ANNUAL REVIEWS

Annual Review of Biophysics

Information Processing in Biochemical Networks

Gašper Tkačik[1] and Pieter Rein ten Wolde[2]

[1]Institute of Science and Technology Austria, Klosterneuburg, Austria; email: gtkacik@ist.ac.at

[2]AMOLF, Amsterdam, The Netherlands; email: tenwolde@amolf.nl

Annu. Rev. Biophys. 2025. 54:249–74

First published as a Review in Advance on February 10, 2025

The *Annual Review of Biophysics* is online at biophys.annualreviews.org

https://doi.org/10.1146/annurev-biophys-060524-102720

Keywords

transcriptional regulation, signaling networks, information theory, optimization

Abstract

Living systems are characterized by controlled flows of matter, energy, and information. While the biophysics community has productively engaged with the first two, addressing information flows has been more challenging, with some scattered success in evolutionary theory and a more coherent track record in neuroscience. Nevertheless, interdisciplinary work of the past two decades at the interface of biophysics, quantitative biology, and engineering has led to an emerging mathematical language for describing information flows at the molecular scale. This is where the central processes of life unfold: from detection and transduction of environmental signals to the readout or copying of genetic information and the triggering of adaptive cellular responses. Such processes are coordinated by complex biochemical reaction networks that operate at room temperature, are out of equilibrium, and use low copy numbers of diverse molecular species with limited interaction specificity. Here we review how flows of information through biochemical networks can be formalized using information-theoretic quantities, quantified from data, and computed within various modeling frameworks. Optimization of information flows is presented as a candidate design principle that navigates the relevant time, energy, crosstalk, and metabolic constraints to predict reliable cellular signaling and gene regulation architectures built of individually noisy components.

Contents

1. INTRODUCTION

It is impossible to talk about life without reference to genetic information and how it is passed from parent to offspring, or without reference to the flow of information in the central dogma of biology, where genetic information is transcribed from DNA into mRNA and later translated into proteins. Similarly central is the notion that organisms collect information about the environment via sensing and respond to it by initiating adaptive changes in their behaviors or internal states. Positional information is the paradigm of developmental biology according to which multicellular organisms establish their precise and reproducible body plans, and if these organisms end up possessing nervous systems, the ongoing information processing in their brains is the key item of the neuroscience agenda. Information is thus central to our biological narrative: It permeates the descriptions of the observed phenomena of life, as well as the teleological rationalizations about why we observe some evolutionary outcomes but not others.

Given how essential the information-centric narrative is to modern biology, we should strive to make it conceptually clear and mathematically precise (Section 2). We should push beyond singular biological examples and toward general principles and (perhaps) a unifying theory of biochemical information processing that can make quantitative and testable predictions. We should realize that the act of information processing in biochemical hardware itself is a nontrivial fundamental physics problem that underpins the information-centric biological narrative.

Physics is well-placed to contribute to this interdisciplinary enterprise, especially if it engages in an ambitious program to go beyond modeling existing data in the service of answering biological questions, and itself drives the next generation of physics-style experiments that rigorously estimate the relevant, theory-informed—and previously often neglected—quantities (Section 3). As soft matter and active matter physics advanced our understanding of the biology of shape, form, and matter flows, so we hope for a productive interaction between physics and the phenomenology of biological information flows.

Why should we be hopeful about the success of this enterprise? Biochemical information processing likely operates under various resource constraints, such as limited energy consumption, limits to molecular diversity and copy numbers, or limited time and space. Resource constraints impose severe limitations on how precisely signals can be represented due to the thermal and number fluctuations, the limited specificity of molecular interactions, and limits due to diffusive transport, yet evolution does not fail to inspire with its inventiveness.[1]

If we react with wonder to evolutionary innovations, we might overlook a subtle yet essential point: Constraints that affect and shape biological information processing are physical and chemical at their core, despite the complexity of the biological processes involved. These constraints are amenable to our theoretical treatment in principle and are quantifiable in physical units in practice; importantly, they are universal (Section 4). This universality of constraints is the first reason why physics should engage and strive toward a unifying framework for understanding biological information processing across different systems and modalities. The second reason is more speculative: Assuming evolution has pushed biological information-processing systems toward optimal function, one can hope that not only the constraints but also the evolved solutions are, in certain aspects, universal and predictable from first principles (Sections 5 and 6).

What is the scope of biochemical information processing? This broad term encompasses sensing, or the detection of input signals of varying modality; possible transformations of how the information present in the inputs is represented by concentrations of signaling molecules (encoding); transmission or recovery of input signals from their internal chemical representations (decoding); cellular memory and copying operations; and the integration of relevant information to drive reliable cellular or organismal responses (decision-making). This functional, signal-processing language can be used, for example, to annotate one of the best-characterized biochemical information-processing circuits, the chemotaxis network of *Escherichia coli*, which enables the bacterium to navigate its environment. In this circuit, one can recognize the sensing module that detects and amplifies chemoattracting signals via fast ligand–receptor binding on the bacterial surface, the adaptation module that recalibrates the receptors via slower methylation reactions so that the receptor signaling activity corresponds to the temporal derivative of the ligand

[1]The vertebrate optic nerve can sustain information rates of ∼1 Mbit/s; it took serious engineering effort to reach similar rates in our modern display devices (at much higher energy expenditure). The fact that the capacity of the optic nerve and the bit rate of high-quality compressed video encoding are similar should not be too surprising after an in-depth reflection. *Escherichia coli*, the model organism of microbiology, replicates its DNA in 40 min using nanoscale machinery, at an information copy rate of ≈2 Kbit/s; this rate is comparable to that of the considerably bulkier modems of the 1980s. As we will see, cells often encode environmental as well as internal states with a typical precision of ∼2 bits using very low concentrations of specialized molecules, ranging up to $\sim 10^1$–10^4 molecules per nucleus in absolute counts; with typical nanometer-scale molecules, 1 Gbit of information could be packed into a volume roughly 10 μm on the side, at an impressive information density (for reference, the current area density of information storage for memory chips would be on the order of 1 Gbit/mm^2). The relevant volume is roughly the size of an eukaryotic nucleus. The DNA molecule itself packs a similar order of magnitude of information more compactly, roughly at 1 nm^3/bit, and stably, but it necessitates dedicated machinery (which may require extra volume) for the genomic information to be read out and influence biological computation.

concentration, and the transmission of this readout to the flagellar motor to bias its rotational mode and implement the run-and-tumble navigation toward the chemoattractant source (138). Much like Marr & Poggio's (68) three levels of analysis in neuroscience, it is common to consider the same system at the level of implementation (a set of chemical reactions), of algorithm (the regulatory network), and of computation (information processing). While it may be difficult to cleanly separate these levels or draw sharp boundaries between various signal-processing stages in evolved systems in practice, the exercise nevertheless forces us to abstract away from the molecular detail and attempt to identify common biochemical network motifs, their information-processing functions, and limitations. We repeatedly refer to these different levels in the sections that follow.

In this review, we cover the application of information-theoretic ideas and quantities (19, 107) to biochemical reaction networks; this covers approximately 20–25 years of work that has been principally directed at intracellular signaling and gene regulatory networks and started roughly with the experimental quantification of biochemical noise in vivo (30). Several previous reviews and perspectives cover similar subject matter from a physics perspective (13, 63, 70, 97, 118, 130, 134, 139); in engineering, this area is often termed molecular communication (32). Biologically inspired information-theoretic considerations have since branched out into cancer- (56) and immunology-related (109) research, the world of lipids (35), and infochemistry (103). It is also possible to take a broader, panoramic view of the role that information flows play in biological systems (9, 124) and evolution (48, 57, 100, 145), which we do not attempt here. Another line of work we do not review but acknowledge for its intellectual heritage is the application of information-theoretic ideas in neuroscience. Only a few years after Shannon's (107) seminal paper, its new formal language was used to consider how individual neurons and neural populations could process information (64). In 1961, Barlow (4) suggested that neural sensory systems may have evolved or adapted to represent information efficiently, which subsequently grew into an optimality hypothesis known as efficient coding or, more generally, the idea that neural representations maximize information under physical constraints (17). These early ideas at the interface of neuroscience, physics, and information theory led to an entire subfield that has productively engaged with experiments (99) and has strongly influenced our thinking about biochemical computations.

2. INFORMATION-THEORETIC QUANTITIES FOR BIOCHEMICAL NETWORKS

A chemical reaction network is a set of reactions between $i = 1, \ldots, M$ chemical species, $\{S_i\}$; theoretical setups often assume that the stoichiometry and reaction rates are known. The state of the network is given by the copy counts (or concentrations) of all the chemical species at each moment in time. There are multiple standard ways to describe the dynamics in such networks. To capture intrinsic stochasticity of reaction events at low molecular numbers in well-mixed systems, we may write down the chemical master equation, a continuous-time discrete-space master equation for the evolution of copy numbers (141); this equation permits explicit solutions only in rare cases, but efficient approximations (83) and Monte Carlo simulation algorithms (37) are available. Alternatively, one can switch to continuous description in terms of concentration dynamics but still keep the ability to capture the intrinsic stochasticity in the Langevin approximation, or one can change to mass-action kinetics where one can conveniently treat extrinsic but not intrinsic noise (115). Treating diffusion explicitly when systems are not well-mixed is much more challenging (111) but has been instrumental for biochemical information processing (cf. Berg–Purcell limit in Section 4). While the choice of mathematical description is context dependent, an essential characteristic of biological regulatory processes is their stochastic nature due to physical constraints alluded to in Section 1.

When thinking of biological regulatory processes, we must equip a reaction network with several additional annotations. We denote some species as inputs to the network (for concreteness, let us select S_1 to be our input and denote by $x(t)$ its concentration[2]): The inputs could correspond to external signals such as a nutrient, hormone, or some other signaling molecule. The externally imposed input dynamics, $x(t)$, drives the network to respond, and we often focus on a small number of biologically motivated network outputs whose concentration(s) we track explicitly (let us select S_M as the output species and denote by $y(t)$ its concentration). Stochasticity means that even at fixed input trajectory $x(t)$, the output trajectory is a random variable drawn from $P(y(t)|x(t))$. In theoretical approaches, this distribution can be either computed, approximated, or sampled from using the mathematical models of the reaction network described above; in experimental approaches, this distribution or its statistics are estimated by measuring samples of input/output response pairs.

For the bacterial chemotaxis example, the input is the time-dependent chemoattractant concentration in the bacterium's environment, and we can think of such concentrations as being drawn from some distribution of inputs, $P_x(x(t))$, that characterizes the statistical structure of the cell's natural habitat (or the experimental setup). The output $y(t)$ is the intracellular concentration of the phosphorylated form of a signaling molecule CheY, which determines bacterial swimming by controlling the rotation of its flagellar motor. These choices now naturally map into the language of information theory, specifically into a problem of information transmission,

$$\text{chemoattractant } x(t) \xrightarrow{\text{network}} y(t) \text{ CheY}_\text{P}$$

$$\text{input } X \sim P_X(X) \xrightarrow[P(Y|X)]{\text{channel}} \text{output } Y.$$

In the bottom equation, channel, the key concept of Shannon's information theory, stochastically maps some input signals X drawn from the distribution $P_X(X)$ into outputs Y via a conditional probability distribution $P(Y|X)$ that captures the noise and possible distortions or transformations in signal transmission. Channels might be abstractions for engineered communication systems such as telegraph, telephone, and optic cables or radio links, but nothing in the theory detracts us from a broader application (e.g., to chemical reaction networks).[3]

The fidelity of input/output mapping can be quantified by mutual information, a nonnegative quantity measured in bits and defined as[4]

$$I(X;Y) = \int P(Y|X)P_X(X)\log_2 \frac{P(Y|X)}{P_Y(Y)}\,\mathrm{d}X\mathrm{d}Y, \qquad 1.$$

where $P_Y(Y)$ is the marginal of the joint distribution, $P_Y(Y) = \int P(Y|X)P_X(X)\mathrm{d}X$. Mutual information is a universal measure of statistical dependence between two random variables X and Y: It is zero if and only if X and Y are independent; i.e., $P(X, Y) = P_X(X)P_Y(Y)$. Shannon (107) showed that any statistical dependence—when carefully utilized by devising encoding and decoding schemes

[2]The network could also have multiple inputs or outputs; the logic would be unchanged but the computations more demanding.

[3]For an interesting historical insight, one can consult Shannon's (107) original text, where he alludes to a possible wide applicability of information theory; yet in his well-known Bandwagon essay (108) seven years later, he calls for a careful and restrained application to topics outside of communications engineering.

[4]Mutual information has a range of attractive mathematical properties: It is symmetric in X and Y; it can be easily defined for discrete distributions by replacing integrals with sums; it is reparameterization invariant (i.e., its value does not change if X and Y are mapped via invertible functions). For details, we recommend Shannon's original text (107) or a standard textbook (19).

that pack information into channel inputs X and unpack it from channel outputs Y—can be used, under appropriate conditions, to transmit information through a noisy channel without error. Note that information depends not only on the channel, $P(Y|X)$, but also on the distribution of inputs we wish to send through it, $P_X(X)$. Given a channel, one can find the best distribution of inputs, $P_X^*(X)$, which maximizes the mutual information; the resulting maximal transmission is then termed the channel capacity $C = \max_{P_X(X)} I(X;Y)$.[5] For discrete X and Y, information maximization to compute capacity or the rate-distortion curves (Section 6) can be done by means of Blahut–Arimoto and related algorithms (10, 50).[6]

Let us now return to the chemotaxis example, where X and Y are concentration trajectories. If we think of trajectories of duration T, the mutual information will typically be extensive in T, so that it makes sense to define an information rate in bits per second as the limit $R = \lim_{T\to\infty} I(x(t); y(t))/T$. Computing or estimating this quantity is in general intractable—it involves integrating over high-dimensional distributions in the space of trajectories—unless we utilize strong prior knowledge. Mattingly and colleagues (69), in a beautiful experiment-theory collaboration, put *E. coli* into weak chemoattractant gradients where the chemotaxis network was in a linear regime, and characterized its response to Gaussian fluctuations. Under these assumptions, Gaussian channel results carry over to entire trajectories and involve measuring only signal and noise power spectra (or alternatively various correlation functions) and integrating them across the frequency range (136, 137). The authors report seemingly low chemotaxis information rates of $R \approx 0.01$ bit/s.[7] We briefly return to the question of how such rates can be formally related to context-specific performance measures (e.g., how quickly they would allow the bacterium to climb the chemoattractant gradient) in Section 6. In parallel with these empirical estimates, Reinhardt and colleagues (96) devised path weight sampling (PWS), an exact Monte Carlo scheme to estimate information rates in biochemical networks when all the reactions and reaction rates are known, by exploiting the properties of the chemical master equation. While we typically do not have sufficient knowledge about real biochemical networks to permit such model-based estimates, the chemotaxis network is an exception:[8] The authors' model-based information rate computations matched the experimentally derived estimates by Mattingly et al. (69).

Information-theoretic analysis of the chemotactic network is one of only a few examples—and to date the only one that captures the full dynamical response—where real network performance has been quantified directly from data and simultaneously computed from a mechanistic model with an excellent match, thereby spanning all three of Marr's levels. For this system, the information rate represents a natural, assumption-free, scalar performance metric. This opens up a plethora of questions: How is the performance affected by changes in the network topology or parameters? How much energy is consumed to maintain the estimated information rates? Has the network evolved to be the best at encoding natural chemotactic stimuli, in line with the efficient

[5]Today's mobile technologies owe much of their incredible success to ever better codes that can quickly encode and decode arbitrary data so as to approach the channel capacity of modern wireless links. Designing such codes is the primary focus of information engineering, although this aspect has not been considered extensively in the case of biochemical networks.

[6]Many other information-theoretic quantities exist and are potentially relevant for biological systems but beyond the scope of this review. We refer the reader to extensive work on information divergences, synergy, redundancy, multi-information, transfer entropy, and various information decompositions for multivariate systems.

[7]If a bacterium's tumble during chemotaxis would be viewed as a binary decision about whether it has been running up or down the gradient, and the typical tumbling frequency is ∼1/s, the expected information rate would be of the order of 1 bit/s.

[8]The entire model consists of 182 coupled chemical reactions specified by ∼20 parameters, essentially all of which are known from the literature.

coding hypothesis in neuroscience? Is the network function robust to various extrinsic fluctuations as has been suggested (1)? These and similar exciting questions are now no longer restricted to verbal speculations in paper discussion sections or to toy models, but can be carried out quantitatively for other real systems and data, as we review in Section 3.

3. QUANTIFYING INFORMATION FLOWS IN REAL NETWORKS

To perform the estimates on which we report below, substantial investment of both experimental and theoretical effort was necessary; specifically, the reader is directed to Section 7 for an overview of information estimation methods.

3.1. Bacteria

Beyond the chemotaxis example reviewed above (69, 74, 96, 136), bacterial biochemical networks serve as one of the paradigms for the application of information-theoretic ideas. Early work looked at quorum sensing, a simple form of extracellular multicomponent chemical communication system used by bacteria to detect the density of other bacteria in the local neighborhood. Model-based estimates suggested capacities of 1.2 to 1.7 bits between two extracellular signals and the internal gene expression state (72); subsequent measurements reported much lower values of $\approx$0.5 bits (88).

Ruiz et al. (102) used fluorescent reporters to measure the precision of *E. coli* tetracycline resistance response to subinhibitory levels of external antibiotic. The authors report that noise suppression via negative feedback in this genetic regulatory circuit enables graded responses and thus channel capacities just short of 2 bits.

Razo-Mejia et al. (95) explored the channel capacity of a simple genetic regulatory element, in which a single transcription factor (TF) controls target gene expression in steady state; this work brings together carefully data-calibrated biophysical models of stochastic gene regulation (54) with early information-theoretic calculations (125) to arrive at realistic estimates. For typical TF copy numbers of 10^1 to 10^3 molecules per cell, the authors compute capacities ranging from 1 to 1.8 bits, close to but roughly 25% higher than their empirical estimates extracted directly from data.[9]

3.2. Yeast and Eukaryotic Signaling Networks

Intracellular signal transduction typically uses energy-consuming covalent modifications (e.g., phosphorylation/dephosphorylation reactions) of signaling proteins to integrate extracellular signals and transduce them from the receptors at the cell membrane into the nucleus. Chemical modification reactions tend to be considerably faster (seconds-to-minutes timescale) than gene regulatory responses where mRNA must be transcribed, processed, and translated (minutes-to-hours timescale); signaling reactions also happen at higher concentrations, thus allowing for increased noise averaging, which typically permits the application of approximate formalisms (e.g., Langevin or linear noise approximation) for computing the mean and covariance dynamics for the concentrations.[10] These properties have made intracellular signaling networks a fertile playground for studying biological information transmission.

[9]It is reassuring to find empirical estimates close to but lower than theoretical computations: Empirical estimates typically include uncontrolled technical sources of noise that can only degrade information transmission. In a well-controlled experiment, these technical noise sources will not dominate the real noise but may not be negligible either (26).

[10]In contrast, gene regulation involves chemical species (such as DNA regulatory sequence elements), which almost by definition must be present at singular copy numbers and are often in binary states of occupied or unoccupied, thus potentially leading to strongly non-Gaussian, nonlinear responses.

A pioneering study by Cheong et al. (18) looked at information transmission from the externally applied tumor necrosis factor (TNF) stimulus to the intracellular NF-κB response target,[11] as quantified by immunochemistry and single-cell quantification. Static Gaussian approximation was used, focusing on response ∼30 min after stimulation, to assess how noise in signaling cascades with different topologies affects information flow. Transmission from TNF to one target was estimated at ≈0.9 bit, just enough to discriminate the presence/absence of the TNF stimulus. Using a combination of modeling and measurements, the authors analyzed transduction also from a single stimulus to two downstream targets, directly or via a common bottleneck node—the latter was more consistent with data and provided a marginal information increase to ∼1 bit. Negative feedback decreased the noise, consistent with theoretical expectations, but also shrank the dynamic range of the output; ultimately, this trade-off provided no net benefit for transmission. This finding stands in contrast to the subsequent study of ERK signaling in mammalian (142) and yeast (2) cells where negative feedback significantly improved the information flow; in the latter case, it furthermore increased efficiency in terms of bits per signaling energy cost. Cheong et al. (18) further estimated that temporal integration over 10 h of TNF exposure could substantially raise the transmission to 1.6 bit, yet this increase was still limited because output fluctuations were predominantly slow, arising from cell-to-cell extrinsic variation that could not be efficiently temporally averaged away. Subsequent work suggested that signaling pathways may be structured to maintain their transmission rates even in the presence of various external perturbations as a form of information-theoretic robustness (140).

Hypotheses on the role of temporal averaging by Cheong et al. (18) provided a backdrop for Selimkhanov et al. (105), who considered the information encoded in full response trajectories of NF-κB, ERK, and Ca^{2+} signaling systems. This and later studies (93) showed that much higher transmissions can be achieved if the output is not restricted to a singular time point but is dynamic. The boost in transmission from below 1 bit to ∼1.5 bit, consistent with previous extrapolations (18), is possible precisely because entire trajectories allow sufficiently powerful downstream decoding circuits to undo cell-to-cell extrinsic variability—which is indeed not possible by simple temporal averaging. An alternative suggestion sees this variability as biologically helpful at the population level: either engaging in bet-hedging against unforeseen environmental changes or acting as a population-level information channel. In the latter case, the channel output is defined as an average over the cell population, for instance, when an external signal should precisely control a fraction of cells deciding between two possible cellular decisions, yet it is irrelevant how individual cells decide (143, 144). Such population averaging can make population-level responses more graded and suppress noise (18, 114).

A decoding-based approach (16) allowed Granados et al. (43) to estimate information about discrete stimulus (stress) conditions encoded in response trajectories of nuclear localization of multiple yeast TFs, such as the well-studied Msn2 and Mig1 TFs, at much improved trajectory sampling. Consistent with previous reports, single-time-point estimates of information yielded below-bit transmission that approached ∼1.5 bit for trajectories. Interestingly, some TFs were better at responding to a wide array of stresses but only coarsely (generalists), whereas specialist TFs responded to a subset of stresses but encoded well their magnitude. Granados et al. (43) also estimated for the first time the information about stress identity as well as magnitude encoded jointly in the temporal responses of six monitored TFs (reaching ≈2.5 bits), showing that distributed intracellular representations substantially increase the capacity compared to individual pathways.

[11] NF-κB is a transcription factor that translocates in and out of the nucleus depending on the stimulus, and the nuclear concentration typically represents the measured output signal.

In related prior work, Hansen & O'Shea (45) interrogated how well temporally structured nuclear TF translocation in yeast can modulate the activity of downstream promoters by estimating the information between various statistical aspects of the TF translocation signal (amplitude and frequency) and promoter activity. The reported transmission for the Msn2 TF of 1 to 1.5 bits is beautifully matched to the upstream capacities from the external signal to Msn2 reported by Granados et al., suggesting that such matching of information capacities may be a general design principle. Complementary computational studies of NF-κB, Msn2, and other systems analyzed the information flow from TF signaling dynamics to downstream gene expression responses (38, 65).

In an elegant synthetic setup, Benzinger et al. (7) controlled Msn2-type activity optogenetically to characterize engineered downstream gene regulatory motifs, each cleanly designed for a specific top-level Marr computation: signal edge detection, signal multiplexing, pulse rejection, and pulse decoding. Of specific interest to us is the demultiplexer: a motif that takes a single dynamic TF signal and differentially regulates two target genes, each with capacities above 1 bit, for a combined 2D output capacity that approaches 3 bits.

3.3. Developmental Patterning

Early developmental patterning of multicellular organisms that underlies the creation of the body plan has long been one of the paradigms for the reproducibility and precision of biological regulatory processes. Patterning brings a new, spatial dimension to the table: Cells with identical genetic material must either break symmetry spontaneously or use external signals to turn on their gene expression programs appropriate for their positions within a developing embryo. The embryo of the fruit fly, *Drosophila melanogaster*, has been a wonderful platform for the application of information-theoretic ideas, as reviewed recently (128) and summarized here; similar ideas are being explored in other developmental systems (120, 150).

About 2 h postfertilization, ∼6,000 nuclei arranged on the surface of a 0.5-mm-long ellipsoid fruit fly egg responded to smooth concentration gradients of special patterning molecules, termed morphogens, established by the fly mother, by expressing complex spatial gene expression patterns that can be quantified with unparalleled precision (26, 44). These gap gene expression patterns instruct later morphogenetic changes that lead to the creation of different body parts.

In an early application of information-theoretic ideas, Tkačik et al. (126) estimated the information flow between the morphogen TF Bicoid and the expression level of *Hunchback*, one of four gap genes, to find the steady-state transmission of $I = 1.5 \pm 0.15$ bits. Contrary to other estimates where the inputs are experimentally controlled but their natural distributions are unknown, here the natural distribution of Bicoid is established by the fly mother and is fully accessible to precise measurements; 1.5 bits is thus the biologically relevant estimate. Intriguingly, this genetic regulatory channel has a maximal transmission (capacity) of 1.7 bits, implying that its utilization is nearly optimal, which was further supported by an excellent parameter-free match between the measured and the predicted (optimal) distribution of *Hunchback* expression levels. Presaging work that was yet to come, Tkačik and colleagues (125, 131, 146) speculated that single input–single output biochemical transmission links may be metabolically limited to capacities in the low-single-bit regime and that higher capacities would be cheaper to achieve by using multiple such lower-capacity channels in parallel.

Follow-up experimental work allowed for imaging of all four *Drosophila* gap genes simultaneously to estimate their position-dependent means and noise covariance matrix (26). Groundbreaking work in developmental biology for which a Nobel Prize was awarded in 1995 showed previously that there are only four gap genes that span the entire anterior–posterior axis of the fruit fly (147), suggesting that these four gap gene spatial profiles form a full representation

of cellular position x along the long axis of the fruit fly embryo. Because x directly determines maternal morphogen profiles, which subsequently regulate these gap genes $\mathbf{g}$, Dubuis et al. (27) realized that one can directly estimate the mutual information $I(\mathbf{g}; x)$ (27, 127). This quantity is a measure for the quality of the fruit fly embryo chemical GPS system that an individual nucleus uses to infer its position; as such, $I(\mathbf{g}; x)$ is literally the mathematical formalization of a central concept of developmental biology, positional information, introduced by Wolpert (148) in his classic paper in 1969.

This direct connection between an information-theoretic quantity and a central concept of developmental biology opened the door for a number of subsequent results (128). Contrary to the textbook paradigm, it was shown that gap genes are not 1 bit binary switches but individually encode ∼2 bits of information about position; together, four genes provide a redundant code that carries 4.3 ± 0.1 bits of positional information, sufficient for individual nuclei to infer their positions with a relative error of ∼1%.[12] This positional error is small and constant along the entire embryo axis, which is a mathematical signature of optimal patterning (27). Positional information in the four gap genes can be decoded or read out into estimates of position (5, 47, 79, 80, 91) that correctly predict the downstream developmental events in wild-type flies as well as mutants without free parameters (91), demonstrating that the >4 bits of positional information are required for long-axis fly embryonic patterning and that these bits are likely optimally used downstream.

4. FUNDAMENTAL LIMITS

Biochemical reaction networks are physical objects that obey fundamental thermodynamic limits, which become increasingly constraining at the nanoscale and at room temperatures.

4.1. Information Processing Out of Equilibrium

A classic example is the Landauer limit (61), which requires at least $k_B T \ln(2)$ joules of energy dissipation for the erasure of 1 bit in computation. It is instructive to bridge the gap from these universal limits to how they are instantiated, how they operate, and how stringent they are for biochemical reaction networks. While the Landauer limit appears irrelevant for today's engineered systems, it may not be far from what cellular systems can achieve, for instance, during polymer-copying protocols (86, 94) or in enzymatic push–pull networks that could theoretically copy information at 99% accuracy for a cost that is within a factor of 10 from the Landauer limit (85). Energy consumption can further keep biochemical networks out of equilibrium and place them into special operating regimes favorable for reliable processing (40, 42, 98, 152).

4.2. Stochasticity at Low Copy Numbers

Another fundamental constraint for biochemical computations stems from their operation at low molecular concentrations. Here, the paradigmatic example deals with the precision by which a microscopic detector of size a (e.g., a receptor protein) can estimate an average concentration $\bar{c}$ of a diffusible chemical (e.g., a ligand with diffusion constant D) in its local vicinity. The fractional error in this estimate is given by the Berg–Purcell limit (8),

$$\left(\frac{\Delta c}{\bar{c}}\right)^2 \gtrsim \frac{1}{Da\bar{c}T}, \qquad 2.$$

[12] Unambiguous, zero-error specification of each of the $N_{\text{row}} = 60$ cell rows along the long axis of the embryo would require a $\log_2(N_{\text{row}}) = 5.9$ bits of information; there is recent evidence that the missing $\log_2(N_{\text{row}}) - 4.3 \approx 1.6$ bits of information may be accessible via spatial correlations in gene expression (14, 71).

where T is the averaging or integration time allowed for the estimation.[13] Much work has been done to clarify and sharpen this limit (55), as reviewed elsewhere (118), and up to prefactors of order 1 the result stands in its generality. Even though the limit was motivated by the problem of sensing extracellular signals at the cell's surface, its broader importance stems from the fact that it also applies to intracellular sensing, for example, in the binding of TFs to enhancers or promoters on the DNA (44, 129).

Berg–Purcell noise represents a kind of input noise that cannot be removed simply by increasing the network gain, much like an amplifier cannot remove the noise at its input, because it amplifies it along with the signal. Input noise can, however, be lowered via the mechanism of spatial averaging (31, 44, 112) or via nonuniform averaging of the input noise when the correlation time of the latter is comparable to the integration time (39). The mechanism of time averaging requires not only energy (40) but also readout molecules to store the ligand binding states of the receptor in the past (41). In fact, receptors and their integration time (the Berg–Purcell limit), readout molecules, and energy each set a fundamental sensing limit, and in an optimally designed system each of these three resource classes is equally limiting so that no resource is in excess and hence wasted (41, 66).

Information processing is constrained by noise sources beyond the Berg–Purcell noise. Intrinsic noise arises from the random firing of chemical reactions in the network and can contribute to Poisson or super-Poisson (bursty) variance.[14] A further switching noise contribution can come from thermal stochasticity by which molecular systems transition among their conformational or activity states; examples include gating of molecular channels, stochastic stepping of molecular motors, or cycling through promoter or enhancer activity states (e.g., 11, 54, 98). Last, extrinsic noise captures the variability between replicate systems, for example, cell-to-cell variability due to noise in their initial conditions, geometry, parameters, or fluctuating environment (115). Noise can enter the system as input noise, before processing and nonlinear response, or during processing, or it can be added at the output, with different consequences for optimal processing architectures.

4.3. Limited Specificity of Molecular Interactions and Crosstalk

Biochemical networks are defined through molecular interactions that, however, are not infinitely specific: With small probability, molecular interactions can occur between noncognate molecules even if these interactions would be deleterious for function.[15] Spending energy to compensate for the limited specificity of molecular recognition dates back to Hopfield's (49) original work that subsequent analyses have generalized (84, 116) and started connecting to gene regulation (12, 42) and information transmission (15). Systems-level crosstalk constraints arising from limited specificity of interactions in entire biochemical networks have recently begun to be explored (34, 90).

[13] Intuitively, this can be understood as stemming from a Poisson counting variance where the detector samples $a^3\bar{c}$ ligands on average per measurement (which is also the variance per measurement); it measures for time T and each independent measurement requires the ligands to clear the volume by diffusion in time a^2/D, for a total of TD/a^2 measurements.

[14] A classic example is bursty gene expression, where even at constant rate birth–death dynamics for mRNA, when each mRNA gives rise to multiple proteins, the steady-state protein distributions will be overdispersed, with the Fano factor bigger than unity.

[15] We often portray biochemical networks graphically—and simplify our life computationally—by assuming that they are infinitely specific and therefore by drawing only functional arrows between network nodes that strongly interact while neglecting noncognate and possibly nonfunctional or even deleterious off-target interactions.

5. BIOCHEMICAL MOTIFS FOR OPTIMAL INFORMATION PROCESSING

Given the fundamental constraints of Section 4, how should biochemical networks be wired together to perform useful signal-processing tasks and transmit information reliably? Theoretical explorations on this topic have proceeded in two directions. Following the terminology of Section 2, we first review static results computed in the small-noise approximation and then focus on dynamic results, which are based mostly on the Gaussian channel formalism.

5.1. Static Results

How much information flows through the simplest biochemical network motif, $X \rightarrow Y$, in which the concentration of a regulator x (e.g., a TF) affects its target y (e.g., the expression level of a regulated gene) in steady state? In simple biophysical models for this process, we can compute the mean input/output mapping, $\bar{y}(x)$, and the noise, $\sigma_y^2(x)$.[16] In early work, Tkačik et al. (125) analyzed the channel capacity for a single gene regulatory element and compared the results to experiments in a companion paper (126), while Ziv et al. (151) looked at the equivalent transmission problem in more general three-node biochemical motifs. In the small-noise limit, this problem has a well-defined channel capacity optimum in terms of its biophysical parameters when transmission is limited by both input Berg–Purcell noise (which becomes limiting at low inputs) and intrinsic output noise (which becomes limiting at high inputs). The optimum emerges as a generic trade-off between increasing the dynamic range of resulting gene expression and keeping total noise as low as possible (113).

What about more complex motifs? Transmission from a single regulator to multiple noninteracting targets, $X \rightarrow (Y_1, Y_2, \ldots, Y_M)$, has two optimal strategies, depending on whether the input or output noise is dominant (131). If the input noise dominates, all M target genes should respond with the same dose-response curve to the input (i.e., they are perfectly redundant). This solution is optimal because it permits the most effective averaging of the (dominant) input noise and the information $I(x; \mathbf{y})$ grows as $0.5\log_2(M)$ in the best case.[17] If the output noise increases, one observes a phase transition to the tiling optimal strategy: Here, the M target genes cover the range of possible inputs x, each responding most sensitively in a different regime of input concentration; $I(x; \mathbf{y})$ can grow up to 1 bit per gene. Similar results for ligand (X) sensing using multiple different receptors ($Y_1, \ldots, Y_M$) have been analyzed (52), including the effect of ligand crosstalk due to limited receptor specificity (59).

If regulated genes can interact in a feed-forward fashion, qualitatively new solutions become optimal (146). Assuming that input x activates downstream targets $\mathbf{y}$, information flow cannot be increased if one regulated gene also activates another, but repression of one output gene, y_i, by another, y_j, does help by opening up the space of effective nonmonotonic responses of individual y to x. If we simplistically think of genes with sigmoid response curves as ON/OFF switches, then M output genes access $M + 1$ combinatorial binary expression states using the tiling strategy; with repression, bump-like activation curves become possible, expanding the output vocabulary (e.g., from 3 to 4 effective output states for $M = 2$). In information-theoretic terms, repression

[16] The model could be, for instance, one where TF molecules bind and unbind from a DNA binding site; when the site is occupied, gene expression takes place at a constant rate. The binding model gives rise to sigmoid mean input/output (induction) curves; noise due to Berg–Purcell input, switching, and intrinsic output contributions can be explicitly computed, yielding a full characterization of this channel.

[17] Observed redundancy in biological networks is often explained in terms of evolutionary robustness against mutations; here, it emerges due to fundamental signal-processing constraints.

in feed-forward networks allows redundancy reduction, in line with decades-older analogous arguments in neuroscience (4).

The simplest case of a feedback loop is one in which an input x regulates a single target y that can also self-regulate (132). Here, too, results crucially depend on the ratio of input noise to output noise. When input noise is not limiting, self-repression can increase information transmission, consistent with previously reported roles of negative feedback in noise suppression. When input noise is dominant, however, subcritical self-activation becomes optimal. Above the critical self-activation strength, the system would become a noisy bistable switch, with a consequent loss of dynamic range (output would be restricted to two possible values) and appearance of branch ambiguity, leading to low information transmission.[18] Just below critical self-activation, however, the dynamics slows down, extending the intrinsic timescales of the system and resulting in more effective temporal (input) noise averaging. This work suggested a previously underappreciated role for subcritical self-activation and a possible explanation for differing accounts of the efficiency of negative feedback (2, 18, 142).

A large and underexplored playing field opens up when one explicitly considers space. Information flow can be optimized via diffusive spatial coupling (31, 44), which should provide sufficient, but not excessive, spatial averaging (112). Spatial partitioning and localization of chemical reactions can decorrelate signals and create parallel information channels that increase the total capacity when the partitioning correctly balances the number of parallel channels against the noise in each (81), or induce multistability when this is beneficial for signal processing (46). And cooperative protein clustering can suppress output noise by digitizing the output, yielding an optimal cluster size that balances noise suppression against the reduction in the number of output states (101).

5.2. Dynamic Results

In many cases, the system of interest responds on a timescale that is similar to the timescale on which the input varies. Pertinent examples are provided by organisms that need to navigate their environment, since the input is shaped by their response during navigation (136). In these cases, the frequency-dependent gain-to-noise ratio, which determines the information rate within the Gaussian framework, becomes the quantity of interest (137). Earlier work has studied how network motifs sculpt such frequency-dependent responses. Feedback and self-regulation can improve information transmission, but only if they occur upstream of the dominant noise source in the cascade; if they are downstream, the gain and noise are affected similarly, leaving the information rate unchanged. In addition, if signals over the entire frequency range have to be transmitted accurately, then (upstream) positive self-regulation is advantageous because it increases the gain-to-noise ratio across all frequencies; if instead the cell is concerned only with low- or high-frequency signals, then positive or negative feedback will be useful (24). Feed-forward motifs can shape the frequency-dependent gain-to-noise ratio: Both a coherent and an incoherent feed-forward loop can act as either a low-pass or a high-pass filter of information, depending on the strength of the coupling between the components (25). Moreover, cooperative activation of the output via a diamond motif can increase the gain-to-noise ratio: It can be useful to split the input signal, relay it over parallel pathways, and recombine the split signals at the output (25).

5.3. Circadian Clocks

Information theory has also been instrumental in elucidating optimal sensing networks for responding to environmental changes. When the variations are highly regular, such as the daily

[18] A bistable circuit could have other, nontransmission-related uses, perhaps as a store of cellular memory, due to its stability and hysteresis.

light–dark cycles, it becomes beneficial to develop a clock from which the time and hence the future environment can be inferred. Interestingly, many organisms, ranging from cyanobacteria to animals, employ a circadian clock that is based on a limit-cycle oscillator that can tick autonomously with a nearly 24 h period. Yet a limit-cycle oscillator is not essential for knowing the time, as illustrated by cyanobacteria that employ an hourglass (i.e., a system that in the presence of oscillatory driving exhibits robust oscillations from which the organism can infer the time but that in the absence of any driving would relax to a stable fixed point rather than a limit cycle). Modeling has revealed that in the limit of low input noise, the accuracy of hourglass clocks is comparable to that of true clocks based on a limit-cycle oscillator. However, in the presence of strong input fluctuations, as arising from, e.g., clouds, the bona fide clocks are far superior because they can lift the trade-off between gain and input noise (77, 92).

Time cannot be uniquely inferred from the oscillations of a single clock output component, because a single readout level maps onto two different time points during the cycle. Interestingly, the clock typically drives the expression of many genes downstream (67). Modeling has shown that the accuracy of telling time increases with the number of downstream oscillations, and that there exists an optimal phase relation between these oscillations, depending on their noise and amplitude (78). Moreover, cross-correlations between the noise can increase the mutual information by improving the tiling of the encoding space, similar to that described in Section 5.1 (131). Information theory has also given tantalizing insights into why many circadian clocks feature a so-called dead zone, during which the clock does not couple to the input (76). Stable clocks benefit from a large dead zone and weak input coupling, while the opposite is true for clocks that are inherently less stable (76).

5.4. Prediction

When variations in the environment are entirely unpredictable, organisms have no choice but to resort to the bet-hedging strategy, by increasing phenotypic diversity, or the detect-and-respond strategy. But there is an interesting regime between the fully unpredictable environment and the fully predictable circadian rhythms: Here, environment is sufficiently predictable to permit initiating a response ahead of time (6). Some signaling networks can reach the fundamental, information-bottleneck bound (121) on predictive information by encoding the relevant features of the past signals into their network state (123). For example, push–pull networks can reach the bound for Markovian signals, while derivative-taking networks can reach the bound for non-Markovian signals. However, reaching this information bound can be prohibitively costly in terms of protein copies and energy. As a result, an optimal system that maximizes the predictive information under an information compression constraint will, in general, be different from a system optimized under a physical resource constraint.

Finally, we note that cellular networks can use temporal dynamics as a degree of freedom to effectively transmit signals. Oscillatory inputs can be used to drive gene promoters (82), with the paradoxical result that an oscillatory input can lead to a more constant output than a constant input (135). Moreover, oscillatory signals can be used to multiplex signals, i.e., transmit multiple signals through a common pathway, and yet respond specifically and reliably to each of them, with minimal crosstalk (22, 23).

6. TOWARD A PRINCIPLED THEORY

We can estimate information flows from data and express them in a common currency across multiple kingdoms of life. We know that these flows are limited by physical constraints that we in principle understand and can sometimes compute or measure. We hypothesize that within those

constraints there exist special operating regimes and reaction network motifs where information flows are maximized. Together, these three statements appear to be reasonable ingredients for a principled theory. But how should these ingredients be combined, are they sufficient, and if so, what precisely is the theory? One is tempted to suggest that biochemical reaction networks studied in living organisms evolved to maximize information transmission between inputs and outputs, subject to fundamental as well as resource constraints (on time, space, energy expenditure, molecular copy numbers, and specificity of molecular interactions). While attractive, this view raises at least two fundamental questions on which we elaborate in Sections 6.1 and 6.2.

6.1. Can the Putative Theory Be Framed as a Constrained Information Optimization Problem?

If the putative theory can be framed as a constrained information optimization problem, it should predict unknown network parameters (e.g., interaction topologies, reaction rates) given the empirical resource constraints (e.g., number of available signaling molecules)—hopefully in a regime where the number of constraints (to be measured) is much smaller than the number of parameters (to be predicted). One important question is, How do we deal with cases for which we know that a constraint should exist on theoretical grounds but for which we cannot directly estimate its value (e.g., metabolic cost of making one signaling molecule)? Another question concerns the notion of optimization itself—we either explicitly assert or implicitly assume that the putative optimization is a result of an evolutionary process, but adaptive evolution has its own fundamental limits.[19] What if the evolutionary optimization is incomplete, for example, when the biochemical network in question is close to but not at optimality, because the adaptive dynamics is still unfolding? Can our plan of action still be framed in a mathematically rigorous way?

There are indications that formal issues relating to optimality theories applied to complex biological systems can be conquered (75).[20] There are two important consequences to this line of research. First, optimization-based theory can be predictive even when the observed system does not sit precisely at the optimal point (89). Second, instead of dealing with point predictions (about the single optimal system), we need to think about optimal ensembles (110), that is, distributions of likely outcomes under selection for functional phenotypes (e.g., high information transmission). There are tantalizing mathematical analogies between the Boltzmann distribution of statistical physics that reweights otherwise neutral configurations by their energy and how fitness reweights mutationally neutral configurations in classic population genetics (48, 106), as well as how probabilistic optimization fits into this picture (20, 75). While we are still working toward a mathematical language to formally connect optimality theories and theoretical evolution using mutually informative quantities, constrained optimization—broadly construed and equipped with novel statistical approaches—seems to be our best bet for the structure of such a theory.

6.2. Why Would Information Flow Be Optimized if It Is Not the Phenotype Directly Under Selection?

While the information rate in chemotaxis might be of theoretical interest, the actual phenotype on which selection can act is likely more concrete and specific, for instance, the average bacterial

[19]The computation of these limits—due to limited selection strength in finite populations, linkage, mutation load, etc.—has been the cornerstone of theoretical population genetics for more than a century.

[20]We also believe that such issues can yield novel statistical problems at the interface of inference and optimization—as a sign of productive interdisciplinary cross-pollination—that are interesting in their own right.

swimming velocity in the direction of chemotactic gradient (drift speed). Given that, in general, all bits of information are neither equally costly nor equally predictive, is the optimal system that maximizes the drift speed under a resource constraint the same as that which maximizes information transmission (123)? In a related vein, while information measures bits per second transmitted from the stimulus, these bits encode different stimulus features, such as its instantaneous value or its derivative, and these features may not be of equal fitness value to the organism (6, 123). Ultimately, isn't an optimal information-processing system a system that discards irrelevant bits[21] while preserving the relevant bits, presumably those that are predictive about future states of the environment that can inform an organism's behavior? If so, maximizing transmission surely is not the full story: Transmission is not computation. We agree. Transmission and reliable information representation are necessary but not sufficient for a full-blown theory.

Promisingly, theoretical frameworks exist that link the reliability of internal representations (subject of information theory) to bounds on achievable performance of any behavior that is guided by such representations. If internal representation of the environmental state is noisy and unreliable, cellular or organismal behaviors based on such representations will be mismatched and possibly maladaptive in the environments in which they arise, leading to a fitness penalty. Low information transmission and, consequently, unreliable internal representations will therefore be selected against; conversely, selection pressure should drive the transmission to be optimized within resource constraints at least to the point where it alone is not limiting behavioral performance. This relationship can be captured mathematically in the context of the rate-distortion theory (3, 21, 36, 69, 133), and there are intriguing connections to reinforcement learning (122) and Bayesian decision-making (58) that follow analogous developments in neuroscience, cognitive sciences, robotics, and machine learning. An information-theoretic framework connecting the accuracy of biochemical processing and self-organized cell fate patterning has also been recently proposed (14). Yet another exciting connection is to the field of control theory, highlighting the benefits (104) and limitations (62) of control paradigms applied to biological contexts, in which sensing, transmission, computation, and actuation are all limited in their speed and accuracy due to fundamental constraints (51). We hope and expect to see exciting work emerging on this interface in the years to come.

6.3. Conclusion

For physicists, evolved systems are a source of fascination and frustration alike. They often operate in murky regimes that we collectively avoid because they are complex: out of equilibrium, away from thermodynamic limits, with disorder that appears essential for function yet is statistically nontrivial. Biochemical networks are paradigmatic examples of such evolved complexity. Even with a partial knowledge of their structure in hand, they are difficult to rationalize or reverse engineer, and an effort toward building any predictive theory may appear hopeless. We appeal against such a defeatist attitude. On the contrary, we argue that the components for the theory are all in place: The fusion of stochastic chemical kinetics, information and rate-distortion theory, optimal decision-making, and, most importantly, evolutionarily plausible optimization should provide the components we need. It is likely that the predictions of an emergent constrained optimization theory will not be unique and neither will the observed systems be fully optimal, so we need to develop a new powerful statistical language for connecting theoretical predictions with partial and incomplete measurements of the high-dimensional network state. There is much work to be done,

[21] One has to pay for the transmission of irrelevant bits, yet they have no bearing on an organism's decisions and therefore on its fitness.

but the overall direction is clear and should be encouraging. In addition to the chemotactic network of *E. coli*, a handful of recent examples of the predictive power of information optimization to quantitatively account for complex evolved regulatory networks should give this direction a significant motivational boost (33, 110) and connect clean, theoretically tractable toy model scenarios of Section 5 to real data.

7. APPENDIX: TECHNICAL DETAILS AND ESTIMATION METHODS

Intuitively, 1 bit of information corresponds roughly to sufficient knowledge gain about variable Y when we are told the value of X, such that we can halve our uncertainty about Y—for instance, by being able to decide whether the value of Y is below or above its median. Expressed in the language of regulation, I bits of mutual information between X and Y imply that Y can be controlled by changing X to access roughly 2^I distinguishable states given its noise and limited dynamic range.[22] Another way to build intuition is to consider the Gaussian case for which information calculations can be done analytically. Let X be a Gaussian signal with variance S^2. The channel corrupts X with additive Gaussian noise with variance N^2 to produce output Y. The capacity of such a channel depends only on the signal-to-noise ratio, SNR $= S/N$, so that $I = \frac{1}{2}\log_2(1 + \mathrm{SNR}^2)$; when SNR $\gg 1$, this reduces to our rough estimate of the logarithm of noise levels that can fit into the output dynamic range. Yet another way to view the Gaussian case is to realize that X and Y will be jointly Gaussian with linear correlation coefficient c, yielding a mutual information of $I = -\frac{1}{2}\log_2(1 - c^2)$; here, a linear correlation of $c = 0.5$ corresponds to a numerically low information value of $I = 0.21$ bits, and values of $c \gtrsim 0.9$ are required to approach or exceed 1 bit of transmission.

Before proceeding, we need to address one more important question: How can information be computed or estimated in practice when we are not in the analytically tractable linear Gaussian regime? Many biological networks, especially involving gene regulation, operate in saturated, nonlinear regimes or may feature non-Gaussian input or noise distributions. Information-theoretic quantities, like mutual information I in Equation 1, require integration or summation over full probability spaces and thus suffer from the curse of dimensionality, such that there are no universal solutions; instead, we schematically organize various existing approaches and approximation schemes in **Table 1**.

7.1. The Information Metric

The first important question is which information metric to use. The answer depends on whether the inputs and outputs are understood as instantaneous states of the reaction network (static case) or full state trajectories (dynamic case). Many biochemical networks may operate in (quasi-)equilibrium: Inputs either change slowly so that the network response equilibrates before the input changes substantially, or the inputs change quickly but rarely, and the relevant response is thought to be encoded in the network steady state for each period of constant input. In these static cases, information is estimated between instantaneous samples of input values and, for instance, instantaneous or delayed steady-state responses. Dimensionality-wise, this is a huge simplification relative to the fully dynamic case. In contrast, when inputs are stochastic trajectories with an internal timescale structure, as would be the case for chemoattractant concentration $x(t)$ drawn, for example, from an Ornstein–Uhlenbeck process with a well-defined autocorrelation time, and the reaction network itself is characterized by response and noise with timescales comparable to the

[22] Thus, if a variable Y takes values in a limited dynamic range between 0 and 1 with a typical noise magnitude of 0.25, and changing X allowed Y to span its entire range, this would correspond to roughly $\log_2(1/0.25) = 2$ bits of information.

Table 1 Approaches to information and information rate estimation in biochemical networks[a]

Regime of validity			Approach	CME model required?	Reference(s)
Discrete	Low-dimensional state space	Static	Solution (perhaps approximate) of CME for systems with simple and finite state spaces, direct summation of joint distributions for information calculations	Yes	Standard textbook (e.g., 19)
			Direct sampling and summation over small state space	No	Literature on empirical information estimates (e.g., 87)
	Full CME/high-dimensional state space	Dynamic	Direct solution of CME for simple, finite systems	Yes	29, 119
Continuous	Gaussian	Static and dynamic	Exact analytical results	No	18, 69, 135
	Non-Gaussian	Static	Density estimation for low-dimensional problems	No	127
			Small-noise approximation	No	126
		Dynamic	kNN estimates over trajectories	No	105
			Direct information estimates in hand-extracted trajectory features	No	45
			Decoding lower bound if either input or output state space is low dimensional and discrete	No	16, 43, 53

[a]Model-based approaches require an explicit reaction model (e.g., stoichiometry, reaction rates) and may include exact calculations, approximations, or Monte Carlo techniques. Data-driven approaches are generic and only rely on samples of channel inputs and outputs; they can always be applied if the model is known, by first drawing samples from the model.

Abbreviations: CME, chemical master equation; kNN, *k*-nearest-neighbor.

input autocorrelation time, the proper approach is dynamic, requiring us to consider the information rate, *R*. Failing to do so can lead to spectacularly wrong results—for instance, the *E. coli* chemotaxis network exhibits perfect adaptation, such that the response to any steady-state level of chemoattractant is zero and *I* would be zero even as the rate, *R*, is not. The mutual information between instantaneous values of time-varying input and output signals also does not properly take into account the effect of the signal (auto)correlations on the information flow per unit amount of time; only the rate, *R*, does (73).

7.2. The Method

The approach for computing the information measure of interest depends on whether the inputs and outputs are discrete or continuous.[23] For discrete unbounded state spaces such as those over which the generic chemical master equation is defined, the options are restricted (see **Table 1**), in contrast to cases where the state spaces of the input or the output are finite and effectively small.

[23]Mixed cases are also possible and relevant, for instance, when an input is a binary variable indicating the presence or absence of a stimulus, while the response is a full dynamic trajectory of a signaling molecule concentration (16, 43).

By small we mean that the states over which the probabilities have nonnegligible mass can be tractably enumerated in the computer and thus information can be obtained by direct summation, for instance, when the input corresponds to binary activation/deactivation of a signaling molecule or a receptor; the presence/absence of a chemical signal; a countable number of internal states of some molecular machinery (e.g., promoter or enhancer); or an absolute count of a molecular species present at very low numbers, such as mRNA in bacterial cells. For some of these cases the chemical master equation can be solved explicitly for model-based estimates; alternatively, it can be simulated or the data can be histogrammed to approximate the channel, $P(Y|X)$, by direct counting, at least in static cases. Dynamic rate computations have been performed only for simple toy model channels or reaction networks (28, 29, 119).

When signals are continuous but non-Gaussian, the channel can be constructed explicitly in static, low-dimensional cases (for instance, when both input and output are 1D, the channel can be represented as a matrix over a finely discretized domain for inputs and outputs) either by explicit computation from a reaction network model or through density estimation from data samples. Mutual information or capacity is then computable by explicit summation or Monte Carlo integration (in somewhat higher dimensions). Alternatively, further assumptions can help: In the Gaussian-noise static approximation, one assumes that the output y is normally distributed given input x, but with mean and covariance that depend on the input x:[24] $y \sim \mathcal{N}(\bar{y}(x); \mathbf{C}(x))$. This is attractive since the first two moments can be obtained by Langevin approximation to reaction network dynamics or can typically be estimated from data (126, 127). If noise is small in addition to being Gaussian, analytic information and capacity computations are sometimes possible (130, 131). For dynamic continuous non-Gaussian cases, not many options are available, although carefully calibrated machine learning–based estimators show early promise (117, 149). In an early attempt, Selimkhanov et al. (105) coarsely sampled output network trajectories and viewed each trajectory as a low-dimensional vector of responses to (a usually small, countable set of) discrete stimuli (105); generic mutual information estimators, such as the k-nearest-neighbor estimator (60), can then be used to extract the information in bits. PWS, discussed in Section 2, is an exact technique to compute the mutual information between input and output trajectories for any stochastic model (96).

Decoding- or feature-based lower-bound estimates for the information are interesting alternatives relevant for discrete as well as continuous cases. Let us assume that the input X is low dimensional but that the output Y is intractable for direct sampling/integration/summation (e.g., Y is high-dimensional discrete or continuous, or is a trajectory). If we can compress Y into low-dimensional $\tilde{Y} = f(Y)$ using some function f, we could estimate $I(X;\tilde{Y})$ directly, and since

$$X \xrightarrow{P(Y|X)} Y \xrightarrow{f} \tilde{Y}$$

is a Markov chain, data processing inequality asserts $I(X;\tilde{Y}) \leq I(X;Y)$ (19), and so a tractable estimate of $I(X;\tilde{Y})$ will lower bound the (intractable) information $I(X;Y)$ of interest. Such lower-bound estimates are attractive because they are conservative; the question is how to select a deterministic function f so that they are also as tight as possible. In decoding-based lower bounds we look for functions f that recover (or decode) the original X as well as possible from channel output Y, which can be done by any regression or classification technique, from linear or logistic to

[24]Note that this does not correspond to a Gaussian additive white noise channel for which we have explicit analytical results, because of the arbitrary dependence of mean and noise on the input. Even if the input distribution $P_X(x)$ were Gaussian, the output marginal $P_Y(y)$ in general would not be. Instead, it would be a Gaussian mixture that is a universal approximator and for which no exact results for entropic quantities would exist.

modern deep learning, depending on training data availability. The idea is that supervised learning of f is possible already at training set sizes that are still way too small to directly sample $P(Y|X)$, and that reconstruction functions f with lower error will yield tighter information estimates (16, 53). Alternatively, instead of decoding-based bounds one can use feature-based bounds; here, f extracts from high-dimensional trajectory samples Y several handcrafted features (e.g., amplitude, frequency, count of pulses) that are deemed biologically relevant. These features are collected into feature vector $\tilde{Y}$, and if $\tilde{Y}$ is sufficiently low dimensional, direct information estimation for $I(X;\tilde{Y})$ can be performed to lower bound the information $I(X;Y)$ of interest.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

G.T. acknowledges the support of the Human Frontiers Science Program (HFSP), the Austrian Science Fund (FWF 10.55776/P34015, 10.55776/P28844), and the European Research Council Synergy DYNATRANS (ERC-2023-SyG 101118866) grant. P.R.t.W. performed his work at the research institute AMOLF and acknowledges support from the Dutch Research Council (NWO) and funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement 885065).

LITERATURE CITED

1. Alon U, Surette MG, Barkai N, Leibler S. 1999. Robustness in bacterial chemotaxis. *Nature* 397(6715):168–71
2. Anders A, Ghosh B, Glatter T, Sourjik V. 2020. Design of a MAPK signalling cascade balances energetic cost versus accuracy of information transmission. *Nat. Commun.* 11(1):3494
3. Andrews BW, Iglesias PA. 2007. An information-theoretic characterization of the optimal gradient sensing response of cells. *PLOS Comput. Biol.* 3(8):e153
4. Barlow HB. 1961. Possible principles underlying the transformation of sensory messages. In *Sensory Communication*, ed. WA Rosenblith, pp. 217–34. Cambridge, MA: MIT Press
5. Bauer M, Petkova MD, Gregor T, Wieschaus EF, Bialek W. 2021. Trading bits in the readout from a genetic network. *PNAS* 118(46):e2109011118
6. Becker NB, Mugler A, Ten Wolde PR. 2015. Optimal prediction by cellular signaling networks. *Phys. Rev. Lett.* 115:258103
7. Benzinger D, Ovinnikov S, Khammash M. 2022. Synthetic gene networks recapitulate dynamic signal decoding and differential gene expression. *Cell Syst.* 13(5):353–64.e6
8. Berg HC, Purcell EM. 1977. Physics of chemoreception. *Biophys. J.* 20(2):193–219
9. Bialek W. 2012. *Biophysics: Searching for Principles*. Princeton, NJ: Princeton Univ. Press
10. Blahut R. 1972. Computation of channel capacity and rate-distortion functions. *IEEE Trans. Inform. Theory* 18(4):460–73
11. Blake WJ, Kærn M, Cantor CR, Collins JJ. 2003. Noise in eukaryotic gene expression. *Nature* 422(6932):633–37
12. Boeger H. 2022. Kinetic proofreading. *Annu. Rev. Biochem.* 91:423–47
13. Bowsher CG, Swain PS. 2014. Environmental sensing, information transfer, and cellular decision-making. *Curr. Opin. Biotechnol.* 28:149–55
14. Brückner DB, Tkačik G. 2024. Information content and optimization of self-organized developmental systems. *PNAS* 121(23):e2322326121
15. Cepeda-Humerez SA, Rieckh G, Tkačik G. 2015. Stochastic proofreading mechanism alleviates crosstalk in transcriptional regulation. *Phys. Rev. Lett.* 115(24):248101

16. Cepeda-Humerez SA, Ruess J, Tkačik G. 2019. Estimating information in time-varying signals. *PLOS Comput. Biol.* 15(9):e1007290
17. Chalk M, Marre O, Tkačik G. 2018. Toward a unified theory of efficient, predictive, and sparse coding. *PNAS* 115(1):186–91
18. Cheong R, Rhee A, Wang CJ, Nemenman I, Levchenko A. 2011. Information transduction capacity of noisy biochemical signaling networks. *Science* 334(6054):354–58
19. Cover TM, Thomas JA. 1999. *Elements of Information Theory*. Hoboken, NJ: Wiley
20. Crisanti A, De Martino A, Fiorentino J. 2018. Statistics of optimal information flow in ensembles of regulatory motifs. *Phys. Rev. E* 97(2):022407
21. De Martino D, Mc Andersson A, Bergmiller T, Guet CC, Tkačik G. 2018. Statistical mechanics for metabolic networks during steady state growth. *Nat. Commun.* 9(1):2988
22. de Ronde W, ten Wolde PR. 2014. Multiplexing oscillatory biochemical signals. *Phys. Biol.* 11(2):026004
23. de Ronde W, Tostevin F, ten Wolde PR. 2011. Multiplexing biochemical signals. *Phys. Rev. Lett.* 107(4):048101
24. de Ronde WH, Tostevin F, ten Wolde PR. 2010. Effect of feedback on the fidelity of information transmission of time-varying signals. *Phys. Rev. E* 82(3 Part 1):031914
25. de Ronde WH, Tostevin F, ten Wolde PR. 2012. Feed-forward loops and diamond motifs lead to tunable transmission of information in the frequency domain. *Phys. Rev. E* 86(2):021913
26. Dubuis JO, Samanta R, Gregor T. 2013. Accurate measurements of dynamics and reproducibility in small genetic networks. *Mol. Syst. Biol.* 9(1):639
27. Dubuis JO, Tkačik G, Wieschaus EF, Gregor T, Bialek W. 2013. Positional information, in bits. *PNAS* 110(41):16301–8
28. Duso L, Zechner C. 2019. Path mutual information for a class of biochemical reaction networks. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pp. 6610–15. Nice, France: IEEE
29. Eckford AW, Thomas PJ. 2018. The channel capacity of channelrhodopsin and other intensity-driven signal transduction receptors. *IEEE Trans. Mol. Biol. Multi-Scale Commun.* 4(1):27–38
30. Elowitz MB, Levine AJ, Siggia ED, Swain PS. 2002. Stochastic gene expression in a single cell. *Science* 297(5584):1183–86
31. Erdmann T, Howard M, Ten Wolde PR. 2009. Role of spatial averaging in the precision of gene expression patterns. *Phys. Rev. Lett.* 103(25):258101
32. Farsad N, Yilmaz HB, Eckford A, Chae CB, Guo W. 2016. A comprehensive survey of recent advancements in molecular communication. *IEEE Commun. Surv. Tutor.* 18(3):1887–919
33. François P, Siggia ED. 2010. Predicting embryonic patterning using mutual entropy fitness and in silico evolution. *Development* 137(14):2385–95
34. Friedlander T, Prizak R, Guet CC, Barton NH, Tkačik G. 2016. Intrinsic limits to gene regulation by global crosstalk. *Nat. Commun.* 7(1):12307
35. Fuchsberger FF, Kim D, Baranova N, Vrban H, Kagelmacher M, et al. 2023. Information transfer in mammalian glycan-based communication. *eLife* 12:e69415
36. Genewein T, Leibfried F, Grau-Moya J, Braun DA. 2015. Bounded rationality, abstraction, and hierarchical decision-making: an information-theoretic optimality principle. *Front. Robot. AI* 2:27
37. Gillespie DT. 2007. Stochastic simulation of chemical kinetics. *Annu. Rev. Phys. Chem.* 58:35–55
38. Givré A, Colman-Lerner A, Ponce Dawson S. 2023. Modulation of transcription factor dynamics allows versatile information transmission. *Sci. Rep.* 13(1):2652
39. Govern CC, Ten Wolde PR. 2012. Fundamental limits on sensing chemical concentrations with linear biochemical networks. *Phys. Rev. Lett.* 109:218103
40. Govern CC, Ten Wolde PR. 2014. Energy dissipation and noise correlations in biochemical sensing. *Phys. Rev. Lett.* 113:258102
41. Govern CC, Ten Wolde PR. 2014. Optimal resource allocation in cellular sensing systems. *PNAS* 111(49):17486–91
42. Grah R, Zoller B, Tkačik G. 2020. Nonequilibrium models of optimal enhancer function. *PNAS* 117(50):31614–22
43. Granados AA, Pietsch JM, Cepeda-Humerez SA, Farquhar IL, Tkačik G, Swain PS. 2018. Distributed and dynamic intracellular organization of extracellular information. *PNAS* 115(23):6088–93

44. Gregor T, Tank DW, Wieschaus EF, Bialek W. 2007. Probing the limits to positional information. *Cell* 130(1):153–64
45. Hansen AS, O'Shea EK. 2015. Limits on information transduction through amplitude and frequency regulation of transcription factor activity. *eLife* 4:e06559
46. Harrington HA, Feliu E, Wiuf C, Stumpf MP. 2013. Cellular compartments cause multistability and allow cells to process more information. *Biophys. J.* 104(8):1824–31
47. Hironaka Ki, Morishita Y. 2012. Encoding and decoding of positional information in morphogen-dependent patterning. *Curr. Opin. Genet. Dev.* 22(6):553–61
48. Hledík M, Barton N, Tkačik G. 2022. Accumulation and maintenance of information in evolution. *PNAS* 119(36):e2123152119
49. Hopfield JJ. 1974. Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *PNAS* 71(10):4135–39
50. Huang J, Meyn SP. 2005. Characterization and computation of optimal distributions for channel coding. *IEEE Trans. Inform. Theory* 51(7):2336–51
51. Iyer KS, Prabhakara C, Mayor S, Rao M. 2023. Cellular compartmentalisation and receptor promiscuity as a strategy for accurate and robust inference of position during morphogenesis. *eLife* 12:e79257
52. Jetka T, Nienałtowski K, Filippi S, Stumpf MP, Komorowski M. 2018. An information-theoretic framework for deciphering pleiotropic and noisy biochemical signaling. *Nat. Commun.* 9(1):4591
53. Jetka T, Nienałtowski K, Winarski T, Błoński S, Komorowski M. 2019. Information-theoretic analysis of multivariate single-cell signaling responses. *PLOS Comput. Biol.* 15(7):e1007132
54. Jones DL, Brewster RC, Phillips R. 2014. Promoter architecture dictates cell-to-cell variability in gene expression. *Science* 346(6216):1533–36
55. Kaizu K, De Ronde WH, Paijmans J, Takahashi K, Tostevin F, Ten Wolde PR. 2014. The Berg-Purcell limit revisited. *Biophys. J.* 106(4):976–85
56. Karolak A, Branciamore S, McCune JS, Lee PP, Rodin AS, Rockne RC. 2021. Concepts and applications of information theory to immuno-oncology. *Trends Cancer* 7(4):335–46
57. Kimura M. 1961. Natural selection as the process of accumulating genetic information in adaptive evolution. *Genet. Res.* 2(1):127–40
58. Kobayashi TJ. 2010. Implementation of dynamic Bayesian decision making by intracellular kinetics. *Phys. Rev. Lett.* 104(22):228104
59. Komorowski M, Tawfik DS. 2019. The limited information capacity of cross-reactive sensors drives the evolutionary expansion of signaling. *Cell Syst.* 8(1):76–85
60. Kraskov A, Stögbauer H, Grassberger P. 2004. Estimating mutual information. *Phys. Rev. E* 69(6):066138
61. Landauer R. 1961. Irreversibility and heat generation in the computing process. *IBM J. Res. Dev.* 5(3):183–91
62. Lestas I, Vinnicombe G, Paulsson J. 2010. Fundamental limits on the suppression of molecular fluctuations. *Nature* 467(7312):174–78
63. Levchenko A, Nemenman I. 2014. Cellular noise and information transmission. *Curr. Opin. Biotechnol.* 28:156–64
64. MacKay DM, McCulloch WS. 1952. The limiting information capacity of a neuronal link. *Bull. Math. Biophys.* 14:127–35
65. Maity A, Wollman R. 2020. Information transmission from NFkB signaling dynamics to gene expression. *PLOS Comput. Biol.* 16(8):e1008011
66. Malaguti G, Ten Wolde PR. 2021. Theory for the optimal detection of time-varying signals in cellular sensing systems. *eLife* 10:e62574
67. Markson JS, Piechura JR, Puszynska AM, O'Shea EK. 2013. Circadian control of global gene expression by the cyanobacterial master regulator RpaA. *Cell* 155(6):1396–408
68. Marr D, Poggio T. 1976. *From understanding computation to understanding neural circuitry*. Tech. Rep., AI Memo 357, MIT, Cambridge, MA
69. Mattingly H, Kamino K, Machta B, Emonet T. 2021. *Escherichia coli* chemotaxis is information limited. *Nat. Phys.* 17(12):1426–31

70. Mc Mahon SS, Sim A, Filippi S, Johnson R, Liepe J, et al. 2014. Information theory and signal transduction systems: from molecular information processing to network inference. *Semin. Cell Dev. Biol.* 35:98–108
71. McGough L, Casademunt H, Nikolić M, Aridor Z, Petkova MD, et al. 2024. Finding the last bits of positional information. *PRX Life* 2(1):013016
72. Mehta P, Goyal S, Long T, Bassler BL, Wingreen NS. 2009. Information processing and signal integration in bacterial quorum sensing. *Mol. Syst. Biol.* 5(1):325
73. Meijers M, Ito S, Ten Wolde PR. 2021. Behavior of information flow near criticality. *Phys. Rev. E* 103(1):L010102
74. Micali G, Endres RG. 2019. Maximal information transmission is compatible with ultrasensitive biological pathways. *Sci. Rep.* 9(1):16898
75. Młynarski W, Hledík M, Sokolowski TR, Tkačik G. 2021. Statistical analysis and optimality of neural systems. *Neuron* 109(7):1227–41
76. Monti M, Lubensky DK, Ten Wolde PR. 2018. Optimal entrainment of circadian clocks in the presence of noise. *Phys. Rev. E* 97(3):032405
77. Monti M, Lubensky DK, Ten Wolde PR. 2018. Robustness of clocks to input noise. *Phys. Rev. Lett.* 121(7):078101
78. Monti M, Ten Wolde PR. 2016. The accuracy of telling time via oscillatory signals. *Phys. Biol.* 13(3):035005
79. Morishita Y, Iwasa Y. 2009. Accuracy of positional information provided by multiple morphogen gradients with correlated noise. *Phys. Rev. E* 79(6):061905
80. Morishita Y, Iwasa Y. 2011. Coding design of positional information for robust morphogenesis. *Biophys. J.* 101(10):2324–35
81. Mugler A, Tostevin F, Ten Wolde PR. 2013. Spatial partitioning improves the reliability of biochemical signaling. *PNAS* 110(15):5927–32
82. Mugler A, Walczak AM, Wiggins CH. 2010. Information-optimal transcriptional response to oscillatory driving. *Phys. Rev. Lett.* 105(5):058101
83. Munsky B, Khammash M. 2006. The finite state projection algorithm for the solution of the chemical master equation. *J. Chem. Phys.* 124(4):044104
84. Murugan A, Huse DA, Leibler S. 2012. Speed, dissipation, and error in kinetic proofreading. *PNAS* 109(30):12034–39
85. Ouldridge TE, Govern CC, Ten Wolde PR. 2017. Thermodynamics of computational copying in biochemical systems. *Phys. Rev. X* 7(2):021004
86. Ouldridge TE, Ten Wolde PR. 2017. Fundamental costs in the production and destruction of persistent polymer copies. *Phys. Rev. Lett.* 118(15):158103
87. Paninski L. 2003. Estimation of entropy and mutual information. *Neural Comput.* 15(6):1191–253
88. Pérez PD, Weiss JT, Hagen SJ. 2011. Noise and crosstalk in two quorum-sensing inputs of *Vibrio fischeri*. *BMC Syst. Biol.* 5(1):153
89. Pérez-Escudero A, Rivera-Alba M, de Polavieja GG. 2009. Structure of deviations from optimality in biological systems. *PNAS* 106(48):20544–49
90. Perkins ML, Crocker J, Tkačik G. 2025. Chromatin enables precise and scalable gene regulation with factors of limited specificity. *PNAS* 122(1):e2411887121
91. Petkova MD, Tkačik G, Bialek W, Wieschaus EF, Gregor T. 2019. Optimal decoding of cellular identities in a genetic network. *Cell* 176(4):844–55
92. Pittayakanchit W, Lu Z, Chew J, Rust MJ, Murugan A. 2018. Biophysical clocks face a trade-off between internal and external noise resistance. *eLife* 7:e37624
93. Potter GD, Byrd TA, Mugler A, Sun B. 2017. Dynamic sampling and information encoding in biochemical networks. *Biophys. J.* 112(4):795–804
94. Poulton JM, Ten Wolde PR, Ouldridge TE. 2019. Nonequilibrium correlations in minimal dynamical models of polymer copying. *PNAS* 116(6):1946–51
95. Razo-Mejia M, Marzen S, Chure G, Taubman R, Morrison M, Phillips R. 2020. First-principles prediction of the information processing capacity of a simple genetic circuit. *Phys. Rev. E* 102(2):022404

96. Reinhardt M, Tkačik G, Ten Wolde PR. 2023. Path weight sampling: exact Monte Carlo computation of the mutual information between stochastic trajectories. *Phys. Rev. X* 13(4):041017
97. Rhee A, Cheong R, Levchenko A. 2012. The application of information theory to biochemical signaling systems. *Phys. Biol.* 9(4):045011
98. Rieckh G, Tkačik G. 2014. Noise and information transmission in promoters with multiple internal states. *Biophys. J.* 106(5):1194–204
99. Rieke F, Warland D, de Ruyter van Steveninck R, Bialek W. 1999. *Spikes: Exploring the Neural Code*. Cambridge, MA: MIT Press
100. Rivoire O, Leibler S. 2011. The value of information for populations in varying environments. *J. Stat. Phys.* 142(6):1124–66
101. Roob E III, Trendel N, Ten Wolde PR, Mugler A. 2016. Cooperative clustering digitizes biochemical signaling and enhances its fidelity. *Biophys. J.* 110(7):1661–69
102. Ruiz R, de la Cruz F, Fernandez-Lopez R. 2018. Negative feedback increases information transmission, enabling bacteria to discriminate sublethal antibiotic concentrations. *Sci. Adv.* 4(11):eaat5771
103. Ryzhkov NV, Nikolaev KG, Ivanov AS, Skorb EV. 2021. Infochemistry and the future of chemical information processing. *Annu. Rev. Chem. Biomol. Eng.* 12:63–95
104. Sarma AA, Li JSL, Stenberg J, Card G, Heckscher ES, et al. 2022. Internal feedback in biological control: architectures and examples. arXiv:2110.05029 [eess.SY]
105. Selimkhanov J, Taylor B, Yao J, Pilko A, Albeck J, et al. 2014. Accurate information transmission through dynamic biochemical signaling networks. *Science* 346(6215):1370–73
106. Sella G, Hirsh AE. 2005. The application of statistical physics to evolutionary biology. *PNAS* 102(27):9541–46
107. Shannon CE. 1949. *The Mathematical Theory of Communication, by CE Shannon (and Recent Contributions to the Mathematical Theory of Communication), W. Weaver*. Champaign: Univ. Ill. Press
108. Shannon CE. 1956. The bandwagon. *IRE Trans. Inform. Theory* 2(1):3
109. Shaper F, Jetka T, Dittrich A. 2022. Decoding cellular communication: an information theoretic perspective on cytokine and endocrine signaling. *Curr. Opin. Endocr. Metab. Res.* 24:100351
110. Sokolowski TR, Gregor T, Bialek W, Tkačik G. 2025. Deriving a genetic regulatory network from an optimization principle. *PNAS* 122(1):e2402925121
111. Sokolowski TR, Paijmans J, Bossen L, Miedema T, Wehrens M, et al. 2019. eGFRD in all dimensions. *J. Chem. Phys.* 150(5):054108
112. Sokolowski TR, Tkačik G. 2015. Optimizing information flow in small genetic networks. IV. Spatial coupling. *Phys. Rev. E* 91(6):062710
113. Sokolowski TR, Walczak AM, Bialek W, Tkačik G. 2016. Extending the dynamic range of transcription factor action by translational regulation. *Phys. Rev. E* 93(2):022404
114. Suderman R, Bachman JA, Smith A, Sorger PK, Deeds EJ. 2017. Fundamental trade-offs between information flow in single cells and cellular populations. *PNAS* 114(22):5755–60
115. Swain PS, Elowitz MB, Siggia ED. 2002. Intrinsic and extrinsic contributions to stochasticity in gene expression. *PNAS* 99(20):12795–800
116. Swain PS, Siggia ED. 2002. The role of proofreading in signal transduction specificity. *Biophys. J.* 82(6):2928–33
117. Tang Y, Adelaja A, Ye FXF, Deeds E, Wollman R, Hoffmann A. 2021. Quantifying information accumulation encoded in the dynamics of biochemical signaling. *Nat. Commun.* 12(1):1272
118. Ten Wolde PR, Becker NB, Ouldridge TE, Mugler A. 2016. Fundamental limits to cellular sensing. *J. Stat. Phys.* 162(5):1395–424
119. Thomas PJ, Eckford AW. 2016. Capacity of a simple intercellular signal transduction channel. *IEEE Trans. Inform. Theory* 62(12):7358–82
120. Thompson MJ, Young CA, Munnamalai V, Umulis DM. 2023. Early radial positional information in the cochlea is optimized by a precise linear BMP gradient and enhanced by SOX2. *Sci. Rep.* 13(1):8567
121. Tishby N, Pereira FC, Bialek W. 2000. The information bottleneck method. arXiv:physics/0004057 [physics.data-an]
122. Tishby N, Polani D. 2010. Information theory of decisions and actions. In *Perception-Action Cycle: Models, Architectures, and Hardware*, ed. V Cutsuridis, A Hussain, JG Taylor, pp. 601–36. New York: Springer

123. Tjalma AJ, Galstyan V, Goedhart J, Slim L, Becker NB, Ten Wolde PR. 2023. Trade-offs between cost and information in cellular prediction. *PNAS* 120(41):e2303078120
124. Tkačik G, Bialek W. 2016. Information processing in living systems. *Annu. Rev. Condens. Matter Phys.* 7:89–117
125. Tkačik G, Callan CG Jr., Bialek W. 2008. Information capacity of genetic regulatory elements. *Phys. Rev. E* 78(1):011910
126. Tkačik G, Callan CG Jr., Bialek W. 2008. Information flow and optimization in transcriptional regulation. *PNAS* 105(34):12265–70
127. Tkačik G, Dubuis JO, Petkova MD, Gregor T. 2015. Positional information, positional error, and readout precision in morphogenesis: a mathematical framework. *Genetics* 199(1):39–59
128. Tkačik G, Gregor T. 2021. The many bits of positional information. *Development* 148(2):dev176065
129. Tkačik G, Gregor T, Bialek W. 2008. The role of input noise in transcriptional regulation. *PLOS ONE* 3(7):e2774
130. Tkačik G, Walczak AM. 2011. Information transmission in genetic regulatory networks: a review. *J. Phys. Condens. Matter* 23(15):153102
131. Tkačik G, Walczak AM, Bialek W. 2009. Optimizing information flow in small genetic networks. *Phys. Rev. E* 80(3):031920
132. Tkacik G, Walczak AM, Bialek W. 2012. Optimizing information flow in small genetic networks. III. A self-interacting gene. *Phys. Rev. E* 85(4 Part 1):041903
133. Tlusty T. 2008. Rate-distortion scenario for the emergence and evolution of noisy molecular codes. *Phys. Rev. Lett.* 100(4):048101
134. Topolewski P, Komorowski M. 2021. Information-theoretic analyses of cellular strategies for achieving high signaling capacity—dynamics, cross-wiring, and heterogeneity of cellular states. *Curr. Opin. Syst. Biol.* 27:100352
135. Tostevin F, De Ronde WH, Ten Wolde PR. 2012. Reliability of frequency and amplitude decoding in gene regulation. *Phys. Rev. Lett.* 108:108104
136. Tostevin F, Ten Wolde PR. 2009. Mutual information between input and output trajectories of biochemical networks. *Phys. Rev. Lett.* 102(21):218101
137. Tostevin F, Ten Wolde PR. 2010. Mutual information in time-varying biochemical systems. *Phys. Rev. E* 81(6):061917
138. Tu Y. 2013. Quantitative modeling of bacterial chemotaxis: signal amplification and accurate adaptation. *Annu. Rev. Biophys.* 42:337–59
139. Uda S. 2020. Application of information theory in systems biology. *Biophys. Rev.* 12(2):377–84
140. Uda S, Saito TH, Kudo T, Kokaji T, Tsuchiya T, et al. 2013. Robustness and compensation of information transmission of signaling pathways. *Science* 341(6145):558–61
141. Van Kampen NG. 1992. *Stochastic Processes in Physics and Chemistry*, Volume 1. Amsterdam: Elsevier
142. Voliotis M, Perrett RM, McWilliams C, McArdle CA, Bowsher CG. 2014. Information transfer by leaky, heterogeneous, protein kinase signaling systems. *PNAS* 111(3):E326–33
143. Wada T, Hironaka K-I, Kuroda S. 2021. Cell-to-cell variability serves as information not noise. *Curr. Opin. Syst. Biol.* 27:100339
144. Wada T, Hironaka Ki, Wataya M, Fujii M, Eto M, et al. 2020. Single-cell information analysis reveals that skeletal muscles incorporate cell-to-cell variability as information not noise. *Cell Rep.* 32(9):108051
145. Wagner A. 2017. Information theory, evolutionary innovations and evolvability. *Philos. Trans. R. Soc. B* 372(1735):20160416
146. Walczak AM, Tkačik G, Bialek W. 2010. Optimizing information flow in small genetic networks. II. Feed-forward interactions. *Phys. Rev. E* 81(4):041905
147. Wieschaus E, Nüsslein-Volhard C. 2016. The Heidelberg screen for pattern mutants of *Drosophila*: a personal account. *Annu. Rev. Cell Dev. Biol.* 32:1–46
148. Wolpert L. 1969. Positional information and the spatial pattern of cellular differentiation. *J. Theor. Biol.* 25(1):1–47
149. Ying T, Alexander H. 2022. Quantifying information of intracellular signaling: progress with machine learning. *Rep. Prog. Phys.* 85(8):086602

150. Zagorski M, Tabata Y, Brandenberg N, Lutolf MP, Tkačik G, et al. 2017. Decoding of position in the developing neural tube from antiparallel morphogen gradients. *Science* 356(6345):1379–83
151. Ziv E, Nemenman I, Wiggins CH. 2007. Optimal signal processing in small stochastic biochemical networks. *PLOS ONE* 2(10):e1077
152. Zoller B, Gregor T, Tkačik G. 2022. Eukaryotic gene regulation at equilibrium, or non? *Curr. Opin. Syst. Biol.* 31:100435