^{SS} W(i,j) \log \frac{W(i,j)}{W(j,i)}, \tag{D.54}$$

where P_i^{SS} is the steady state probability distribution of state i while $W(i,j)$ denotes the transition probability from state i to state j. Concretely, P_i^{SS} is the fraction of reactants participating in the transition reaction starting from state i while $W(i,j)$ can be calculated from the relevant reaction fluxes. For example, P_3^{SS} is the molar fraction of cytoplasmic NTF2-RanGDP ($\sim [X_3]$) whereas $W(3,10)$ is the transition probability of of NTF2-RanGDP into the nucleus: $W(3,10) = (k_6^+[X_3])/(k_6^+[X_3] + k_1^-[X_3])$ (See FIG.D·1C). Note that in principle the summation in Eq.(D.54) to obtain the entropy production is taken over all links in FIG.D·1B. It can be separated, however, into reactions 1-6 that represent the Ran futile cycle (i.e. FIG. D·1C) and the remaining reactions 7-11 that do not explicitly involve Ran. The latter are essentially passive and could be expected to satisfy detailed balance at steady state. We have confirmed numerically that the contributions of reactions 7-11 in Eq.(D.54) cancel to zero, so the total entropy production is equal that evaluated in the futile cycle alone. We can also inspect the reactions qualitatively. Reaction 9 is trivially in detailed balance because the concentrations X_4 and X_{11} are equal in steady state. These represent the importin-cargo complex in cytoplasm and nucleus, respectively. Clearly the net cargo binding/unbinding to importin in the cytoplasm must balance that in the nucleus as well, so the contributions of reactions 7 and 10 cancel. Finally, the free receptors importin and NTF2 exchange passively across the nuclear envelope (reactions 8 and 11). Again in steady state their cycle fluxes must balance, so their contributions to the entropy production sum also cancel.

D.6 Weak sensitivity to the GTP:GDP ratio θ

Ultimately the (chemical) free-energetic fuel driving the transport cycle is the ratio of GTP to GDP, θ, which is held out of equilibrium by cellular metabolism. We find that the nuclear localization ratio, as well as biased concentrations of transport receptors, is not strongly dependent on θ. This reflects the counterbalancing effects of RanGEF and RanGAP as described in Section II above.

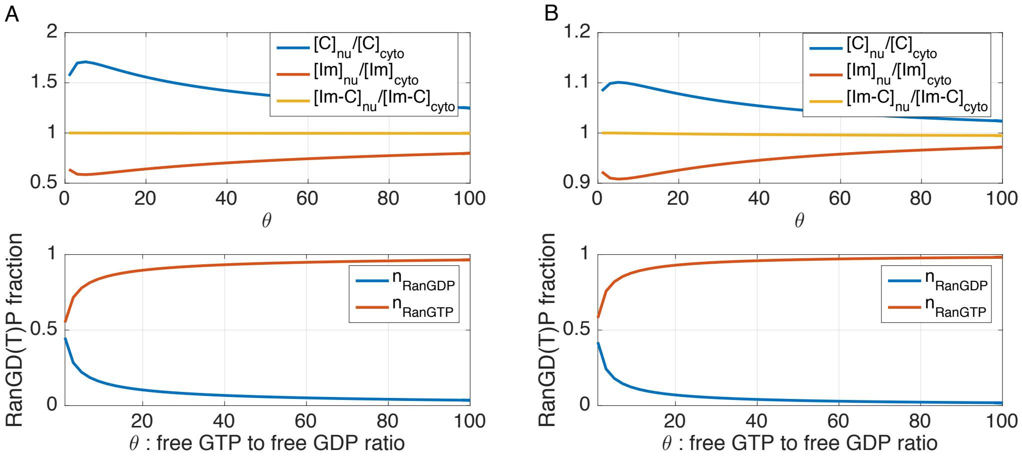


Figure D·4: Effect of free GTP to free GDP ratio θ on nuclear localization. (A) [NTF2]$_{tot}$=100 nM (B) [NTF2]$_{tot}$=10 nM. Other parameters are the same for both panels: [C]$_{tot}$=10 nM, [Ran]$_{tot}$=75 nM and [Im]$_{tot}$=100 nM. Kinetics rate constants used are given in SM Section II.

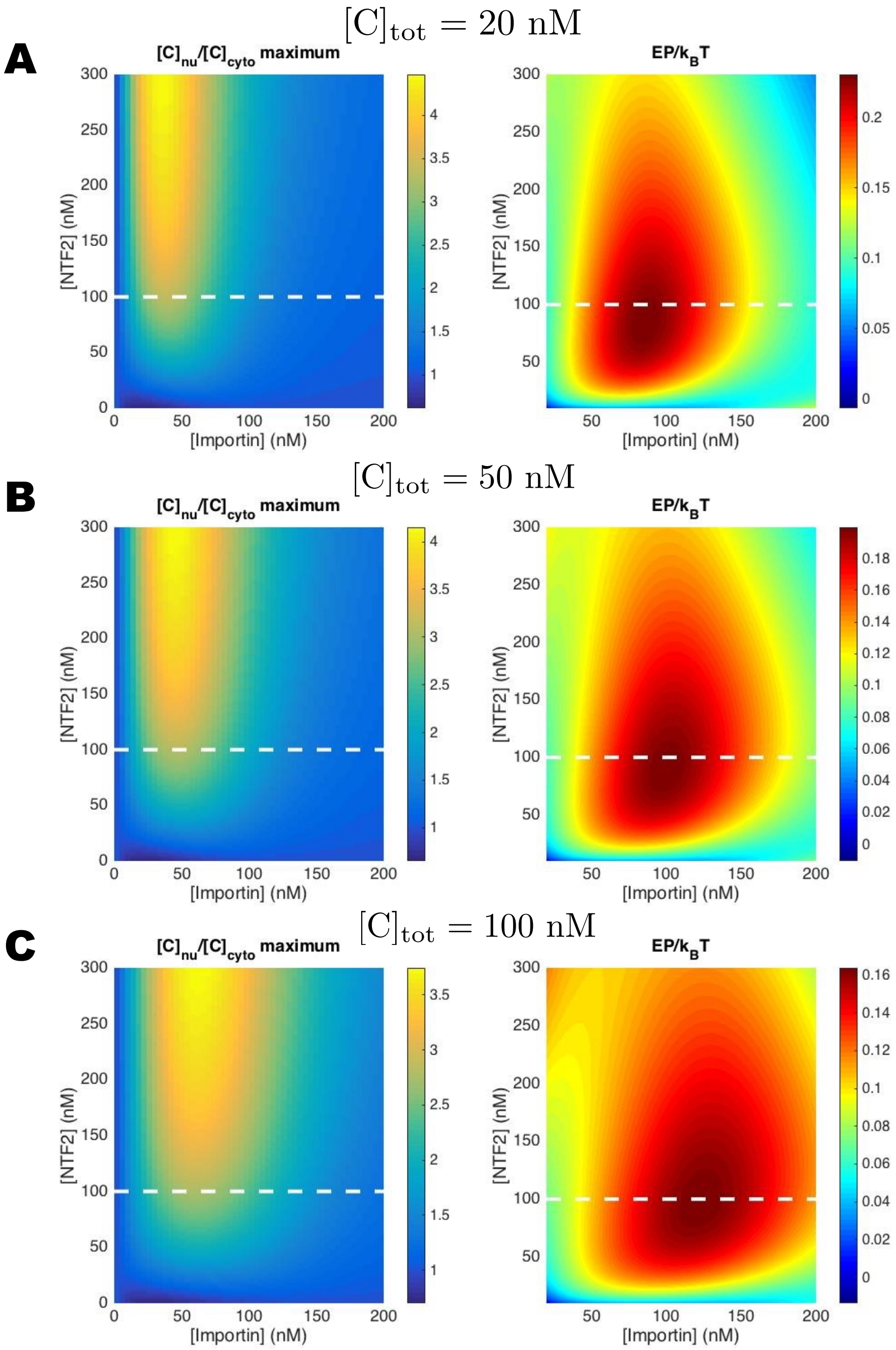


Figure D·5: Phase diagram of nuclear localization and entropy production. (A) total cargo concentration $[\mathrm{C}]_{\mathrm{tot}} = 20$ nM (B) $[\mathrm{C}]_{\mathrm{tot}} = 50$ nM and (C) $[\mathrm{C}]_{\mathrm{tot}} = 100$ nM. Other parameters used are the same as in FIG.3 and 4: $[\mathrm{Ran}]_{\mathrm{tot}} = 75$ nM.

References

See Supplemental Material, which includes Refs. (Kim and Elbaum, 2013b; Smith et al., 2002; Harreman et al., 2003; Lebowitz and Spohn, 1999).

Ackers, G. K., Johnson, A. D., and Shea, M. A. (1982). Quantitative model for gene regulation by lambda phage repressor. *Proceedings of the National Academy of Sciences*, 79(4):1129–1133.

Alon, U. (2007). Network motifs: theory and experimental approaches. *Nature Reviews Genetics*, 8(6):450–461.

Anastas, J. N. and Moon, R. T. (2013). Wnt signalling pathways as therapeutic targets in cancer. *Nature Reviews Cancer*, 13(1):11.

Angers, S. and Moon, R. T. (2009). Proximal events in wnt signal transduction. *Nature reviews Molecular cell biology*, 10(7):468.

Anjum, R. and Blenis, J. (2008). The rsk family of kinases: emerging roles in cellular signalling. *Nature reviews. Molecular cell biology*, 9(10):747.

Barabasi, A.-L. and Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nature reviews genetics*, 5(2):101.

Barato, A. C. and Seifert, U. (2015). Thermodynamic uncertainty relation for biomolecular processes. *Physical review letters*, 114(15):158101.

Bashor, C. J. and Collins, J. J. (2018). Understanding biological regulation through synthetic biology. *Annual review of biophysics*, 47:399–423.

Bashor, C. J., Horwitz, A. A., Peisajovich, S. G., and Lim, W. A. (2010). Rewiring cells: synthetic biology as a tool to interrogate the organizational principles of living systems. *Annual review of biophysics*, 39:515.

Bennett, C. H. (1982). The thermodynamics of computation—a review. *International Journal of Theoretical Physics*, 21(12):905–940.

Berg, H. C. and Purcell, E. M. (1977). Physics of chemoreception. *Biophysical journal*, 20(2):193.

Bintu, L., Buchler, N. E., Garcia, H. G., Gerland, U., Hwa, T., Kondev, J., Kuhlman, T., and Phillips, R. (2005a). Transcriptional regulation by the numbers: applications. *Current opinion in genetics & development*, 15(2):125–135.

Bintu, L., Buchler, N. E., Garcia, H. G., Gerland, U., Hwa, T., Kondev, J., and Phillips, R. (2005b). Transcriptional regulation by the numbers: models. *Current opinion in genetics & development*, 15(2):116–124.

Bintu, L., Buchler, N. E., Garcia, H. G., Gerland, U., Hwa, T., Kondev, J., and Phillips, R. (2005c). Transcriptional regulation by the numbers: models. *Current opinion in genetics & development*, 15(2):116–124.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer-Verlag Berlin.

Blais, A. and Dynlacht, B. D. (2005). Constructing transcriptional regulatory networks. *Genes & development*, 19(13):1499–1511.

Bos, J. L., Rehmann, H., and Wittinghofer, A. (2007). Gefs and gaps: critical elements in the control of small g proteins. *Cell*, 129(5):865–877.

Breitkreutz, A., Choi, H., Sharom, J. R., Boucher, L., Neduva, V., Larsen, B., Lin, Z.-Y., Breitkreutz, B.-J., Stark, C., Liu, G., et al. (2010). A global protein kinase and phosphatase interaction network in yeast. *Science*, 328(5981):1043–1046.

Brennan, M. D., Cheong, R., and Levchenko, A. (2012). How information theory handles cell signaling and uncertainty. *Science*, 338(6105):334–335.

Cheong, R., Rhee, A., Wang, C. J., Nemenman, I., and Levchenko, A. (2011). Information transduction capacity of noisy biochemical signaling networks. *science*, page 1204553.

Chung, I., Akita, R., Vandlen, R., Toomre, D., Schlessinger, J., and Mellman, I. (2010). Spatial control of egf receptor activation by reversible dimerization on living cells. *Nature*, 464(7289):783.

Cover, T. M. (1965). Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Transactions on Electronic Computers*, EC-14(3):326–334.

Cover, T. M. and Thomas, J. A. (2012). *Elements of information theory*. John Wiley & Sons.

Cui, Q. and Karplus, M. (2008). Allostery and cooperativity revisited. *Protein science*, 17(8):1295–1307.

Dayarian, A., Chaves, M., Sontag, E. D., and Sengupta, A. M. (2009). Shape, size, and robustness: feasible regions in the parameter space of biochemical networks. *PLoS Comput Biol*, 5(1):e1000256.

Deribe, Y. L., Pawson, T., and Dikic, I. (2010). Post-translational modifications in signal integration. *Nature structural & molecular biology*, 17(6):666.

Detwiler, P. B., Ramanathan, S., Sengupta, A., and Shraiman, B. I. (2000). Engineering aspects of enzymatic signal transduction: photoreceptors in the retina. *Biophysical Journal*, 79(6):2801–2817.

Dijkstra, M. and Frenkel, D. (1994). Evidence for entropy-driven demixing in hard-core fluids. *Physical review letters*, 72(2):298.

D'Angelo, M. A. and Hetzer, M. W. (2008). Structure, dynamics and function of nuclear pore complexes. *Trends in cell biology*, 18(10):456–466.

Einav, T., Mazutis, L., and Phillips, R. (2016). Statistical mechanics of allosteric enzymes. *The Journal of Physical Chemistry B*, 120(26):6021–6037.

Elowitz, M. B. and Leibler, S. (2000). A synthetic oscillatory network of transcriptional regulators. *Nature*, 403(6767):335–338.

Endres, R. G. and Wingreen, N. S. (2009). Maximum likelihood and the single receptor. *Physical review letters*, 103(15):158101.

Engel, A. and Van den Broeck, C. (2001). *Statistical mechanics of learning*. Cambridge University Press.

Fischbach, M. A., Bluestone, J. A., and Lim, W. A. (2013). Cell-based therapeutics: the next pillar of medicine. *Science translational medicine*, 5(179):179ps7–179ps7.

Garcia, H. G. and Phillips, R. (2011). Quantitative dissection of the simple repression input–output function. *Proceedings of the National Academy of Sciences*, 108(29):12173–12178.

Garcia, H. G., Sanchez, A., Kuhlman, T., Kondev, J., and Phillips, R. (2010). Transcription by the numbers redux: experiments and calculations that surprise. *Trends in cell biology*, 20(12):723–733.

Garrett, R. and Grisham, C. (2010). Biochemistry, brooks. *Cole, Cengage learning*, pages 321–323.

Garrett, R. H. and Grisham, C. M. (2008). *Biochemistry*. Brooks Cole.

Gerhart, J. and Kirschner, M. (1997). *Cells, embryos, and evolution: Toward a cellular and developmental understanding of phenotypic variation and evolutionary adaptability*. Number 575.21 GER. Blackwell Science Malden.

Gingrich, T. R., Horowitz, J. M., Perunov, N., and England, J. L. (2016). Dissipation bounds all steady-state current fluctuations. *Physical review letters*, 116(12):120601.

Glynn, P. D. and Reardon, E. J. (1990). Solid-solution aqueous-solution equilibria: thermodynamic theory and representation. *American Journal of Science*, 290(2):164–201.

Görlich, D. and Kutay, U. (1999). Transport between the cell nucleus and the cytoplasm. *Annual review of cell and developmental biology*, 15(1):607–660.

Görlich, D., Seewald, M. J., and Ribbeck, K. (2003). Characterization of ran-driven cargo transport and the rangtpase system by kinetic measurements and computer simulation. *The EMBO journal*, 22(5):1088–1100.

Granados, A. A., Pietsch, J. M., Cepeda-Humerez, S. A., Farquhar, I. L., Tkačik, G., and Swain, P. S. (2018). Distributed and dynamic intracellular organization of extracellular information. *Proceedings of the National Academy of Sciences*, 115(23):6088–6093.

Grote, M., Kubitscheck, U., Reichelt, R., and Peters, R. (1995). Mapping of nucleoporins to the center of the nuclear pore complex by post-embedding immunogold electron microscopy. *Journal of Cell Science*, 108(9):2963–2972.

Grünwald, D., Singer, R. H., and Rout, M. (2011). Nuclear export dynamics of RNA-protein complexes. *Nature*, 475(7356):333–341.

Hao, N. and O'shea, E. K. (2012). Signal-dependent dynamics of transcription factor translocation controls gene expression. *Nature structural & molecular biology*, 19(1):31–39.

Harreman, M. T., Cohen, P. E., Hodel, M. R., Truscott, G. J., Corbett, A. H., and Hodel, A. E. (2003). Characterization of the auto-inhibitory sequence within the n-terminal domain of importin α. *Journal of Biological Chemistry*, 278(24):21361–21369.

Heinrich, R., Neel, B. G., and Rapoport, T. A. (2002). Mathematical models of protein kinase signal transduction. *Molecular cell*, 9(5):957–970.

Hill, S. M. (1998). Receptor crosstalk: communication through cell signaling pathways. *The Anatomical Record: An Official Publication of the American Association of Anatomists*, 253(2):42–48.

Hill, T. L. (2012). *Free energy transduction and biochemical cycle kinetics*. Springer Science & Business Media.

Hill, T. L. (2013). *Cooperativity theory in biochemistry: steady-state and equilibrium systems*. Springer Science & Business Media.

Hunter, T. (2000). Signaling—2000 and beyond. *Cell*, 100(1):113–127.

Hunter, T. (2007). The age of crosstalk: phosphorylation, ubiquitination, and beyond. *Molecular cell*, 28(5):730–738.

Jin, J. and Pawson, T. (2012). Modular evolution of phosphorylation-based signalling systems. *Phil. Trans. R. Soc. B*, 367(1602):2540–2555.

Johnson, H. A. (1970). Information theory in biology after 18 years. *Science*, 168(3939):1545–1550.

Jordan, J. D., Landau, E. M., and Iyengar, R. (2000). Signaling networks: the origins of cellular multitasking. *Cell*, 103(2):193–200.

Karsenti, E. (2008). Self-organization in cell biology: a brief history. *Nature Reviews Molecular Cell Biology*, 9(3):255–262.

Keminer, O. and Peters, R. (1999). Permeability of Single Nuclear Pores. *Biophysical Journal*, 77(1):217–228.

Keymer, J. E., Endres, R. G., Skoge, M., Meir, Y., and Wingreen, N. S. (2006). Chemosensing in escherichia coli: two regimes of two-state receptors. *Proceedings of the National Academy of Sciences*, 103(6):1786–1791.

Khalil, A. S. and Collins, J. J. (2010). Synthetic biology: applications come of age. *Nature Reviews Genetics*, 11(5):367–379.

Kiel, C., Yus, E., and Serrano, L. (2010). Engineering signal transduction pathways. *Cell*, 140(1):33–47.

Kim, S. and Elbaum, M. (2013a). Enzymatically driven transport: A kinetic theory for nuclear export. *Biophysical journal*, 105(9):1997–2005.

Kim, S. and Elbaum, M. (2013b). A simple kinetic model with explicit predictions for nuclear transport. *Biophysical journal*, 105(3):565–569.

Kimura, M. and Imamoto, N. (2014). Biological significance of the importin-β family-dependent nucleocytoplasmic transport pathways. *Traffic*, 15(7):727–748.

Kinney, J. B. and Atwal, G. S. (2014). Equitability, mutual information, and the maximal information coefficient. *Proceedings of the National Academy of Sciences*, page 201309933.

Kinney, J. B., Murugan, A., Callan, C. G., and Cox, E. C. (2010). Using deep sequencing to characterize the biophysical mechanism of a transcriptional regulatory sequence. *Proceedings of the National Academy of Sciences*.

Koch, C. A., Anderson, D., Moran, M. F., Ellis, C., and Pawson, T. (1991). Sh2 and sh3 domains: elements that control interactions of cytoplasmic signaling proteins. *Science*, 252(5006):668–674.

Kolch, W., Halasz, M., Granovskaya, M., and Kholodenko, B. N. (2015). The dynamic control of signal transduction networks in cancer cells. *Nature reviews. Cancer*, 15(9):515.

Kontogeorgaki, S., Sánchez-García, R. J., Ewing, R. M., Zygalakis, K. C., and Mac Arthur, B. D. (2017). Noise-processing by signaling networks. *Scientific Reports*, 7(1):532.

Kopito, R. B. and Elbaum, M. (2007). Reversibility in nucleocytoplasmic transport. *Proceedings of the National Academy of Sciences*, 104(31):12743–12748.

Kopito, R. B. and Elbaum, M. (2009). Nucleocytoplasmic transport: a thermodynamic mechanism. *HFSP journal*, 3(2):130–141.

Koshland, D. E. (2002). The seven pillars of life. *Science*, 295(5563):2215–2216.

Kubitscheck, U., Grünwald, D., Hoekstra, A., Rohleder, D., Kues, T., Siebrasse, J. P., and Peters, R. (2005). Nuclear transport of single molecules dwell times at the nuclear pore complex. *The Journal of cell biology*, 168(2):233–243.

Kuchroo, V. K., Das, M. P., Brown, J. A., Ranger, A. M., Zamvil, S. S., Sobel, R. A., Weiner, H. L., Nabavi, N., and Glimcher, L. H. (1995). B7-1 and b7-2 costimulatory molecules activate differentially the th1/th2 developmental pathways: application to autoimmune disease therapy. *Cell*, 80(5):707–718.

Ladbury, J. E. and Arold, S. (2000). Searching for specificity in sh domains. *Chemistry & biology*, 7(1):R3–R8.

Ladbury, J. E. and Arold, S. T. (2011). Energetics of src homology domain interactions in receptor tyrosine kinase-mediated signaling. In *Methods in enzymology*, volume 488, pages 147–183. Elsevier.

Ladbury, J. E. and Arold, S. T. (2012). Noise in cellular signaling pathways: causes and effects. *Trends in biochemical sciences*, 37(5):173–178.

Lang, A. H., Fisher, C. K., Mora, T., and Mehta, P. (2014). Thermodynamics of statistical inference by cells. *Physical review letters*, 113(14):148103.

Lebowitz, J. L. and Spohn, H. (1999). A gallavotti–cohen-type symmetry in the large deviation functional for stochastic dynamics. *Journal of Statistical Physics*, 95(1-2):333–365.

Lee, T. I., Rinaldi, N. J., Robert, F., Odom, D. T., Bar-Joseph, Z., Gerber, G. K., Hannett, N. M., Harbison, C. T., Thompson, C. M., Simon, I., et al. (2002). Transcriptional regulatory networks in saccharomyces cerevisiae. *science*, 298(5594):799–804.

Letsou, W. and Cai, L. (2016). Noncommutative biology: Sequential regulation of complex networks. *PLoS Comput Biol*, 12(8):e1005089.

Levy, E. D., Landry, C. R., and Michnick, S. W. (2010). Signaling through cooperation. *Science*, 328(5981):983–984.

Lim, W., Mayer, B., and Pawson, T. (2014). *Cell signaling: principles and mechanisms*. Taylor & Francis.

Lim, W. A. (2002). The modular logic of signaling proteins: building allosteric switches from simple binding domains. *Current opinion in structural biology*, 12(1):61–68.

Lim, W. A. (2010). Designing customized cell signalling circuits. *Nature reviews Molecular cell biology*, 11(6):393–403.

Lim, W. A. and Pawson, T. (2010). Phosphotyrosine signaling: evolving a new cellular communication system. *Cell*, 142(5):661–667.

Lolodi, O., Yamazaki, H., Otsuka, S., Kumeta, M., and Yoshimura, S. H. (2016). Dissecting in vivo steady-state dynamics of karyopherin-dependent nuclear transport. *Molecular biology of the cell*, 27(1):167–176.

MacArthur, B. D., Ma'ayan, A., and Lemischka, I. R. (2009). Systems biology of stem cell fate and cellular reprogramming. *Nature Reviews Molecular Cell Biology*, 10(10):672.

MacDonald, B. T., Tamai, K., and He, X. (2009). Wnt/β-catenin signaling: components, mechanisms, and diseases. *Developmental cell*, 17(1):9–26.

Mangan, S., Zaslaver, A., and Alon, U. (2003). The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks. *Journal of molecular biology*, 334(2):197–204.

Martello, G. and Smith, A. (2014). The nature of embryonic stem cells. *Annual review of cell and developmental biology*, 30:647–675.

Massagué, J. (2012). Tgfβ signalling in context. *Nature reviews Molecular cell biology*, 13(10):616.

Maul, G. G. and Deaven, L. (1977). Quantitative determination of nuclear pore complexes in cycling cells with differing dna content. *The Journal of cell biology*, 73(3):748–760.

Mehta, P., Goyal, S., Long, T., Bassler, B. L., and Wingreen, N. S. (2009). Information processing and signal integration in bacterial quorum sensing. *Molecular systems biology*, 5(1):325.

Mehta, P., Lang, A. H., and Schwab, D. J. (2016). Landauer in the age of synthetic biology: energy consumption and information processing in biochemical networks. *Journal of Statistical Physics*, 162(5):1153–1166.

Mehta, P. and Schwab, D. J. (2012). Energetic costs of cellular computation. *Proceedings of the National Academy of Sciences*.

Mezard, M. and Montanari, A. (2009). *Information, physics, and computation*. Oxford University Press.

Milo, R. and Phillips, R. (2015). *Cell biology by the numbers*. Garland Science.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002). Network motifs: simple building blocks of complex networks. *Science*, 298(5594):824–827.

Mistry, K. H., McGovern, R. K., Thiel, G. P., Summers, E. K., Zubair, S. M., and Lienhard, J. H. (2011). Entropy generation analysis of desalination technologies. *Entropy*, 13(10):1829–1864.

Mok, J., Kim, P. M., Lam, H. Y., Piccirillo, S., Zhou, X., Jeschke, G. R., Sheridan, D. L., Parker, S. A., Desai, V., Jwa, M., et al. (2010). Deciphering protein kinase specificity through large-scale analysis of yeast phosphorylation site motifs. *Sci. Signal.*, 3(109):ra12–ra12.

Monod, J., Wyman, J., and Changeux, J.-P. (1965). On the nature of allosteric transitions: a plausible model. *Journal of molecular biology*, 12(1):88–118.

Mora, T. and Wingreen, N. S. (2010). Limits of sensing temporal concentration changes by single cells. *Physical review letters*, 104(24):248101.

Morsut, L., Roybal, K. T., Xiong, X., Gordley, R. M., Coyle, S. M., Thomson, M., and Lim, W. A. (2016). Engineering customized cell sensing and response behaviors using synthetic notch receptors. *Cell*, 164(4):780–791.

Moustakas, A. and Heldin, P. (2014). Tgfβ and matrix-regulated epithelial to mesenchymal transition. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1840(8):2621–2634.

Nédélec, F., Surrey, T., Maggs, A. C., and Leibler, S. (1997). Self-organization of microtubules and motors. *Nature*, 389(6648):305–308.

Niopek, D., Benzinger, D., Roensch, J., Draebing, T., Wehler, P., Eils, R., and Di Ventura, B. (2014). Engineering light-inducible nuclear localization signals for precise spatiotemporal control of protein dynamics in living cells. *Nature Communications*, 5:4404.

Parmeggiani, A., Jülicher, F., Ajdari, A., and Prost, J. (1999). Energy transduction of isothermal ratchets: Generic aspects and specific examples close to and far from equilibrium. *Physical Review E*, 60(2):2127.

Parrondo, J., de Cisneros, B., and Brito, R. (2000). Thermodynamics of isothermal brownian motors. *Stochastic Processes in Physics, Chemistry, and Biology*, pages 38–49.

Pasparakis, M. (2009). Regulation of tissue homeostasis by nf-κb signalling: implications for inflammatory diseases. *Nature Reviews Immunology*, 9(11):778–788.

Pawson, T. and Nash, P. (2003). Assembly of cell regulatory systems through protein interaction domains. *science*, 300(5618):445–452.

Peters, R. (1984). Nucleo-cytoplasmic flux and intracellular mobility in single hepatocytes measured by fluorescence microphotolysis. *The EMBO Journal*, 3(8):1831.

Peters, R. (2003). Optical single transporter recording: transport kinetics in microarrays of membrane patches. *Annual review of biophysics and biomolecular structure*, 32(1):47–67.

Phillips, R. (2015). Napoleon is in equilibrium. *Annu. Rev. Condens. Matter Phys.*, 6(1):85–111.

Phillips, R., Theriot, J., Kondev, J., and Garcia, H. (2012). *Physical biology of the cell*. Garland Science.

Purvis, J. E. and Lahav, G. (2013). Encoding and decoding cellular information through signaling dynamics. *Cell*, 152(5):945–956.

Razo-Mejia, M., Barnes, S. L., Belliveau, N. M., Chure, G., Einav, T., Lewis, M., and Phillips, R. (2017). Tuning transcriptional regulation through signaling: A predictive theory of allosteric induction. *arXiv preprint arXiv:1702.07460*.

Rexach, M. and Blobel, G. (1995). Protein import into nuclei: association and dissociation reactions involving transport substrate, transport factors, and nucleoporins. *Cell*, 83(5):683–692.

Ribbeck, K. and Görlich, D. (2001). Kinetic analysis of translocation through nuclear pore complexes. *The EMBO Journal*, 20(6):1320–1330.

Riddick, G. and Macara, I. G. (2005). A systems analysis of importin-{alpha}-{beta} mediated nuclear protein import. *J. Cell Biol.*, 168(7):1027–1038.

Roybal, K. T., Williams, J. Z., Morsut, L., Rupp, L. J., Kolinko, I., Choe, J. H., Walker, W. J., McNally, K. A., and Lim, W. A. (2016). Engineering t cells with customized therapeutic response programs using synthetic notch receptors. *Cell*, 167(2):419–432.

Samudram, A., Mangalassery, B. M., Kowshik, M., Patincharath, N., and Varier, G. K. (2016). Passive permeability and effective pore size of HeLa cell nuclear membranes. *Cell Biol. Int.*

Sartori, P., Granger, L., Lee, C. F., and Horowitz, J. M. (2014). Thermodynamic costs of information processing in sensory adaptation. *PLoS Computational Biology*, 10(12):e1003974.

Schlessinger, J. (2000). Cell signaling by receptor tyrosine kinases. *Cell*, 103(2):211–225.

Schwartz, M. A. and Ginsberg, M. H. (2002). Networks and crosstalk: integrin signalling spreads. *Nature cell biology*, 4(4):E65.

Scott, J. D. and Pawson, T. (2009). Cell signaling in space and time: where proteins come together and when they're apart. *Science*, 326(5957):1220–1224.

Seet, B. T., Dikic, I., Zhou, M.-M., and Pawson, T. (2006). Reading protein modifications with interaction domains. *Nature reviews Molecular cell biology*, 7(7):473.

Selimkhanov, J., Taylor, B., Yao, J., Pilko, A., Albeck, J., Hoffmann, A., Tsimring, L., and Wollman, R. (2014). Accurate information transmission through dynamic biochemical signaling networks. *Science*, 346(6215):1370–1373.

Shannon, C. E. (2001). A mathematical theory of communication. *ACM SIGMOBILE mobile computing and communications review*, 5(1):3–55.

Shen-Orr, S. S., Milo, R., Mangan, S., and Alon, U. (2002). Network motifs in the transcriptional regulation network of escherichia coli. *Nature genetics*, 31(1):64.

Shoval, O. and Alon, U. (2010). Snapshot: network motifs. *Cell*, 143(2):326–326.

Smith, A. E., Slepchenko, B. M., Schaff, J. C., Loew, L. M., and Macara, I. G. (2002). Systems analysis of ran transport. *Science*, 295(5554):488–491.

Stewart, M. (2007). Molecular mechanism of the nuclear protein import cycle. *Nature Reviews Molecular Cell Biology*, 8(3):195–208.

Surrey, T., Nédélec, F., Leibler, S., and Karsenti, E. (2001). Physical properties determining self-organization of motors and microtubules. *Science*, 292(5519):1167–1171.

Taganov, K. D., Boldin, M. P., Chang, K.-J., and Baltimore, D. (2006). Nf-κb-dependent induction of microrna mir-146, an inhibitor targeted to signaling proteins of innate immune responses. *Proceedings of the National Academy of Sciences*, 103(33):12481–12486.

Talcott, B. and Moore, M. S. (1999). Getting across the nuclear pore complex. *Trends in cell biology*, 9(8):312–318.

Tareen, A., Wingreen, N. S., and Mukhopadhyay, R. (2018). Modeling evolution of crosstalk in noisy signal transduction networks. *Physical Review E*, 97(2):020402.

Tay, S., Hughey, J. J., Lee, T. K., Lipniacki, T., Quake, S. R., and Covert, M. W. (2010). Single-cell nf-κb dynamics reveal digital activation and analog information processing in cells. *Nature*, 466(7303):267.

Taylor, S. S. and Radzio-Andzelm, E. (1997). Protein kinase inhibition: natural and synthetic variations on a theme. *Current opinion in chemical biology*, 1(2):219–226.

ten Wolde, P. R., Becker, N. B., Ouldridge, T. E., and Mugler, A. (2016). Fundamental limits to cellular sensing. *Journal of Statistical Physics*, 162(5):1395–1424.

Teng, S.-W., Schaffer, J. N., Tu, K. C., Mehta, P., Lu, W., Ong, N., Bassler, B. L., and Wingreen, N. S. (2011). Active regulation of receptor ratios controls integration of quorum-sensing signals in vibrio harveyi. *Molecular systems biology*, 7(1):491.

Terry, L. J., Shows, E. B., and Wente, S. R. (2007). Crossing the nuclear envelope: hierarchical regulation of nucleocytoplasmic transport. *Science*, 318(5855):1412–1416.

Tkačik, G. and Walczak, A. M. (2011). Information transmission in genetic regulatory networks: a review. *Journal of Physics: Condensed Matter*, 23(15):153102.

Tkačik, G., Walczak, A. M., and Bialek, W. (2009). Optimizing information flow in small genetic networks. *Physical Review E*, 80(3):031920.

Tostevin, F. and Ten Wolde, P. R. (2009). Mutual information between input and output trajectories of biochemical networks. *Physical review letters*, 102(21):218101.

Voliotis, M., Perrett, R. M., McWilliams, C., McArdle, C. A., and Bowsher, C. G. (2014). Information transfer by leaky, heterogeneous, protein kinase signaling systems. *Proceedings of the National Academy of Sciences*, 111(3):E326–E333.

Vovk, A., Gu, C., Opferman, M. G., Kapinos, L. E., Lim, R. Y., Coalson, R. D., Jasnow, D., and Zilman, A. (2016). Simple biophysics underpins collective conformations of the intrinsically disordered proteins of the nuclear pore complex. *eLife*, 5:e10785.

Wagner, R. S., Kapinos, L. E., Marshall, N. J., Stewart, M., and Lim, R. Y. (2015). Promiscuous binding of karyopherinβ1 modulates fg nucleoporin barrier function and expedites ntf2 transport kinetics. *Biophysical journal*, 108(4):918–927.

Walczak, A. M., Tkačik, G., and Bialek, W. (2010). Optimizing information flow in small genetic networks. ii. feed-forward interactions. *Physical Review E*, 81(4):041905.

Weinert, F. M., Brewster, R. C., Rydenfelt, M., Phillips, R., and Kegel, W. K. (2014). Scaling of gene expression with transcription-factor fugacity. *Physical review letters*, 113(25):258101.

Wente, S. R. and Rout, M. P. (2010). The nuclear pore complex and nuclear transport. *Cold Spring Harbor perspectives in biology*, 2(10):a000562.

Yang, W., Gelles, J., and Musser, S. M. (2004). Imaging of single-molecule translocation through nuclear pore complexes. *Proceedings of the National Academy of Sciences of the United States of America*, 101(35):12887–12892.

Yedidia, J. (2001). An idiosyncratic journey beyond mean field theory. *Advanced mean field methods: Theory and practice*, pages 21–36.

Yedidia, J. S., Freeman, W. T., and Weiss, Y. (2003). Understanding belief propagation and its generalizations. *Exploring artificial intelligence in the new millennium*, 8:236–239.

Zahn, R., Osmanović, D., Ehret, S., Callis, C. A., Frey, S., Stewart, M., You, C., Görlich, D., Hoogenboom, B. W., and Richter, R. P. (2016). A physical model describing the interaction of nuclear transport receptors with FG nucleoporin domain assemblies. *eLife*, 5:e14119.

CURRICULUM VITAE

[redacted]

[redacted]

[redacted]