PNAS RESEARCH ARTICLE | BIOPHYSICS AND COMPUTATIONAL BIOLOGY

Trade-offs between cost and information in cellular prediction

Age J. Tjalma[a], Vahe Galstyan[a], Jeroen Goedhart[a], Lotte Slim[a], Nils B. Becker[b], and Pieter Rein ten Wolde[a,1]

Edited by Ned S. Wingreen, Princeton University, Princeton, NJ; received February 23, 2023; accepted August 23, 2023 by Editorial Board Member Mehran Kardar

Living cells can leverage correlations in environmental fluctuations to predict the future environment and mount a response ahead of time. To this end, cells need to encode the past signal into the output of the intracellular network from which the future input is predicted. Yet, storing information is costly while not all features of the past signal are equally informative on the future input signal. Here, we show for two classes of input signals that cellular networks can reach the fundamental bound on the predictive information as set by the information extracted from the past signal: Push–pull networks can reach this information bound for Markovian signals, while networks that take a temporal derivative can reach the bound for predicting the future derivative of non-Markovian signals. However, the bits of past information that are most informative about the future signal are also prohibitively costly. As a result, the optimal system that maximizes the predictive information for a given resource cost is, in general, not at the information bound. Applying our theory to the chemotaxis network of *Escherichia coli* reveals that its adaptive kernel is optimal for predicting future concentration changes over a broad range of background concentrations, and that the system has been tailored to predicting these changes in shallow gradients.

prediction | chemotaxis | sensing | information theory | resource allocation

Significance

Organisms, ranging from bacteria to animals, can predict changes in their environment. To this end, they need to compress the past signal into the dynamics of the internal system from which they predict the future. Yet, storing information is costly while not all features of the past signal are equally predictive. Here, we show for cellular systems that the bits of past information that are most informative about the future are also prohibitively costly, in terms of physical resources like protein copies and energy. Consequently, systems that maximize predictive power under a resource constraint differ from those that maximize this under an information-compression constraint. These findings have implications for the optimal design of both natural and man-made information-processing systems.

Single-celled organisms live in a highly dynamic environment to which they continually have to respond and adapt. To this end, they employ a range of response strategies, tailored to the temporal structure of the environmental variations. When these variations are highly regular, such as the daily light variations, it becomes beneficial to develop a clock from which the time and hence the current and future environment can be inferred (1, 2). In the other limit, when the fluctuations are entirely unpredictable, cells have no choice but to resort to either the strategy of detect-and-respond or the bet-hedging strategy of stochastic switching between different phenotypes (3). Yet arguably the most fascinating strategy lies in between these two extremes. When the environmental fluctuations happen with some regularity, then it becomes feasible to predict the future environment and initiate a response ahead of time. While it is commonly believed that only higher organisms can predict the future, experiments have vividly demonstrated that even single-cell organisms can leverage temporal correlations in environmental fluctuations in order to predict, e.g., future nutrient levels (4, 5).

The ability to predict future signals can provide a fitness benefit (6). The capacity to anticipate changes in oxygen levels (4), or the arrival of sugars or stress signals (5), can increase the growth rate of single-celled organisms; modeling has revealed that prediction can enhance bacterial chemotaxis (7). Yet, a predict-and-anticipate strategy is only advantageous if the cell can reliably predict the future on timescales that are longer than the time it takes to mount a response. What fundamentally limits the accuracy of cellular prediction remains, however, poorly understood.

While the cell needs to predict the future environment, it can only sense the present and remember the past (Fig. 1*A*). Consequently, for a given amount of information, the cell can store about the present and past signal, there is a maximum amount of information it can possibly have about the future (8, 9) (Fig. 1 *D*, I). This *information bound* is determined by the temporal structure of the environmental fluctuations (9, 10) (Fig. 1*B*).

How close cells can come to this bound depends on the design of the intracellular biochemical network that senses and processes the environmental signals (Fig. 1*C*). To maximize the predictive power the cell must use its memory effectively: It should extract only those characteristics from the present and past signal that are most informative about the future (7, 10–12). Whether it can do so, is determined by the topology of the signaling network. Moreover, like any information processing device, biochemical networks require resources to be built and run. Molecular components are needed to construct the network, space is required to accommodate the components, time is needed

Author affiliations: [a]AMOLF, Science Park 104, 1098 XG Amsterdam, The Netherlands; and [b]Theoretical Systems Biology, German Cancer Research Center, 69120 Heidelberg, Germany

Author contributions: P.R.t.W. designed research; A.J.T., V.G., J.G., and L.S. performed research; A.J.T. and V.G. analyzed data; N.B.B. contributed ideas; and A.J.T. and P.R.t.W. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. N.S.W. is a guest editor invited by the Editorial Board.

[1]To whom correspondence may be addressed. Email: tenwolde@amolf.nl.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2303078120/-/DCSupplemental.

Published October 4, 2023.

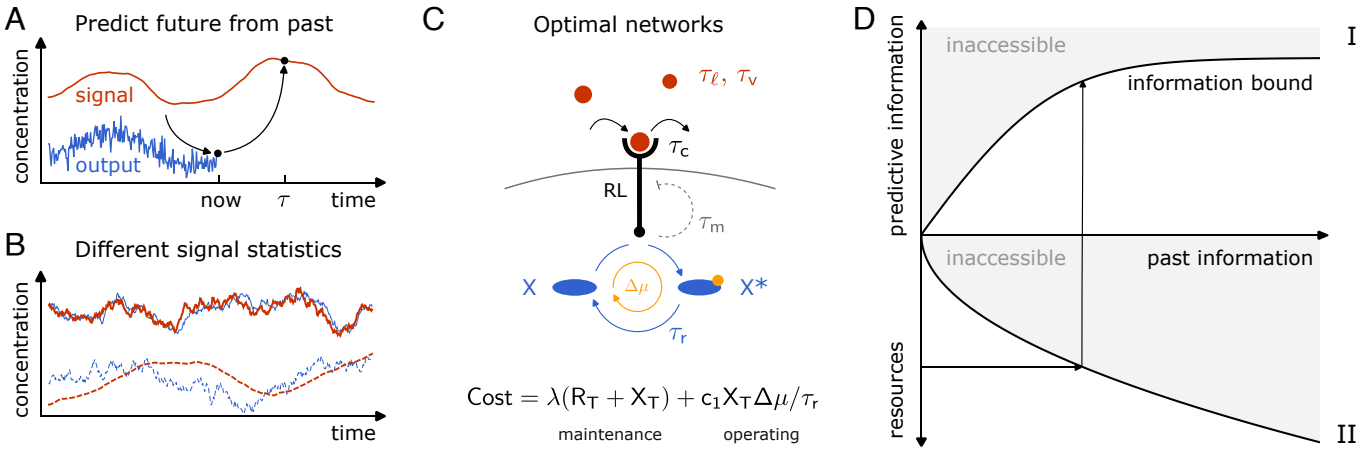

Fig. 1. Cells use biochemical networks to remember the past and predict the future. (*A*) Cells compress the past input into the dynamics of the signaling network from which the future input is then predicted. (*B*) We study Markovian (*Top*, red) and non-Markovian signals (below, red), with outputs (blue) generated via a push–pull and a derivative-taking network, respectively. (*C*) The optimal topology of the network for predicting the future signal depends on the temporal statistics of the input signal. Push–pull networks, consisting of chemical modification cycles or GTPase cycles, can optimally predict the future value of Markovian signals (*B*, *Top*), with correlation time τ_ℓ; derivative-taking networks, like the *E. coli* chemotaxis system, can optimally predict the future derivative of non-Markovian signals (*B*, *Bottom*), with correlation time τ_v. The push–pull network consists of a receptor that drives a downstream phosphorylation cycle. The ligand binds the receptor with a correlation time τ_c. The push–pull network, driven by ATP hydrolysis with free energy $\Delta\mu$, integrates the receptor with an integration time τ_r. The chemotaxis system is a push–pull network, yet augmented with negative feedback on the receptor activity via methylation on a timescale τ_m, as indicated by the dashed gray line. The total resource cost consists of a maintenance cost of receptor and readout synthesis at the growth rate λ, and an operating cost of driving the cycle. (*D*) The predictive information on the future signal I_{pred} is fundamentally bounded by how much information I_{past} it has about the past signal (panel I), which in turn is limited by the resources necessary to build and operate the biochemical network (panel II) (8).

to process the information, and energy is required to synthesize the components and operate the network (13). These resources constrain the design and performance of any biochemical network, and the capacity to sense and process information is no exception (Fig. 1 *D*, II).

The costs and benefits of prediction are connected via the past and predictive information, as first pointed out in the context of neuronal systems (8). For these systems, tremendous progress has been made in characterizing the coding of information (12), and in quantifying the relation between past and predictive information (11). Yet, understanding how the predictive information depends on resource cost is much harder for neuronal systems (14).

In contrast, cellular signaling systems provide a unique opportunity for revealing the resource requirements for prediction. For these systems, we can readily quantify the information processing capacity as a function of the resources that are necessary to build and run them—protein copies, time, and energy (13, 15). Moreover, cells live in a highly dynamic environment, with temporal statistics that are expected to vary markedly. Indeed, signaling networks have distinct topologies, which are likely tailored to the temporal statistics of the environment (7). Cellular systems are thus ideal for elucidating the relationships between future and past information, system design (network topology), and resource constraints. Here, we derive the bound on the prediction precision as set by the information extracted from the past signal for two types of input signals (Fig. 1*B*). We determine how close cellular networks can come to this bound, and how this depends on the topology of the network and the resources available to build and run it.

We find for the two classes of input signals studied, cellular networks exist that can reach the information bound, yet reaching the bound is exceedingly costly. The first class of input signals consists of Markovian signals. Using the Information Bottleneck Method (IBM) (9, 16), we first show that, to the reach the information bound, the system must copy the most recent input signal into the output from which the future input is predicted. Push–pull networks consisting of chemical modification or GTPase cycles, which are ubiquitous in prokaryotic and eukaryotic cells (17, 18), should be able to do so, because they are at heart copying devices (13, 15). Yet, copying the most recent input into the output is extremely costly, because the operating cost, as set by the chemical power to drive the cycle, diverges at high copying speed. More surprisingly, even when the operating cost is negligible, and the total resource cost is dominated by protein synthesis (Fig. 1*C*), the bits of past information that are most informative about the future, are also prohibitively costly. As a result, the past and predictive information can be raised simultaneously by moving away from the bound. The optimal system that maximizes the predictive information for a given resource constraint is, in general, not at the information bound.

Living cells that navigate their environment typically experience signals that are influenced by their own persistent motion, which motivated us to study a simple class of non-Markovian signals. Interestingly, the range of background concentrations that cells can navigate is typically much larger than the concentration change over the signal correlation time, as set by their motion. Our analysis reveals that in such a scenario the optimal kernel that allows the system to reach the information bound on predicting future concentration changes is a perfectly adaptive, derivative-taking kernel, precisely as the bacterium *E. coli* employs (19). We again find, however, that reaching the information bound would be prohibitively costly, because it requires taking an instantaneous derivative. The optimal system that maximizes the predictive information under a resource constraint is therefore not at the information bound, emerging from a trade-off between taking a derivative that is recent and one that is reliable. Finally, computing the past and predictive information directly from experimental data (20) reveals that the *E. coli* chemotaxis system is indeed markedly away from the information bound and that it is optimally designed to predict future concentration changes

in shallow gradients. In this regime, it uses its collected past information for prediction most efficiently.

Results

We focus on cellular signaling systems that respond linearly to changes in the input signal (13, 15, 20–22). These systems not only allow for analytical results, but also describe information transmission often remarkably well (22–25). The output of these systems can be written as

$$x(t) = \int_{-\infty}^{t} dt' k(t-t')\ell(t') + \eta_x(t), \qquad [1]$$

where $k(t)$ is the linear response function, $\ell(t)$ the input signal, and $\eta_x(t)$ describes the noise in the output. We will consider stationary signals with different temporal correlations, obeying Gaussian statistics.

In what follows, we first assume that the cell needs to predict the value of the input $\ell_\tau \equiv \ell(t+\tau)$ at a time τ into the future, although later we will revisit this assumption. We expect the forecast interval τ to be set by the time to mount a response. For accurate prediction, it cannot be much longer than the correlation time of the input signal.

Any prediction about the future state of the environment must be based on information obtained from its past (Fig. 1*A*). In particular, the cell needs to predict ℓ_τ from the current output $x_0 \equiv x(t)$, which itself depends on the input signal trajectory in the past, $\boldsymbol{L}_p \equiv \{\ell(t')\}_{t'<t}$. The (qualitative) shape of the integration kernel $k(t)$, e.g., exponential, adaptive, or oscillatory, is determined by the topology of the signaling network (7). The kernel shape describes how the past signal is mapped onto the current output, and hence which characteristics of the past signal the cell uses to predict the future signal. To maximize the prediction accuracy, the cell should extract those features that are most informative about the future signal. These depend on the signal statistics.

Deriving the upper bound on the predictive information as set by the past information is an optimization problem, which can be solved using the IBM (9). It entails the maximization of an objective function $\mathcal{L}$:

$$\max_{P(x_0|\boldsymbol{L}_p)} \left[\mathcal{L} \equiv I(x_0;\ell_\tau) - \gamma I(x_0;\boldsymbol{L}_p)\right]. \qquad [2]$$

Here, $I_{\text{pred}} \equiv I(x_0;\ell_\tau)$ is the predictive information, which is the mutual information between the system's current output x_0 and the future ligand concentration ℓ_τ. It quantifies the degree to which knowledge of the output x_0 reduces the uncertainty about the future input ℓ_τ, and, within the Gaussian framework, it is related to the mean-squared prediction error as derived, e.g., using Wiener filtering (7). The past information $I_{\text{past}} \equiv I(x_0;\boldsymbol{L}_p)$ is the mutual information between x_0 and the trajectory of past ligand concentrations $\boldsymbol{L}_p$. The Lagrange multiplier γ sets the relative cost of storing past over obtaining predictive information. Given a value of γ, the objective function in Eq. **2** is maximized by optimizing the conditional probability distribution of the output given the past input trajectory, $P(x_0|\boldsymbol{L}_p)$. For the linear systems considered here, this corresponds to optimizing the mapping of the past input signal onto the current output via the integration kernel $k(t)$; in fact, this linear-response strategy is optimal for signals that obey Gaussian statistics (16). We use the Gaussian IBM to derive the optimal kernel $k^{\text{opt}}(t)$ and the *information bound*, defined to be the maximum predictive information as set by the past information (16) (*SI Appendix*, C).

Markovian Signals.

Optimal prediction of Markovian signals: Biochemical copying. Arguably the most elementary signal type is a Markovian signal $\ell(t)$ with correlation time τ_ℓ (Fig. 1*B*). While the signal is Markovian, its fluctuations remain correlated on the finite timescale set by τ_ℓ, which means that its future value ℓ_τ can be predicted with some accuracy. The deviations $\delta\ell(t) = \ell(t) - \bar{\ell}$ from its mean $\bar{\ell}$ follow an Ornstein–Uhlenbeck process:

$$\delta\dot{\ell} = -\delta\ell(t)/\tau_\ell + \eta_\ell(t), \qquad [3]$$

where $\eta_\ell(t)$ is Gaussian white noise, $\langle\eta_\ell(t)\eta_\ell(t')\rangle = 2\sigma_\ell^2/\tau_\ell\,\delta(t-t')$, with σ_ℓ^2 the amplitude of the signal fluctuations. This input signal obeys Gaussian statistics, characterized by $\langle\delta\ell(0)\delta\ell(t)\rangle = \sigma_\ell^2\exp(-t/\tau_\ell)$. Employing the Gaussian IBM framework (16), we find that the optimal integration kernel is given by (*SI Appendix*, C2)

$$k^{\text{opt}}(t-t') = a\delta(t-t'). \qquad [4]$$

This kernel corresponds to a signaling system that copies the current input into the output. This is intuitive, since for a Markovian signal there is no additional information in the past signal that is not already contained in the present one. The prefactor a determines the gain $\partial\bar{x}/\partial\bar{\ell}$, which together with the noise strength $\sigma_{\eta_x}^2$ (Eq. **1**) and the signal amplitude σ_ℓ^2 set the magnitude of the past and predictive information, I_{past} and I_{pred}, respectively (*SI Appendix*, C1).

Fig. 2, I shows the maximum predictive information as set by the past information. This information bound applies to any system that needs to predict a Markovian signal obeying Gaussian statistics. How close can biochemical systems come to this bound?

Push–Pull Network Can be at the Information Bound, yet Increase the Predictive and Past Information by Moving Away fom It. Although the upper bound on the accuracy of prediction is determined by the signal statistics, how close cells can come

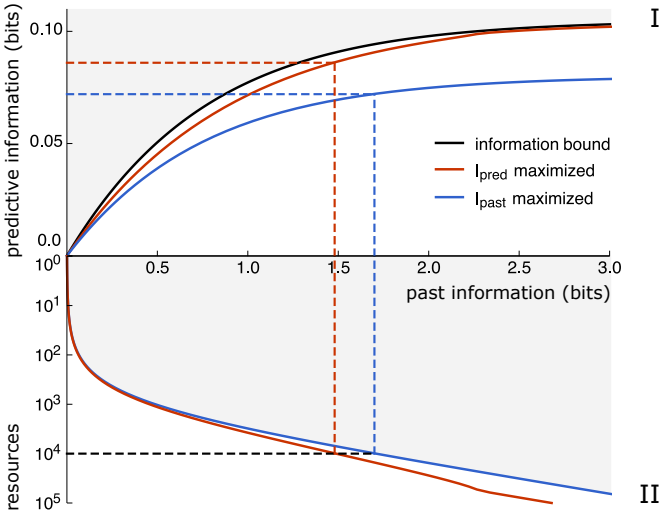

Fig. 2. The optimal push–pull network is not at the information bound. Panel I: The black line is the information bound that maximizes the predictive information $I_{\text{pred}} = I(x_0;\ell_\tau)$ for a given past information $I_{\text{past}} = I(x_0;\boldsymbol{L}_p)$. The blue curve shows I_{pred} versus I_{past} for systems where I_{past} has been maximized for a given resource cost $C = R_T + X_T$. The red curve shows I_{pred} against I_{past} for systems in which I_{pred} has been maximized for a given C. Panel II shows I_{past} against C for the corresponding systems. The dashed lines indicate how the resource cost C sets I_{past} and hence I_{pred}. The forecast interval is $\tau = \tau_\ell$. The optimization parameters are the ratio X_T/R_T, τ_R, p, and f (*SI Appendix*, E). Parameter values: $(\sigma_\ell/\bar{\ell})^2 = 10^{-2}$, $\tau_c/\tau_\ell = 10^{-2}$.

to this bound depends on the topology of the cellular signaling system, and the resources devoted to building and operating it. A network motif that may reach the information bound for Markovian signals is the push–pull network (Fig. 2), because it is at heart a copying device: It samples the input by copying the state of the input, e.g., the ligand-binding state of a receptor or the activation state of a kinase, into the activation state of the output, e.g., phosphorylation state of the readout (13, 15, 26).

We model the push–pull network in the linear-noise approximation (*SI Appendix*, E):

$$\delta\dot{RL} = b\delta\ell(t) - \delta RL(t)/\tau_c + \eta_{RL}(t), \quad [5]$$

$$\delta\dot{x}^* = \gamma\,\delta RL(t) - \delta x^*(t)/\tau_r + \eta_x(t). \quad [6]$$

Here, δRL and δx^* are, respectively, the deviations of the number of ligand-bound receptors and modified readout molecules from their mean values; b and γ are parameters that depend on the number of receptor and readout molecules, R_T and X_T, respectively, the fraction of ligand-bound receptors p and active readouts f; η_{RL} and η_x are Gaussian white noise terms (*SI Appendix*, Eqs. **S68** and **S69**). Key parameters are the correlation time of receptor–ligand binding, τ_c, and the relaxation time of x^*, τ_r. The latter determines for how long x^* carries information on the ligand-binding state of the receptor and thus sets the integration time. The readout-modification dynamics yield an exponential integration kernel $k(t) \propto \exp(-t/\tau_r) - \exp(-t/\tau_c)$ (*SI Appendix*, Eq. **S74**), which in the limit $\tau_r, \tau_c \to 0$ reduces to a δ-function, hinting that the system may reach the information bound.

How much information cells can extract from the past signal depends on the resources devoted to building and operating the network (Fig. 2, II). We define the total resource cost as:

$$C = \lambda(R_T + X_T) + c_1 X_T \Delta\mu/\tau_r. \quad [7]$$

The first term expresses the fact that over the course of the cell cycle all components need to be duplicated, which means that they have to be synthesized at a speed that is at least the growth rate λ. The second term describes the chemical power that is necessary to run the push–pull network (13, 15); it depends on the flux through the network, X_T/τ_r, and the free-energy drop $\Delta\mu$ over a cycle, e.g., the free energy of ATP hydrolysis in the case of a phosphorylation cycle. The coefficient c_1 describes the relative energetic cost of synthesizing the components during the cell cycle versus that of running the system. For simplicity, we first consider the scenario that the cost is dominated by that of protein synthesis, setting $c_1 \to 0$. While in this scenario $R_T + X_T$ is constrained, X_T/R_T and other system parameters are free for optimization.

The available resources C put a hard bound on the information I_{past} that can be extracted from the past signal, which in turn sets a hard limit on the predictive information I_{pred} (Fig. 1*D*). However, this does not imply that the optimal system that maximizes the predictive information I_{pred} per resource cost C, $I_{pred}/C = \left(I_{pred}/I_{past}\right)\left(I_{past}/C\right)$, is also a maximally predictive system, maximizing I_{pred}/I_{past}, or a parsimonious system, maximizing I_{past}/C. To elucidate the interplay between I_{pred}, I_{past}, and C, we first maximize I_{past} for a given resource constraint C. We find a unique optimal design for the push–pull network (blue line in Fig. 2, II), which implies that not all bits of past information are equally costly. We then compute the corresponding predictive information for the systems along this line, which is the blue line in Fig. 2, I. Strikingly, the resulting information curve lies far below the information bound (black line, Fig. 2, I), demonstrating that parsimonious systems, which maximize I_{past}/C, are not maximally predictive. This is because not all bits of past information are equally predictive.

Precisely because bits of past information are neither equally predictive nor equally costly, it is paramount to directly maximize the predictive information for a given resource cost, I_{pred}/C, in order to obtain the most efficient prediction device. This yields the red lines in panels I and II in Fig. 2. It can be seen that compared to parsimonious systems (blue lines), the predictive information is higher while the past information is lower. While the bound on the predictive information as set by the resource cost (red line panel I) is close to the bound on the predictive information as set by the past information (black line), it does remain lower. This is surprising, because the push–pull network is a copying device (13, 26), which can, as we will also show below, reach the latter bound.

Taken together, our results imply that the most cost-efficient prediction systems are neither parsimonious nor maximally predictive, but instead trade off those bits of past information that are most informative about the future, maximizing I_{pred}/I_{past}, against those that are cheapest, maximizing I_{past}/C. We will now make these observations more concrete.

Trade-Off between Cost and Predictive Power per Bit. To understand the connection between predictive and past information, and resource cost, we map out the region in the information plane that can be reached given a resource constraint C (Fig. 3*A*, green region). We immediately make two observations. First, the system can indeed reach the information bound. Second, the system can increase both the past and the predictive information by moving away from it. To elucidate these two observations, we investigate the system along the isocost line of $C = 10^4$, which together with the information bound envelopes the accessible region for the maximum resource cost $C \leq 10^4$.

Along the line $C = 10^4$, the ratio of the number of readout over receptor molecules is $X_T/R_T = 2\sqrt{p/(1-p)}\sqrt{1+\tau_r/\tau_c}$ (*SI Appendix*, E4). Systems that do not obey this relation are inside the accessible region (*SI Appendix*, Fig. S2), as are the systems with $C < 10^4$. The relation can be understood intuitively using the optimal resource allocation principle (13). It states that in a sensing system that employs its proteins optimally, the total number of independent concentration measurements at the level of the receptor during the integration time τ_r, $R_T(1 + \tau_r/\tau_c)$, equals the number of readout molecules X_T that store these measurements, so that neither the receptors nor the readout molecules are in excess. This design principle specifies, for a given integration time τ_r, the ratio X_T/R_T at which the readout molecules sample each receptor molecule roughly once every receptor correlation time τ_c.

While the optimal allocation principle gives the optimal ratio X_T/R_T of the number of readouts over receptors for a *given* integration time τ_r, it does not prescribe what the optimal integration time τ_r^{opt}, and hence the (globally) optimal ratio X_T^{opt}/R_T^{opt}, is that maximizes I_{pred} for a given resource constraint $C = R_T + X_T$. Fig. 3*B* shows that as the distance θ along the isocost line is increased, τ_r and hence X_T/R_T increase monotonically. For $\theta \to 0$, the integration time τ_r is zero and the number of readout molecules equals the number of receptor molecules: $X_T = R_T$. In this limit, the push–pull network is an instantaneous responder, with an integration kernel given by Eq. **4**. The system is indeed very close to the information bound;

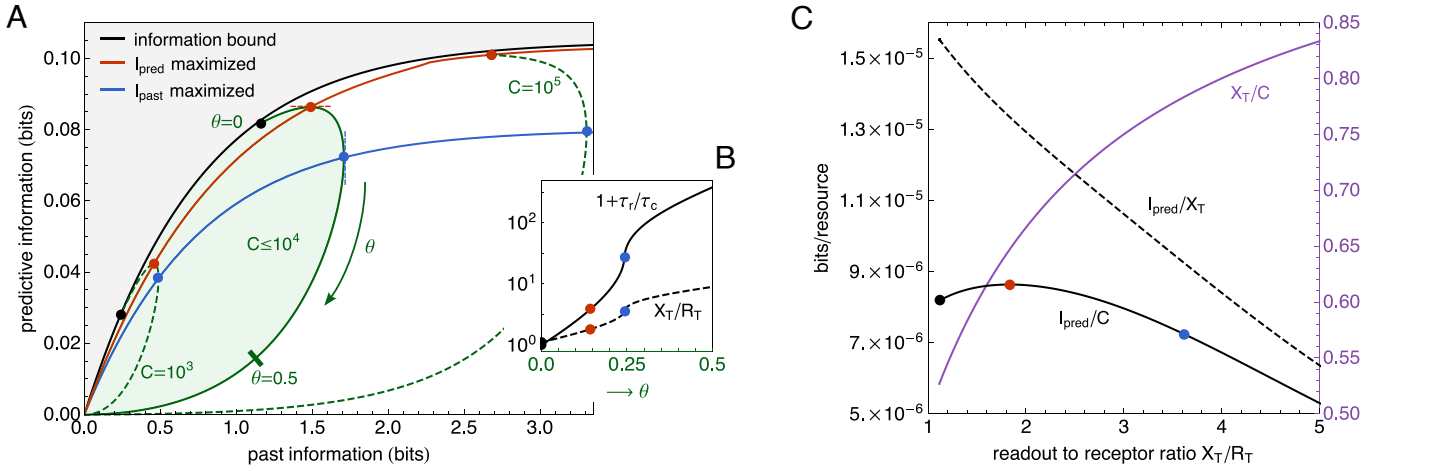

Fig. 3. The push–pull network maximizes the predictive power under a resource constraint by moving away from the information bound. (*A*) The region of accessible predictive information $I_{pred} = I(x_0; \ell_\tau)$ and past information $I_{past} = I(x_0; \boldsymbol{L}_p)$ in the push–pull network under a resource constraint $C \leq R_T + X_T$, for the Markovian signals specified by Eq. **3** (green). The black line is the information bound at which I_{pred} is maximized for a given I_{past}. The push–pull network can be at the information bound (black points), but maximizing I_{pred} for a resource constraint C moves the system away from it. The red and blue lines connect, respectively, the points where I_{pred} and I_{past} are maximized along the green isocost lines (the contourlines of constant C); they correspond to the red and blue lines in Fig. 2, respectively. The accessible region of I_{pred} and I_{past} for a given C has been obtained by optimizing over τ_r, p, f, and X_T/R_T. The forecast interval is $\tau = \tau_\ell$. (*B*) The integration time τ_r over the receptor correlation time τ_c, τ_r/τ_c, and the ratio of the number of readout and receptor molecules, X_T/R_T, as a function of the distance θ along the isocost line corresponding to $C = 10^4$ in panel *A*; the red and blue points denote where I_{pred} and I_{past} are maximized along the contourline, respectively. For $\theta \to 0$, $\tau_r \to 0$: the system is an instantaneous responder, and only the finite receptor correlation time τ_c prevents the system from fully reaching the information bound; as predicted by the optimal resource allocation principle, $X_T = R_T$ (13). The system can increase I_{pred} and I_{past} (*A*) by increasing τ_r and X_T/R_T. (*C*) While this decreases the predictive information I_{pred} per physical bit of past information, I_{pred}/X_T (dashed line), increasing X_T/R_T does increase the number of physical bits per resource cost, X_T/C (purple line). This trade-off gives rise to an optimal predictive information per resource cost, I_{pred}/C (red dot on solid black line). Parameter values unless specified: $(\sigma_\ell/\bar{\ell})^2 = 10^{-2}$, $\tau_c/\tau_\ell = 10^{-2}$.

only the finite receptor correlation time τ_c prevents the system from fully reaching it. Yet, as θ increases and the system moves away from the bound, the predictive and past information first rise along the contour, and thus with X_T/R_T and τ_r, before they eventually both fall.

To understand why the predictive and past information first rise and then fall with X_T/R_T and τ_r, we note that each readout molecule constitutes 1 physical bit and that its binary state (phosphorylated or not) encodes at most 1 bit of information on the ligand concentration. The number of readout molecules X_T thus sets a hard upper bound on the sensing precision and hence the predictive information. To raise this bound, X_T must be increased. For a given resource constraint $C = R_T + X_T$, X_T can only be increased if the number of receptors R_T is simultaneously decreased. However, the cell infers the concentration not from the readout molecules directly, but via the receptor molecules: A readout molecule is a sample of the receptor that provides at most 1 bit of information about the ligand-binding state of a receptor molecule, which in turn provides at most 1 bit of information about the input signal. To raise the lower bound on the predictive information, the information on the input must increase at both the receptor and the readout level.

To elucidate how this can be achieved, we note that the maximum number of independent receptor samples and hence concentration measurements is given by $N_I^{max} = \min(X_T, R_T(1 + \tau_r/\tau_c))$ (13). For $\theta > 0$, the system can increase N_I^{max} if, and only if, X_T and $R_T(1 + \tau_r/\tau_c)$ can be raised simultaneously. This can be achieved, while obeying the constraint $C = X_T + R_T$, by decreasing R_T yet increasing τ_r (Fig. 3*B*). This is the mechanism of time averaging, which makes it possible to increase the number of independent receptor samples (15), and explains why both I_{pred} and I_{past} initially increase (Fig. 3*C*). However, as τ_r is raised further, the receptor samples become older: The readout molecules increasingly reflect receptor states in the past that are less informative about the future ligand concentration. The collected bits of past information have become less predictive about the future (Fig. 3*C*). For a given resource cost, the cell thus faces a trade-off between maximizing the number of physical bits of past information (i.e., receptor samples X_T) and the predictive information per bit. This antagonism gives rise to an optimal integration time τ_r^{opt} and ratio X_T^{opt}/R_T^{opt} (Fig. 3 *A* or *B*) that maximizes the total predictive information I_{pred} (Fig. 3*C*).

In *SI Appendix*, E, we present the signal-to-noise ratio (SNR), and elucidate how the optimal integration time τ_r^{opt} can be understood in terms of a trade-off between time-averaging the receptor noise and, on the other hand, maximizing the dynamic gain to not only lift the signal above the receptor-switching and readout-modification noise but also to minimize the dynamical error (15). This analysis also naturally explains why the system moves toward the information bound as C is increased (Fig. 3*A*): Raising C allows for more instantaneous concentration measurements, which makes time averaging less important, yielding a smaller τ_r^{opt}. We also note here that while the magnitude of I_{pred} depends on the forecast interval τ, the optimal design does not, because the signal is Markovian.

Interestingly, while I_{pred} decreases beyond τ_r^{opt}, the past information I_{past} first continues to rise because N_I^{max} still increases (Fig. 3*A*). However, when the integration time becomes longer than the input signal correlation time, the correlation between input and output will be lost and I_{past} will fall too.

Chemical Power Prevents the System from Reaching the Information Bound. So far, we have only considered the cost of maintaining the cellular system, the protein cost $C = R_T + X_T$. Yet, driving a push–pull network also requires energy. In fact, for

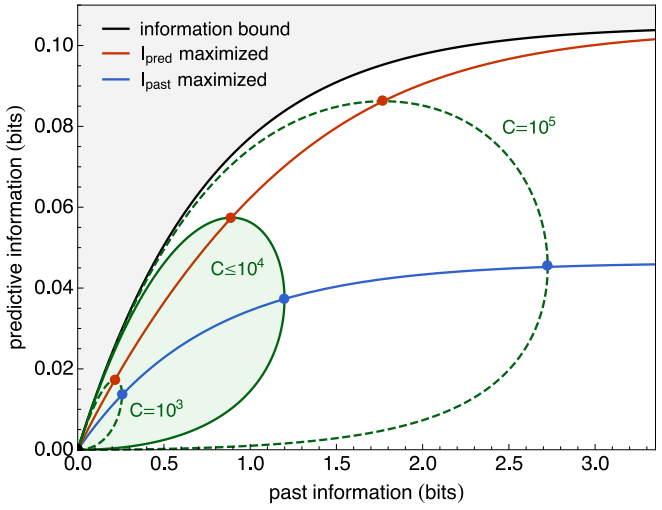

Fig. 4. The cost of operating the network moves the system away from the information bound. The region of accessible predictive information $I_{\text{pred}} = I(x_0;\ell_\tau)$ and past information $I_{\text{past}} = I(x_0;\boldsymbol{L}p)$ in the push–pull network under a resource constraint that is not only given by a protein cost but also a running cost, $C \leq \lambda(R_T + X_T) + c_1 X_T \Delta\mu/\tau_r$, for the Markovian signals specified by Eq. **3** (green). The black line is the information bound at which I_{pred} is maximized for a given I_{past}. The cost of operating the network, $\sim X_T\Delta\mu/\tau_r$, moves the system away from the information bound (compare to Fig. 3). The red and blue lines correspond to systems that maximize I_{pred} and I_{past}, respectively, under the constraint C. The accessible region of I_{pred} and I_{past} for a given C has been obtained by optimizing over τ_r, p, f, and X_T/R_T. Parameter values: $\tau = \tau_\ell$, $(\sigma_\ell/\bar{\ell})^2 = 10^{-2}$, $\tau_c/\tau_\ell = 10^{-2}$, $\tau_\ell = 1$s, $\lambda^{-1} = 1$h, $c_1 = 10^{-4}/\Delta\mu$ (see also *SI Appendix*, Fig. S3).

E. coli at a typical cell doubling time $\lambda^{-1} \simeq 1$h, the operating cost is comparable to the maintenance cost (13, 27). As Eq. **7** shows, the operating cost scales with the flux around the phosphorylation cycle, which is proportional to the inverse of the integration time, ${\tau_r}^{-1}$. The power thus diverges for $\tau_r \to 0$. Since the information bound is reached precisely in this limit, it is clear that the chemical power prevents the push–pull network from reaching the bound (Fig. 4).

Non-Markovian Signals.

Predicting the future change. The push–pull network can optimally predict Markovian signals, yet not all signals are expected to be Markovian. Especially organisms that navigate through an environment with directional persistence will sense a non-Markovian signal, as generated by their own motion (Fig. 1*B*). Moreover, when these organisms need to climb a concentration gradient, as *E. coli* during chemotaxis, then knowing the change in the concentration is arguably more useful than knowing the concentration itself. Indeed, it is well known that the kernel of the *E. coli* chemotaxis system detects the (relative) change in the ligand concentration by taking a temporal derivative of the concentration (19). However, as we will show here, the converse statement is more subtle. If the system needs to predict the (future) change in the signal, v_τ, then the optimal kernel is not necessarily one that is based on the derivative only: In general, the optimal kernel uses a combination of the signal value and its derivative. However, the *E. coli* chemotaxis system can respond to concentrations that vary between the dissociation constants of the inactive and active state of the receptors, which differ by several orders of magnitude (28). This range of possible background concentrations is much larger than the typical concentration change over the orientational correlation time of the bacterium. As our analysis below reveals, in this regime the optimal kernel is a perfectly adaptive, derivative-taking kernel that is insensitive to the current signal value, precisely like that of the *E. coli* chemotaxis system (19, 29–31). Our analysis thus predicts that this system has an adaptive kernel, because this is the optimal kernel for predicting concentration derivatives over a broad range of background concentrations. We note that our explanation for the origin of the adaptive kernel differs from others, which are based on requirements for sensing and motility (32–34). It also differs from the argument that perfect adaption enables the system to keep the CheY_p concentration in the working range of the motor over a wide range of background concentrations.

To reveal the signal characteristics that control the shape of the optimal integration kernel, we will consider the family of stationary Gaussian signals generated by (10, 11)

$$\delta\dot{\ell} = v(t), \qquad [8]$$

$$\dot{v} = -\omega_0^2\delta\ell(t) - v(t)/\tau_v + \eta_v(t), \qquad [9]$$

where $\delta\ell$ is the deviation of the ligand concentration from its mean $\bar{\ell}$, v its derivative, τ_v a relaxation time, and η_v a Gaussian white noise; $\omega_0^2 = \sigma_v^2/\sigma_\ell^2$ sets the ratio of the variance in the derivative of the concentration, σ_v^2, to that of its value σ_ℓ^2, which allows us to control the range of input concentrations as discussed below. This model covers a range of signal behaviors, allowing us to elucidate the signal characteristics that control the optimal shape of the integration kernel. Moreover, we will later show that this model can describe the biologically relevant regime of chemotaxis in shallow gradients.

Applying the IBM framework to Eq. **2** with ℓ_τ replaced by v_τ, it can be shown that the optimal encoding that enables the system to reach the information bound on predicting the future concentration change v_τ, is based on a linear combination of the current concentration $\ell(t)$ and its derivative $v(t)$ (*SI Appendix*, C3) (10, 11):

$$x(t) = a\frac{\delta\ell(t)}{\sigma_\ell} + b\frac{v(t)}{\sigma_v} + \eta_x(t). \qquad [10]$$

This can be understood by noting that while the signal of Eqs. **8** and **9** is non-Markovian in the space of ℓ, the concentration *E. coli* actually senses, it is Markovian in ℓ and v: all the information on the future signal is contained in the current concentration and its derivative. To reach the information bound, the coefficients must obey

$$a^{\text{opt}} = G\frac{\langle\delta\ell(0)\delta v(\tau)\rangle}{\sigma_\ell\sigma_v} \equiv G\rho_{\ell_0 v_\tau}, \qquad [11]$$

$$b^{\text{opt}} = G\frac{\langle\delta v(0)\delta v(\tau)\rangle}{\sigma_v^2} \equiv G\rho_{v_0 v_\tau}. \qquad [12]$$

Here, G is the gain, which together with the noise $\sigma_{\eta_x}^2$ sets the scale of I_{pred} and I_{past}, $\rho_{\ell_0 v_\tau}$ is the cross-correlation coefficient between the current concentration value ℓ_0 and the future concentration derivative v_τ and $\rho_{v_0 v_\tau}$ that between the current and future derivative (*SI Appendix*, C3). These expressions can be understood intuitively: If the future signal derivative that needs to be predicted is correlated with the current signal derivative, it is useful to include in the prediction strategy the latter, leading to a nonzero value of b^{opt}. Perhaps more surprisingly, if the future signal derivative is also correlated with the current signal *value*, then the system can enhance the prediction accuracy by also including the current signal value, yielding a nonzero a^{opt}.

Clearly, in general, to optimally predict the future signal change, the system should base its prediction on both the current signal value and its derivative.

The degree to which the optimal system bases its prediction on the current signal value versus the current derivative depends on the relative magnitudes of $\rho_{\ell_0 v_\tau}$ and $\rho_{v_0 v_\tau}$, respectively. In *SI Appendix*, B2 we show that when the concentration change over the timescale τ_v is much smaller than the range of possible concentrations σ_ℓ that the bacterium can experience, i.e., when $\sigma_v \tau_v \ll \sigma_\ell$ and hence $\omega_0 \ll {\tau_v}^{-1}$, the cross-correlation coefficient $\rho_{\ell_0 v_\tau}$ vanishes. In this case, a^{opt} becomes zero (Eq. **11**): The optimal kernel has become a perfectly adaptive, derivative-taking kernel. We emphasize that while we have derived this result for the class of signals defined by Eqs. **8** and **9**, the idea is far more generic. In particular, while we do not know the temporal structure of the ligand statistics that *E. coli* experiences in general, we do know it can detect concentration changes over a range of background concentrations that is much larger than the typical concentration change over a run (28). In this regime, the correlation between the concentration value and its future change is likely to be very small. As our analysis shows, a perfectly adaptive kernel then emerges naturally from the requirement to predict the future concentration change.

While the class of signals specified by Eqs. **8** and **9** is arguably limited, it does describe the biologically important regime of chemotaxis in shallow gradients. In this regime, the stimulus only weakly affects the swimming behavior, such that the statistics of the concentration change v is to a good approximation given by the steepness of the concentration gradient and the swimming statistics of the bacterium in the absence of a gradient (20). In particular, in an exponential gradient $\ell(x) \propto e^{gx}$ with steepness g, noting that $v = g\bar{\ell} v_x$ with v_x the velocity of the bacterium in the direction of the gradient, the statistics of the signal v is given by

$$\langle \delta v(0) \delta v(\tau) \rangle = g^2 \bar{\ell}^2 \langle \delta v_x(0) \delta v_x(\tau) \rangle, \qquad [13]$$

where the latter is the autocorrelation function of the swimming velocity of the bacterium in the absence of a gradient. It is a characteristic of the bacterium, not of the environment, and has been measured to decay exponentially with a correlation time τ_{v_x} (20):

$$\langle \delta v_x(0) \delta v_x(\tau) \rangle \simeq \sigma_{v_x}^2 e^{-\tau/\tau_{v_x}}. \qquad [14]$$

Above, we argued that, because *E. coli* can sense concentrations over a range of background concentrations that is much larger than the concentration change over a run, the relevant limit of our model (Eqs. **8** and **9**) is $\omega_0 \ll {\tau_v}^{-1}$. In this limit, the concentration change v decays exponentially, as in Eq. **14**. With $\tau_v = \tau_{v_x}$ and $\sigma_v^2 = g^2 \bar{\ell}^2 \sigma_{v_x}^2$ our model then describes the input statistics in shallow gradients as measured experimentally (20). We also note here that while in this limit the statistics of v becomes Markovian, that of the concentration ℓ remains non-Markovian.

Finite Resources Prevent the Chemotaxis System from Taking An Instantaneous Derivative and Reaching the Information Bound. The above analysis indicates that the chemotaxis system seems ideally designed to predict the future concentration change, because its integration kernel is nearly perfectly adaptive (19, 29–31, 35). But how close can this system come to the information bound for the non-Markovian signals specified by Eqs. **8** and **9**?

To address this question, we consider a molecular model that can accurately describe the response of the chemotaxis system to a wide range of time-varying signals (36–39). In this model, the receptors are partitioned into clusters. Each cluster is described via a Monod–Wyman–Changeux model (40). While each receptor can switch between an active and an inactive conformational state, the energetic cost of having different conformations in the same cluster is prohibitively large. Each cluster is thus either active or inactive. Ligand binding favors the inactive state while methylation does the opposite, enabling adaptation. This model allows for a cooperative response, with the degree of cooperativity set by the cluster size N. Last, active receptor clusters can via the associated kinase CheA phosphorylate the downstream messenger protein CheY.

Linearizing around the steady state, we obtain (*SI Appendix*, F):

$$\delta a_i(t) = \alpha \delta m_i(t) - \beta \delta \ell(t), \qquad [15]$$

$$\delta \dot{m}_i = -\delta a_i(t)/(\alpha \tau_{\text{m}}) + \eta_{m_i}(t), \qquad [16]$$

$$\delta \dot{x}^* = \gamma \sum_{i=1}^{R_{\text{T}}} \delta a_i(t) - \delta x^*(t)/\tau_{\text{r}} + \eta_x(t). \qquad [17]$$

Here, $\delta a_i(t)$ and $\delta m_i(t)$ are the deviations of the activity and methylation level of receptor cluster i from their steady-state values, and R_{T} is the total number of receptor clusters, each consisting of N receptor molecules; $\delta \ell(t)$ and $\delta x^*(t)$ are, respectively, the deviations of the ligand and CheY_p concentration from their steady-state values; τ_{m} and τ_{r} are the timescales of receptor methylation and CheY_p (de)phosphorylation; η_{m_i} and η_x are independent Gaussian white noise sources; the parameters α and β depend on the cluster size N and, respectively, the free-energy of methylation and ligand binding (*SI Appendix*, F1). In Eq. **15**, we have assumed that ligand binding is much faster than the other timescales in the system, so that it can be integrated out. There is thus no need to time average receptor–ligand binding noise, which means that, in the absence of running costs, the optimal integration time $\tau_{\text{r}}^{\text{opt}}$ is zero. In what follows, we set τ_{r} to the value measured experimentally, $\tau_{\text{r}} \approx 100$ ms (13, 41). We consider the non-Markovian signals specified by Eqs. **8** and **9** in the physiologically relevant limit $\omega_0 \to 0$, such that the optimal kernel is perfectly adaptive, like that of *E. coli*. For these signals, we determine the accessible region of I_{past} and I_{pred} under the resource constraint $C \leq N R_{\text{T}} + X_{\text{T}}$ (Fig. 5) by optimizing over the remaining free parameters, the methylation time τ_{m} and the ratio of readout over receptor molecules $X_{\text{T}}/R_{\text{T}}$. The forecast interval τ is set to τ_v, but we emphasize that the optimal design is independent of the value of τ (*SI Appendix*, F6).

Fig. 5*A* shows that the optimal system that maximizes the predictive information I_{pred} under a resource constraint C is markedly away from the information bound. Yet, as the resource constraint is relaxed and C is increased, the optimal system moves toward it. Panel *B* shows that the methylation time τ_{m} rises along the three respective isocost lines of panel *A*. It highlights that there exists an optimal methylation time $\tau_{\text{m}}^{\text{opt}}$ that maximizes the predictive information I_{pred}. Moreover, $\tau_{\text{m}}^{\text{opt}}$ decreases as the resource constraint is relaxed. Along the isocost lines $X_{\text{T}}/R_{\text{T}}$ varies only mildly (*SI Appendix*, Fig. S6).

These observations can be understood by noting that the system faces a trade-off between taking a derivative that is recent versus one that is robust. All the information on the future derivative, which the cell aims to predict, is contained

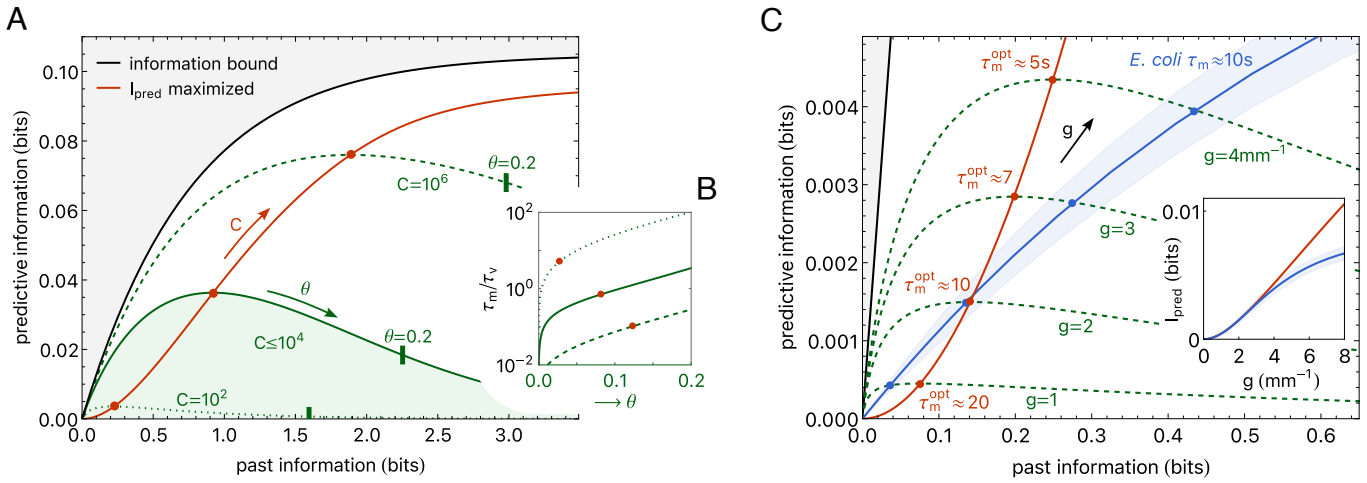

Fig. 5. Finite resources prevent the chemotaxis system from reaching the information bound. (*A*) The region of accessible predictive information $I_{\rm pred} = I(x_0; v_\tau)$ and past information $I_{\rm past} = I(x; \boldsymbol{L}_p)$ for the chemotaxis system of Eqs. **15–17** under a resource constraint $C \le NR_{\rm T} + X_{\rm T}$, for the non-Markovian signals specified by Eqs. **8** and **9** (green). The black line shows the information bound at which $I_{\rm pred}$ is maximized for a given $I_{\rm past}$. The chemotaxis system is not at the information bound, but it does move toward it as C is increased. The red line connects the red points where $I_{\rm pred}$ is maximized for a given resource cost C. The accessible region of $I_{\rm pred}$ and $I_{\rm past}$ under a given resource constraint C is obtained by varying the methylation time $\tau_{\rm m}$ and optimizing the ratio of readout over receptor molecules $X_{\rm T}/R_{\rm T}$; $g = 4\ {\rm mm}^{-1}$, $N = 12$. (*B*) The methylation time $\tau_{\rm m}$ over the input correlation time τ_v as a function of the distance θ along the three respective isocost lines shown in panel *A*. The methylation time $\tau_{\rm m}$ increases along the isocost line, but there exists an optimal $\tau_{\rm m}$ that maximizes the predictive information, marked by the red points; $\theta \to 0$ corresponds to the origin of panel *A*, $(I_{\rm pred}, I_{\rm past}) = (0, 0)$; the points where $\theta = 0.2$ along the isocost lines of panel *A* are marked with a bar. As the resource constraint is relaxed (higher C), the optimal $\tau_{\rm m}$ decreases: the system moves toward the information bound, where it takes an instantaneous derivative, corresponding to $\tau_r, \tau_{\rm m} \to 0$. (*C*) Comparison of the *E. coli* chemotaxis system (blue line) to the optimal system (red line) as a function of the steepness g of an exponential ligand concentration gradient $\ell(x) = \bar{\ell}e^{gx}$. For the *E. coli* system, $I_{\rm pred}$ and $I_{\rm past}$ have been computed directly from the measured response kernel and signal and noise correlation functions (20) in a model-free manner. The optimal system is based on the model of Eqs. **15–17**, in which the resource cost $C = NR_{\rm T} + X_{\rm T}$ is fixed, with $X_{\rm T} = 10^4$, $N = 12$, and $R_{\rm T} = 8$ as inferred from experimental data (20, 27) (*SI Appendix*, F8), yet $\tau_{\rm m}$ has been optimized; the dashed lines mark the boundaries of the accessible region of $(I_{\rm pred}, I_{\rm past})$ of this model upon varying $\tau_{\rm m}$, with the red dot corresponding to the system with the optimal $\tau_{\rm m}$. The predictive information increases with gradient steepness g, yet the optimal methylation time decreases with g. The inset shows that $I_{\rm pred}$ of the *E. coli* system is very close to that of the optimal system for shallow gradients $g \lesssim 4{\rm mm}^{-1}$, but deviates from it at steeper gradients, suggesting the system has been optimized for sensing shallow gradients. Parameter values: Forecast interval $\tau = \tau_v$, $\tau_r = 100$ ms (13, 20, 41); $\tau_v^{-1} = 0.9\ {\rm s}^{-1}$ and $\sigma_v^2 = g^2\bar{\ell}^2\sigma_{v_x}^2$, with $\bar{\ell} = 100\ \mu$M and $\sigma_{v_x}^2 = 157.1\ \mu{\rm m}^2{\rm s}^{-2}$ (20); $f = 0.5$, $p = 0.3$ (20).

in the current derivative of the signal; measuring the current derivative would allow the system to reach the information bound. However, computing the recent derivative is extremely costly. The cell takes the temporal derivative of the ligand concentration at the level of the receptor via two antagonistic reactions that occur on two distinct timescales: Ligand binding rapidly deactivates the receptor, while methylation slowly reactivates it (37). The receptor ligand-occupancy thus encodes the current concentration, the methylation level stores the average concentration over the past $\tau_{\rm m}$, and the receptor activity reflects the difference between the two—the temporal derivative of the signal over the timescale $\tau_{\rm m}$. To obtain an instantaneous derivative, $\tau_{\rm m}$ must go to zero. However, the gain $\tilde{g}$, i.e., change in output due to a change in input, scales as $\tilde{g} \sim \tau_{\rm m}/(1+\tau_{\rm m}/\tau_v)$ (*SI Appendix*, Eq. **S131**). Clearly, in the limit $\tau_{\rm m} \to 0$, the gain becomes zero because the receptor activity instantly adapts to the ligand concentration change. Since the push–pull network downstream of the receptor is a device that samples the receptor stochastically (13, 15), the gain must be raised to lift the signal above this sampling noise. This requires a finite methylation time $\tau_{\rm m}$. The trade-off between a recent derivative and a reliable one gives rise to an optimal methylation time $\tau_{\rm m}^{\rm opt}$ that maximizes the predictive information for a given resource cost (see also *SI Appendix*, F6).

The same analysis also explains why the optimal methylation time $\tau_{\rm m}^{\rm opt}$ decreases and the predictive information increases when the resource constraint is relaxed (Fig. 5). The sampling noise in estimating the average receptor activity decreases as the number of readout molecules increases (13, 15). A smaller gain is then required to lift the signal above this noise. In addition, a larger number of receptors decreases the noise in the methylation level, which also allows for a smaller gain, and hence a shorter methylation time. These two effects together explain why the optimal methylation decreases with C, scaling as $\tau_{\rm m}^{\rm opt} \sim 1/\sqrt{C}$ (*SI Appendix*, Eq. **S138**), and $I_{\rm pred}$ increases with C (*SI Appendix*, F5).

Fig. 5*A* also shows that the past information $I_{\rm past} = I(x_0; \boldsymbol{L}_p)$ does not return to zero along the contourline of constant resource cost. Along the contourline, the methylation time $\tau_{\rm m}$ rises (Fig. 5*B*). While the predictive information $I_{\rm pred}$ exhibits an optimal methylation time $\tau_{\rm m}^{\rm opt}$, the past information $I_{\rm past}$ continues to rise with $\tau_{\rm m}$ because the system increasingly becomes a copying device, rather than one that takes a temporal derivative.

We have seen that the operating cost moves the push–pull network away from the information bound (Fig. 4). Similarly, the operating cost of the methylation cycle (31, 35) pushes the system further away from the information bound (*SI Appendix*, Fig. S7*A*). Yet, for biologically realistic values of C, the effect is very small: While the free energy of SAM hydrolysis, driving the methylation cycle, is comparable to that of ATP hydrolysis, driving the phosphorylation cycle (42), the optimal methylation time $\tau_{\rm m}^{\rm opt}$ is about two orders of magnitude longer than the relaxation time $\tau_{\rm r}$ of the phosphorylation cycle (13, 41). Only for much higher values of C, when $\tau_{\rm m}^{\rm opt}$ becomes substantially shorter, does the methylation cost significantly move the system further away from the bound, reducing the accessible region of $(I_{\rm pred}, I_{\rm past})$ noticeably (*SI Appendix*, Fig. S7*A*). Indeed,

in the biologically relevant regime, the phosphorylation cost reduces this region more than the methylation cost (*SI Appendix*, Fig. S7).

Comparison with Experiment. To study how close the *E. coli* chemotaxis system comes to the information bound, we compute the predictive and past information directly from experimental data. Within the Gaussian framework, the predictive and past information can be obtained from the signal correlation function, noise correlation function, and network response kernel (*SI Appendix*, F8). Interestingly, these functions have recently been measured experimentally (20). This indeed makes it possible to compute the predictive and past information from experimental data without the need to invoke a biochemical model. The only assumption required is that the system obeys Gaussian statistics, but recent work indicates that this is the case, to an excellent approximation (39).

The resulting $I_{\rm pred}$ and $I_{\rm past}$ between the signal and the CheA kinase activity for cells swimming in an exponential concentration gradient for different values of its steepness g are shown in Fig. 5*C* (blue line). It is seen that the *E. coli* chemotaxis system is remarkably far away from the information bound. This raises the question how close this system is to an optimal system that maximizes the predictive information under a resource constraint. To answer this question, we compare the predictive and past information of the *E. coli* system to those of an optimal system with the same topology and resource constraint. The latter system is based on the model of Eqs. **15**–**17**, with the methylation time $\tau_{\rm m}$ optimized to maximize $I_{\rm pred}/C$ for each gradient steepness g, but with the other key parameters—the number of receptor clusters $R_{\rm T}$, cluster size N, number of readout molecules $X_{\rm T}$, and relaxation time $\tau_{\rm r}$—fixed to values inferred from experimental data (*SI Appendix*, F8). Crucially, when $\tau_{\rm m}$ equals the measured value, $\tau_{\rm m} \approx 10$ s (41), this model reproduces the measured response kernel and noise correlation function, such that the predicted $I_{\rm pred}$ and $I_{\rm past}$ agree with those directly inferred from the experimental data.

Fig. 5*C* shows that both in the optimal system (red line) and the chemotaxis system (blue line) $I_{\rm pred}$ and $I_{\rm past}$ increase as the gradient steepness g increases. A steeper gradient increases the signal strength, $\sigma_v^2 \sim g^2$ (Eq. **13**), which increases the predictive and past information (*SI Appendix*, F). Yet, the figure also shows that in the optimal system $I_{\rm pred}$ rises much faster with $I_{\rm past}$ than in the *E. coli* system. The stronger signal received in steeper concentration gradients raises the signal above the receptor sampling noise more, permitting a smaller gain. This allows the optimal system to take a more recent derivative, with a smaller $\tau_{\rm m}$, which is more informative about the future: the optimal methylation time scales as $\tau_{\rm m}^{\rm opt} \sim 1/g$ (*SI Appendix*, Fig. S5). In contrast, the methylation time $\tau_{\rm m}$ of the *E. coli* chemotaxis system is fixed—the system cannot tune $\tau_{\rm m}$ to g. As the *Inset* in Fig. 5*C* shows, this methylation time is close to optimal for detecting shallow gradients, $g \lesssim 4\ {\rm mm}^{-1}$: $I_{\rm pred}$ of the *E. coli* system is nearly identical to that of the optimal system. Moreover, in this regime, not only $I_{\rm pred}$ but also $I_{\rm past}$ is fairly similar for the two systems. For steeper gradients $I_{\rm past}$ becomes much higher in the *E. coli* system than in the optimal one, even though $I_{\rm pred}$ remains lower. The bacterium increasingly collects information that is less informative about the future. Taken together, these results strongly suggest that the system has been optimized to predict future concentration changes in shallow gradients, which necessitate a relatively long methylation time.

Discussion

Cellular systems need to predict the future signal by capitalizing on information that is contained in the past signal. To this end, they need to encode the past signal into the dynamics of the intracellular biochemical network from which the future input is inferred. To maximize the predictive information for a given amount of information that is extracted, the cell should store those signal characteristics that are most informative about the future signal. For a Markovian signal this is the current signal value, while for the non-Markovian signal corresponding to an underdamped particle in a harmonic well (Eqs. **8** and **9**), this is the current signal value and its derivative. As we have seen here, cellular systems are able to extract these signal characteristics: The push–pull network can copy the current input into the output, while the chemotaxis network can take an instantaneous derivative. We have thus demonstrated that at least for two classes of signals, cellular systems are in principle able to extract the most predictive information, allowing them to reach the information bound.

Yet, our analysis also shows that extracting the most relevant information can be exceedingly costly. To copy the most recent input signal into the output, the integration time of the push–pull network needs to go to zero, which means that the chemical power diverges. Moreover, taking an instantaneous derivative reduces the gain to zero, such that the signal is no longer lifted above the inevitable intrinsic biochemical noise of the signaling system. Also the chemical power cost to drive the adaptation cycle (31, 35) pushes the system away from the information bound, although this effect is fairly small.

So far, bounds on predictive information have been studied using information-compression constraints (10–12). Yet, while information is a resource—the cell cannot predict the future without extracting information from the past signal—the principal resources that have a direct cost are time, building blocks and energy. The predictive information per protein and energy cost is therefore most likely a more relevant fitness measure than the predictive information per past information.

Our analysis reveals that, in general, it is not optimal to operate at the information bound: Cells can increase the predictive information for a given resource constraint by moving away from the bound. Increasing the integration time in the push–pull network reduces the chemical power and makes it possible to take more concentration measurements per protein copy. And increasing the methylation time in the chemotaxis system increases the gain. Both enable the system to extract more information from the past signal. Yet, increasing the integration time or the methylation time also means that the information that has been collected is less informative about the future signal. This interplay gives rise to an optimal integration and methylation time, which maximize the predictive information for a given resource constraint. This argument also explains why the respective systems move toward the information bound when the resource constraint is relaxed: Increasing the number of receptor and readout molecules allows the system to take more instantaneous concentration measurements, which makes time averaging less important, thus reducing the integration time. Increasing the number of readout molecules also reduces the error in sampling the receptor state. This makes it easier to detect a change in the receptor activity resulting from the signal, thus allowing for a smaller gain and hence a shorter methylation time.

While the predictive information of the *E. coli* chemotaxis system in shallow gradients is very close to that of the optimal chemotaxis system with the same topology and resource con-

straint, it is much further away from the information bound than other information processing systems, in neuronal signaling (11) and embryonic development (43). We believe this is because the chemotaxis systems take a temporal derivative only. This computation is inherently costly because it is based on signal subtraction, which dramatically reduces the gain. A small gain is particularly detrimental in the noise-dominated regime of low past and predictive information, where the signal strength σ_v^2 and the resource availability C are low. To maximize the predictive information in this regime, it becomes paramount to raise the gain by increasing the time interval over which the derivative is taken, as set by τ_m. However, this moves the system inevitably away from the bound.

Any system that needs to predict the future environment, be it a living organism or a man-made device, must base its prediction on information it has collected from the past. In many cases, information from the recent past is likely to be more predictive than that from the distant past, as for the signals studied here. It means that the system must continually and rapidly update its information on the environment. Yet, the laws of thermodynamics imply that there is a fundamental trade-off between precision, power, and speed (44): Obtaining information rapidly is inherently costly. We thus expect that our principal result, namely that there is a trade-off between obtaining information that is predictive versus that which is cheap, is generic, applying to a broad class of systems. For these systems, the optimal design that maximizes the predictive power under a resource constraint will differ from the design that maximizes predictive power under an information compression constraint.

Information theory shows that the amount of transmitted information depends not only on the characteristics of the information processing system, but also on the statistics of the input signal. While much progress has been made in characterizing cellular signaling systems, the statistics of the input signal is typically not known, with a few notable exceptions (45). Here, we have focused on two classes of input signals, but it seems likely that the signals encountered by natural systems are much more diverse. It will be interesting to extend our analysis to signals with a richer temporal structure (10), and see whether cellular systems exist that can optimally encode these signals for prediction.

Finally, while we have analyzed the design of cellular signaling networks to optimally predict future signals, we have not addressed the utility of information for function or behavior. It is clear that many functional or behavioral tasks, like chemotaxis (20), require information, but what the relevant bits of information are is poorly understood (7). Moreover, cells ultimately employ their resources—protein copies, time, and energy—for function or behavior, not for processing information per se. Here, we have shown that maximizing predictive information under a resource constraint, $C \rightarrow I_{\text{past}} \rightarrow I_{\text{pred}}$, does not necessarily imply maximizing past information. This hints that optimizing a functional or behavioral task under a resource constraint, $C \rightarrow I_{\text{pred}} \rightarrow$ function, may not imply maximizing the predictive information necessary to carry out this task.

Data, Materials, and Software Availability. All study data are included in the article and/or *SI Appendix*.

ACKNOWLEDGMENTS. We thank Jeroen van Zon, Tom Shimizu, and Avishek Das for a careful reading of the manuscript, and Jenny Poulton, Manuel Reinhardt, Michael Vennettilli, and Daan de Groot for many useful discussions. This work is part of the Dutch Research Council (NWO) and was performed at the research institute AMOLF. This project has received funding from the European Research Council under the European Union's Horizon 2020 research and innovation program (grant agreement No. 885065).

1. M. Monti, D. K. Lubensky, P. R. ten Wolde, Robustness of clocks to input noise. *Phys. Rev. Lett.* **121**, 078101 (2018).
2. W. Pittayakanchit, Z. Lu, J. Chew, M. J. Rust, A. Murugan, Biophysical clocks face a trade-off between internal and external noise resistance. *eLife* **7**, e37624 (2018).
3. E. Kussell, S. Leibler, Phenotypic diversity, population growth, and information in fluctuating environments. *Science* **309**, 2075–2078 (2005).
4. I. Tagkopoulos, Y. C. Liu, S. Tavazoie, Predictive behavior within microbial genetic networks. *Science* **320**, 1313–1317 (2008).
5. A. Mitchell *et al.*, Adaptive prediction of environmental changes by microorganisms. *Nature* **460**, 220–224 (2009).
6. W. Bialek, *Biophysics: Searching for Principles* (Princeton Editorial Associates, Inc., Princeton University Press, Woodstock, Oxfordshire, 2012).
7. N. B. Becker, A. Mugler, P. R. ten Wolde, Optimal prediction by cellular signaling networks. *Phys. Rev. Lett.* **115**, 1–5 (2015).
8. W. Bialek, R. R. De Ruyter Van Steveninck, N. Tishby, "Efficient representation as a design principle for neural coding and computation" in *2006 IEEE International Symposium on Information Theory* (2006), pp. 659–663.
9. N. Tishby, F. C. Pereira, W. Bialek, "The information bottleneck method in *Proceedings of the 37th Allerton Conference on Communication, Control, and Computing* (1999).
10. V. Sachdeva, T. Mora, A. M. Walczak, S. E. Palmer, Optimal prediction with resource constraints using the information bottleneck. *PLoS Comput. Biol.* **17**, e1008743 (2021).
11. S. E. Palmer, O. Marre, M. J. Berry, W. Bialek, Predictive information in a sensory population. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 6908–6913 (2015).
12. M. Chalk, O. Marre, G. Tkačik, Toward a unified theory of efficient, predictive, and sparse coding. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 186–191 (2018).
13. C. C. Govern, P. R. ten Wolde, Optimal resource allocation in cellular sensing systems. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 17486–17491 (2014).
14. S. B. Laughlin, R. R. de Ruyter van Steveninck, J. C. Anderson, The metabolic cost of neural information. *Nat. Neurosci.* **1**, 36–41 (1998).
15. G. Malaguti, P. R. ten Wolde, Theory for the optimal detection of time-varying signals in cellular sensing systems. *eLife* **10**, 1–26 (2021).
16. G. Chechik, A. Globerson, N. Tishby, Y. Weiss, Information bottleneck for Gaussian variables. *J. Mach. Learn. Res.* **6**, 165–188 (2005).
17. A. Goldbeter, D. E. Koshland, An amplified sensitivity arising from covalent modification in biological systems. *Proc. Natl. Acad. Sci. U.S.A.* **78**, 6840–6844 (1981).
18. U. Alon, *An Introduction to Systems Biology: Design Principles of Biological Circuits* (Chapman and Hall/CRC, New York, NY, 2006).
19. J. E. Segall, S. M. Block, H. C. Berg, Temporal comparisons in bacterial chemotaxis. *Proc. Natl. Acad. Sci. U.S.A.* **83**, 8987–8991 (1986).
20. H. H. Mattingly, K. Kamino, B. B. Machta, T. Emonet, Escherichia coli chemotaxis is information limited. *Nat. Phys.* **17**, 1426–1431 (2021).
21. M. Hinczewski, D. Thirumalai, Cellular signaling networks function as generalized Wiener–Kolmogorov filters to suppress noise. *Phys. Rev. X* **4**, 3–15 (2014).
22. T. L. Wang, B. Kuznets-Speck, J. Broderick, M. Hinczewski, The price of a bit: Energetic costs and the evolution of cellular signaling. bioRxiv [Preprint] (2022). https://doi.org/10.1101/2020.10.06.327700 (Accessed 22 February 2023).
23. S. Tanase-Nicola, P. B. Warren, P. R. ten Wolde, Signal detection, modularity, and the correlation between extrinsic and intrinsic noise in biochemical networks. *Phys. Rev. Lett.* **97**, 068102 (2006).
24. E. Ziv, I. Nemenman, C. H. Wiggins, Optimal signal processing in small stochastic biochemical networks. *PLoS ONE* **2**, e1077 (2007).
25. W. De Ronde, F. Tostevin, P. R. ten Wolde, Effect of feedback on the fidelity of information transmission of time-varying signals. *Phys. Rev. E* **82**, 031914 (2010).
26. T. E. Ouldridge, C. C. Govern, P. R. ten Wolde, Thermodynamics of computational copying in biochemical systems. *Phys. Rev. X* **7**, 021004 (2017).
27. M. Li, G. L. Hazelbauer, Cellular stoichiometry of the components of the chemotaxis signaling complex. *J. Bacteriol.* **186**, 3687–3694 (2004).
28. J. E. Keymer, R. G. Endres, M. Skoge, Y. Meir, N. S. Wingreen, Chemosensing in *Escherichia coli*: Two regimes of two-state receptors. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 1786–1791 (2006).
29. N. Barkai, S. Leibler, Robustness in simple biochemical networks. *Nature* **387**, 913–917 (1997).
30. R. G. Endres, N. S. Wingreen, Precise adaptation in bacterial chemotaxis through "assistance neighborhoods". *Proc. Natl. Acad. Sci. U.S.A.* **103**, 13040–13044 (2006).
31. G. Lan, P. Sartori, S. Neumann, V. Sourjik, Y. Tu, The energy-speed-accuracy trade-off in sensory adaptation. *Nat. Phys.* **8**, 422–428 (2012).
32. P. G. de Gennes, Chemotaxis: The role of internal delays. *Eur. Biophys. J.* **33**, 691–693 (2004).
33. D. A. Clark, L. C. Grant, The bacterial chemotactic response reflects a compromise between transient and steady-state behavior. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 9150–9155 (2005).
34. A. Celani, M. Vergassola, Bacterial strategies for chemotaxis response. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 1391–1396 (2010).
35. P. Sartori, Y. Tu, Free energy cost of reducing noise while maintaining a high sensitivity. *Phys. Rev. Lett.* **115**, 118102 (2015).

36. T. S. Shimizu, Y. Tu, H. C. Berg, A modular gradient-sensing network for chemotaxis in *Escherichia coli* revealed by responses to time-varying stimuli. *Mol. Syst. Biol.* **6**, 1–14 (2010).
37. Y. Tu, T. S. Shimizu, H. C. Berg, Modeling the chemotactic response of *Escherichia coli* to time-varying stimuli. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 14855–14860 (2008).
38. F. Tostevin, P. R. ten Wolde, Mutual information between input and output trajectories of biochemical networks. *Phys. Rev. Lett.* **102**, 1–4 (2009).
39. M. Reinhardt, G. T. kačik, P. R. ten Wolde , Path weight sampling: Exact monte carlo computation of the mutual information between stochastic trajectories. arXiv [Preprint] (2022). http://arxiv.org/abs/2203.03461 (Accessed 22 February 2023).
40. J. Monod, J. Wyman, J. P. Changeux, On the nature of allosteric transitions: A plausible model. *J. Mol. Biol.* **12**, 88–118 (1965).
41. V. Sourjik, H. C. Berg, Binding of the *Escherichia coli* response regulator CheY to its target measured in vivo by fluorescence resonance energy transfer. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12669–12674 (2002).
42. C. Walsh, *Posttranslational Modification of Proteins: Expanding Nature's Inventory* (Roberts & Company Publishers, 2006).
43. M. Bauer, M. D. Petkova, T. Gregor, E. F. Wieschaus, W. Bialek, Trading bits in the readout from a genetic network. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2109011118 (2021).
44. N. Shiraishi, K. Funo, K. Saito, Speed limit for classical stochastic processes. *Phys. Rev. Lett.* **121**, 070601 (2018).
45. J. O. Dubuis, G. Tkacik, E. F. Wieschaus, T. Gregor, W. Bialek, Positional information, in bits. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 16301–16308 (2013).