The price of a bit: energetic costs and the evolution of cellular signaling

Teng-Long Wang,[1] Benjamin Kuznets-Speck,[1,2] Joseph Broderick,[1] and Michael Hinczewski[1]

[1] *Department of Physics, Case Western Reserve University, Cleveland, OH 44106.*

[2] *Biophysics Graduate Group, University of California, Berkeley, CA 94720, USA.*

Recent experiments have uncovered a fundamental information scale for cellular signaling networks: the correlation between input and output concentrations of molecules in a signaling pathway corresponds to at most 1-3 bits of mutual information. Our understanding of the physical constraints and evolutionary pressures that determine this scale remains incomplete. By focusing on a basic element of signaling pathways, the kinase-phosphatase enzymatic push-pull loop, we highlight the pivotal role played by energy resources available for signaling and their expenditure: the chemical potential energy of ATP hydrolysis, and the rate of ATP consumption. Scanning a broad range of reaction parameters based on enzymatic databases, we find that ATP chemical potentials in modern organisms are just above the threshold necessary to achieve empirical mutual information values. We also derive an analytical relation for the minimum ATP consumption required to maintain a certain signal fidelity across a range of input frequencies, where we quantify fidelity either through instantaneous or time-delayed mutual information. Attempting to increase signal fidelity beyond a few bits lowers the bandwidth, the maximum characteristic signal frequency that the network can handle at a given energy cost. The observed information scale thus represents a balancing act between fidelity and the ability to process fast-changing environmental signals. Our analytical relation defines a performance limit for kinase-phosphatase networks, and we find evidence that a component of the yeast osmotic shock pathway may be close to the optimality line. By quantifying the evolutionary pressures that operate on these networks, we argue that this is not a coincidence: natural selection on energy expenditures is capable of pushing signaling systems toward optimality, particularly in unicellular organisms. Our theoretical framework is directly verifiable using existing experimental techniques, and predicts that more examples of such optimality should exist in nature.

I. INTRODUCTION

Survival for living cells depends in part on accurate and responsive signaling: the ability to collect enough information about the micro-environment to make decisions in response to external stimuli such as nutrients, hormones, and toxic agents [1]. This capacity to react to extracellular cues developed early in evolutionary history, and is now seen at all levels of biological organization, from chemotaxis in unicellular organisms [2–4] to the pathways that regulate cell differentiation and disease in multicellular life [5–8]. Despite the resulting

diversity of biochemical networks that implement this signaling, information theory provides a powerful universal framework to quantify the amount of information transferred through a network, allowing comparisons between different systems [9].

Over the last decade a remarkable experimental consensus has emerged from such comparisons: studies of both prokaryotic and eukaryotic signaling pathways have found they can transmit at most ~ 1 to 3 bits of information [10–17]. These values refer to mutual information (MI) between pathway input (concentrations of a molecule representing the signal) and output (concentrations of a downstream molecule produced by the network, sampled either at a single or multiple time points). MI is a measure of signal fidelity, representing the degree of correlation between input and output. Experiments have typically focused on a closely related quantity known as the channel capacity [18, 19]: the maximum MI achievable among all input distributions.

The consistently small channel capacities observed in cellular signaling pathways seem to indicate that cells operate with a fairly coarse representation of their surroundings: n bits of MI corresponds to being able to reliably distinguish between 2^n levels of the input, so a 1 bit pathway can only discriminate between "high" versus "low" signal concentrations. Though 1 bit is typical for MI measured at single time points, one can achieve higher MIs by focusing on output responses collected over several time points [14, 15], or by designing the experiment to isolate single-cell responses (as opposed to estimating MI from the responses of a population of cells) [17]. But these enhancements, which can push values to the 2-3 bit range, do not change the fundamental order of magnitude of the MI.

The central question we explore in this work is to what extent this fundamental information scale is shaped by the energy requirements of the underlying biochemical signaling networks. In order to transmit information, these networks necessarily need to operate out of equilibrium, fueled by processes like ATP hydrolysis that consume energetic resources. Recent research highlights these costs as an essential factor in understanding constraints on signaling [2–4, 20–23], often focusing on the ATP hydrolysis chemical potential difference $\Delta\mu = \mu_{\mathrm{ATP}} - \mu_{\mathrm{P}_i} - \mu_{\mathrm{ADP}}$ between the reactant (ATP) and products (ADP and inorganic phosphate, P_i), quantifying the free energy available to drive the system per ATP. Crossing a certain minimum threshold of $\Delta\mu$ is a prerequisite for a variety of signaling functions: accurate read-out of ligand-bound receptors [2, 3, 23], maintaining the phase coherence of oscillations in circadian clocks [20], or preserving the integrity of methylation-based "memory" to facilitate adaptation in chemotaxis [4]. This threshold is typically a few times larger (i.e. by a factor of $\sim 3-4$ [2, 23]) than the energy scale of thermal fluctuations, k_BT, where k_B is the Boltzmann constant and T the temperature. And indeed cells across the various domains of life maintain a sufficiently high $\Delta\mu \approx 21-29\ k_BT$ [24] to enable such functions.

The large value and remarkably narrow range of $\Delta\mu$ observed in modern organisms opens

up additional questions. The metabolic cycles that sustain $\Delta\mu$, constantly replenishing ATP as it is hydrolyzed, must almost necessarily have been far more inefficient and wasteful in the earliest stages of evolutionary history [25]. To what degree could organisms operating with smaller $\Delta\mu$ still process information about their environment? What kinds of evolutionary pressures might have driven $\Delta\mu$ to its modern range? And if the costs of individual signaling systems are non-trivial [2, 4], could natural selection have driven these networks toward optimized, energy-efficient solutions?

To investigate these issues, we focus on one of the canonical signaling circuits in biology, the kinase-phosphatase "push-pull loop", which often forms a basic unit of more complicated signaling cascades [26–29]. An active kinase enzyme instigates the "push", chemically modifying a substrate protein via phosphorylation (consuming ATP in the process), while a phosphatase enzyme provides the "pull", dephosphorylating the modified substrate, reverting it to its original state. We derive the relationships between three facets of the system: i) the MI between the input (active kinase) and output (phosphorylated substrate) molecular populations; ii) the timescales over which the input signal varies; and iii) the energy requirements, expressed in terms of $\Delta\mu$ and the rate of ATP consumption. Exploring the entire spectrum of kinase/phosphatase enzymatic parameters from bioinformatic databases, we find that physiological $\Delta\mu$ values are just large enough to enable a MI of 1-2 bits for the widest possible parameter range. However to achieve this MI for signals that vary rapidly in time becomes more challenging, requiring both precise fine-tuning of parameters and a certain minimum rate of ATP consumption. In fact, taking advantage of results from optimal noise filter theory [30, 31], we derive a remarkably simple analytical relationship that describes the tradeoffs between minimum ATP rate, the MI, and the maximum characteristic signal frequency (the so-called bandwidth) which the push-pull network can handle. Verified via extensive numerical simulations across the whole gamut of enzymatic parameters, this relation is a novel theoretical prediction that can be directly tested in future experiments. We also demonstrate that analogous constraints hold even in the general case of time-delayed MI, where we look at correlation between the output and input signals separated by a time offset. The analytical relation rationalizes the observed range of MI by showing that values much higher than 1-2 bits would require sacrificing the ability to process fast-changing signals. Finally we explore the question of whether there exist evolutionary pressures that would push such a system to be energy efficient, optimizing the ATP consumption for a given target MI and bandwidth. Using a recently developed formalism relating metabolic costs to the strength of natural selection [32, 33], we show that these pressures can indeed be significant, particularly for single-celled organisms. We highlight a kinase-phosphatase loop in the yeast Hog1 signaling pathway as a system that may have been optimized by such

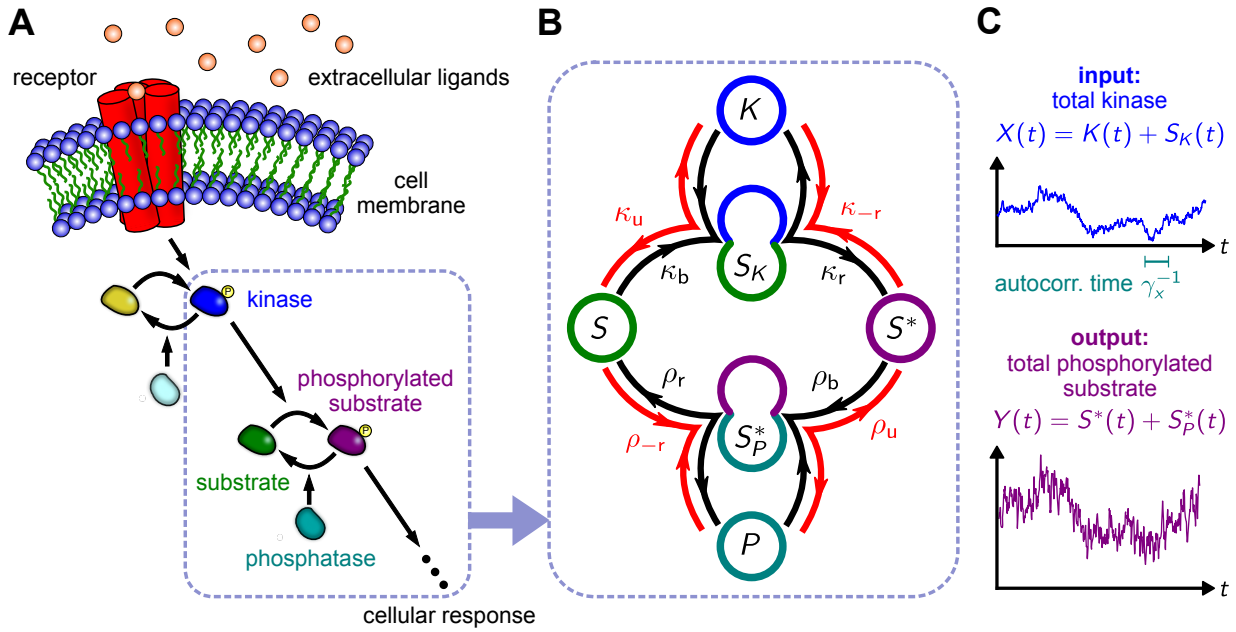

FIG. 1. **(A)** A schematic signaling pathway involving cascades of kinase phosphorylation, initiated by a receptor embedded in the cell membrane that responds to extracellular ligands. The system we focus on will be one stage of the pathway, a kinase-phosphatase push-pull loop, highlighted in the dashed box. **(B)** The molecular species and reaction parameters of the push-pull loop. The kinase (K) binds to the substrate (S), forming the complex (S_K) that catalyzes the production of phosphorylated substrate (S^*). Phosphatase (P) binds to S^*, forming a complex (S_P^*) that catalyzes the dephosporylation of the substrate. Forward reaction / binding rates are labeled in black, while reverse reaction / unbinding rates are in red. **(C)** The loop serves to transduce an input signal, defined as the total population of kinase (bound or unbound), $X(t) = K(t) + S_K(t)$, into an output, defined as the total population of phosphorylated substrate, $Y(t) = S^*(t) + S_P^*(t)$. The input signal has a characteristic autocorrelation time γ_x^{-1}.

pressures.

II. THEORY

A. Modeling an enzymatic push-pull loop

This push-pull network consists of two opposing reactions: a kinase enzyme instigates the "push", chemically modifying a substrate protein via phosphorylation, while a phosphatase enzyme provides the "pull", dephosphorylating the modified substrate, reverting it to its original state [26–29]. Since a single kinase can catalyze the phosphorylation of many substrate proteins, this loop can effectively act like an amplifier [28], translating a weaker signal (a small cellular population of an active kinase) into a stronger one (a large population of a phosphorylated substrate). Often the substrate itself is a kinase that can exist in catalyti-

cally inactive and active states, with activation triggered by phosphorylation. In this case one can have multi-tiered signaling cascades enhancing the amplification (as shown schematically in Fig. 1A) with the active substrate produced by one loop serving as the kinase for a downstream loop [34]. More complex signaling networks are also possible, with multiple cascades connected by crosstalk through shared components [35], feedback from downstream to upstream populations [34], or activation requiring multisite phosphorylation [36]. However, the starting point for understanding any of these more complex signaling topologies is the behavior of a single loop, with a substrate activated / deactivated through a single phosphorylation site.

The reaction scheme of a single push-pull loop is shown in Fig. 1B. Binding of free kinase (population $K(t)$ at time t) to substrate (population $S(t)$) occurs with rate constant κ_b, forming a kinase-substrate complex (population $S_K(t)$). Phosphorylation of the substrate and its subsequent release constitutes the catalytic step, with rate κ_r, yielding free phosphorylated substrates (population $S^*(t)$). A phosphatase can subsequently bind, with rate ρ_b, forming a phosphatase-substrate complex (population $S_P^*(t)$), and catalyzing the dephosphorylation / release of the substrate with rate ρ_r. These reactions also can occur in reverse: kinase-substrate unbinding (rate κ_u), reverse kinase catalysis (rate κ_{-r}), phosphatase-substrate unbinding (rate ρ_u) and reverse phosphatase catalysis (rate ρ_{-r}). Under physiological conditions some of these reverse rates may be negligible compared to their forward counterparts, but accounting for them is crucial to enforce thermodynamic consistency. In fact the product of the ratios of the reverse rates relative to the forward ones must satisfy a key thermodynamic relation arising from the principle of local detailed balance (closely related to the Haldane relation for enzymes) [37, 38],

$$\frac{\kappa_{-r}\rho_u\rho_{-r}\kappa_u}{\kappa_r\rho_b\rho_r\kappa_b} = e^{-\beta\Delta\mu}. \tag{1}$$

This relation is derived in the Supplementary Information (SI), and reflects the fact that for every complete traversal of the loop along the forward direction (clockwise along the black arrows in Fig. 1B) a single ATP molecule is removed from the environment, hydrolyzed, and the products ADP and inorganic phosphate $\mathrm{P_i}$ released back into the surroundings. $\Delta\mu$ depends on the concentrations [ATP], [ADP], and [$\mathrm{P_i}$] through $\Delta\mu = \Delta\mu_0 + k_BT\ln([\mathrm{ATP}](1\,\mathrm{M})/([\mathrm{ADP}][\mathrm{P_i}]))$, where $\Delta\mu_0$ is the standard free energy of ATP hydrolysis ($\Delta\mu_0 \approx 12\ k_BT$ at room temperature [24]). Living systems expend energetic resources to maintain an imbalance of [ATP] relative to [ADP] and [$\mathrm{P_i}$], making $\Delta\mu$ in physiological conditions larger than $\Delta\mu_0$. Despite the wide variety of metabolic pathways used to achieve this, measured $\Delta\mu$ values in organisms from *E. coli* to humans lie within a relatively narrow range, $\Delta\mu \approx 21-29\ k_BT$ [24]. This means reverse rates are sufficiently slow that the

numerator in Eq. (1) is 9-12 orders of magnitude smaller than the denominator. One of the questions we tackle below is the significance of this disparity for transmitting information through the loop.

To quantitatively measure this information transfer, it is useful to explicitly describe the network behavior in terms of transducing an input signal into an amplified output, with degradation of the signal due to the stochastic nature of the reactions that mediate this process. We take the time-dependent input $X(t) = K(t) + S_K(t)$ to be the population of active kinases (both free and substrate-bound), and the corresponding output signal $Y(t) = S^*(t) + S_P^*(t)$ as the population of phosphorylated substrates (free and phosphatase-bound). For any specific system, the input kinases would be activated through a particular upstream signaling network. Here, however, we are interested in a more general problem: what is the effectiveness of this loop in processing a variety of possible input signals, spanning different amplitudes and timescales. The simplest mechanism that allows us to tune the dynamical characteristics of the input is to imagine the kinases activated at a constant rate F and deactivated at a constant rate γ_K. We focus on the long-time limit where a stationary state has been achieved, and so F allows us to regulate the amplitude of the input signal while γ_K controls the autocorrelation time of the input fluctuations. While the analysis below could be done for other, system-specific models of the input, our choice allows us to explore a broad range of possible inputs to establish general bounds on information processing through the loop. With this input model, the reaction network model is fully specified. For a given set of parameters (drawn from distributions based on kinase/phosphatase biochemical information collected in enzymatic databases, as described below) we can derive analytical results for dynamical quantities using the linearized chemical Langevin approximation [39]. As shown in the SI, this provides excellent agreement with the exact kinetic Monte Carlo [40] simulation results in the parameter ranges of interest.

In focusing on how $X(t)$ is transduced to $Y(t)$, we frame our analysis in terms of three properties of the system. The first is the autocorrelation time of the input, γ_x^{-1}, defined through $\overline{\delta X(t+\tau)\delta X(t)} = \overline{\delta X^2}\exp(-\gamma_x|\tau|)$, where the bar denotes an average over an ensemble of trajectories in the stationary state and $\delta X(t) \equiv X(t) - \overline{X}$. Note that instantaneous averages like $\overline{X} \equiv \overline{X(t)}$ and $\overline{\delta X^2} \equiv \overline{\delta X^2(t)}$ are independent of t in the stationary state. γ_x^{-1} is the characteristic timescale of the input fluctuations, and we will denote its inverse, γ_x, as the effective "frequency" of the input. The second property is related to the mean rate at which phosphorylated substrates are produced through the catalytic reaction step, $\kappa_r\overline{S_K}$, relative to the mean total number of activated kinases $\overline{X}$. We define the gain parameter $R_0 \equiv \kappa_r\overline{S_K}/\overline{X}$ as a measure of the production of output for a given input level. Both γ_x and R_0 can be expressed, to a good approximation, in terms of the reaction rates as follows

(see SI for derivation):

$$\gamma_x = \frac{C_1}{C_1 + C_2}\gamma_K, \quad R_0 = \frac{C_2}{C_1 + C_2}\kappa_r, \tag{2}$$

where $C_1 \equiv \kappa_- \overline{P}\gamma_K \rho_b \rho_r + F\kappa_{-r}\kappa_u \rho_-$, $C_2 \equiv \overline{S}[F\kappa_b\kappa_{-r}\rho_- + \overline{P}\gamma_K(\kappa_b\rho_b\rho_r + \kappa_{-r}\rho_{-r}\rho_u)]$. Here $\kappa_- \equiv \kappa_u + \kappa_r$, $\rho_- \equiv \rho_u + \rho_r$. Note the dependence on mean unmodified substrate $\overline{S}$ and free phosphatase $\overline{P}$ populations: these two numbers are free parameters that (along with the reaction rates) determine the network dynamics.

The final property of interest is the MI between input and output. Unless otherwise noted, we focus on the instantaneous stationary MI between $X(t)$ and $Y(t)$ at the same time point t, which we denote as $\mathcal{I}$. However in Sec. III C we also consider the MI between $X(t)$ and $Y(t+\alpha)$, which we will denote with a subscript, $\mathcal{I}_\alpha$, where α is the time offset. This allows us to investigate whether including a time delay $\alpha > 0$ (compensating for the finite propagation time of the signal) substantially changes the observed input-output correlation. For the instantaneous ($\alpha = 0$) case, the MI is defined in terms of the joint probability $P(X, Y)$ of observing input value X and output value Y at the same moment of time, and the corresponding marginal probabilities $P(X)$ and $P(Y)$,

$$\mathcal{I} = \sum_{X,Y} P(X, Y) \log_2 \frac{P(X, Y)}{P(X)P(Y)}. \tag{3}$$

The value of $\mathcal{I}$ is non-negative in all cases, and is measured in bits, with larger values translating to a greater degree of correlation between input and output. For our parameter ranges, $P(X, Y)$ can be approximated as a bivariate Gaussian, and so we use an expression for $\mathcal{I}$ valid in this limit that is more convenient to evaluate [19]:

$$\mathcal{I} \approx -\frac{1}{2}\log_2 E, \quad \text{where} \quad E \equiv 1 - \frac{(\overline{XY} - \overline{X}\,\overline{Y})^2}{\left(\overline{X^2} - \overline{X}^2\right)\left(\overline{Y^2} - \overline{Y}^2\right)}. \tag{4}$$

Here $E = 1 - \rho^2$, where ρ is the Pearson correlation coefficient, and hence lies in the range $0 \leq E \leq 1$. For $E = 0$ (or equivalently $\mathcal{I} = \infty$) we have perfect correlation between the input and output signal, while $E = 1$ ($\mathcal{I} = 0$) corresponds to an output that is completely independent of the input. For the $\alpha \neq 0$ case, the expressions for $\mathcal{I}_\alpha$ would be analogous, except we use $P_\alpha(X, Y)$, the joint probability of observing X at time t and Y at time $t + \alpha$. All averages, for example the cross-correlation $\overline{XY}$, are then calculated with respect to $P_\alpha(X, Y)$.

B. Determining the enzymatic parameter range

Once the input signal is specified through F and γ_K, there are ten parameters related to the kinase, phosphatase, and substrate that determine the observables of interest γ_x, R_0, and $\mathcal{I}$ discussed above. These parameters are: κ_b, κ_u, κ_r, κ_{-r}, ρ_b, ρ_u, ρ_r, ρ_{-r}, $\bar{S}$, $\bar{P}$. We know from surveys of enzymatic parameters that each of these quantities can span several orders of magnitude among different systems, often with an approximately log-normal distribution [41, 42]. To understand the performance limits of enzymatic loops in general, it makes sense to explore the entire range of biologically realistic parameters, rather than focus on a single choice of parameters. Existing online databases are excellent resources for this purpose, and Fig. 2 shows the resulting histograms of kinase / phosphatase parameters (full extraction details are available in the SI). For the substrate protein (which we take as a kinase) and the phosphatase, the concentrations $[S]$ and $[P]$ in Fig. 2A are derived from the PaxDb protein abundance database [43], using UnitProt gene ontology associations to identify kinases and phosphatases [44]. Enzymatic reaction parameters are available in the Sabio-RK database [45]. The reaction rates κ_r and ρ_r (Fig. 2D) are typically listed directly, but the others are most often in specific combinations: the Michaelis constants $K_M^{\text{kin}} = (\kappa_r + \kappa_u)/\kappa_b$, $K_M^{\text{pho}} = (\rho_r + \rho_u)/\rho_b$ for kinase/phosphatase respectively (Fig. 2B) and the specificity ratios $\kappa_r/K_M^{\text{kin}}$, ρ_r/K_M^{pho} (Fig. 2C). For all of these parameters there is a paucity of data on phosphatases relative to kinases, but the phosphatase ranges seem to largely overlap with those of kinases. Thus for simplicity we take kinase and phosphatase parameters to have the same distributions (log-normal) and use a numerical fitting procedure to find an overall log-normal joint probability distribution for the eight underlying model parameters represented in the data: κ_b, κ_u, κ_r, ρ_b, ρ_u, ρ_r, $\bar{S}$, $\bar{P}$ (see SI). Note that data in concentration units (like $[S]$ and $[P]$ in molars) is converted to mean abundance ($\bar{S}$ and $\bar{P}$) by assuming a volume of 30 fL (comparable to the cytoplasmic volume of yeast [24, 46]). This procedure is designed so that the resulting joint distribution yields marginal probability densities (solid curves in Fig. 2) that exhibit good agreement with the histogram data for any of the measured parameter combinations. Despite this agreement, we note that the joint distribution likely spans a portion of the parameter space larger than the true distribution of biological values: this is because it cannot fully capture correlations between different parameters. (Such correlations are difficult to reconstruct since many database entries are incomplete, containing some but not all of the enzymatic parameters.) For our purposes, having a distribution that effectively acts like a superset of the biological distribution is fine: whatever performance bounds we infer from the whole distribution will then also apply to the subset of the distribution that corresponds to current real-world systems. Moreover this also allows us to explore a larger enzymatic design space, which may have been accessible

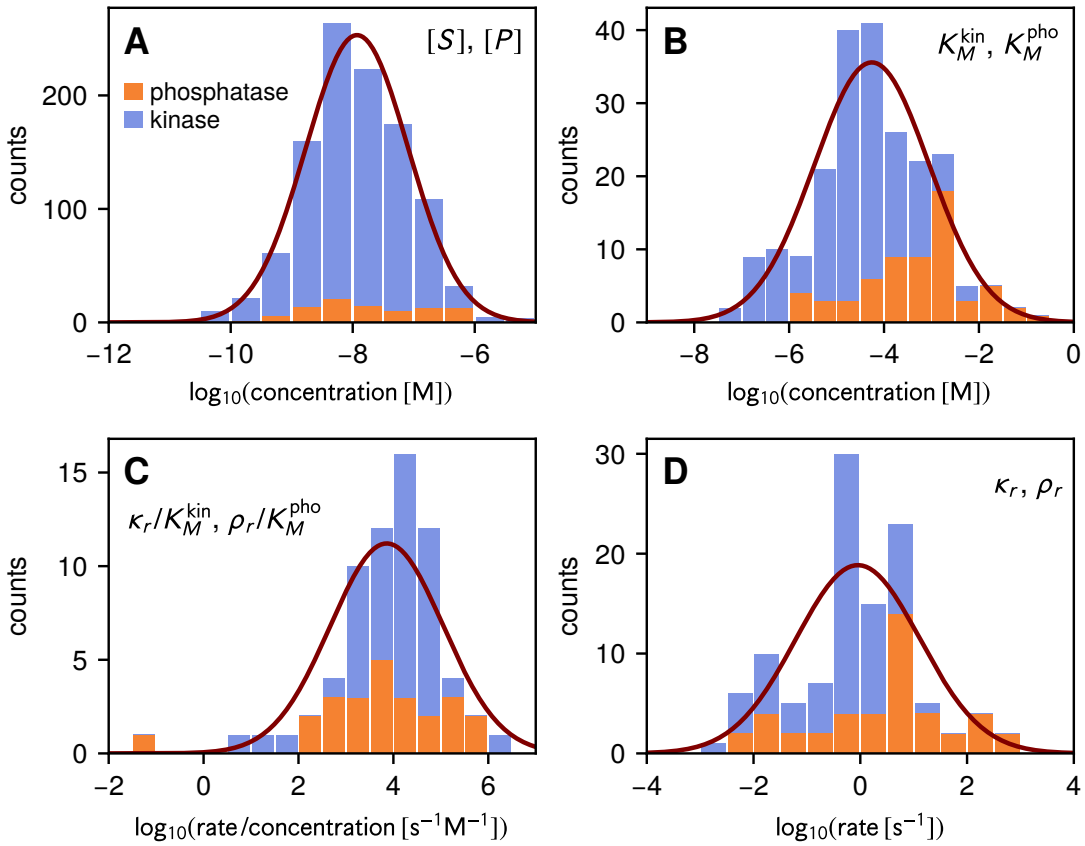

FIG. 2. Enzymatic parameter ranges for kinases/phosphatases based on the PaxDb [43] and Sabio-RK [45] databases. Because of the relative lack of phosphatase data (orange histograms) relative to kinases (blue histograms), we fit an overall log-normal joint probability to the total data set including both kinases and phosphatases. The marginal distributions from that global fit are plotted as purple curves. The parameters are as follows: **(A)** kinase substrate $[S]$ and phosphatase $[P]$ concentrations; **(B)** kinase/phosphatase Michaelis constants K_M^{kin}, K_M^{pho}; **(C)** the corresponding specificity ratios $\kappa_r/K_M^{\text{kin}}$, ρ_r/K_M^{pho}; **(D)** kinase/phosphatase catalytic rates κ_r and ρ_r.

at earlier points in evolutionary history.

Two of the model parameters are still unaccounted for: the reverse reaction rates κ_{-r} and ρ_{-r}. Though usually small in magnitude and typically not measured in enzyme kinetic assays, we also know that they are crucially related to $\Delta\mu$ through the local detailed balance relation of Eq. (1). Thus, as explained in the next section, these become important free parameters that we can vary to explore signaling efficiency and its dependence on $\Delta\mu$.

III. RESULTS

A. Minimum cost of transmitting information

Given the model described above, with a parameter set drawn at random from the empirical joint distribution, we can ask a basic first question: what is the minimum chemical potential difference $\Delta\mu$ required to achieve a certain mutual information $\mathcal{I}$? The answer will depend on the nature of the input signal $X(t)$, and thus we would like to test different effective input frequencies γ_x. To do this we will fix the mean free kinase concentration at the level of a low amplitude input, $[K] = 5$ nM, and vary γ_K, which varies γ_x according to Eq. (2) with $F = \gamma_K \bar{K}$ for fixed $\bar{K}$. In the SI we also show the same analysis for $[K] = 0.5$ and 50 nM, with results qualitatively similar to those described below. After drawing enzyme parameters from the joint distribution and specifying γ_x at a given $[K]$, the only two free parameters are the reverse reaction rates κ_{-r} and ρ_{-r}.

Fig. 3A shows a contour diagram of $\mathcal{I}$ as a function of κ_{-r} and ρ_{-r} for a sample enzyme parameter set and value of γ_x. Superimposed are dotted lines of constant $\Delta\mu$ from Eq. (1). If one were interested in achieving a particular $\mathcal{I}$ value, for example $\mathcal{I} = 1$ bit, one can then numerically determine the κ_{-r} and ρ_{-r} point along the $\mathcal{I} = 1$ bit contour where $\Delta\mu$ is smallest. For this specific enzyme parameter set and γ_x, the value turns out to be $\Delta\mu = 6.72$ $k_B T$, which would then be recorded as the minimum necessary $\Delta\mu$ to achieve 1 bit of MI. Note that it is not guaranteed that a minimum $\Delta\mu$ solution exists for every parameter set sampled from the joint distribution. If the $\mathcal{I}$ contours plateau at a maximum less than 1 bit, no possible $\Delta\mu$ will allow that particular system to achieve the desired MI target. We will return to this important point below.

If one keeps the enzyme parameters (other than κ_{-r} and ρ_{-r}) fixed, and just varies γ_x, an interesting trend appears in the minimum $\Delta\mu$ results. Fig. 3B shows two examples of minimum $\Delta\mu$ curves, for target $\mathcal{I}$ values of 1 and 2 bits respectively. For a given $\mathcal{I}$ target, the minimum $\Delta\mu$ is nearly constant at low input frequencies, but then increases rapidly and diverges at a maximum frequency which we will dub the "bandwidth" of the system. This intuitively makes sense: the higher the input frequency, the more rapid the catalytic reaction rates needed to accurately transmit the signal through the system, increasing the required $\Delta\mu$ threshold. However, there is an inherent limit, given finite enzyme catalysis rates. Above the bandwidth, whose value depends on the enzyme parameters, the system can no longer achieve the target $\mathcal{I}$. The higher the informational burden (i.e. increasing the target $\mathcal{I}$ from 1 to 2 bits) the lower the bandwidth: if one desires higher fidelity transmission, the range of transmissible signal frequencies will suffer.

To make more sense of these results, it is useful to look at a broad sample of enzyme

parameters rather than a single set. To visualize global behaviors, we will calculate two numerical results for each set drawn from our joint distribution. The procedure is as follows: i) Sample an enzyme parameter set from the distribution; ii) Determine if it can achieve our target $\mathcal{I}$ for any input frequency; iii) If the answer is yes, vary γ_K until one finds the maximum possible value $\gamma_K^{\max}$ where one can still achieve the $\mathcal{I}$ target. iv) Calculate the minimum $\Delta\mu$ for an input signal very near the bandwidth frequency, where $\gamma_K = 0.95\gamma_K^{\max}$. We will call this result $\Delta\mu^{\text{high}}$. The corresponding input frequency is γ_x^{high}. v) Analogously, calculate the minimum $\Delta\mu$ for an input signal with a frequency much lower than the bandwidth, where $\gamma_K = 0.01\gamma_K^{\max}$. This set of results we denote as $\Delta\mu^{\text{low}}$ and γ_x^{low}. Fig. 3B shows the two points $(\gamma_x^{\text{high}}, \Delta\mu^{\text{high}})$ and $(\gamma_x^{\text{low}}, \Delta\mu^{\text{low}})$ as blue and green dots respectively for that particular parameter set at $\mathcal{I} = 1$ bit. These two points encapsulate several key features of the minimum $\Delta\mu$ versus γ_x curve: $\Delta\mu^{\text{low}}$ roughly corresponds to an "entry level" price, the minimum ATP hydrolysis chemical potential necessary to transmit the signal at any frequency, while the difference $\Delta\mu^{\text{high}} - \Delta\mu^{\text{low}}$ is the premium one has to pay to transmit signals near the highest possible frequencies. The value γ_x^{high} approximately corresponds to the bandwidth.

If one were to make numerous draws from the parameter distribution, and plot $(\gamma_x^{\text{high}}, \Delta\mu^{\text{high}})$ and $(\gamma_x^{\text{low}}, \Delta\mu^{\text{low}})$ for each draw, one would get a cloud of blue and green dots. These are shown in Fig. 3D-F for target $\mathcal{I}$ of 1, 1.5, and 2 bits respectively. As mentioned above, not every draw will lead to a parameter set that can achieve the target, and the plots are labeled by the fraction of draws that are capable of reaching that particular value of $\mathcal{I}$. That fraction decreases with $\mathcal{I}$, from 13% for $\mathcal{I} = 1$ bit down to only 2% for $\mathcal{I} = 2$ bits. As $\mathcal{I}$ increases not only does it become progressively more difficult to find enzymatic parameters compatible with higher fidelity, but the accessible frequency range becomes more restricted. Fig. 3C shows the percentage of the parameter space that can achieve bandwidths higher than a given frequency for different $\mathcal{I}$. For example let us consider the frequency 1.22×10^{-3} s^{-1}, which is the $\mathcal{I} = 1$ bit bandwidth for the yeast Pbs2/Hog1 system described in detail in Sec. IIIC. (This system is part of the Hog1 osmotic stress response pathway, whose overall bandwidth has been experimentally estimated to be of a similar scale [47]). From Fig. 3C it is evident that only about 0.41% of the draws from the parameter distribution have $\gamma_x^{\text{high}} \geq 1.22 \times 10^{-3}$ s^{-1} for a target $\mathcal{I} = 1$ bit. If one were to attempt to transmit signals at such high frequencies for $\mathcal{I} = 2$ bits, the fraction of compatible parameter space shrinks to a miniscule 9×10^{-3}%. This reflects the exquisite fine-tuning required to put together a set of enzymatic loops capable of responding to quick, life-or-death variations of the external environment on time scales of a couple of minutes. Going much beyond $\mathcal{I} = 1$ bit and maintaining fast response times for a single push-pull loop is extremely difficult, and hence it makes sense that biology settles for $\mathcal{I}$ in the vicinity of 1 bit in many circumstances.

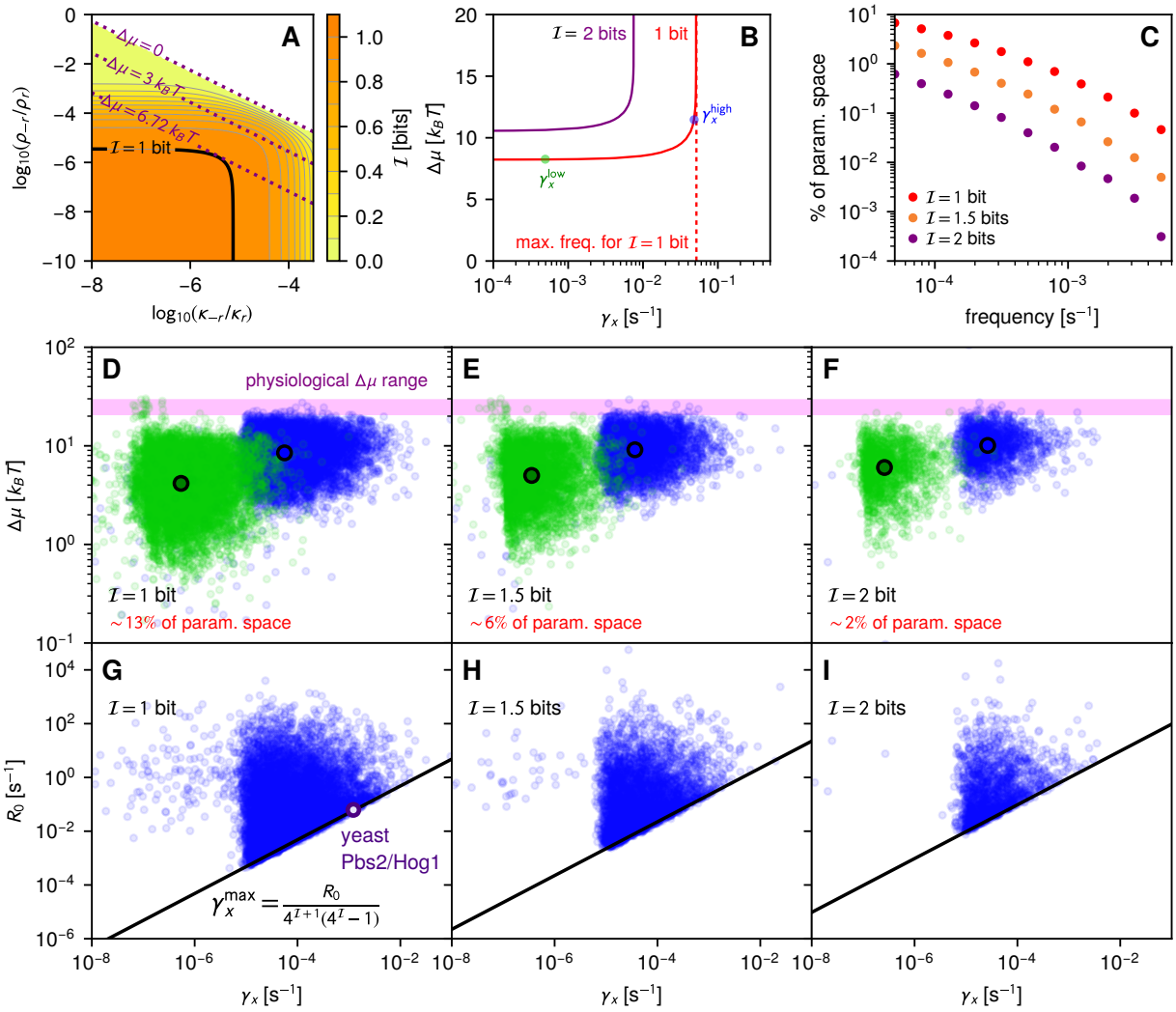

FIG. 3. **(A)** A representative contour diagram of $\mathcal{I}$ (solid curves) as a function of κ_{-r} and ρ_{-r} for a parameter set drawn randomly from the joint distribution. Dotted lines denote contours of constant $\Delta\mu$. In this case $\Delta\mu = 6.72\ k_BT$ is the smallest value at which the system can achieve $\mathcal{I} = 1$ bit. **(B)** For a sample parameter set, the minimum $\Delta\mu$ needed to achieve $\mathcal{I} = 1$, 2 bits as a function of input frequency γ_x. For the 1 bit case, the dashed line represents γ_x^{high}, the maximum γ_x compatible with $\mathcal{I} = 1$ bit for this parameter set. As described in the text, we highlight two points along the curve: one at a frequency γ_x^{high} at roughly 95% of the bandwidth, and the other at frequency γ_x^{low} at roughly 1% of the bandwidth. The points will be plotted for a many random draws of the enzyme parameters from the joint distribution in the lower panels of the figue. **(C)** For each target value of $\mathcal{I} = 1$, 1.5, 2 bits, the percentage probability of randomly drawing a parameter set that has a γ_x^{high} higher than a given frequency. **(D-F)** The distribution of γ_x^{high} (blue) and γ_x^{low} (green) for many random parameter draws, keeping only those that can achieve $\mathcal{I} = 1$ bit (D), 1.5 bits (E), or 2 bits (F). The probabilities of successfully drawing such a set are shown in red in each panel. The blue and green circles denote the median of each distribution respectively. **(G-I)** The same γ_x^{high} distributions as in panels (D-F), except plotted in terms of gain R_0 on the vertical axis. The solid line is the analytical maximum bandwidth bound $\gamma_x^{\max}$ of Eq. (5). The purple circle in panel G shows the estimated result for the near-optimal yeast Pbs2/Hog1 system.

Going much below 1 bit poses another set of difficulties, since such systems would not even be able to reliably transmit the difference between high and low values of input signal. For signaling that can occur over longer timescales (hours instead of minutes) it becomes much easier to find compatible parameter sets, with the median of the distribution of γ_x^{high} for $\mathcal{I} = 1$ bit around $\sim 6 \times 10^{-5}$ s^{-1}.

From the perspective of costs, the bulk of the distribution of entry level prices $\Delta\mu^{\text{low}}$ for $\mathcal{I} = 1$ bit is $\gtrsim 1\ k_BT$. Any system much below this would be too close to equilibrium (reverse rates comparable to forward rates) for effective information transfer to occur. The median of the $\Delta\mu^{\text{low}}$ distribution in Fig. 3C is 4 k_BT, increasing to about 6 k_BT for $\mathcal{I} = 2$ bits in Fig. 3E. These values are on the same scale as estimates of minimum $\Delta\mu \sim 4$ $k_BT \ln 2$ required for 99% accurate readout of a ligand-bound receptor via the activation of a downstream molecule, assuming an arbitrarily slow readout process [23]. In that system (as in ours), processing information at faster time scales requires large $\Delta\mu$. Indeed we find that the median values for $\Delta\mu^{\text{high}}$ range between $8 - 10\ k_BT$ for $\mathcal{I} = 1 - 2$ bits. The minimum $\Delta\mu$ near the bandwidth is typically shifted up by about 4 k_BT, reflecting the premium necessary to transmit near the frequency limit. Paying this premium is worthwhile: frequencies γ_x^{low} accessible at $\Delta\mu^{\text{low}}$ prices are likely far too low to have biological relevance, with the distributions of γ_x^{low} largely below 10^{-5} s^{-1}. To get the ability to respond to signals at more biologically reasonable time scales thus means being capable of transmitting closer to the bandwidth, making $\Delta\mu^{\text{high}}$ a more useful measure of minimum biological costs.

The $\Delta\mu^{\text{high}}$ distributions show that it is possible to have signaling systems that transmit at least 1 bit of MI and operate at $\Delta\mu$ lower than the current physiological range ($\Delta\mu \approx 21-29$ k_BT [24], indicated in pink in Fig. 3D-F). This is true even for systems with the fastest responses (large γ_x^{high} near the right edges of the distribution). This means that one can imagine enzymatic signaling systems in the earliest stages of evolutionary history that can reliably distinguish high and low inputs even before ATP metabolism (maintaining high ATP concentrations relative to ADP and P_i) reached its modern levels of efficiency.

In fact a fascinating universal feature of the distributions is that the physiological $\Delta\mu$ range lies just above the top edge of the distributions. Naively it would seem as if the physiological values are just high enough to allow these signaling loops to transmit $\mathcal{I} = 1-2$ bits across the broadest possible parameter subset. This gives evolution the largest possible space in which to tweak tradeoffs between fidelity and response times without running into chemical potential limitations. Of course $\Delta\mu$ influences not just signaling networks but the entire range of cellular functions, so it is impossible to say with certainty what factors played the largest role in determining the values of $\Delta\mu$ we see in present-day organisms. But at least from the perspective of signaling at the level of a push-pull loop, it is clear that $\Delta\mu \approx 21-29\ k_BT$ is more than good enough for basic information transfer needs, and there

would be no benefit in having a system with substantially higher $\Delta\mu$. To maintain $\Delta\mu = 40$ or 50 $k_B T$ for example, would require significant additional metabolic resources, with little payoff in terms of either $\mathcal{I}$ or bandwidth.

B. Analytical bound describes tradeoff between bandwidth and information

The results above already illustrated the tradeoff between bandwidth and MI, with parameter sets that achieve very large γ_x^{high} becoming progressively harder to find as the target $\mathcal{I}$ increases. Can we understand this relationship in more detail? For this purpose we take advantage of optimal noise filter theories, originally developed in the context of signal processing [48–50], and in recent years applied to a variety of biological signaling networks [30, 31, 51–54]. The original motivation involved designing a filter for a signal corrupted by noise, such that the output matched the uncorrupted input signal as closely as possible. In the biological context, this same framework allows us to put bounds on the maximum MI achievable between input and output signals for given input and enzymatic parameters. As shown in the SI, our enzymatic push-pull loop can be approximately mapped onto an effective two-species input-output system, which is then amenable to analytical treatment using the Wiener-Kolmogorov optimal filter theory [30, 48–50].

The end result is a remarkably simple analytical relation between the maximum possible bandwidth $\gamma_x^{\max}$ achievable given a target value of $\mathcal{I}$,

$$\gamma_x^{\max} = \frac{R_0}{4^{\mathcal{I}+1}(4^{\mathcal{I}} - 1)}. \tag{5}$$

The only other enzymatic parameter that appears in the relation is the gain R_0, a measure of output production relative to the input. Fig. 3G-I shows the same parameter set distribution as the $(\gamma_x^{\text{high}}, \Delta\mu^{\text{high}})$ points in Fig. 3D-F, except replotted in terms of $(\gamma_x^{\text{high}}, R_0)$, where R_0 is the gain for each parameter set. The solid line is the bound of Eq. (5). Even though this bound is based on an approximation of the full enzymatic system, and hence is not guaranteed to be exact, it still provides an excellent cutoff for the distribution of $(\gamma_x^{\text{high}}, R_0)$ points. For systems at a certain R_0, we see that as $\mathcal{I}$ is increased and the denominator in Eq. (5) gets larger, the maximum bandwidth $\gamma_x^{\max}$ shifts to lower values. If we are interested in a fast response time, increasing $\mathcal{I}$ systematically reduces the compatible parameter space, since we are forced to rely on cases with larger and larger R_0. Thus Eq. (5) rationalizes the earlier observation of limited options for networks that can simultaneously respond to signals fluctuating on minute time scales and achieve $\mathcal{I}$ significantly larger than 1 bit.

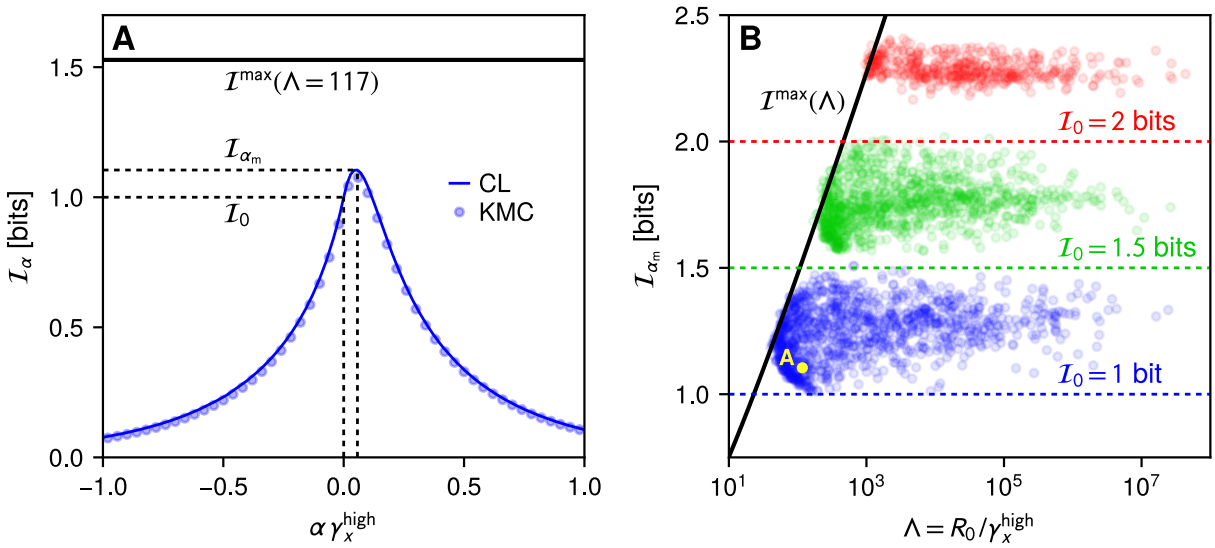

FIG. 4. **(A)** $\mathcal{I}_\alpha$, the MI between input and output with a time offset α, for a sample parameter set where the system achieves instantaneous MI $\mathcal{I}_0 = 1$ bit. The maximum $\mathcal{I}_{\alpha_m}$ is achieved at a finite time delay $\alpha_m > 0$. Results from the chemical Langevin approximation (CL) agree with kinetic Monte Carlo simulations (KMC). The analytical bound $\mathcal{I}^{\max}(\Lambda)$ from Eq. (6) is shown as a horizontal line, with $\Lambda = 117$ for this parameter set. **(B)** The three sets of points correspond to the parameter sets distributions used in Fig. 3G-I, with $\mathcal{I}_0 = 1$ bit (blue), 1.5 bits (green), and 2 bits (red). For each parameter set, the maximum time-delayed MI $\mathcal{I}_{\alpha_m}$ is plotted versus $\Lambda = R_0/\gamma_x^{\text{high}}$. The bound $\mathcal{I}^{\max}(\Lambda)$, based on the approximate two species optimal filter theory, is shown as a black curve.

C. Time-delayed mutual information

Up to now we have focused on the instantaneous MI $\mathcal{I}$ between input and output, which is constrained not just by noise in signal transduction, but also by the fact that changes in the input take a finite time to propagate into the output. We can compensate for the finite propagation speed by using the time-delayed MI $\mathcal{I}_\alpha$, which describes the correlation between the input $X(t)$ and the output $Y(t+\alpha)$ at a later time [55, 56]. Indeed some of the gains in MI observed in experiments collecting outputs over multiple time points [14, 15] reflect the fact that later output time points can have higher correlations with the current input. Fig. 4A shows the typical behavior of $\mathcal{I}_\alpha$ versus α, for a parameter set that achieves instantaneous MI $\mathcal{I}_0 = 1$ bit at $\alpha = 0$, with $\Lambda = R_0/\gamma_x^{\text{high}} = 117$. The MI reaches a maximum value $\mathcal{I}_{\alpha_m} = 1.1$ bits at time delay $\alpha_m > 0$, which is usually a fraction of the characteristic time scale $1/\gamma_x^{\text{high}}$. The results for the chemical Langevin approach (solid curve) are consistent with kinetic Monte Carlo simulations (circles). For completeness we also show the behavior of $\mathcal{I}_\alpha$ for $\alpha < 0$, which describes the extent to which the output can anticipate future input signals. As expected, $\mathcal{I}_\alpha < \mathcal{I}_0$ for $\alpha < 0$, and $\mathcal{I}_\alpha$ decreases as the time offset $|\alpha|$ becomes

larger, reflecting the difficulty of forecasting the input as one looks further into the future.

The approximate two species Wiener-Kolmogorov optimal filter theory described in the previous section easily generalizes to incorporate time delay [54] (details in the SI). It yields an analytical bound $I^{\max}(\Lambda)$ on the MI regardless of time offset α:

$$\mathcal{I}_\alpha \leq \mathcal{I}^{\max}(\Lambda) = \frac{\ln\left[\frac{1}{2}\left(1+\sqrt{1+\Lambda/(\ln 2)^2}\right)\right]}{2\ln 2}. \tag{6}$$

The value $I^{\max}(\Lambda) = 1.53$ for $\Lambda = 117$ is shown as a horizontal line in Fig. 4A, so for this parameter set the actual maximum $\mathcal{I}_{\alpha_m}$ is not near the bound.

However there are systems that can achieve $I^{\max}(\Lambda)$. Fig. 4B shows parameter sets from Fig. 3G (blue), 3H (green), and 3I (red), corresponding to systems with instantaneous MI values of $\mathcal{I}_0 = 1$, 1.5, or 2 bits respectively. For each parameter set, the maximum time-delayed MI $\mathcal{I}_{\alpha_m}$ is plotted versus $\Lambda = R_0/\gamma_x^{\text{high}}$. The bound $I^{\max}(\Lambda)$ is plotted as a solid curve, and there is a subset of systems with performance near the bound. As we noted earlier, the two species optimal filter theory is an approximation for the full enzymatic system, and hence the bound does not have to be exact. But it still provides a close estimate of the MI limits for the full system, with the largest violations of the bound only $\sim 4\%$. The benefits of time delay are particularly evident for parameter sets with lower $\mathcal{I}_0$: some cases with $\mathcal{I}_0 = 1$ bit exhibit $\mathcal{I}_{\alpha_m}$ approaching 1.5 bits.

D. Optimality and the yeast Pbs2/Hog1 push-pull loop

Returning to the discussion of instantaneous MI, there is an alternative way of thinking about the R_0 versus γ_x^{high} results in Fig. 3G-I. Imagine a system working at γ_x^{high} with a certain gain parameter R_0 and achieving a target value $\mathcal{I}$. Comparing other parameter sets with the same bandwidth γ_x^{high} and target $\mathcal{I}$ (taking a vertical slice of one of the panels in Fig. 3G-I), they will have a variety of different R_0 values, but all of these will be bounded from below by the minimum value

$$R_0^{\min} = 4^{\mathcal{I}+1}(4^{\mathcal{I}}-1)\gamma_x^{\text{high}}. \tag{7}$$

When $R_0 = R_0^{\min}$, the system sits on the optimality line of Eq. (5), with $\gamma_x^{\text{high}} = \gamma_x^{\max}$.

The discrepancy between R_0 and $R_0^{\min}$ for a given system allows us to see how close the signaling behavior is to optimality. Let us take a concrete biological example: the Pbs2/Hog1 enzymatic push-pull loop from yeast, part of the Hog1 signaling pathway that allows the organism to respond to osmotic stress. As described in the SI, key parameters for this system

can be estimated based on an earlier model [46] fit to microfluidic experimental data where yeast was exposed to periodic salt shocks [57]. The results for the bandwidth and gain for $\mathcal{I} = 1$ bit are: $\gamma_x^{\text{high}} = (1.22 \pm 0.04) \times 10^{-3}$ s^{-1} and $R_0 = 0.0621 \pm 0.0001$ s^{-1}, with the error bars reflecting uncertainties due to unknown parameters (where we used priors based on the log-normal distributions in Fig. 2.) The scale of the predicted bandwidth γ_x^{high} is consistent with microfluidic estimates. Ref. [47] found a steep dropoff in the mean amplitude of the Hog1 response to periodic step-like changes in external osmolyte concentrations when the frequencies of the changes increased from 10^{-3} s^{-1} to 10^{-2} s^{-1}. At frequencies beyond the dropoff the Hog1 output can no longer reproduce the osmolyte input at high fidelity. Though the form of the input in this case is different than in our model, and the experiment probes the entire pathway rather than just the Pbs2/Hog1 component, the similarity in scales to our γ_x^{high} value suggests that the Pbs2/Hog1 system may play a major role in determining the bandwidth of the whole pathway (since the bandwidth of the whole is constrained by the bandwidths of the components).

Intriguingly, the estimated gain R_0 is very close to the minimum possible value $R_0^{\min}$ for signaling at the bandwidth γ_x^{high} with $\mathcal{I} = 1$ bit, as seen in Fig. 3G. Using Eq. (7), we find $R_0^{\min} = 0.059 \pm 0.002$ s^{-1}. This naturally leads to the question: is the fact that this system lies so close to optimality a coincidence, or are there reasons why natural selection might favor minimizing R_0 in this case? To answer this question, we first have to consider the relationship between gain and ATP consumption.

E. Minimum ATP consumption to achieve a certain signaling fidelity and bandwidth

This bound on the gain parameter in Eq. (7) is directly related to the metabolic cost of signaling, since higher production of the output per given input level will generally require a higher rate of phosphorylation events. We can roughly quantify the average rate of phosphorylation: in the stationary state this is just the mean rate of the kinase-catalyzed reaction step, $A = \kappa_r \overline{S_K}$. Assuming one ATP hydrolyzed per reaction, A is the mean rate at which ATP is consumed by the system, and is related to R_0 through $A = \kappa_r R_0 \overline{K}/(\kappa_r - R_0)$, as shown in the SI. In the enzymatic parameter ranges we consider, κ_r is typically much larger than R_0, so we can approximate this relation as $A \approx R_0 \overline{K}$. Using Eq. (7) we can then estimate the minimum possible ATP consumption rate given a target $\mathcal{I}$ and bandwidth γ_x^{high}:

$$A_{\min} \approx R_0^{\min} \bar{K} = 4^{\mathcal{I}+1}(4^{\mathcal{I}} - 1)\gamma_x^{\text{high}} \overline{K}. \tag{8}$$

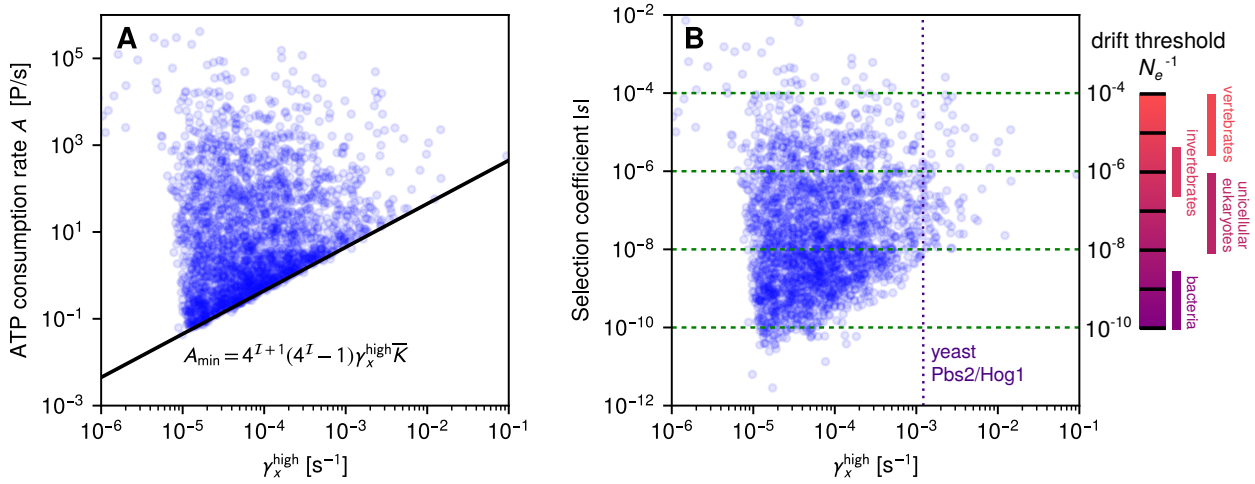

FIG. 5. **(A)** The same $(\gamma_x^{\text{high}}, \Delta\mu^{\text{high}})$ point distribution as in Fig. 3C for $\mathcal{I} = 1$ bit, except plotted in terms of ATP consumption rate A on the vertical axis. The solid line is the approximate lower bound A_{min} on ATP consumption given by Eq. (8). **(B)** This distribution replotted with selection coefficient $|s|$ on the vertical axis. $|s|$ quantifies the fitness cost associated with a system that achieves the target $\mathcal{I} = 1$ bit but is sub-optimal in ATP consumption, relative to an optimal variant where $A = A_{\text{min}}$. The value of $|s|$ becomes evolutionarily significant when it is higher than a "drift threshold" N_e^{-1}, where N_e is the effective population of the organism (a measure of genetic diversity). The ranges of N_e^{-1} for different classes of organisms are shown on the right [32, 58]. The vertical dotted line corresponds to the estimated γ_x^{high} for the yeast Pbs2/Hog1 system.

Fig. 5A shows the same parameter set values as the $(\gamma_x^{\text{high}}, \Delta\mu^{\text{high}})$ points in Fig. 3D for $\mathcal{I} = 1$ bit, except plotted in terms of $(\gamma_x^{\text{high}}, A)$. The A values are exact, but the approximate relation of Eq. (8) provides an excellent lower bound on the distribution. Qualitatively, the individual elements of Eq. (8) all make intuitive sense. An increase in any of the constituent factors (the mean free input kinase population $\overline{K}$, the target information $\mathcal{I}$, the bandwidth γ_x^{high}) puts greater demands on the signaling system, requiring more catalytic activity and hence faster ATP consumption. Note that the above results are easily generalized if the reaction step consumes more than one ATP: for example the effective model for yeast Pbs2/Hog1 discussed above involves phosphorylation at two sites, which would lead to expressions for A and A_{min} getting a prefactor of two.

F. Evolutionary pressure on the metabolic costs of signaling

It is clear from Fig. 5A that for many parameter set choices the ATP consumption rate A is significantly larger than for a system near optimality ($A \approx A_{\text{min}}$) given the same $\mathcal{I}$ and γ_x^{high}. Let us consider a specific scenario where the bandwidth γ_x^{high} and the target $\mathcal{I}$ are sufficient for the biological function of the signaling i.e. there are rapidly diminishing fitness returns in going to higher bandwidth and signal fidelity. In this scenario a system with

$A > A_{\min}$ has no significant adaptive advantage over one with $A \approx A_{\min}$, but instead incurs a fitness penalty because of the superfluous ATP consumption. Would there be evolutionary pressure on this sub-optimal system to move toward optimality?

The answer to this question has practical ramifications, because it will allow us to predict whether we should expect to see natural enzymatic push-pull loops cluster around the optimality line (as we saw in the yeast Pbs2-Hog1 example). The alternative, in the absence of strong evolutionary pressure to optimize, is a wider dispersion, more similar to Fig. 5A where the points are drawn at random from the enzymatic parameter distribution. Note that this is a question that is directly amenable to future kinetic experiments: for systems where we can fully characterize the enzymatic parameters of the push-pull loop (for both the kinase and phosphatase), all the relevant quantities like γ_x^{high}, A, and $\mathcal{I}$ can be calculated.

Naively one might expect evolution to always drive systems to optimality due to natural selection, but genetic drift can play a significant competing role, allowing sub-optimal variants to flourish and even fix in a population [59]. To be specific, let us consider a unicellular organism that reproduces via binary fission, and two genetic variants of that organism that differ in the enzymatic parameters of a push-pull signaling loop: both variants achieve the same γ_x^{high} and $\mathcal{I}$, but one has $A > A_{\min}$ and one has $A = A_{\min}$. Let us denote the relative fitness of the sub-optimal versus the optimal type as $1 + s$, defining a selection coefficient s. In other words the sub-optimal variant will have on average $1 + s$ offspring relative to the optimal one during the generation time of the optimal type. In the scenario described above, where the extra production does not confer any adaptive advantage and only imposes a metabolic cost, we will have $s < 0$, because the superfluous ATP consumption will lead to slower growth.

The magnitude of s determines the degree of selective pressure on the sub-optimal variant. The key quantity that sets the relevant scale for s is the effective population N_e of the organism, the size of an idealized population that exhibits the same changes in genetic diversity per generation due to drift as the actual population [58]. When $s < 0$ and $|s| \gg N_e^{-1}$, natural selection dominates drift, exponentially suppressing the probability of a sub-optimal mutant fixing in a population of optimal organisms. On the other hand if $|s| \ll N_e^{-1}$, drift dominates, and the fixation probability of sub-optimal mutants is roughly the same as for a neutral ($s = 0$) mutation [60]. In this case it would be difficult to maintain optimality in a population over the long term. N_e for organisms is typically smaller than their actual population in the wild, and varies by several orders of magnitude among different classes: for unicellular species it can be as high as $\sim 10^9 - 10^{10}$ in bacteria down to $\sim 10^6 - 10^8$ in single-celled eukaryotes [32, 58]. (It becomes even smaller among higher eukaryotes, going down to $\sim 10^4$ in vertebrates.) The corresponding ranges for the "drift threshold" N_e^{-1} [32] are shown on the right in Fig. 5B.

The question then becomes: how do we estimate s and how does it compare to the relevant N_e^{-1} for the class of interest? For the case where a variant imposes metabolic costs but no adaptive advantage, there is a very useful relation that posits $s \sim -\delta C_T/C_T$ [32, 61, 62]. Here C_T is the total resting metabolic expenditure of an organism during a generation time, measured for example in units of P, where 1 P = one phosphate bond hydrolyzed (ATP or ATP equivalent consumed). δC_T is the extra expenditure incurred by the more costly mutant. This relation has already been used to explore selective pressures in yeast [62], unicellular prokaryotes and eukaryotes [32], and viral infections [63]. It was recently derived from first principles through a general bioenergetic growth model [33], where the relation was refined with a more accurate prefactor: $s \approx -\ln(R_b)\delta C_T/C_T$. Here R_b is the mean number of offspring per individual (i.e. $R_b = 2$ for binary fission).

The value of C_T can be readily estimated for single-celled organisms, where it scales roughly with cell volume [32, 33]. Given the 30 fL cell volume used in our calculations, and assuming a generation time (cell division time) $t_r = 1$ hr, we find $C_T \approx 7 \times 10^{11}$ P (see details in the SI), comparable in magnitude to experimental estimates for yeast [32]. Since δC_T reflects the extra ATP consumed by the costly mutant (with consumption rate A) versus the optimal variant (rate $A_{\min}$) over one generation time, we can write $\delta C_T = (A - A_{\min})t_r$. We can thus calculate s for all the near-bandwidth $\mathcal{I} = 1$ bit parameter sets represented in Fig. 5A. The results for $|s|$ versus γ_x^{high} are plotted in Fig. 5B. Because increased ATP consumption is required to achieve larger bandwidths (as seen in Eq. (8)), the distribution of selective penalties $|s|$ for being sub-optimal is pushed to larger values with greater γ_x^{high}. In other words, higher bandwidths make the energetic stakes more significant.

We can now rationalize why the yeast Pbs2/Hog1 loop might be close to optimality. The bandwidth for that system (indicated by a vertical dashed line in Fig. 5B) is near the higher end of the spectrum. Suboptimal parameter values that achieve approximately the same bandwidth at $\mathcal{I} = 1$ bit span a range of $|s|$ values between 10^{-8} and 10^{-4}. Given $N_e = 10^6 - 10^8$ for single-celled eukaryotes [32, 58], and estimates of $N_e \approx 10^7$ for wild yeast populations [64], these suboptimal systems likely have $|s|$ near or above the drift threshold N_e^{-1}. Thus we would expect yeast to be under evolutionary pressure to optimize the energy expenditures associated with the enzymatic loop.

IV. DISCUSSION AND CONCLUSIONS

The kinase-phosphatase push-pull signaling network, which maintains a certain value of mutual information $\mathcal{I}$ between input and output, incurs energetic costs in the form of ATP consumption. These costs have two related facets: (i) the free energy expenditure $\Delta\mu$ for each hydrolysis reaction, and (ii) the number of such reactions A per unit time. Achieving

empirical values like $\mathcal{I} = 1 - 2$ bits requires satisfying both aspects of the cost. There is a minimal price in terms of $\Delta\mu$ to achieve any given $\mathcal{I}$, and this price increases if one demands either greater fidelity (larger $\mathcal{I}$) or the ability to process faster signals (larger γ_x). Modern cells are more than willing to pay this part of the price, with $\Delta\mu$ sufficiently high to meet the minimal requirements for any enzymatic parameter set that hits a target $\mathcal{I}$ on the order of 1 bit. However, as the distributions in Fig. 3D-F illustrate, there are certainly options for signaling systems that work at similar fidelities under conditions of smaller $\Delta\mu$, the presumptive scenario earlier in evolutionary history. In all cases we require some degree of fine-tuning of enzymatic parameters: the higher the fidelity or frequency demands, the smaller the fraction of parameter space that satisfies them. This leaves vanishingly small room to achieve networks that operate at $\mathcal{I}$ significantly larger than the known empirical range.

For particular parameter combinations the system is optimal, exhibiting the maximum possible bandwidth ($\gamma_x^{\max}$ of Eq. (5)) with the minimal ATP consumption ($A_{\min}$ of Eq. (8)). These theoretical bounds can also be generalized to the case of time-delayed MI, giving the simple expression for the overall maximum MI shown in Eq. (6). Is such optimality widely realized in nature? Analyzing the selective pressures due to superfluous ATP expenditures indicates that this is a worthwhile question to pursue. We have already identified one near-optimal candidate in the yeast Hog1 signaling pathway. Based on the results of the previous section, we predict that the best place to look for others is among signaling pathways with high bandwidths, for example $\sim 10^{-3} - 10^{-2}$ s^{-1} at the extremes of the current biological distribution. Here the metabolic costs of being suboptimal would be significant for single-celled organisms.

More broadly, strong selective pressure on the costs of running signaling networks in single-celled organisms is likely to be a widespread phenomenon. To give another example, the expenditure of running the chemotaxis machinery in *E. coli* has been estimated to be about $\sim 10^7$ P per ~ 1 hr cell cycle [2, 4]. Compared to a value of $C_T \approx 2 \times 10^{10}$ P for *E. coli* [32, 33], we get an $|s| \sim 10^{-4}$, which is definitely significant for a bacterial population. We have barely begun to understand the kinds of optimization that such selective pressure has induced. Our approach readily generalizes beyond the kinase-phosphatase system, setting the stage for exploring these issues in a much wider array of biochemical networks.

Data and code availability: The code for our analysis, along with the data used to generate the figures, is available at: `https://github.com/hincz-lab/cell-signaling`.

Acknowledgments: We thank Shishir Adhikari for useful discussions. Parts of the numerical analysis were carried out using the High Performance Computing Resource in the Core Facility for Advanced Research Computing at Case Western Reserve University.

---

[1] Balázsi, G., van Oudenaarden, A. & Collins, J. J. Cellular decision making and biological noise: from microbes to mammals. *Cell* **144**, 910–925 (2011).

[2] Govern, C. C. & ten Wolde, P. R. Optimal resource allocation in cellular sensing systems. *Proc. Natl. Acad. Sci.* **111**, 17486–17491 (2014).

[3] ten Wolde, P. R., Becker, N. B., Ouldridge, T. E. & Mugler, A. Fundamental limits to cellular sensing. *J. Stat. Phys.* **162**, 1395–1424 (2016).

[4] Lan, G. & Tu, Y. Information processing in bacteria: memory, computation, and statistical physics: a key issues review. *Rep. Prog. Phys.* **79**, 052601 (2016).

[5] Logan, C. Y. & Nusse, R. The Wnt signaling pathway in development and disease. *Annu. Rev. Cell Dev. Biol.* **20**, 781–810 (2004).

[6] Parsons, D. W. *et al.* Colorectal cancer: mutations in a signalling pathway. *Nature* **436**, 792 (2005).

[7] Sarkar, F. H., Li, Y., Wang, Z. & Kong, D. Cellular signaling perturbation by natural products. *Cell. Sign.* **21**, 1541 – 1547 (2009).

[8] Riera, C. E., Merkwirth, C., De Magalhaes Filho, C. D. & Dillin, A. Signaling networks determining life span. *Annu. Rev. Biochem.* **85**, 35–64 (2016).

[9] Uda, S. Application of information theory in systems biology. *Biophys. Rev.* 1–8 (2020).

[10] Tkačik, G., Callan, C. G. & Bialek, W. Information flow and optimization in transcriptional regulation. *Proc. Natl. Acad. Sci.* **105**, 12265–12270 (2008).

[11] Cheong, R., Rhee, A., Wang, C. J., Nemenman, I. & Levchenko, A. Information transduction capacity of noisy biochemical signaling networks. *Science* **334**, 354–358 (2011).

[12] Uda, S. *et al.* Robustness and compensation of information transmission of signaling pathways. *Science* **341**, 558–561 (2013).

[13] Voliotis, M., Perrett, R. M., McWilliams, C., McArdle, C. A. & Bowsher, C. G. Information transfer by leaky, heterogeneous, protein kinase signaling systems. *Proc. Natl. Acad. Sci.* **111**, E326–E333 (2014).

[14] Selimkhanov, J. *et al.* Accurate information transmission through dynamic biochemical signaling networks. *Science* **346**, 1370–1373 (2014).

[15] Potter, G. D., Byrd, T. A., Mugler, A. & Sun, B. Dynamic sampling and information encoding in biochemical networks. *Biophys. J.* **112**, 795–804 (2017).

[16] Suderman, R., Bachman, J. A., Smith, A., Sorger, P. K. & Deeds, E. J. Fundamental trade-offs between information flow in single cells and cellular populations. *Proc. Natl. Acad. Sci.* **114**, 5755–5760 (2017).

[17] Keshelava, A. *et al.* High capacity in g protein-coupled receptor signaling. *Nat. Commun.* **9**, 1–8 (2018).

[18] Shannon, C. E. A mathematical theory of communication. *Bell Systems Tech. J.* **27**, 379–423 (1948).

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

[20] Cao, Y., Wang, H., Ouyang, Q. & Tu, Y. The free-energy cost of accurate biochemical oscillations. *Nat. Phys.* **11**, 772–778 (2015).

[21] Hasegawa, Y. Optimal temporal patterns for dynamical cellular signaling. *New J. Phys.* **18**, 113031 (2016).

[22] Mehta, P., Lang, A. H. & Schwab, D. J. Landauer in the age of synthetic biology: energy consumption and information processing in biochemical networks. *J. Stat. Phys.* **162**, 1153–1166 (2016).

[23] Ouldridge, T. E., Govern, C. C. & ten Wolde, P. R. Thermodynamics of computational copying in biochemical systems. *Phys. Rev. X* **7**, 021004 (2017).

[24] Milo, R. & Phillips, R. *Cell biology by the numbers* (Garland Science, 2015).

[25] Lane, N. & Martin, W. F. The origin of membrane bioenergetics. *Cell* **151**, 1406–1416 (2012).

[26] Stadtman, E. & Chock, P. Superiority of interconvertible enzyme cascades in metabolic regulation: analysis of monocyclic systems. *Proc. Natl. Acad. Sci.* **74**, 2761–2765 (1977).

[27] Goldbeter, A. & Koshland, D. E. An amplified sensitivity arising from covalent modification in biological systems. *Proc. Natl. Acad. Sci.* **78**, 6840–6844 (1981).

[28] Detwiler, P. B., Ramanathan, S., Sengupta, A. & Shraiman, B. I. Engineering aspects of enzymatic signal transduction: photoreceptors in the retina. *Biophys. J.* **79**, 2801–2817 (2000).

[29] Heinrich, R., Neel, B. G. & Rapoport, T. A. Mathematical models of protein kinase signal transduction. *Molecular cell* **9**, 957–970 (2002).

[30] Hinczewski, M. & Thirumalai, D. Cellular signaling networks function as generalized wiener-kolmogorov filters to suppress noise. *Phys. Rev. X* **4**, 041017 (2014).

[31] Hinczewski, M. & Thirumalai, D. Noise control in gene regulatory networks with negative feedback. *J. Phys. Chem. B* **120**, 6166–6177 (2016).

[32] Lynch, M. & Marinov, G. K. The bioenergetic costs of a gene. *Proc. Natl. Acad. Sci.* **112**, 15690–15695 (2015).

[33] Ilker, E. & Hinczewski, M. Modeling the growth of organisms validates a general relation between metabolic costs and natural selection. *Phys. Rev. Lett.* **122**, 238101 (2019).

[34] Sturm, O. E. *et al.* The mammalian MAPK/ERK pathway exhibits properties of a negative feedback amplifier. *Sci. Signal.* **3**, ra90–ra90 (2010).

[35] Saxena, M., Williams, S., Taskén, K. & Mustelin, T. Crosstalk between camp-dependent kinase and map kinase through a protein tyrosine phosphatase. *Nat. Cell Biol.* **1**, 305 (1999).

[36] Salazar, C. & Höfer, T. Multisite protein phosphorylation–from molecular mechanisms to kinetic models. *FEBS J.* **276**, 3177–3198 (2009).

[37] Fersht, A. *Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding* (Macmillan, 1999).

[38] Qian, H. Cooperativity and specificity in enzyme kinetics: a single-molecule time-based perspective. *Biophys. J.* **95**, 10–17 (2008).

[39] Gillespie, D. T. The chemical Langevin equation. *J. Chem. Phys.* **113**, 297–306 (2000).

[40] Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81**, 2340–2361 (1977).

[41] Bar-Even, A. *et al.* The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* **50**, 4402–4410 (2011).

[42] Liebermeister, W. & Klipp, E. Biochemical networks with uncertain parameters. *IEE Proc.-Syst. Biol.* **152**, 97–107 (2005).

[43] Wang, M., Herrmann, C. J., Simonovic, M., Szklarczyk, D. & von Mering, C. Version 4.0 of PaxDb: protein abundance data, integrated across model organisms, tissues, and cell-lines. *Proteomics* **15**, 3163–3168 (2015).

[44] UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515 (2018).

[45] Wittig, U. *et al.* SABIO-RK—database for biochemical reaction kinetics. *Nucleic Acids Res.* **40**, D790–D796 (2011).

[46] Zi, Z., Liebermeister, W. & Klipp, E. A quantitative study of the Hog1 MAPK response to fluctuating osmotic stress in Saccharomyces cerevisiae. *PLOS ONE* **5**, e9522 (2010).

[47] Hersen, P., McClean, M. N., Mahadevan, L. & Ramanathan, S. Signal processing by the hog map kinase pathway. *Proc. Natl. Acad. Sci.* **105**, 7165–7170 (2008).

[48] Wiener, N. *Extrapolation, Interpolation and Smoothing of Stationary Times Series* (Wiley, New York, 1949).

[49] Kolmogorov, A. N. Interpolation and extrapolation of stationary random sequences. *Izv. Akad. Nauk SSSR., Ser. Mat.* **5**, 3–14 (1941).

[50] Bode, H. W. & Shannon, C. E. A simplified derivation of linear least square smoothing and prediction theory. *Proc. Inst. Radio. Engin.* **38**, 417–425 (1950).

[51] Becker, N. B., Mugler, A. & ten Wolde, P. R. Optimal prediction by cellular signaling networks.

Phys. Rev. Lett. **115** (258103, 2015).

[52] Zechner, C., Seelig, G., Rullan, M. & Khammash, M. Molecular circuits for dynamic noise filtering. *Proc. Natl. Acad. Sci.* **113**, 4729–4734 (2016).

[53] Samanta, H. S., Hinczewski, M. & Thirumalai, D. Optimal information transfer in enzymatic networks: A field theoretic formulation. *Phys. Rev. E* **96**, 012406 (2017).

[54] Hathcock, D., Sheehy, J., Weisenberger, C., Ilker, E. & Hinczewski, M. Noise filtering and prediction in biological signaling networks. *IEEE Trans. Mol. Biol. Multi-Scale Commun.* **2**, 16–30 (2016).

[55] Nemenman, I. Gain control in molecular information processing: lessons from neuroscience. *Phys. Biol.* **9**, 026003 (2012).

[56] Mancini, F., Wiggins, C. H., Marsili, M. & Walczak, A. M. Time-dependent information transmission in a model regulatory circuit. *Phys. Rev. E* **88**, 022708 (2013).

[57] Mettetal, J. T., Muzzey, D., Gómez-Uribe, C. & van Oudenaarden, A. The frequency dependence of osmo-adaptation in Saccharomyces cerevisiae. *Science* **319**, 482–484 (2008).

[58] Charlesworth, B. Effective population size and patterns of molecular evolution and variation. *Nature Rev. Genet.* **10**, 195 (2009).

[59] Gillespie, J. H. *Population genetics: a concise guide* (JHU Press, 2010).

[60] Kimura, M. On the probability of fixation of mutant genes in a population. *Genetics* **47**, 713–719 (1962).

[61] Orgel, L. E. & Crick, F. H. Selfish DNA: the ultimate parasite. *Nature* **284**, 604 (1980).

[62] Wagner, A. Energy constraints on the evolution of gene expression. *Mol. Biol. Evol.* **22**, 1365–1374 (2005).

[63] Mahmoudabadi, G., Milo, R. & Phillips, R. Energetic cost of building a virus. *Proc. Natl. Acad. Sci.* **114**, E4324–E4333 (2017).

[64] Tsai, I. J., Bensasson, D., Burt, A. & Koufopanou, V. Population genomics of the wild yeast Saccharomyces paradoxus: quantifying the life cycle. *Proc. Natl. Acad. Sci.* **105**, 4957–4962 (2008).

Supplementary Information for:
The price of a bit: energetic costs and the evolution of cellular signaling

CONTENTS

I. Derivation of the local detailed balance relation 2

II. Chemical Langevin approach for the kinase-phosphatase push-pull loop 2

A. Dynamical equations 3

B. Second moments 4

C. Comparison to kinetic Monte Carlo simulations for mutual information 6

III. Characteristic frequency γ_x, gain R_0, and the conditions for Wiener-Kolmogorov noise filter optimality 6

A. Deriving the γ_x and R_0 expressions in main text Eq. (2) 6

B. Validation through kinetic Monte Carlo simulations 8

C. Relating maximum bandwidth, minimum ATP consumption rate, and mutual information via Wiener-Kolmogorov optimal noise filter theory 8

D. Optimality with respect to time-offset mutual information 10

IV. Enzymatic parameter distribution 11

V. Results for alternative input kinase concentrations 13

VI. Analysis of the Pbs2-Hog1 push-pull loop in yeast 13

A. Parameter estimation based on earlier literature 14

B. Estimation of remaining parameters 16

C. Bandwidth and gain 17

VII. Estimation of total resting metabolic expenditure 17

References 18

I. DERIVATION OF THE LOCAL DETAILED BALANCE RELATION

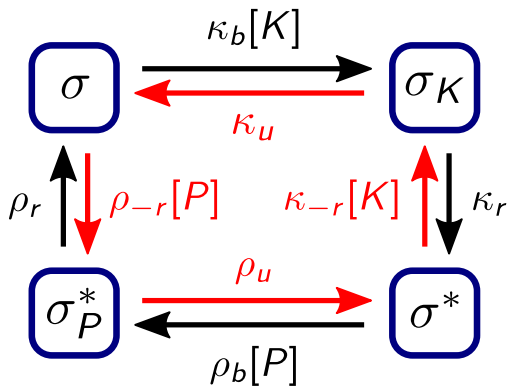

FIG. S1. The enzymatic push-pull loop from the perspective of an individual substrate molecule. The protein can exist in one of four states: unmodified substrate (σ), bound to kinase (σ_K), phosphorylated (σ^*), and bound to phosphatase while phosphorylated (σ_P^*). The forward (clockwise) transition rates between these states are indicated in black, while the reverse (counterclockwise) rates are in red.

To derive the local detailed balance relation of main text Eq. (1), it is convenient to focus on the reactions from the perspective of an individual substrate molecule [1]. A given molecule in our model can be in one of four states, indicated in Fig. S1 with corresponding forward and reverse transition rates. For example if the molecule is an unmodified substrate (state σ) it can transition to a kinase-bound substrate (state σ_K) with rate $\kappa_b[K]$, proportional to the surrounding concentration $[K]$ of kinase molecules. It can revert from σ_K to σ with rate κ_u. The other transitions in Fig. S1 are defined analogously, with forward rates colored black and reverse rates in red. Local detailed balance entails that product of reverse rates divided by the product of forward rates is equal to $\exp(\beta\Delta G)$, where ΔG is the free energy change of the system associated with a single forward traversal of the loop and $\beta = (k_B T)^{-1}$ [1]. Since after one loop from σ to σ the substrate is back in the same state (as well as the kinase and phosphatase), there is no contribution to ΔG from these molecules. However a single loop leads to the hydrolysis of a single molecule of ATP, so $\Delta G = -\Delta\mu$, as defined in the main text. Putting everything together, the local detailed balance relation reads

$$e^{-\beta\Delta\mu} = \frac{\kappa_u}{\kappa_b[K]}\frac{\kappa_{-r}[K]}{\kappa_r}\frac{\rho_u}{\rho_b[P]}\frac{\rho_{-r}[P]}{\rho_r} = \frac{\kappa_{-r}\rho_u\rho_{-r}\kappa_u}{\kappa_r\rho_b\rho_r\kappa_b}, \tag{S1}$$

yielding main text Eq. (1).

II. CHEMICAL LANGEVIN APPROACH FOR THE KINASE-PHOSPHATASE PUSH-PULL LOOP

In this section we derive the stationary state properties of the kinase-phosphatase push-pull loop via the chemical Langevin approximation. The derivation will follow analogously to Ref. [2], except here the system is more complicated due to the inclusion of reverse enzymatic reactions. The end goal will be a method to estimate the instantaneous mutual information

$\mathcal{I}$, given by main text Eq. (4),

$$\mathcal{I} \approx -\frac{1}{2}\log_2 E, \quad \text{where} \quad E \equiv 1 - \frac{(\overline{XY} - \overline{X}\,\overline{Y})^2}{\left(\overline{X^2} - \overline{X}^2\right)\left(\overline{Y^2} - \overline{Y}^2\right)}, \tag{S2}$$

which requires evaluating the variances of the input and output, $\mathrm{var}(X) = \overline{X^2} - \overline{X}^2$, $\mathrm{var}(Y) = \overline{Y^2} - \overline{Y}^2$, as well as the covariance $\mathrm{cov}(X,Y) = \overline{XY} - \overline{X}\,\overline{Y}$. The quantity E here will be referred to as the "error" in signal propagation between input and output, and can be equivalently expressed as $E = 1 - \rho^2$, where ρ is the Pearson correlation coefficient between X and Y.

A. Dynamical equations

Our starting point is the full system of reactions for the enzymatic push-pull loop,

$$\begin{aligned}
&\varnothing \underset{\gamma_K}{\overset{F}{\rightleftharpoons}} K, \\
&K + S \underset{\kappa_u}{\overset{\kappa_b}{\rightleftharpoons}} S_K \underset{\kappa_{-r}}{\overset{\kappa_r}{\rightleftharpoons}} K + S^*, \\
&P + S^* \underset{\rho_u}{\overset{\rho_b}{\rightleftharpoons}} S_P^* \underset{\rho_{-r}}{\overset{\rho_r}{\rightleftharpoons}} P + S,
\end{aligned} \tag{S3}$$

where $\varnothing$ represents the void (upstream deactivated kinase which does not enter into our model). The corresponding chemical Langevin equations [3] are given by:

$$\begin{aligned}
\frac{dK}{dt} &= F - \gamma_K K - \kappa_b K S + (\kappa_u + \kappa_r) S_K - \kappa_{-r} K S^* + n_1 + n_2 + n_3, \\
\frac{dS_K}{dt} &= \kappa_b K S - (\kappa_u + \kappa_r) S_K + \kappa_{-r} K S^* - n_2 - n_3, \\
\frac{dS^*}{dt} &= \kappa_r S_K - \rho_b P S^* + \rho_u S_P^* - \kappa_{-r} K S^* + n_3 + n_4, \\
\frac{dS_P^*}{dt} &= \rho_b P S^* - (\rho_u + \rho_r) S_P^* + \rho_{-r} S P - n_4 + n_5, \\
\frac{dP}{dt} &= -\frac{dS_P^*}{dt}, \quad \frac{dS}{dt} = -\frac{dS_K}{dt} - \frac{dS^*}{dt} - \frac{dS_P^*}{dt},
\end{aligned} \tag{S4}$$

where the last line ensures that the total populations of free or bound phosphatase $(P + S_P^*)$ and free or bound substrate in all its forms $(S + S_K + S^* + S_P^*)$ remain constant. The noise terms $n_i(t) = \sqrt{\Pi_i}\eta_i(t)$, where $\eta_i(t)$ are Gaussian noise functions with zero mean and correlations $\overline{\eta_i(t)\eta_j(t')} = \delta_{ij}\delta(t - t')$. The five noise terms are associated with reactions in the system, and the corresponding prefactors represent the sum of the mean production (forward) and deactivation/unbinding (backward) contributions to each reaction:

$$\begin{aligned}
&\Pi_1 = F + \gamma_K \overline{K}, \quad \Pi_2 = \kappa_b \overline{K}\,\overline{S} + \kappa_u \overline{S}_K, \quad \Pi_3 = \kappa_r \overline{S}_K + \kappa_{-r} \overline{K}\,\overline{S}^*, \\
&\Pi_4 = \rho_b \overline{P}\,\overline{S}^* + \rho_u \overline{S}_P^*, \quad \Pi_5 = \rho_r \overline{S}_P^* + \rho_{-r} \overline{S}\,\overline{P}.
\end{aligned} \tag{S5}$$

As described in the next section, we will be linearizing Eq. (S4), keeping terms up to first order in deviations from the stationary state values. In this linearized approach, the stationary state populations are given by:

$$\overline{K} = \frac{F}{\gamma_K}, \quad \overline{S}_K = \frac{FC_2}{\gamma_K C_1}, \quad \overline{S}^* = \frac{C_3}{C_1}, \quad \overline{S}^*_P = \frac{C_4}{C_1}. \tag{S6}$$

with the following definitions:

$$\begin{aligned}
&\kappa_- \equiv \kappa_u + \kappa_r, \quad \rho_- \equiv \rho_u + \rho_r, \\
&C_1 \equiv \kappa_- \overline{P}\gamma_K \rho_b \rho_r + F\kappa_{-r}\kappa_u \rho_-, \\
&C_2 \equiv \overline{S}\left[F\kappa_b\kappa_{-r}\rho_- + \overline{P}\gamma_K(\kappa_b\rho_b\rho_r + \kappa_{-r}\rho_{-r}\rho_u)\right], \\
&C_3 \equiv \overline{S}(\overline{P}\gamma_K\kappa_-\rho_{-r}\rho_u + F\kappa_b\kappa_r\rho_-), \\
&C_4 \equiv \overline{P}\overline{S}\left[\overline{P}\gamma_K\kappa_-\rho_{-r}\rho_b + F(\kappa_b\rho_b\rho_r + \kappa_{-r}\rho_{-r}\kappa_u)\right].
\end{aligned} \tag{S7}$$

The input (total kinase) is $X = K + S_K$ and the output (total activated substrate) is $Y = S^* + S^*_P$, and hence Eq. (S6) can be used to calculate the stationary values $\overline{X}$ and $\overline{Y}$.

B. Second moments

In order to calculate the variance and covariance of the input and output, we also need to know $\overline{X^2}$, $\overline{Y^2}$, $\overline{XY}$. To estimate these quantities, the first step is to switch variables in Eq. (S4) to focus on deviations from the stationary state values: $\delta K \equiv K - \overline{K}$, $\delta S_K \equiv S_K - \overline{S_K}$, $\delta S^* \equiv S^* - \overline{S}^*$, $\delta S^*_P \equiv S^*_P - \overline{S}^*_P$. We can in turn rewrite these four variables in terms of the input and output deviations $\delta X = X - \overline{X}$ and $\delta Y = Y - \overline{Y}$:

$$\begin{aligned}
\delta K &= \frac{C_1}{C_1 + C_2}\delta X + \delta X_q, \\
\delta S_K &= \frac{C_2}{C_1 + C_2}\delta X - \delta X_q, \\
\delta S^* &= \frac{C_3}{C_3 + C_4}\delta Y + \delta Y_q, \\
\delta S^*_P &= \frac{C_4}{C_3 + C_4}\delta Y - \delta Y_q,
\end{aligned} \tag{S8}$$

where we have introduced two additional auxiliary variables δX_q and δY_q. Plugging Eq. (S8) into Eq. (S4), we simplify the system through linearization, ignoring any terms of second order or higher in the deviations. As demonstrated below in comparisons with kinetic Monte Carlo (KMC) simulations [4] of the original system, this linearized chemical Langevin approximation works well for our parameter ranges. Finally, we Fourier transform the

linearized Eq. (S4), and the resulting system of equations takes the form

$$M(\omega)\begin{pmatrix}\widetilde{\delta X}\\ \widetilde{\delta X}_q\\ \widetilde{\delta Y}\\ \widetilde{\delta Y}_q\end{pmatrix}=\begin{pmatrix}-\widetilde{n}_1\\ \widetilde{n}_2+\widetilde{n}_3\\ -\widetilde{n}_3-\widetilde{n}_5\\ \widetilde{n}_4-\widetilde{n}_5\end{pmatrix} \tag{S9}$$

where $\widetilde{Q}(\omega)$ denotes the Fourier transform of quantity $Q(t)$. The matrix M is given by:

$$M(\omega)=\begin{pmatrix} i\omega-\frac{C_1\gamma_K}{C_1+C_2} & \gamma_K & 0 & 0\\ \frac{C_1(\kappa_b\overline{S}+\kappa_{-r}\overline{S}^*)-C_2(\kappa_b\overline{K}+\kappa_--i\omega)}{C_1+C_2} & \kappa_b(\overline{K}+\overline{S})+\kappa_{-r}\overline{S}^*+\kappa_--i\omega & \frac{C_3\overline{K}\kappa_{-r}}{C_3+C_4}-\kappa_b\overline{K} & \overline{K}\kappa_{-r}\\ \frac{C_2\kappa_r-C_1\kappa_{-r}\overline{S}^*}{C_1+C_2} & -\kappa_r-\kappa_{-r}\overline{S}^* & -\frac{C_3\overline{K}\kappa_{-r}+C_4\rho_r}{C_3+C_4}+i\omega & \rho_r-\kappa_{-r}\overline{K}\\ 0 & 0 & \frac{C_3\overline{P}\rho_b-C_4(\overline{S}^*\rho_b+\rho_--i\omega)}{C_3+C_4} & \overline{P}\rho_b+\overline{S}^*\rho_b+\rho_--i\omega \end{pmatrix}. \tag{S10}$$

The Fourier-space system of equations Eq. (S9)-(S10) can be solved for $\widetilde{\delta X}(\omega)$ and $\widetilde{\delta Y}(\omega)$. The expressions are complicated, but take the form of a linear combination of Fourier-space noise terms:

$$\widetilde{\delta X}(\omega)=\sum_{i=1}^{5}a_i^X(\omega)\widetilde{n}_i,\quad \widetilde{\delta Y}(\omega)=\sum_{i=1}^{5}a_i^Y(\omega)\widetilde{n}_i, \tag{S11}$$

where $a_i^X(\omega)$ and $a_i^Y(\omega)$ are some prefactors which can be expressed as rational functions of ω. The prefactors have the property $a_i^X(-\omega)=(a_i^X(\omega))^*$, $a_i^Y(-\omega)=(a_i^Y(\omega))^*$. In Fourier space the correlations among the noise terms take the form $\overline{\widetilde{n}_i(\omega)\widetilde{n}_j(\omega')}=2\pi\delta_{ij}\Pi_i\delta(\omega+\omega')$. Hence we can calculate the input power spectral density (PSD) $P_X(\omega)$, the output PSD $P_Y(\omega)$ and the cross PSD $P_{XY}(\omega)$, defined via

$$\begin{aligned}&\overline{\widetilde{\delta X}(\omega)\widetilde{\delta X}(\omega')}=2\pi P_X(\omega)\delta(\omega+\omega'),\quad \overline{\widetilde{\delta Y}(\omega)\widetilde{\delta Y}(\omega')}=2\pi P_Y(\omega)\delta(\omega+\omega'),\\ &\overline{\widetilde{\delta X}(\omega)\widetilde{\delta Y}(\omega')}=2\pi P_{XY}(\omega)\delta(\omega+\omega').\end{aligned} \tag{S12}$$

Plugging Eq. (S11) into Eq. (S12), we find expressions for the PSDs in terms of the prefactor functions:

$$P_X(\omega)=\sum_{i=1}^{5}|a_i^X(\omega)|^2\Pi_i,\quad P_Y(\omega)=\sum_{i=1}^{5}|a_i^Y(\omega)|^2\Pi_i,\quad P_{XY}(\omega)=\sum_{i=1}^{5}a_i^X(\omega)a_i^Y(-\omega)\Pi_i. \tag{S13}$$

The final step is to calculate the second moments from integrals of the PSDs, using the inverse Fourier transform of Eq. (S13) evaluated at $t=t'$:

$$\overline{X^2}=\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}P_X(\omega),\quad \overline{Y^2}=\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}P_Y(\omega),\quad \overline{XY}=\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}P_{XY}(\omega). \tag{S14}$$

Given the explicit expressions for the prefactor functions in Eq. (S13) (which are available as part of the *Mathematica* notebooks in the Github repository associated with the manuscript), one can numerically evaluate the integrals in Eq. (S14) to get the moments.

C. Comparison to kinetic Monte Carlo simulations for mutual information

The chemical Langevin calculation of the second moments allows us to use Eq. (S2) to estimate the mutual information $\mathcal{I}$. We can then check whether this estimate is consistent with the results we would get from KMC simulations of the full system. Fig. S2 shows this comparison for two sample parameter sets drawn from the enzymatic parameter distribution described in Sec. IV. Since we are interested in exploring the full range of chemical potentials $\Delta\mu$, in each case we calculate $\mathcal{I}$ varying the reverse-to-forward rate ratio κ_{-r}/κ_r, keeping all other parameters constant. Through main text Eq. (1), increasing κ_{-r}/κ_r corresponds to decreasing the magnitude of $\Delta\mu$. At very large $\Delta\mu$ (small κ_{-r}/κ_r) the $\mathcal{I}$ curves saturate at the maximum possible mutual information for that parameter set, while at small $\Delta\mu$ (large κ_{-r}/κ_r) the mutual information approaches zero, the equilibrium limit. Across the whole range we see that the chemical Langevin theoretical prediction is in close agreement with the KMC results.

III. CHARACTERISTIC FREQUENCY γ_x, GAIN R_0, AND THE CONDITIONS FOR WIENER-KOLMOGOROV NOISE FILTER OPTIMALITY

A. Deriving the γ_x and R_0 expressions in main text Eq. (2)

Since the effective frequency γ_x of the input and the gain R_0 play central roles in the analysis, having simple closed form approximations for them [main text Eq. (2)] is useful. The original definitions of these two variables, as described in the main text, are as follows: (i) γ_x is related to the autocorrelation of input fluctuations, $\overline{\delta X(t+\tau)\delta X(t)} = \overline{\delta X^2}\exp(-\gamma_x|\tau|)$; (ii) $R_0 \equiv \kappa_r\overline{S_K}/\overline{X}$ measures output production for a given level of input. As demonstrated in the next section, both of these can be calculated from KMC simulations (at significant computational expense for each different set of parameters). Alternatively, the chemical Langevin approximation of SI Sec. II can be used to derive somewhat cumbersome analytical expressions.

However the most convenient option is to take advantage of the meaning of γ_x and R_0 in an effective, two-species description of the kinase-phosphatase reaction network. Imagine a system with an input species population $X(t)$, output $Y(t)$, and a simplified chemistry with only four reactions: production of input at rate F, deactivation of input at rate $\gamma_x X(t)$, production of output at rate $R_0 X(t)$, and deactivation of output at rate $\gamma_y Y(t)$. In this two-species system the inverse input autocorrelation time is given by the deactivation rate parameter γ_x, and the coefficient R_0 in the output production rate is also the gain parameter. To relate this simplified model to the full reaction network of SI Sec. II, we compare analogous quantities in the simplified and full schemes. For example, let us take the mean input population $\overline{X}$. In the simplified scheme this is given by

$$\overline{X} = \frac{F}{\gamma_x}. \tag{S15}$$

In the full network $\overline{X} = \overline{K} + \overline{S}_K$ can be calculated from Eq. (S6) as

$$\overline{X} = \frac{F(C_1 + C_2)}{C_1\gamma_K}, \tag{S16}$$

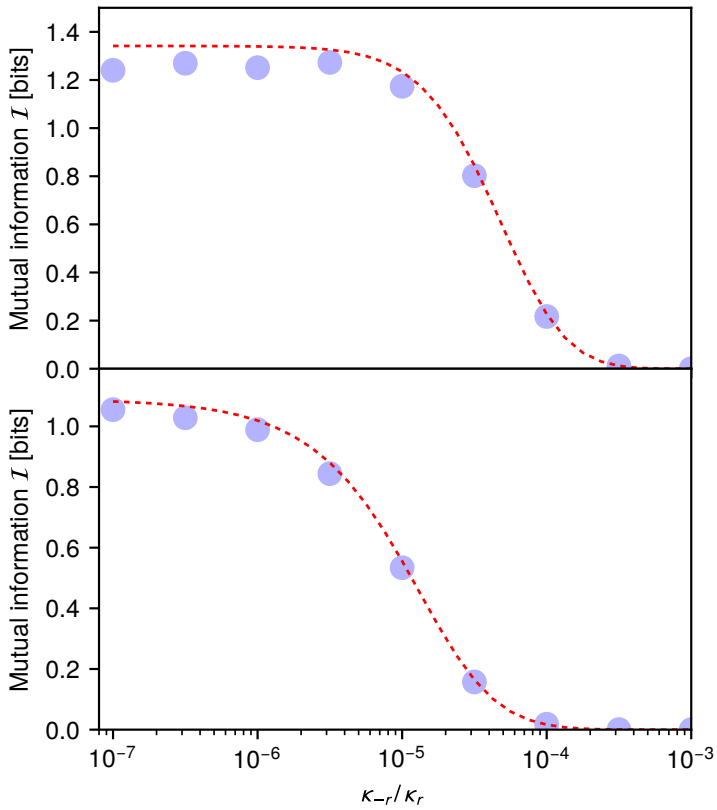

FIG. S2. The mutual information $\mathcal{I}$ for the enzymatic push-pull loop as a function of the reverse-forward rate ratio $\frac{\kappa_{-r}}{\kappa_r}$. The predictions from the chemical Langevin approach (dashed line) are compared against the corresponding KMC simulation results (circles). The parameters sets are as follows (all units are s^{-1} except for the mean populations; molar units have been converted to populations by assuming a cell volume of 30 fL): (top) $\kappa_b = 2.94 \times 10^{-6}$, $\rho_b = 3.68 \times 10^{-7}$, $\kappa_u = 1.58 \times 10^{-2}$, $\rho_u = 4.42 \times 10^{-4}$, $\kappa_r = 12.8$, $\rho_r = 1.34$, $\rho_{-r} = 2.50 \times 10^{-5}$, $F = 2.49 \times 10^{-3}$, $\gamma_k = 2.68 \times 10^{-5}$, $\bar{S} = 614$, and $\bar{P} = 45$; (bottom) $\kappa_b = 2.32 \times 10^{-5}$, $\rho_b = 1.46 \times 10^{-4}$, $\kappa_u = 6.94 \times 10^{-2}$, $\rho_u = 5.48$, $\kappa_r = 0.994$, $\rho_r = 5.05 \times 10^{-2}$, $\rho_{-r} = 2.06 \times 10^{-8}$, $F = 2.46 \times 10^{-2}$, $\gamma_k = 2.65 \times 10^{-4}$, $\bar{S} = 2380$, and $\bar{P} = 127$.

where the C_i are expressed in terms of full network parameters in Eq. (S7). Comparing Eqs. (S15) and (S16) we see that γ_x should be given by

$$\gamma_x = \frac{C_1 \gamma_K}{C_1 + C_2}, \tag{S17}$$

which is the first expression in main text Eq. (2). Similarly the mean production rate of the output in the simplified scheme is $R_0 \bar{X}$. In the full system the mean output production is the average rate at which new phosphorylated substrate is produced via catalysis by the kinase-substrate complex,

$$\kappa_r \bar{S}_K = \kappa_r \frac{F C_2}{\gamma_K C_1} = \kappa_r \frac{C_2}{C_1 + C_2} \bar{X}, \tag{S18}$$

where we have again used Eqs. (S6)-(S7). Comparing Eq. (S18) to $R_0\bar{X}$, we see that R_0 should correspond to

$$R_0 = \kappa_r \frac{C_2}{C_1 + C_2}, \tag{S19}$$

which is the second expression in main text Eq. (2).

B. Validation through kinetic Monte Carlo simulations

To verify that the expressions for γ_x and R_0 derived above are good approximations, we ran KMC simulations for various parameter sets drawn at random from the enzymatic parameter distribution detailed in the SI Sec. IV. For each parameter set the simulation was run long enough after reaching the stationary state to collect sufficient statistics for both the mean population values and the input autocorrelation function. As described above, these allow us to calculate γ_x and R_0. The simulation results are compared against the approximation from Eqs. (S17) and (S19) in Fig. S3. The agreement is excellent for both quantities, across the entire range of γ_x and R_0 values. Thus we can confidently use the simple analytical expressions of Eqs. (S17) and (S19) to predict γ_x and R_0 for any given parameter set.

C. Relating maximum bandwidth, minimum ATP consumption rate, and mutual information via Wiener-Kolmogorov optimal noise filter theory

One of the benefits of the approximate relation between the full system and the two-species model described in Sec. IIIA is that it allows us to use results from the two-species case to make predictions for the behavior of the kinase-phosphatase push-pull loop. The two-species model has been analyzed in detail in Refs. [2, 5], where it was shown to be able to map onto a Wiener-Kolmogorov optimal noise filter. The error E from Eq. (S2) for the two-species case can be evaluated in closed form as [2]:

$$E = 1 - \frac{\gamma_y^2 R_0}{(\gamma_x + \gamma_y)^2} \left[\gamma_y + R_0 \frac{\gamma_y}{\gamma_y + \gamma_x} \right]^{-1}. \tag{S20}$$

It achieves its minimum value (hence maximizing the mutual information $\mathcal{I}$) when the following condition is fulfilled:

$$\gamma_y = \gamma_x \sqrt{1 + \Lambda}, \tag{S21}$$

where $\Lambda = R_0/\gamma_x$. The corresponding minimum E, where the system behaves like an optimal Wiener-Kolmogorov (WK) noise filter is given by:

$$E_{\mathrm{WK}} = \frac{2}{1 + \sqrt{1 + \Lambda}}. \tag{S22}$$

Interestingly, this remains the bound even if we generalize the output production term $R_0 X(t)$ to be nonlinear in $X(t)$ [2]. Using the relation between E and $\mathcal{I}$ in Eq. (S2), we can translate the bound $E \geq E_{\mathrm{WK}}$ into an equivalent statement that $\gamma_x \leq \gamma_x^{\max}$ at a given value

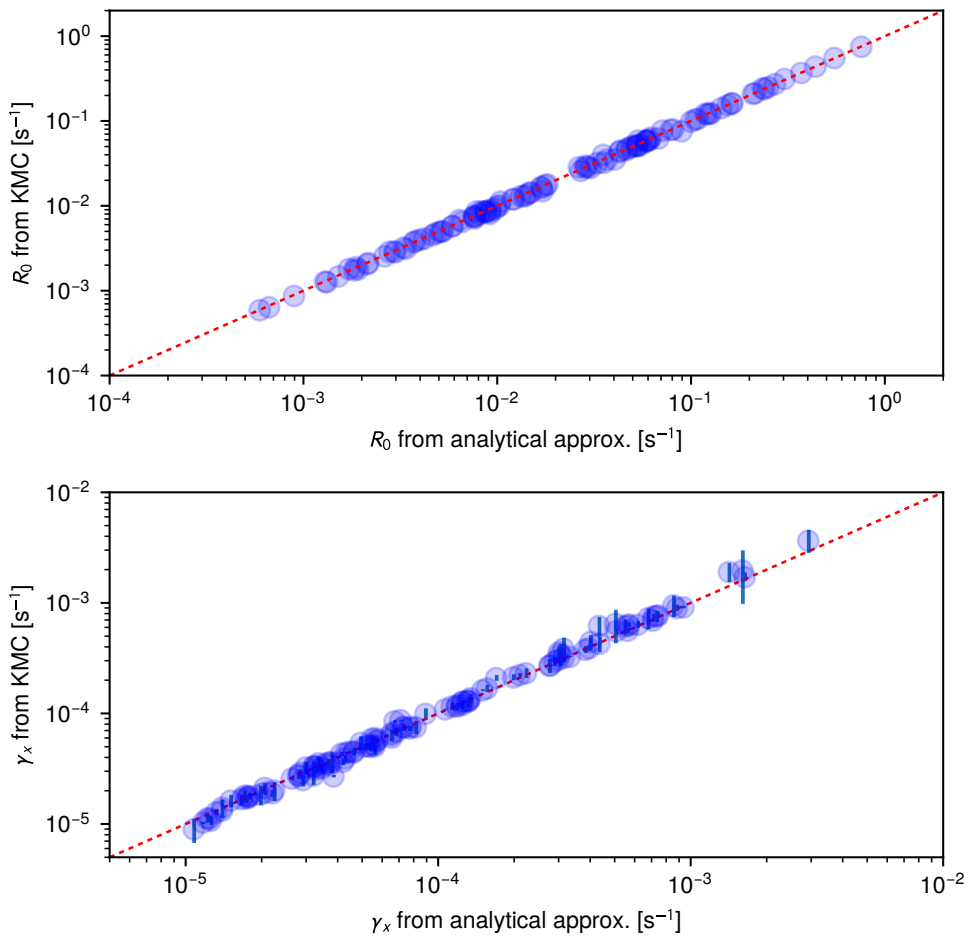

FIG. S3. Comparison of the simple analytical approximations for R_0 from Eq. (S19) (top) and γ_x from Eq. (S17) (bottom) versus KMC simulation results. Each point corresponds to a parameter set drawn randomly from the enzymatic parameter distribution described in SI Sec. IV. The red dashed line corresponds to perfect agreement. Error bars for R_0 are smaller than the symbol size, and hence not indicated in the figure.

of mutual information $\mathcal{I}$. The value of $\gamma_x^{\max}$ is shown in main text Eq. (5):

$$\gamma_x^{\max} = \frac{R_0}{4^{\mathcal{I}+1}(4^{\mathcal{I}}-1)}. \tag{S23}$$

As shown in main text Figs. 3G-I, the above $\gamma_x^{\max}$ expression provides an excellent approximate upper bound on the γ_x^{high} values calculated for the full enzymatic system. Even though the effective two-species model lacks reverse rates, it provides a useful tool for deriving this bound, since the maximum bandwidth is achieved when the reverse rates are negligible (large $\Delta\mu$).

As mentioned in the discussion around main text Eq. (7), the expression for $\gamma_x^{\max}$ in Eq. (S23) also has an alternative interpretation. This gives the minimum production rate

$R_0^{\min}$ necessary to achieve mutual information $\mathcal{I}$ at a certain bandwidth γ_x^{high}:

$$R_0^{\min} = 4^{\mathcal{I}+1}(4^{\mathcal{I}}-1)\gamma_x^{\text{high}}. \tag{S24}$$

By relating R_0 in turn to the ATP consumption rate $A = \kappa_r \overline{S}_K$, we can convert Eq. (S24) into an expression for the minimum necessary ATP consumption rate $A_{\min}$. To accomplish this, note that A can be rewritten as:

$$A = \kappa_r \overline{K}\frac{C_2}{C_1} = \kappa_r \overline{K}\frac{R_0}{\kappa_r - R_0}, \tag{S25}$$

where we have used Eqs. (S6) and (S19). Finally, taking advantage of the fact that typically $\kappa_r \gg R_0$ for the parameter distributions of interest, we make the approximation $A \approx R_0\overline{K}$. This allows us to derive main text Eq. (8):

$$A_{\min} \approx R_0^{\min}\bar{K} = 4^{\mathcal{I}+1}(4^{\mathcal{I}}-1)\gamma_x^{\text{high}}\overline{K}. \tag{S26}$$

D. Optimality with respect to time-offset mutual information

The two-species optimality calculations of the previous section can be generalized to investigate mutual information $\mathcal{I}_\alpha$ between input $X(t)$ and output $Y(t+\alpha)$ with a time offset α. Following the formalism of Ref. [5], the Wiener-Kolmogorov minimal error $E_{\text{WK},\alpha}$ for arbitrary α is

$$E_{\text{WK}}(\alpha) = \begin{cases} 1 - \frac{e^{-2\alpha(\sqrt{\Lambda+1}+1)\gamma_x}\left(\Lambda e^{\alpha\sqrt{\Lambda+1}\gamma_x} - 2\left(\sqrt{\Lambda+1}-1\right)e^{\alpha\gamma_x}\right)^2}{\Lambda\left(\sqrt{\Lambda+1}-1\right)^2} & \alpha \geq 0 \\ \frac{2e^{\alpha\gamma_x}\left(\cosh(\alpha\gamma_x) - \sqrt{\Lambda+1}\sinh(\alpha\gamma_x)\right)}{\sqrt{\Lambda+1}+1} & \alpha < 0 \end{cases}. \tag{S27}$$

For $\alpha = 0$ this expression reduces to E_{WK} from Eq. (S22). The associated mutual information $\mathcal{I}_{\text{WK}}(\alpha) = -(1/2)\ln_2 E_{\text{WK}}(\alpha)$ varies with α in a manner qualitatively similar to main text Fig. 4A. It peaks at a value $\alpha = \alpha_m^{\text{WK}} > 0$ given by:

$$\alpha_m^{\text{WK}} = \frac{\ln\left(\frac{2\Lambda - 2\sqrt{\Lambda+1}+2}{\Lambda}\right)}{\left(\sqrt{\Lambda+1}-1\right)\gamma_x}. \tag{S28}$$

The mutual information $\mathcal{I}_{\text{WK}}(\alpha_m^{\text{WK}})$ at the peak can be expressed as

$$\mathcal{I}_{\text{WK}}(\alpha_m^{\text{WK}}) = -\frac{1}{2\ln 2}\ln\left(1 - \frac{4^{\frac{1}{1-\sqrt{\Lambda+1}}}\Lambda\left(\frac{1}{\sqrt{\Lambda+1}}+1\right)^{\frac{2\left(\sqrt{\Lambda+1}+1\right)}{\Lambda}}}{\Lambda+1}\right). \tag{S29}$$

Since the expression in Eq. (S29) is somewhat unwieldy, we can simplify it using a close upper bound,

$$\mathcal{I}_{\text{WK}}(\alpha_m^{\text{WK}}) \leq \mathcal{I}^{\max}(\Lambda) \equiv \frac{\ln\left[\frac{1}{2}\left(1+\sqrt{1+\Lambda/(\ln 2)^2}\right)\right]}{2\ln 2}. \tag{S30}$$

Parameter x	Unit	$\log_{10}\tilde{x}$	σ_x	Data source
direct fits to database values:				
kinase/phosphatase concentrations [S], [P]	[M]	-7.93	0.84	PaxDb [8]
Michaelis constants $K_M^{\rm kin}$, $K_M^{\rm pho}$	[M]	-4.26	1.21	Sabio-RK [9]
specificity ratios $\kappa_r/K_M^{\rm kin}$, $\rho_r/K_M^{\rm pho}$	[M^{-1} s^{-1}]	3.86	1.19	Sabio-RK [9]
reaction rates κ_r, ρ_r	[s^{-1}]	-0.04	1.16	Sabio-RK [9]
results of joint fitting:				
reaction rates κ_r, ρ_r	[s^{-1}]	-0.06	1.18	joint fit
binding rates κ_b, ρ_b	[M^{-1} s^{-1}]	3.94	1.12	joint fit
dissociation constants $K_D^{\rm kin}$, $K_D^{\rm rho}$	[M]	-7.00	1.31	joint fit

TABLE S1. Results of log-normal fits to various kinase/phosphatase enzymatic parameters. For each fit the mean $\log_{10}\tilde{x}$ and standard deviation σ_x are listed. The top rows of the table correspond to individual fits to parameters collected from the PaxDb and Sabio-RK databases. The bottom rows show the results of a joint fit, described in the text of SI Sec. IV.

The bound $\mathcal{I}^{\max}(\Lambda)$ is never more than ≈ 0.04 bits above $\mathcal{I}_{\rm WK}(\alpha_m^{\rm WK})$ for any Λ, and converges to $\mathcal{I}_{\rm WK}(\alpha_m^{\rm WK})$ in the limit $\Lambda \to \infty$. Hence for all practical purposes we can use Eq. (S30) in place of Eq. (S29) to estimate the bound on the mutual information, as we do in main text Eq. (6).

IV. ENZYMATIC PARAMETER DISTRIBUTION

Earlier surveys of enzymatic kinetic parameters in Refs. [6, 7], over broader classes than just kinases and phosphatases, showed that their distributions could be approximately described by log-normal distributions. For a given parameter x, we will denote this as $\log_{10} x \sim \mathcal{N}(\log_{10}\tilde{x}, \sigma_x^2)$, or in other words that the base-10 logarithm of x is distributed according to a normal distribution with mean $\log_{10}\tilde{x}$ and standard deviation σ_x. The value $\tilde{x}$ is the median of the resulting log-normal distribution for x.

For our work the focus is on kinases and phosphatases, and we are interested in looking at the push-pull loop signaling behavior over the entire distribution of biologically plausible parameters. The parameter data we collected, summarized in the histograms of main text Fig. 2, had far more representation of kinases than phosphatases, which is a well known limitation of the existing experimental literature. Despite this sampling issue, the orders of magnitude spanned by phosphatase parameters were comparable to those of the kinases. For each parameter type, we thus decided to fit both types of enzyme with a single overall distribution, based on pooling of all the available kinase and phosphatase data together. The data available from the databases took the forms listed below (all raw data and the files used to process it are included in the Github repository associated with the manuscript). The mean $\log_{10}\tilde{x}$ and standard deviation σ_x values from the log-normal fits for the different parameter classes are listed in the first four rows of Table S1.

Enzymatic data:

- Mean substrate $[S]$ and phosphatase $[P]$ concentrations, where the substrate is taken

to be a kinase [main text Fig. 2A]. These numbers were derived from the PaxDb protein abundance database [8], taking advantage of UnitProt gene ontology associations to focus on just kinases and phosphatases in signal transduction pathways [10]. Each PaxDb data entry is in terms of ppm (parts per million) of abundance, relative to the total number of proteins in the cell. To convert from ppm to molar concentrations, we looked at data from human cells (which had the best representation in the database), and used the estimated total concentration of 2.7×10^6 proteins per μm^3 for human cells [11]. The latter concentration corresponds to 4.48×10^{-3} M. If y is the abundance in ppm units, then $4.48(y/10^6) \times 10^{-3}$ M is the corresponding molar concentration. Note that total concentrations are very similar across many different types of species [11], so there should not be a strong species-dependence in the analysis. For example the same analysis in mouse cells rather than human ones yields quantitatively similar results: a mean kinase/phosphatase concentration $10^{-8.31}$ M (versus $10^{-7.93}$ M in human cells), and a log-normal standard deviation of 1.03 (versus 0.84 in human cells).

- Reaction parameters [main text Fig. 2B-D]. These values were taken from the Sabio-RK database [9], where they were most often available in the following forms: Michaelis constants $K_M^{\text{kin}} = (\kappa_r + \kappa_u)/\kappa_b$, $K_M^{\text{pho}} = (\rho_r + \rho_u)/\rho_b$ for the kinase/phosphatase (main text Fig. 2B), the corresponding specificity ratios $\kappa_r/K_M^{\text{kin}}$, ρ_r/K_M^{pho} (main text Fig. 2C), and the reaction rates κ_r and ρ_r (main text Fig. 2D). The resulting distributions were entirely consistent (though slightly narrower) with the distributions for the same parameter types analyzed in Ref. [6], which considered all enzymes (not just kinases and phosphatases).

Note that the six reaction parameter types that were collected from the Sabio-RK database (K_M^{kin}, K_M^{pho}, $\kappa_r/K_M^{\text{kin}}$, ρ_r/K_M^{pho}, κ_r, ρ_r) are not directly in the form that we need to calculate push-pull loop signaling properties. For the latter we would like to know (κ_b, ρ_b, κ_u, ρ_u, κ_r, ρ_r), or equivalently (κ_b, ρ_b, K_D^{kin}, K_D^{pho}, κ_r, ρ_r). Here the dissociation constants are defined as $K_D^{\text{kin}} = \kappa_u/\kappa_b$ and $K_D^{\text{pho}} = \rho_u/\rho_b$. Let us denote the parameter vector (κ_b, ρ_b, K_D^{kin}, K_D^{pho}, κ_r, ρ_r) as $\mathbf{v}$, with components v_α, $\alpha = 1, \ldots, 6$. We would like to find a joint distribution for $\mathbf{v}$ that is self-consistent with the individual log-normal distributions for the alternative parameter types fitted directly from the database values (first 4 rows of Table S1). We will assume the simplest form for the joint distribution Φ: a product of individual log-normal distributions for each parameter v_α, with median values $\tilde{v}_\alpha$ and standard deviations σ_α:

$$\Phi(\mathbf{v}) = \prod_{\alpha=1}^{6} \frac{1}{v_\alpha \ln(10)\sqrt{2\pi\sigma_\alpha^2}} \exp\left(-\frac{(\log_{10} v_\alpha - \log_{10} \tilde{v}_\alpha)^2}{2\sigma_\alpha^2}\right). \tag{S31}$$

Note that the $v_\alpha \ln(10)$ term in the denominator of the prefactor comes from the Jacobian due to the variable change between $\log_{10} v_\alpha$ and v_α. This ensures that the probability is properly normalized: $\int_0^\infty \prod_\alpha dv_\alpha\, \Phi(\mathbf{v}) = 1$. As explained above, kinase and phosphatase parameters are assumed to be drawn from the same distributions, so we enforce that $\tilde{v}_1 = \tilde{v}_2$, $\tilde{v}_3 = \tilde{v}_4$, $\tilde{v}_5 = \tilde{v}_6$, and analogously for the standard deviations σ_α. This leaves six distinct values that determine the distribution: $\tilde{v}_1$, $\tilde{v}_3$, $\tilde{v}_5$, σ_1, σ_3, σ_5.

To estimate these six distribution parameters, we use the following iterative numerical fitting procedure. We start with a guess for ($\tilde{v}_1$, $\tilde{v}_3$, $\tilde{v}_5$, σ_1, σ_3, σ_5) and then draw 10^4

parameter sets $\mathbf{v}$ from the resulting distribution $\Phi(\mathbf{v})$. For each parameter set we can calculate the alternative parameter types ($K_M^{\rm kin}$, $K_M^{\rm pho}$, $\kappa_r/K_M^{\rm kin}$, $\rho_r/K_M^{\rm pho}$, κ_r, ρ_r). We then fit the resulting 10^4 values for these alternative types to individual log-normal distributions, and compare the means and standard deviations to the empirical results in the top half of Table S1. The sum of the relative absolute errors between the new joint fit values and the empirical results for the means / standard deviations is our overall goodness-of-fit measure. We perturb our guess for ($\tilde{v}_1$, $\tilde{v}_3$, $\tilde{v}_5$, σ_1, σ_3, σ_5) and accept the perturbation if it improves the goodness-of-fit. This procedure is iterated until convergence. The results of this joint fit are shown in the bottom half of Table S1. The joint fit predictions for the binding rate (κ_b, ρ_b) and dissociation constant ($K_D^{\rm kin}$, $K_D^{\rm pho}$) distributions are consistent with earlier estimates of these parameters in specific kinase/phosphatase systems [12]. As another consistency check, the joint fit distribution for the reaction rates (κ_r, ρ_r) is nearly identical to the individual empirical fit based on the Sabio-RK database values.

Finally we note that the simple joint distribution $\Phi(\mathbf{v})$ in Eq. (S31) is by construction too broad: it may produce the correct marginal distributions for quantities collected from the Sabio-RK database, but it ignores any correlations between those individual parameters that may be present in natural systems. Estimating these correlations from the existing database entries is quite challenging, because relatively few entries have a complete list of all the parameters of interest. Hence, as explained in the main text, we take $\Phi(\mathbf{v})$ to be effectively a superset: it should contain the true, presumably narrower, biological distribution plus parameter sets that are less likely to be observed in nature. A convenient aspect of this interpretation is that any collective conclusion we draw from the entire distribution $\Phi(\mathbf{v})$ should also be true for the subset of biological parameters. Moreover we can thus explore a larger design space (potentially available for evolution) than what we currently observe in modern biological systems.

V. RESULTS FOR ALTERNATIVE INPUT KINASE CONCENTRATIONS

The results in main text Fig. 3D-F were for a mean input kinase concentration $[K] = 5$ nM. In Fig. S4 we show the analogous results for two different choices: $[K] = 0.5$ nM (left column) and $[K] = 50$ nM (right column). The main conclusions remain unchanged: the physiological $\Delta\mu$ range (highlighted in pink) is always just above the upper edge of the $\gamma_x^{\rm high}$ cloud, and the number of available parameter sets decreases rapidly as the mutual information $\mathcal{I}$ is increased.

VI. ANALYSIS OF THE PBS2-HOG1 PUSH-PULL LOOP IN YEAST

To illustrate our theoretical framework in a concrete biological example, let us consider a kinase-phosphatase loop from one of the most extensively studied signaling pathways: the Hog1 mitogen-activated protein kinase (MAPK) pathway that allows yeast to adapt to extracellular osmotic changes [13–15]. We will focus in particular on the final portion of the pathway, where the active (phosphorylated) kinase Pbs2pp catalyzes the conversion of inactive Hog1 into phosphorylated Hog1pp. The latter protein is interchanged quickly between cytoplasm and nucleus, where it regulates a variety of responses to osmotic stress. Hog1pp is dephosphorylated by a combination of phosphatases Ptp2 (mainly in the nucleus) and Ptp3 in the cytoplasm [16]. Thus Pbs2pp will play the role of K in our model, Hog1

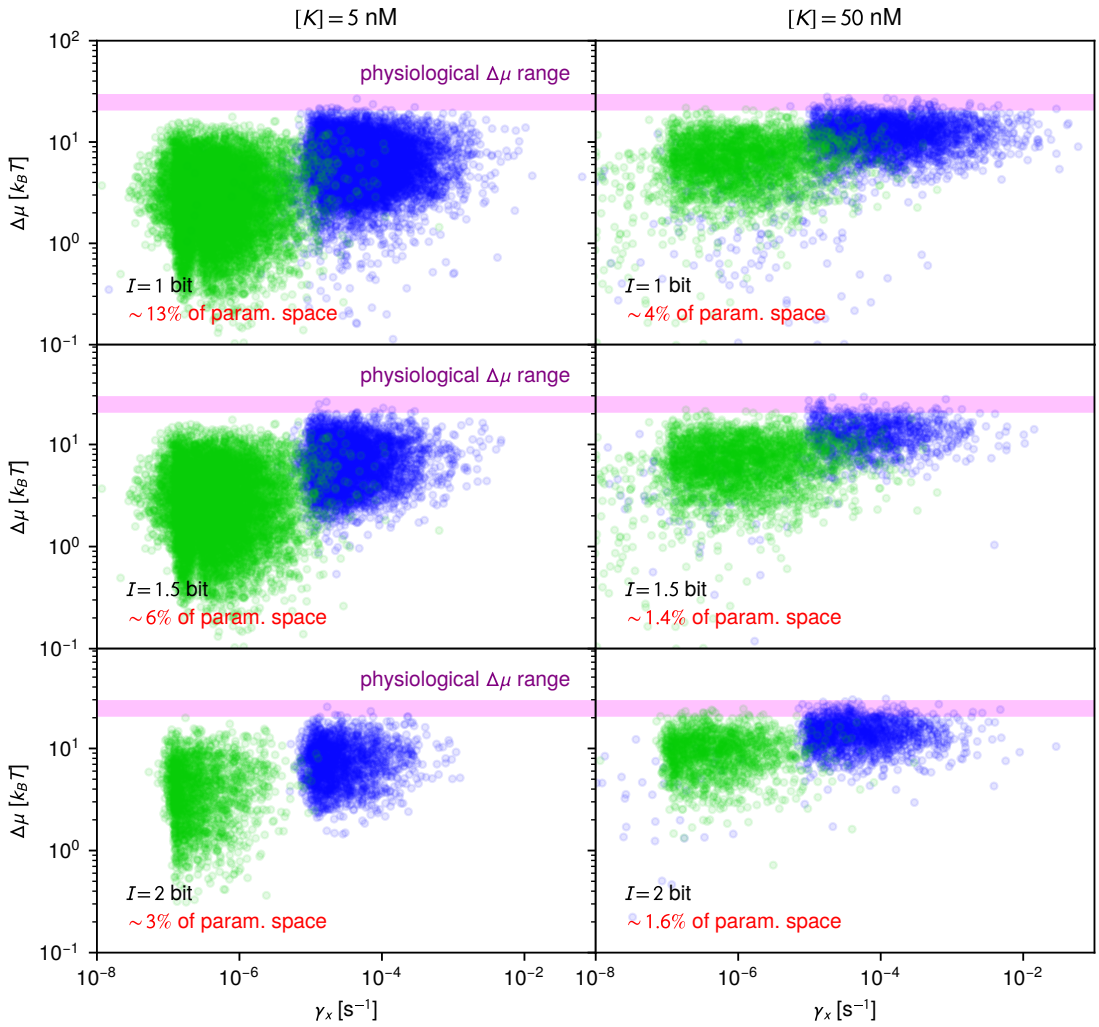

FIG. S4. Analogous to main Figure 3D-F, except for input kinase concentration $[K] = 5$ nM (left column) and 50 nM (right column). The rows correspond to mutual information $\mathcal{I} = 1$, 1.5, and 2 bits respectively. The probabilities of successfully drawing a parameter set that achieves the specified $\mathcal{I}$ value are shown in red in panel.

will be S, Hog1pp will be S^*, and Ptp2/Ptp3 will be P. To parameterize our model, we start with a more detailed theoretical description of the entire pathway developed by Zi *et al.* [13]. A key appeal of this work is that its parameters were carefully fit to extensive experimental data from yeast cells exposed to different time series of external salt shocks in microfluidic experiments [14]. However since the parameters of Zi *et al.* are not expressed in the same form as the enzymatic reaction rates of our model, we do have to convert from their framework to ours, as described below.

A. Parameter estimation based on earlier literature

Ref. [13] explicitly distinguishes between the concentration of Hog1 and Hog1pp in the cytoplasm and nucleus, denoted with c and n superscripts respectively: $[\text{Hog1}^{\text{c}}]$, $[\text{Hog1}^{\text{n}}]$,

[Hog1pp$^{\rm c}$], [Hog1pp$^{\rm n}$]. If we are interested in the average concentrations overall, we can denote these as:

$$[S] \equiv \frac{[\text{Hog1}^{\rm c}]V_c + [\text{Hog1}^{\rm n}]V_n}{V_c + V_n}, \quad [S^*] \equiv \frac{[\text{Hog1pp}^{\rm c}]V_c + [\text{Hog1pp}^{\rm n}]V_n}{V_c + V_n}, \tag{S32}$$

where V_c and V_n are the volumes of the cytoplasm and nucleus respectively, taken to have a ratio of $V_n/V_c = 0.14$ [13]. Eq. (S32) also implies:

$$\frac{d[S]}{dt} = \frac{d[\text{Hog1}^{\rm c}]}{dt} f + \frac{d[\text{Hog1}^{\rm n}]}{dt}(1-f), \quad \frac{d[S^*]}{dt} = \frac{d[\text{Hog1pp}^{\rm c}]}{dt} f + \frac{d[\text{Hog1pp}^{\rm n}]}{dt}(1-f), \tag{S33}$$

where $f = V_c/(V_c + V_n) = 0.88$. As a simplification of Eq. (S32), we note in Ref. [13] import and export of the Hog1 proteins is fast relative to other reactions, and for a given input level the system rapidly reaches a stationary state with [Hog1$^{\rm n}$]≈[Hog1$^{\rm c}$]≈ $[S]$, [Hog1pp$^{\rm n}$]≈[Hog1pp$^{\rm c}$]≈ $[S^*]$.

We can now look at individual reactions that contribute to the time derivatives on the right-hand sides of Eq. (S33) and find their analogues in our model. For example the phosphorylation step that converts Hog1$^{\rm c}$ to Hog1pp$^{\rm c}$ is expressed in Ref. [13] as an effective second order reaction of the form $K_{\rm pho}^{\rm Hog1}$[Pbs2pp][Hog1$^{\rm c}$], with rate constant $K_{\rm pho}^{\rm Hog1} = 11.2\ \mu\text{M}^{-1}\cdot\text{min}^{-1}$. This contributes positively to $d[\text{Hog1pp}^{\rm c}]/dt$ and with a minus sign to $d[\text{Hog1}^{\rm c}]/dt$, and so leads to contributions magnitude $fK_{\rm pho}^{\rm Hog1}$[Pbs2pp][Hog1$^{\rm c}$] to the right-hand sides of Eq. (S33). Note that even though activation of Hog1 is actually a double phosphorylation (of a threonine and tyrosine residue), the entire process in this case can be well approximated through a single rate constant.

In our model the conversion of S to S^* occurs through the intermediate state S_K. However if we want to compare to the phosphorylation step of Ref. [13] in order to match parameters, we can look at the deterministic contribution to the dynamics (ignoring fluctuations) in the Michaelis-Menten approximation for enzyme kinetics [17]. In this picture the phosphorylation reaction contributes to $d[S]/dt$ and $d[S^*]/dt$ through a term of magnitude $\kappa_r[K][S]/(K_M^{\rm kin} + [S]) \approx (\kappa_r/K_M^{\rm kin})[K][S]$, where the last simplification is valid when $K_M^{\rm kin} \gg [S]$. If we compare $(\kappa_r/K_M^{\rm kin})[K][S]$ to $fK_{\rm pho}^{\rm Hog1}$[Pbs2pp][Hog1$^{\rm c}$], noting that $[K] = $[Pbs2pp] and $[S] \approx$ [Hog1$^{\rm c}$], we can make the following identification:

$$\frac{\kappa_r}{K_M^{\rm kin}} \approx fK_{\rm pho}^{\rm Hog1} = 1.64 \times 10^5\ \text{M}^{-1}\text{s}^{-1}. \tag{S34}$$

The dephosphorylation steps in Ref. [13] are modeled as two pseudo-first-order reactions: conversion of Hog1pp$^{\rm c}$ to Hog1$^{\rm c}$ with rate $K_{\rm depho}^{\rm Hog1pp^c}$[Hog1pp$^{\rm c}$], and the conversion of Hog1pp$^{\rm n}$ to Hog1$^{\rm n}$ with rate $K_{\rm depho}^{\rm Hog1pp^n}$[Hog1pp$^{\rm n}$]. The pseudo-first-order rate constants are given by: $K_{\rm depho}^{\rm Hog1pp^c} = 0.0906\ \text{min}^{-1}$ and $K_{\rm depho}^{\rm Hog1pp^c} = 4.14\ \text{min}^{-1}$. These reactions will lead to contributions of magnitude $(fK_{\rm depho}^{\rm Hog1pp^c}[\text{Hog1pp}^{\rm c}] + (1-f)K_{\rm depho}^{\rm Hog1pp^n}[\text{Hog1pp}^{\rm n}])$ to the right-hand sides of Eq. (S33). In our model (using a similar Michaelis-Menten approximation to the one described above, with $K_M^{\rm pho} \gg [P]$), the analogous expression for dephosphorylation is effectively a second-order reaction with rate $(\rho_r/K_M^{\rm pho})[P][S^*]$. Comparison of the two ex-

Parameter	Value	Data source
substrate concentration $[S]$	0.38 μM	Hog1 abundance from Ref. [18]
phosphatase concentration $[P]$	0.058 μM	Ptp2/Ptp3 abundance from Ref. [18]
kinase specificity ρ_r/K_M^{kin}	1.64×10^5 M^{-1}s^{-1}	analysis of Ref. [13] model fit to experiments of Ref. [14]
phosphatase specificity ρ_r/K_M^{pho}	1.69×10^5 M^{-1}s^{-1}	analysis of Ref. [13] model fit to experiments of Ref. [14]

TABLE S2. Summary of parameters for the yeast Pbs2/Hog1 system estimated from earlier literature.

pressions, using the approximation [Hog1pp^{n}]≈[Hog1pp^{c}]≈ $[S^*]$, leads to the identification:

$$\frac{\rho_r}{K_M^{\text{pho}}} \approx [P]^{-1} \left(f K_{\text{depho}}^{\text{Hog1pp}^{\text{c}}} + (1-f) K_{\text{depho}}^{\text{Hog1pp}^{\text{n}}} \right) = 1.69 \times 10^5 \text{ M}^{-1}\text{s}^{-1}. \tag{S35}$$

Here we set $[P] = 0.058$ μM as an average measure of phosphatase concentrations, to facilitate the conversion from pseudo-first-order to second-order rate constants. The value of $[P]$ is based on estimates of the concentrations of the two phosphatases in yeast from Ref. [18]: 0.049 μM for Ptp3 in the cytoplasm, and 0.067 μM for Ptp2 in the nucleus, where we have used $V_c = f(V_c + V_n)$, $V_n = (1-f)(V_c + V_n)$ and $V_c + V_n \approx 30$ fL [13, 19] to convert from populations to concentrations. Since the concentrations were of similar scale, we let $[P]$ be the mean of the two values.

As a consistency check to make sure the final estimates of the specificity ratios $\kappa_r/K_M^{\text{kin}}$ and ρ_r/K_M^{pho} in Eqs. (S34)-(S35) are biologically plausible, we can compare them with the distribution of these ratios among kinases/phosphatases from the Sabio-RK database in main text Fig. 2C. The values for the Hog1/Pbs2 system are not unusual, and lie near the higher end of the range, at about the 0.87 quantile. The final parameter value we can estimate from the literature is the mean Hog1 concentration $[S] = 0.38$ μM, based on the abundance reported in Ref. [18].

B. Estimation of remaining parameters

Based on the above analysis, we have estimates for four quantities in the Pbs2/Hog1 system drawn from the earlier literature: $\kappa_r/K_M^{\text{kin}}$, ρ_r/K_M^{pho}, $[S]$, $[P]$. These are summarized in Table S2. The relationship of the enzymatic reaction/binding/unbinding rate parameters

to the estimated values then takes the form:

$$\begin{aligned}
\kappa_r &= \left(\frac{\kappa_r}{K_M^{\rm kin}}\right) K_M^{\rm kin} = (1.64\times 10^5 {\rm M}^{-1}{\rm s}^{-1}) K_M^{\rm kin}, \\
\kappa_b &= \left(\frac{\kappa_r}{K_M^{\rm kin}}\right) \frac{K_M^{\rm kin}}{K_M^{\rm kin}-K_D^{\rm kin}} = (1.64\times 10^5 {\rm M}^{-1}{\rm s}^{-1}) \frac{K_M^{\rm kin}}{K_M^{\rm kin}-K_D^{\rm kin}}, \\
\kappa_u &= \left(\frac{\kappa_r}{K_M^{\rm kin}}\right) \frac{K_M^{\rm kin}K_D^{\rm kin}}{K_M^{\rm kin}-K_D^{\rm kin}} = (1.64\times 10^5 {\rm M}^{-1}{\rm s}^{-1}) \frac{K_M^{\rm kin}K_D^{\rm kin}}{K_M^{\rm kin}-K_D^{\rm kin}}, \\
\rho_r &= \left(\frac{\rho_r}{K_M^{\rm pho}}\right) K_M^{\rm pho} = (1.69\times 10^5 {\rm M}^{-1}{\rm s}^{-1}) K_M^{\rm pho}, \\
\rho_b &= \left(\frac{\rho_r}{K_M^{\rm rho}}\right) \frac{K_M^{\rm rho}}{K_M^{\rm rho}-K_D^{\rm rho}} = (1.69\times 10^5 {\rm M}^{-1}{\rm s}^{-1}) \frac{K_M^{\rm rho}}{K_M^{\rm rho}-K_D^{\rm rho}}, \\
\rho_u &= \left(\frac{\rho_r}{K_M^{\rm rho}}\right) \frac{K_M^{\rm rho}K_D^{\rm rho}}{K_M^{\rm rho}-K_D^{\rm rho}} = (1.69\times 10^5 {\rm M}^{-1}{\rm s}^{-1}) \frac{K_M^{\rm rho}K_D^{\rm rho}}{K_M^{\rm rho}-K_D^{\rm rho}},
\end{aligned} \tag{S36}$$

The above parameters depend on the values of $K_M^{\rm kin}$, $K_M^{\rm pho}$, $K_D^{\rm kin}$, $K_D^{\rm kin}$. While we do not know what these are for the Pbs2/Hog1 system, we can draw their values from the corresponding empirical log-normal distributions described in Table S1. By repeating the draw many times, we can check how our final optimality analysis (see below) depends on the precise values of the unknown parameters. As it turns out the dependence of R_0, $\gamma_x^{\rm high}$ and $R_0^{\rm min}$ on the unknown values is quite weak, and we will be able to make robust estimates for these quantities. In the cases of $K_M^{\rm kin}$ and $K_M^{\rm pho}$, we constrain the random draw from their log-normal distributions to enforce $K_M^{\rm kin} \geq 100[S]$ and $K_M^{\rm pho} \geq 100[P]$. This ensures self-consistency with the assumptions $K_M^{\rm kin} \gg [S]$ and $K_M^{\rm pho} \gg [P]$, which were used in the previous subsection to match the form of the phosphorylation / dephosphorylation reactions between Ref. [13] and our model. The final two parameters are the reverse reaction rates κ_{-r} and ρ_{-r}. Since we do not have any experimental estimates of these for the Pbs2/Hog1 system, we assume that the physiological value of $\Delta\mu$ in yeast (around 21 k_BT [19]) is sufficiently high that κ_{-r} and ρ_{-r} are negligible under normal conditions.

C. Bandwidth and gain

Given the parameter estimation procedure described above, we can calculate $\gamma_x^{\rm high}$, R_0, $R_0^{\rm min}$ for each draw of the unknown parameters. The results remain within a narrow distribution, relatively insensitive to the values of the unknown parameters. The mean and standard deviations for 50 draws are: $\gamma_x^{\rm high} = (1.22 \pm 0.04)\times 10^{-3}$ s^{-1}, $R_0 = 0.0621 \pm 0.0001$ s^{-1}, $R_0^{\rm min} = 0.059 \pm 0.002$ s^{-1}.

VII. ESTIMATION OF TOTAL RESTING METABOLIC EXPENDITURE

For single-celled organisms, the total resting metabolic expenditure C_T can be estimated by the approach outlined in Ref. [20]. C_T has two contributions: $C_T = C_G + t_r C_M$. Here C_G is the expenditure involved in growth during one generation time t_r, and C_M is the maintenance cost per unit time. Using a large collection of metabolic data from Ref. [21],

covering both prokaryotes and single-celled eukaryotes, one can observe that both C_M and C_T scale approximately linearly with cell volume V, agreeing with the prediction of the bioenergetic growth model of Ref. [20]. The expression for C_T based on the results of these linear fits is [20]:

$$C_T = (2.3 \times 10^{10}\,\mathrm{P/fL})V + (9.2 \times 10^{4}\,\mathrm{P/(s \cdot fL)})t_r V. \tag{S37}$$

where the unit P corresponds to the hydrolysis of a phosphate bond (i.e. the consumption of one ATP or ATP equivalent). Using the main text values of $V = 30$ fL and $t_r = 3600$ s, we get $C_T = 7.0 \times 10^{11}$ P.

---

[1] Qian, H. Cooperativity and specificity in enzyme kinetics: a single-molecule time-based perspective. *Biophys. J.* **95**, 10–17 (2008).

[2] Hinczewski, M. & Thirumalai, D. Cellular signaling networks function as generalized Wiener-Kolmogorov filters to suppress noise. *Phys. Rev. X* **4**, 041017 (2014).

[3] Gillespie, D. T. The chemical Langevin equation. *J. Chem. Phys.* **113**, 297–306 (2000).

[4] Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81**, 2340–2361 (1977).

[5] Hathcock, D., Sheehy, J., Weisenberger, C., Ilker, E. & Hinczewski, M. Noise filtering and prediction in biological signaling networks. *IEEE Trans. Mol. Biol. Multi-Scale Commun.* **2**, 16–30 (2016).

[6] Bar-Even, A. *et al.* The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* **50**, 4402–4410 (2011).

[7] Liebermeister, W. & Klipp, E. Biochemical networks with uncertain parameters. *IEE Proc.-Syst. Biol.* **152**, 97–107 (2005).

[8] Wang, M., Herrmann, C. J., Simonovic, M., Szklarczyk, D. & von Mering, C. Version 4.0 of PaxDb: protein abundance data, integrated across model organisms, tissues, and cell-lines. *Proteomics* **15**, 3163–3168 (2015).

[9] Wittig, U. *et al.* SABIO-RK—database for biochemical reaction kinetics. *Nucleic Acids Res.* **40**, D790–D796 (2011).

[10] UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515 (2018).

[11] Milo, R. What is the total number of protein molecules per cell volume? a call to rethink some published values. *Bioessays* **35**, 1050–1055 (2013).

[12] Schoeberl, B., Eichler-Jonsson, C., Gilles, E. D. & Müller, G. Computational modeling of the dynamics of the map kinase cascade activated by surface and internalized EGF receptors. *Nat. Biotech.* **20**, 370–375 (2002).

[13] Zi, Z., Liebermeister, W. & Klipp, E. A quantitative study of the Hog1 MAPK response to fluctuating osmotic stress in Saccharomyces cerevisiae. *PLOS ONE* **5**, e9522 (2010).

[14] Mettetal, J. T., Muzzey, D., Gómez-Uribe, C. & van Oudenaarden, A. The frequency dependence of osmo-adaptation in saccharomyces cerevisiae. *Science* **319**, 482–484 (2008).

[15] Hersen, P., McClean, M. N., Mahadevan, L. & Ramanathan, S. Signal processing by the HOG MAP kinase pathway. *Proc. Natl. Acad. Sci.* **105**, 7165–7170 (2008).

[16] Mattison, C. P. & Ota, I. M. Two protein tyrosine phosphatases, Ptp2 and Ptp3, modulate the subcellular localization of the Hog1 MAP kinase in yeast. *Genes Dev.* **14**, 1229–1235 (2000).

[17] Fersht, A. *Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding* (Macmillan, 1999).

[18] Ghaemmaghami, S. *et al.* Global analysis of protein expression in yeast. *Nature* **425**, 737–741 (2003).

[19] Milo, R. & Phillips, R. *Cell biology by the numbers* (Garland Science, 2015).

[20] Ilker, E. & Hinczewski, M. Modeling the growth of organisms validates a general relation between metabolic costs and natural selection. *Phys. Rev. Lett.* **122**, 238101 (2019).

[21] Lynch, M. & Marinov, G. K. The bioenergetic costs of a gene. *Proc. Natl. Acad. Sci.* **112**, 15690–15695 (2015).