Predicting concentration changes via discrete receptor sampling

Age J. Tjalma and Pieter Rein ten Wolde*
AMOLF, Science Park 104, 1098 XG Amsterdam, The Netherlands

(Received 9 February 2024; accepted 20 June 2024; published 10 July 2024)

To successfully navigate chemical gradients, microorganisms need to predict how the ligand concentration changes in space. Due to their limited size, they do not take a spatial derivative over their body length but rather a temporal derivative, comparing the current signal with that in the recent past over the so-called adaptation time. This strategy is pervasive in biology, but it remains unclear what determines the accuracy of such measurements. Using a generalized version of the previously established sampling framework, we investigate how resource limitations and the statistics of the input signal set the optimal design of a well-characterized network that measures temporal concentration changes: the *Escherichia coli* chemotaxis network. Our results show how an optimal adaptation time arises from the trade-off between the sampling error, caused by the stochastic nature of the network, and the dynamical error, caused by uninformative fluctuations in the input. A larger resource availability reduces the sampling error, which allows for a smaller adaptation time, thereby simultaneously decreasing the dynamical error. Similarly, we find that the optimal adaptation time scales inversely with the gradient steepness, because steeper gradients lift the signal above the noise and reduce the sampling error. These findings shed light on the principles that govern the optimal design of the *E. coli* chemotaxis network specifically, and any system measuring temporal changes more broadly.

DOI: 10.1103/PhysRevResearch.6.033049

I. INTRODUCTION

Organisms ranging from bacteria to mammals have learned to navigate their environment in order to find food and avoid threats. Successful navigation requires an organism to predict the spatial structure of its surroundings, which necessitates measuring and storing relevant environmental properties. Therefore, how accurately these signals are sensed can fundamentally limit the success of navigation [1]. This in turn raises the question how accurately such signals can be transduced.

Microorganisms that navigate chemical gradients need to determine the correct direction to move in, which entails predicting the change in concentration that they will encounter, rather than its value. Because these organisms are typically small relative to the gradient length, the measurement error is large compared with the concentration difference over their body length [2]. Therefore, they cannot directly measure the gradient. Instead, these microorganisms only have access to the local concentration. Yet, they can also store past concentrations. How these cells should integrate the current and past information to predict the concentration change remains, however, unclear. In principle, cells can combine the concentration value with its derivative to predict the concentration change, and the optimal strategy for combining this information depends on the statistics of the environment. If the range of background concentrations is large compared with the typical concentration change over the signal correlation time as set by the organism's own motion, then the optimal system for predicting the concentration change is one that exhibits perfect adaptation [3]. It means that the organism bases its prediction on the concentration change only.

Interestingly, various organisms have indeed been shown to employ this strategy. A canonical example is the bacterial chemotaxis system, which is widely conserved across species [4–8]. But also eukaryotic sperm cells measure temporal changes when navigating towards an egg [9–11], and even the multicellular nematode *Caenorhabditis elegans* depends on temporal derivatives in a range of taxis behaviors [12].

Even though measuring temporal changes appears to be a common and important function, it is not clear what sets the accuracy of such measurements. The fundamental information-processing devices that allow living cells to measure concentration changes are biochemical signaling networks. Like any device, the accuracy of such networks is limited by the physical resources required to build and operate them, such as energy, components, and time. Here, we investigate how these resources limit the accuracy with which cells can predict changes in the encountered concentration during navigation. Specifically, we ask what determines the optimal design of the signaling network under limited resource availability.

To measure a temporal change, cells subtract from the most recent signal the signal further back into the past. The latter is performed via the adaptation system. Crucially, to yield a response of nonzero amplitude, which is necessary to lift the signal above the inevitable biochemical noise, the system cannot adapt instantly; it therefore cannot take an instantaneous derivative. On the other hand, the adaptation time should not

*Contact author: tenwolde@amolf.nl

be too long, because then the temporal derivative is taken over a larger window stretching further back into the past, which is less informative about the current or future derivative that the cell needs to predict. We thus expect that there exists an optimal adaptation time that arises from this trade-off between a derivative that is most recent and one that is most reliable [3]. However, what precisely controls the optimal adaptation time and how this depends on the statistics of the input and the available resources such as receptor and readout copies, remains unknown.

An intuitive perspective that is ideally suited to answer these questions is the previously established sampling framework [13–15]. This framework views the signaling network downstream of the receptor as a device that discretely samples the state of the receptor. From this starting point, it enables identification of the different contributions that comprise the full sensing error: the sampling error, caused by fluctuations in the number of samples, the binary nature of the receptor state, and receptor-level noise; and the dynamical error, resulting from uninformative fluctuations in the input. While previous work has used the sampling framework to investigate sensing the current signal, we here generalize and extend it to include the prediction of signal properties a specified time into the future. We then apply this generalized sampling framework to the *Escherichia coli* chemotaxis network, a well-characterized example of a network which measures temporal changes. We model the input signal after the experimentally measured input for *E.coli* chemotaxis in shallow gradients [1].

Our results distinctly show how an optimum for the adaptation time arises from its opposing effects on the sampling error and the dynamical error. While the former decreases with adaptation time, the latter increases with it. Given the adaptation time, a larger number of receptor and readout molecules reduces the sampling error, shifting the balance between the sampling and dynamical error. Therefore, increasing the resource availability reduces the optimal adaptation time. Similarly, we find that the optimal adaptation time scales inversely with the steepness of the chemical gradient in which the organism navigates. The reason is that, in a steeper gradient, the signal is more easily distinguished from the noise under the same resource availability. This again means that the sampling error decreases relative to the dynamical error, reducing the optimal adaptation time to decrease the latter. Finally, if the dynamics of the concentration change are Markovian, the optimal adaptation time is independent of the prediction interval. These findings likely extend well beyond *E. coli* and have implications for the optimal design of any system that measures temporal changes, be it natural or man-made.

II. RESULTS

A. Theory: Sampling framework

In general, the function of a biochemical signaling network is to estimate the value of a signal of interest, which typically varies in time. Sensing entails estimating the value of the signal at the current time t_0, while predicting the future state of the environment, implies estimating the value a time τ into the future. To extend the sampling framework to be applicable to prediction as well as sensing we define the signal of interest as $s_\tau \equiv s(t_0+\tau)$ with $\tau \geqslant 0$. In this work we consider a time-varying input signal described by stationary Gaussian statistics (see Sec. II B).

In biochemical signaling networks, the activity state of receptor proteins is altered by the ligand molecules that bind them. In turn, downstream readout proteins stochastically sample the receptor state $n \in \{0, 1\}$. From these samples the signal of interest must then be inferred. A canonical motif that samples the activity state of upstream receptor proteins is the push-pull network [16]. In this network a sample of the receptor state is stored in the chemical modification state of a readout protein, which decorrelates from the receptor state over the response time τ_r [Fig. 1(a)].

To estimate the signal value s_τ a time τ into the future, the cell integrates the receptor activity over a time τ_r, leading to an estimate $\hat{p}_{\tau_r}$ of the average receptor activity p_{τ_r} over the integration time τ_r [Fig. 1(b)]. However, during this past time τ_r, the input signal varies over its own timescale τ_v, which leads to changes in the receptor activity on this timescale as well [14,15]. On top of variation on the timescale of the input dynamics, the receptor activity fluctuates on the timescale of ligand binding and unbinding, and on the timescale of the adaptation mechanism τ_m. In the linear regime, the dynamic input-output relation between the average receptor activity p_{τ_r} and the signal of interest s_τ is given by

$$p_{\tau_\mathrm{r}}(s_\tau) \equiv \mathbb{E}[\langle n(t_i)|s_\tau\rangle]_{t_i} = p + \tilde{g}s_\tau, \tag{1}$$

where the angle brackets denote an ensemble average over all receptors, $\mathbb{E}[\dots]_{t_i}$ is an average over all sampling times t_i, which are exponentially distributed over the integration time τ_r [Eq. (A4)], and $p \equiv \mathbb{E}[\langle n(t_i)\rangle]_{t_i}$ is the average receptor activity over all signal values. The dynamic input-output relation thus gives the average receptor activity p_{τ_r} over the response time τ_r given that the future signal is s_τ; p_{τ_r} is thus an average over all sources of noise, arising from receptor-ligand binding and receptor methylation, readout activation, and fluctuations in the past input that are not informative because they map onto the same future signal s_τ (see Fig. 2). The slope of the mapping between s_τ and p_{τ_r} is the dynamic gain $\tilde{g}$ [Fig. 1(c)] [17].

The accuracy of any signaling device can be quantified using the signal-to-noise ratio (SNR), which is a measure for the number of distinct signal values the system can distinguish. For systems with Gaussian statistics, as studied here, the SNR is given by the ratio of the signal variance $\sigma^2_{s_\tau}$ over the error in the cell's estimate of the signal $(\delta\hat{s}_\tau)^2 \equiv \mathbb{E}[\mathrm{Var}(\hat{s}_\tau|s_\tau)]_{s_\tau}$, i.e., the variance of the cell's signal estimate $\hat{s}_\tau$ under a fixed signal s_τ, averaged over all s_τ:

$$\mathrm{SNR} \equiv \frac{\sigma^2_{s_\tau}}{(\delta\hat{s}_\tau)^2} = \frac{\tilde{g}^2\sigma^2_{s_\tau}}{\sigma^2_{\hat{p}_{\tau_\mathrm{r}}}}. \tag{2}$$

The cell estimates the signal s_τ from the average receptor activity over the integration time, p_{τ_r}, via the dynamic input-output relation, see Eq. (1) and Fig. 1(c). Using the rules of error propagation, the error in the signal estimate is thus given by

$$(\delta\hat{s}_\tau)^2 = \sigma^2_{\hat{p}_{\tau_\mathrm{r}}}/\tilde{g}^2, \tag{3}$$

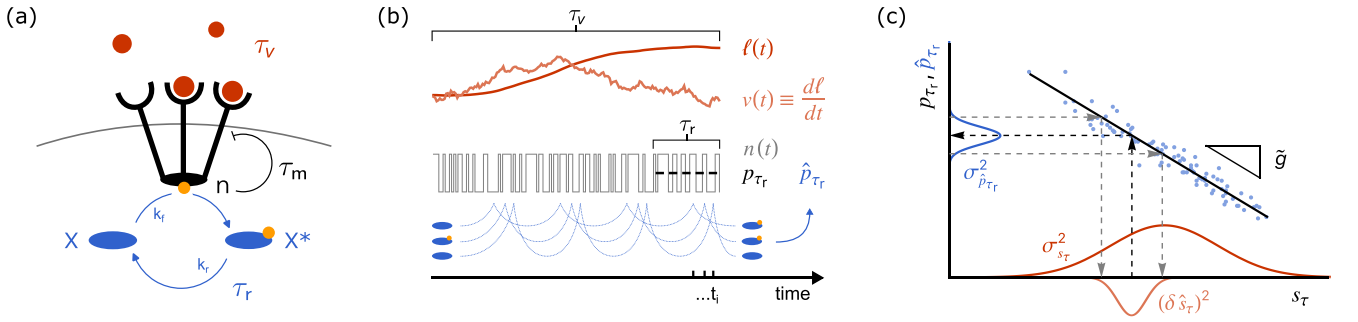

FIG. 1. A push-pull motif samples the binary state of the chemotaxis receptor cluster. (a) Ligand binding affects the probability of a chemotaxis receptor cluster to reside in its active or inactive conformation. This binary cluster state n controls the methylation dynamics of its constituent receptors, leading to negative feedback on the adaptation timescale τ_m. The cluster state is sampled by the readout molecules X on the response timescale τ_r. (b) We consider an input signal defined by its concentration $\ell(t)$ and concentration derivative $v(t)$, with correlation time τ_v [Eqs. (12)–(14)]. The instantaneous cluster activity $n \in \{0, 1\}$ switches fast relative to the input correlation time, response time τ_r, and adaptation time τ_m. Due to the negative feedback, the mean cluster activity reflects the change in concentration over the past adaptation time τ_m. The network makes an estimate $\hat{p}_{\tau_r}$ of the cluster activity over the past response time τ_r by discretely sampling the instantaneous cluster state via the push-pull motif [panel (a)]. The estimate $\hat{p}_{\tau_r} = x^*/\overline{N}$ is given by the current number of active readout molecules x^*, reflecting the number of samples of active receptor clusters during the past integration time τ_r, over the mean number of samples $\overline{N}$ during this time τ_r [Eq. (5c)]. For linear Gaussian systems the future signal s_τ maps onto a current mean cluster activity over the response time p_{τ_r} via the dynamic input-output relation of Eq. (1). The variance in the estimate $\hat{p}_{\tau_r}$ given a signal value s_τ is the prediction error $\sigma^2_{\hat{p}_{\tau_r}}$. Mapping the prediction error back onto the signal gives the network's error in the signal estimate $(\delta \hat{s}_\tau)^2$. The ratio between the total variance in the signal $\sigma^2_{s_\tau}$ and the error in the signal estimate $(\delta \hat{s}_\tau)^2$ is the signal-to-noise ratio [Eq. (2)].

where the error in the estimate of the receptor activity $\hat{p}_{\tau_r}$ over the integration time τ_r is defined as

$$\sigma^2_{\hat{p}_{\tau_r}} \equiv \mathbb{E}[\mathrm{Var}(\hat{p}_{\tau_r}|s_\tau)]_{s_\tau}. \tag{4}$$

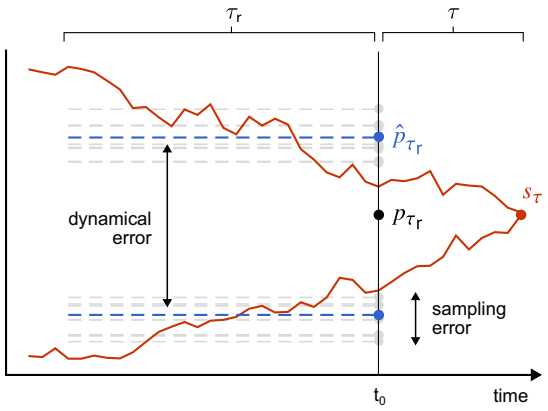

FIG. 2. The total error in the cell's estimate of the receptor activity can be decomposed into the dynamical error and the sampling error. For linear signaling systems, a given current or future signal s_τ (red dot) maps onto a single mean receptor activity p_{τ_r} at the current time t_0 (black dot) via the dynamic input-output relation of Eq. (1) [Fig. 1(c)]. However, the past input and thus receptor activity on which the estimate $\hat{p}_{\tau_r}$ (blue dot) is based varies in time, leading to a dynamical error. This error arises because different past trajectories of the signal map onto a common future value s_τ, leading to uninformative variations in $\hat{p}_{\tau_r}$. Even for a given input trajectory the receptor noise, which in this work is only caused by receptor methylation, and the stochastic nature of the sampling process downstream of the receptor, lead to deviations in the estimate $\hat{p}_{\tau_r}$ (gray dots) which constitute the sampling error.

The signal-to-noise ratio of Eq. (2) also specifies the Gaussian mutual information between the signal and the network output [18].

To quantify the error in the cell's estimate of the receptor activity [Eq. (4)], we have to consider how the cell makes this estimate. As a model system to investigate networks that measure changes in the input we use the *E. coli* chemotaxis network. In this network, the activity of a receptor cluster reflects the change in signal concentration over the past adaptation time τ_m (see Sec. II C for details). Downstream of the cluster, its activity state is sampled via a push-pull motif [Fig. 1(a)] [16]. The cell's estimate of the fraction of active clusters is given by (also see Ref. [13])

$$\hat{p}_{\tau_r} = \frac{1}{\overline{N}} \sum_{i=1}^{N} n_i(t_i) = \frac{x^*}{\overline{N}}, \tag{5}$$

where $n_i(t_i) \in \{0, 1\}$ is the outcome of sample i at sampling time t_i, which is set by the binary activity state of the receptor cluster that was sampled at time t_i [Fig. 1(b)]. The physical readout of the network is the number of active readout molecules $x^* = \sum_{i=1}^{N} n_i(t_i)$, which have been phosphorylated by an active receptor cluster. Since readout phosphorylation is driven by ATP hydrolysis, we consider the sampling process in the irreversible limit.

The number N of samples is set by the rate r of sampling and the timescale over which samples remain correlated with the receptor state, which is set by the integration, or response time τ_r. In the push-pull motif the sampling rate is set by the forward rate constant k_f, the number R_T of receptor clusters, and the number X of available readout molecules: $r = k_f x R_T$ [Fig. 1(a)]. We assume that N is Poisson distributed with mean $\overline{N} = \bar{r}\tau_r$. This mean number of samples can be expressed in terms of the steady-state fraction of phosphorylated readouts

$f = k_f p R_T \tau_r$ and the total number of readouts X_T [15],

$$\overline{N} = f(1-f)X_T/p. \tag{6}$$

The steady-state flux of readout molecules is given by $\bar{r}p = f(1-f)X_T/\tau_r$.

Using the definition of the cell's estimate of the receptor activity [Eq. (5)] the error in this estimate [Eq. (4)] can be decomposed into independent parts in a very general manner. We set out this decomposition in the section that follows. After the decomposition of the error we describe the dynamics and statistics of the input of the chemotaxis network (Sec. II B). Subsequently we introduce the chemotaxis network in more detail and compute the dynamic gain $\tilde{g}$ [see Eq. (1)], and the different contributions to the error in terms of the parameters of the system (Sec. II C). We then compute the full expression for the SNR and investigate its behavior as a function of the prediction interval, the resource availability, and the adaptation time (Secs. II D–II F). We compare the predictions of our theory to available experimental data on the *E. coli* chemotaxis network in Sec. II G. Finally, we illustrate the effect of the signal statistics on our results by computing the relative error for a different set of statistics (Sec. II H).

1. The error in the estimate of the receptor activity

We can derive a general expression for the prediction error, $\sigma^2_{\hat{p}_{\tau_r}}$, which shows how the complete error decomposes into independent parts. We start from the definition of the error [Eq. (4)], which we rewrite using the law of total variance:

$$\begin{aligned}\sigma^2_{\hat{p}_{\tau_r}} &= \mathrm{Var}(\hat{p}_{\tau_r}) - \mathrm{Var}(\mathbb{E}[\hat{p}_{\tau_r}|s_\tau]) \\ &= \mathrm{Var}(\mathbb{E}[\hat{p}_{\tau_r}|N]) + \mathbb{E}[\mathrm{Var}(\hat{p}_{\tau_r}|N)] - \mathrm{Var}(\mathbb{E}[\hat{p}_{\tau_r}|s_\tau]),\end{aligned} \tag{7}$$

where in the first line we use that the total variance in the estimate of the activity, $\mathrm{Var}(\hat{p}_{\tau_r})$, is the sum of the variance in the mean of $\hat{p}_{\tau_r}$ given s_τ, $\mathrm{Var}(\mathbb{E}[\hat{p}_{\tau_r}|s_\tau])$, and the mean of the variance in $\hat{p}_{\tau_r}$ conditional on s_τ, $\mathbb{E}[\mathrm{Var}(\hat{p}_{\tau_r}|s_\tau)]$, which is the error $\sigma^2_{\hat{p}_{\tau_r}}$ [Eq. (4)]. Indeed, the error in the estimate is its total variance minus the part which is informative about the signal of interest s_τ. Subsequently, in the second line, we split the total variance in the estimate $\hat{p}_{\tau_r}$ into a part that arises from fluctuations in the number of samples N, the first right-hand side (RHS) term, and the mean variance in $\hat{p}_{\tau_r}$ when N is fixed, the second RHS term.

In Appendix A we show how each term of Eq. (7) can be simplified further using the definition of the cell's estimate $\hat{p}_{\tau_r}$ [Eq. (5)]. The first term, the error caused by fluctuations in the number of samples, is given by

$$\mathrm{Var}(\mathbb{E}[\hat{p}_{\tau_r}|N]) = \frac{p^2}{\overline{N}}, \tag{8}$$

with the average cluster activity $p \equiv \mathbb{E}[\langle n(t_i)\rangle]_{t_i}$. As shown in previous work, this error would be zero if the sampled cluster functions bidirectionally, i.e., if inactive clusters would dephosphorylate readout molecules [13]. In contrast, in the chemotaxis network deactivation is not driven by inactive receptor clusters but rather by an enzyme (CheZ) independent of the receptor state, and then this term is nonzero. The fluctuations under a fixed number of samples, the second RHS term of Eq. (7), can be decomposed further into three parts:

$$\begin{aligned}\mathbb{E}[\mathrm{Var}(\hat{p}_{\tau_r}|N)] = \frac{p(1-p)}{\overline{N}} &+ \mathbb{E}[\mathrm{Cov}(n_i(t_i), n_j(t_j)|\boldsymbol{s})]_{t_i,t_j,\boldsymbol{s}} \\ &+ \mathrm{Var}(\mathbb{E}[\langle n(t_i)|\boldsymbol{s}\rangle]_{t_i}),\end{aligned} \tag{9}$$

where the first part reflects the instantaneous variance of each sampled cluster, the second part is the cluster covariance under a fixed past signal trajectory $\boldsymbol{s} \equiv \{s(t)\}_{t\leqslant t_0}$, and the third part quantifies the effect of the signal history $\boldsymbol{s}$ on the activity of the cluster. Finally, the variance that is informative of the future signal value, i.e., the third RHS term of Eq. (7), is given by

$$\mathrm{Var}(\mathbb{E}[\hat{p}_{\tau_r}|s_\tau]) = \mathrm{Var}(\mathbb{E}[\langle n(t_i)|s_\tau\rangle]_{t_i}) = \tilde{g}^2\sigma^2_{s_\tau}, \tag{10}$$

which follows directly from the dynamic input output relation in Eq. (1). Substituting Eqs. (8)–(10) into Eq. (7) yields the full prediction error:

$$\sigma^2_{\hat{p}_{\tau_r}} = \underbrace{\frac{p^2}{\overline{N}} + \frac{p(1-p)}{\overline{N}} + \mathbb{E}[\mathrm{Cov}(n_i(t_i), n_j(t_j)|\boldsymbol{s})]_{t_i,t_j,\boldsymbol{s}}}_{\text{sampling error}} + \underbrace{\mathrm{Var}(\mathbb{E}[\langle n(t_i)|\boldsymbol{s}\rangle]_{t_i}) - \tilde{g}^2\sigma^2_{s_\tau}}_{\text{dynamical error}}. \tag{11}$$

The first three terms together make up the sampling error. This error arises due to the stochastic nature of the sampling process downstream of the receptor, receptor-ligand binding and unbinding, and the adaptation mechanism. In this work we integrate out ligand binding, and we therefore find that receptor methylation constitutes the only noise source on the receptor level. The sampling error quantifies all variability in the output under a constant input, as in Ref. [13] (see Fig. 2). The final two terms constitute the dynamical error; this is the error that arises from fluctuations in $\hat{p}_{\tau_r}$ that are caused by differences between past signal trajectories that map onto the same future signal of interest. These fluctuations contribute to the error in $\hat{p}_{\tau_r}$ because they do not provide any information on the future signal of interest [14] (Fig. 2).

Equation (11) reflects that, in general, the error for a linear sensing system can be decomposed into a contribution that arises from the stochastic sampling of the signal and a contribution that comes from the fact that not all signal fluctuations in the past correspond to the signal which the cell aims to predict. More specifically, Eq. (11) holds for any cellular sensing system in which the signal is inferred from the receptor activity, estimated using the downstream signaling system as a sampling device, as in Eq. (5). Yet, to derive the sensing error SNR^{-1} [see Eqs. (2)–(4)] for the chemotaxis network, we need to evaluate the sampling error

and the dynamical error, as well as the dynamic gain $\tilde{g}$ [see Eq. (3)]. These quantities depend on the specific characteristics of the sensing system and the signal statistics, discussed next.

B. Signal statistics

In general, it is hard to know what the natural input statistics are that an organism experiences, and these may vary widely. We can start from the observation that microorganisms in dilute environments are faced with chemical gradients that are exceedingly shallow compared with their own length. In such environments, the only signal property that the cell can measure is the local concentration. But to determine if it is moving in the right direction, the cell must predict the change in concentration over time. So, while the cell can only measure concentrations, it is interested in the concentration's temporal derivative.

An ideal model system to study networks that can predict temporal changes is the *E. coli* chemotaxis network. *E. coli* swims in its environment with a speed which exhibits persistence. This leads to an autocorrelation function for the concentration change which does not decay instantaneously [1]. To model a signal which is characterized by both the concentration and its derivative, and in which correlations in the derivative persist over the correlation time set by the motion of the cell, we use the classical model of a particle in a harmonic well [3],

$$\delta\dot{\ell} = v(t), \tag{12}$$

$$\dot{v} = -\omega_0^2 \delta\ell(t) - v(t)/\tau_v + \eta_v(t). \tag{13}$$

Here, $\delta\ell(t) \equiv [c(t) - c_0]/c_0$ is the relative deviation of the concentration $c(t)$ from its background value c_0. The derivative of this relative concentration is $v(t)$, and $\eta_v(t)$ is a Gaussian white-noise process that drives the stochastic fluctuations in the signal. The parameter ω_0 sets the variance in the concentration σ_ℓ^2 relative to that in its derivative $\sigma_\ell^2 = \sigma_v^2/\omega_0^2$, where the variance in the derivative σ_v^2 is set by the swimming behavior of the cell. The relaxation time τ_v is set by the run duration because this is the timescale over which the input fluctuations decorrelate.

The range of ligand concentrations which *E. coli* might encounter is very large, based on the dissociation constants of the inactive and active receptor conformations. For the Tar-MeAsp receptor ligand combination these are, respectively, $K_D^I = 18$ μM and $K_D^A = 2900$ μM [19–21]. This suggests that the total variance in the ligand concentration is much larger than the concentration change over the course of a run, i.e., $\sigma_\ell \gg \tau_v \sigma_v$ and thus $\omega_0 \ll \tau_v^{-1}$. In this regime, the correlation function of $v(t)$ becomes a simple exponential with variance σ_v^2 and decay time τ_v:

$$\langle \delta v(t) \delta v(t') \rangle = \sigma_v^2 e^{-|t-t'|/\tau_v}. \tag{14}$$

The correlation function of Eq. (14) corresponds to what has been observed experimentally for *E. coli* cells swimming in shallow exponential concentration gradients [1]. When cells swim in shallow gradients, i.e., with a characteristic length much longer than the length of a run, they swim as if there is no gradient. The correlation function of the positional velocity $v_x(t)$ in the absence of a gradient has been measured to be an exponential with variance $\sigma_{v_x}^2$ and decay time τ_v set by the duration of a run [1]. This can be mapped onto the correlation function of Eq. (14), where $v(t) \equiv c_0^{-1} dc/dt$, when we consider that the concentration gradient is given by $c(t) = c_0 \exp[gx(t)]$ with the gradient steepness g. We find for the absolute concentration change over time $dc/dt = (dc/dx)(dx/dt) = gc(t)v_x(t)$, and thus we have for variance of the relative concentration change $v(t)$:

$$\sigma_v^2 = g^2 \sigma_{v_x}^2. \tag{15}$$

Experimental measurements provide the relaxation time $\tau_v^{-1} = 0.86\ \mathrm{s}^{-1}$ and the variance of the positional derivative $\sigma_{v_x}^2 = 157\ \mu\mathrm{m}^2\ \mathrm{s}^{-2}$ [1].

C. Chemotaxis model

In the *E. coli* chemotaxis network, receptors cooperatively control the activity of the kinase CheA, which controls the phosphorylation of the readout protein CheY [Fig. 1(a)] [22–25]. The receptor cooperativity has been successfully described using the Monod-Wyman-Changeux (MWC) model, where individual receptors are assumed to form clusters in which all receptors must reside in the same activity state [1,21,24,26–30]. Furthermore, inactive receptors are methylated by the enzyme CheR, which increases the probability for the cluster to be active, and active receptors are demethylated by CheB. These methylation dynamics ensure that the network exhibits perfect adaptation with respect to the background concentration [5,20,31–34]. Therefore, the activity state of the cluster only responds transiently to changes in the input, and reflects the recent change in concentration.

Because both ligand binding and switching between the active and inactive state of the cluster are fast compared with the input, methylation, and phosphorylation dynamics, it is instructive to take a quasi-equilibrium approach and consider the average cluster activity given the methylation level of the cluster and the extracellular ligand concentration. In the linear noise approximation we have for the activity (see Appendix B)

$$a(t) \equiv \langle n(t) | \delta m, \delta\ell \rangle = p + \alpha \delta m(t) - \beta \delta\ell(t), \tag{16}$$

where p is the mean activity, $\delta m(t)$ represents the methylation level of the cluster, and $\delta\ell(t)$ represents the ligand concentration, both defined as deviations from their mean. The constants α and β respectively depend on the free-energy cost of methylation, $\tilde{\alpha}$, and on the dissociation constants K_D^I and K_D^A and background concentration c_0. The methylation dynamics are given by,

$$\dot{\delta m} = -\delta a(t)/(\alpha \tau_m) + \eta_m(t), \tag{17}$$

where τ_m is the adaptation time, and η_m is Gaussian white noise [see Eq. (B7)].

The dynamic gain of the network maps the signal of interest onto the receptor activity [Eq. (1), Fig. 1(c)]. For the purpose of navigation, we define the signal of interest to be the change in concentration $v_\tau \equiv v(t_0 + \tau)$ some time $\tau \geqslant 0$ into the future. The autocorrelation of the change in concentration is given by Eq. (14). The dynamic gain of the chemotaxis

network with respect to this signal of interest is [Eqs. (B10)–(B13)]

$$\tilde{g} = \frac{g_{v\to p} e^{-\tau/\tau_v}}{(1+\tau_\mathrm{m}/\tau_v)(1+\tau_\mathrm{r}/\tau_v)} = \frac{-\tau_\mathrm{m}\beta e^{-\tau/\tau_v}}{(1+\tau_\mathrm{m}/\tau_v)(1+\tau_\mathrm{r}/\tau_v)}, \tag{18}$$

where τ_v is the signal correlation time, τ_r is the network response time, τ_m is the adaptation time, and the static gain from the input signal derivative v to the steady-state activity p is given by

$$g_{v\to p} \equiv \partial_v p = -\tau_\mathrm{m}\beta. \tag{19}$$

Equation (18) shows that the dynamic gain $\tilde{g}$ is maximized for a fast response $\tau_\mathrm{r} \ll \tau_v$, and slow adaptation $\tau_\mathrm{m} \gg \tau_v$. A longer adaptation time increases the dynamic gain via the static gain [Eq. (19)], because the absolute difference between sequential inputs is on average larger over this longer time. Yet, the dynamic gain saturates as τ_m increases:

$$\lim_{\tau_\mathrm{m}\to\infty} \tilde{g} = \frac{-\tau_v\beta e^{-\tau/\tau_v}}{1+\tau_\mathrm{r}/\tau_v}. \tag{20}$$

In this limit, considering that typically $\tau \leqslant \tau_v$ and $\tau_\mathrm{r} \ll \tau_v$, the dynamic gain is approximately proportional to the signal correlation time τ_v. The reason is that fluctuations further than τ_v in the past cannot affect the mapping from the current signal, which is most correlated to the signal of interest v_τ, to the current receptor state. Finally, increasing the prediction interval τ reduces the dynamic gain because the correlation between future signal and sensed input decreases.

To determine the sampling error of the chemotaxis network, we require the cluster covariance under a fixed input signal, which is the third RHS term in Eq. (11). In our chemotaxis model, this covariance is a consequence of the methylation noise only and it is given by [Eqs. (B14)–(B19)]

$$\mathbb{E}[\mathrm{Cov}(n_i(t_i), n_j(t_j)|\boldsymbol{s})]_{t_i,t_j,\boldsymbol{s}} = \frac{\alpha p(1-p)}{R_\mathrm{T}(1+\tau_\mathrm{r}/\tau_\mathrm{m})} \approx \alpha p(1-p)/R_\mathrm{T}. \tag{21}$$

Here, p is the mean cluster activity and R_T is the total number of independent receptor clusters. Substitution of Eq. (21) into Eq. (11) yields the full sampling error of the chemotaxis network,

$$\sigma^{2,\mathrm{samp}}_{\hat{p}_{\tau_\mathrm{r}}} = \frac{p^2}{\overline{N}} + \frac{p(1-p)}{\overline{N}_\mathrm{I}}, \tag{22}$$

with the number of independent samples

$$\overline{N}_\mathrm{I} \equiv f_\mathrm{I}\overline{N} = \frac{\overline{N}}{1+\overline{N}/R_\mathrm{I}}, \tag{23}$$

where $f_\mathrm{I} = 1/(1+\overline{N}/R_\mathrm{I})$ is the fraction of independent samples and

$$R_\mathrm{I} = R_\mathrm{T}(1+\tau_\mathrm{r}/\tau_\mathrm{m})/\alpha \tag{24}$$

is the number of independent receptor states during an integration time τ_r.

In contrast with previous work [14,15], the sampling error does not depend on the correlation time τ_c of receptor-ligand binding because here we have assumed that ligand binding is much faster than the response time τ_r. Still, the cluster state remains correlated over time due to receptor methylation, and this means that the expression for the sampling error of the chemotaxis network studied here is almost identical to that of the push-pull network studied in Refs. [14,15]. However, unlike ligand binding noise, the methylation noise cannot be averaged out because the methylation timescale τ_m is longer than the response time τ_r, i.e., $1+\tau_\mathrm{r}/\tau_\mathrm{m} \approx 1$ in Eq. (21). Moreover, because the methylation noise affects the receptor activity via the factor α [Eqs. (16) and (17)], which controls how strongly methylation changes the receptor activity, the cluster covariance also increases with α, since it increases the temporal covariance within each cluster.

Equation (23) reflects that the number of receptor samples $\overline{N}$, which is proportional to the number of readouts X_T [Eq. (6)], and the number of independent receptor states R_I, proportional to the number of receptor clusters R_T [Eq. (24)], are fundamental resources that limit the sensing accuracy like weak links in a chain [13]: when $\overline{N} \gg R_\mathrm{I}$ the number of independent samples is limited by the number of receptor states and $\overline{N}_\mathrm{I} \approx R_\mathrm{I}$, and vice versa, when $R_\mathrm{I} \gg \overline{N}$ the total number of samples is limiting and $\overline{N}_\mathrm{I} \approx \overline{N}$ (also see Fig. 5 and Appendix C).

Finally, to compute the dynamical error, we derive the variation in the network output that is caused by the past input trajectory [Eqs. (B20)–(B25)]:

$$\mathrm{Var}(\mathbb{E}[\langle n(t_i)|\boldsymbol{s}\rangle]_{t_i}) = \frac{g^2_{v\to p}\sigma_v^2}{(1+\tau_\mathrm{m}/\tau_v)(1+\tau_\mathrm{r}/\tau_v)} \times \left(1+\frac{\tau_\mathrm{m}\tau_\mathrm{r}}{\tau_v(\tau_\mathrm{m}+\tau_\mathrm{r})}\right), \tag{25}$$

with the static gain $g_{v\to p}$ given by Eq. (19). Just like the dynamic gain [Eq. (18)] this variation is maximized for a fast response $\tau_\mathrm{r} \ll \tau_v$ and slow adaptation $\tau_\mathrm{m} \gg \tau_v$. Indeed, in the regime that $\tau_\mathrm{m} \gg \tau_v$ we have $\mathrm{Var}(\mathbb{E}[\langle n(t_i)|\boldsymbol{s}\rangle]_{t_i}) \propto \tau_\mathrm{m}$. Therefore, unlike the dynamic gain, Eq. (25) does not saturate for an increasing adaptation time. The reason is that more and more values of the historical input contribute to the variance in the output as long as the system does not adapt. Clearly, not all of the variation quantified by Eq. (25) will carry information about the signal of interest v_τ. Substituting Eq. (25) in Eq. (11) yields the dynamical error, which is the total uninformative variation caused by the past input trajectory

$$\sigma^{2,\mathrm{dyn}}_{\hat{p}_{\tau_\mathrm{r}}} = \tilde{g}^2\sigma_v^2\left[e^{2\tau/\tau_v}\left(1+\frac{\tau_\mathrm{m}}{\tau_v}\right)\left(1+\frac{\tau_\mathrm{r}}{\tau_v}\right) \times \left(1+\frac{\tau_\mathrm{m}\tau_\mathrm{r}}{\tau_v(\tau_\mathrm{m}+\tau_\mathrm{r})}\right) - 1\right], \tag{26}$$

with the dynamic gain $\tilde{g}$ of Eq. (18). Even though Eq. (26) is the dynamical error in predicting the concentration change (rather than the value), its form is strikingly similar to the dynamical error of a push-pull network that predicts the current concentration, derived in Refs. [14,15]. The reason for this similar form is that both the concentration considered in these previous works and the concentration derivative considered here are Markovian signal properties. Indeed, if we consider a non-Markovian signal concentration and derivative, the dynamic gain and dynamical error change (see Sec. II H).

D. Relative prediction error

The central result of this work is the relative error, i.e., SNR^{-1}, made by the *E. coli* chemotaxis network when it predicts the future concentration change. Using the definition of the signal-to-noise ratio [Eq. (2)], with the dynamic gain given in Eq. (18), and the prediction error $\sigma^2_{\hat{p}_{\tau_r}} = \sigma^{2,\mathrm{samp}}_{\hat{p}_{\tau_r}} + \sigma^{2,\mathrm{dyn}}_{\hat{p}_{\tau_r}}$ given by Eqs. (22) and (26), we obtain

$$\mathrm{SNR}^{-1} = \underbrace{\frac{e^{2\tau/\tau_v}}{\tau_m^2\beta^2\sigma_v^2}\left(1+\frac{\tau_m}{\tau_v}\right)^2\left(1+\frac{\tau_r}{\tau_v}\right)^2\left(\frac{p^2}{\overline{N}}+\frac{p(1-p)}{\overline{N}_I}\right)}_{\text{sampling error}} + \underbrace{e^{2\tau/\tau_v}\left(1+\frac{\tau_m}{\tau_v}\right)\left(1+\frac{\tau_r}{\tau_v}\right)\left(1+\frac{\tau_m\tau_r}{\tau_v(\tau_m+\tau_r)}\right)-1}_{\text{dynamical error}}. \quad (27)$$

This expression is similar in structure to the relative error of the push-pull network without adaptation, which was derived in earlier work [Eq. (6) of Ref. [14]]. The reason is that, while the adaptation system affects the receptor dynamics, the downstream push-pull motif still acts as a device that discretely samples the receptor state. As a result, the relative error has two contributions: the sampling error, which arises from the stochasticity in sampling the state of the receptor, and the dynamical error, which arises from the dynamics of the input signal (see Fig. 2). However, while this expression for the relative error has a form that is similar to that for the push-pull network, there are also key differences.

First of all, both the sampling and the dynamical error depend on the forecast interval. In general, the dynamical error arises because while the system aims to predict the current or future derivative, it measures the change in concentration over the timescale τ_m on the level of the receptor, and reads out the receptor activity over the timescale τ_r (Fig. 2). The network thus only measures an instantaneous concentration change when both τ_m and τ_r go to zero. Still, even in this limit, the dynamical error remains finite as long as the forecast interval is larger than zero, due to the inherent unpredictability of the future signal.

Perhaps surprisingly, the relative sampling error also depends on the forecast interval τ. While the absolute sampling error of the network is independent of the forecast interval [Eq. (22)], the dynamic gain does depend on it [Eq. (18)]. When the forecast interval increases, the dynamic gain decreases, reducing the effect of the signal of interest on the receptor activity. Therefore, while the absolute sampling error remains constant, the relative sampling error increases with the forecast interval. In short, for a larger forecast interval it becomes harder to lift the signal above the sampling noise.

The second notable difference with the result on the push-pull network concerns the role of adaptation. It reflects the fact that the chemotaxis system takes a temporal derivative at the receptor level on a timescale set by the adaptation time. The dynamical error increases monotonically with the adaptation time τ_m, because for a longer adaptation time the system compares the current concentration to concentrations further in the past. Consequently, this change in concentration is less informative about the current derivative, which is the signal property most correlated to the future derivative [Eq. (14)]. In fact, when $\tau_m \to \infty$, the system does not adapt and the chemotaxis network therefore reduces to a push-pull network, which does not measure the derivative but rather the signal value. In this limit, the dynamical error diverges because the signal value is not correlated with the future derivative that the cell aims to predict. In contrast, the sampling error decreases monotonically with τ_m, because a longer adaptation time increases the dynamic gain [Eq. (18)]. How the optimal adaptation time that arises from these antagonistic effects depends on other parameters, such as the gradient steepness and the resource availability, is discussed in Sec. II F.

A third difference resides in the number of independent samples $\overline{N}_I$ [Eq. (23)]. For a push-pull network driven by a simple receptor, the number of independent samples is given by $\overline{N}_I^{\mathrm{PPN}} = f_I^{\mathrm{PPN}}\overline{N}_{\mathrm{eff}}$, where the number of effective samples $\overline{N}_{\mathrm{eff}} = \overline{N}$ in the irreversible limit, as we also study here [13–15]. For the push-pull network the fraction of independent samples can be expressed as $f_I^{\mathrm{PPN}} = 1/(1+\overline{N}/R_I^{\mathrm{PPN}})$ with the number of independent receptor states during an integration time $R_I^{\mathrm{PPN}} = R_T(1+\tau_r/\tau_c)$, where τ_c is the correlation time of the receptor binding state [13–15]. In our treatment of the chemotaxis model we consider the limit where $\tau_c \ll \tau_r$, in which case R_I^{PPN} diverges and $f_I^{\mathrm{PPN}} \approx 1$. However, the fraction of independent samples does not become unity for the chemotaxis network because the receptor state remains correlated due to the slow methylation dynamics, $\tau_m \gg \tau_r$. Therefore, the number of independent receptor states becomes limited by the number of receptor clusters and their covariance $R_I \approx R_T/\alpha$ [Eq. (23)].

The sampling error can be mitigated in a number of ways. One is to increase the number of receptors per cluster N_r, because this increases the magnitude of the static gain [see Eq. (19) where $\beta \propto N_r$]. Another is to simultaneously increase the total number $\overline{N}$ of samples and the number $\overline{N}_I$ of independent samples, which requires increasing both the number X_T of readout molecules and the number R_T of receptor clusters [Eqs. (6) and (23)]. Indeed, as observed for the push-pull network in earlier work [13,14], these resources limit sensing and hence prediction like weak links in a chain [also see Fig. 5] and Appendix C. However, increasing the cluster size, the number of clusters, or the number of readout molecules all require a larger number of proteins to be used by the network, which are resources that come at a physical cost.

E. Optimal resource allocation

To investigate how resources should be optimally allocated to minimize the sampling error [Eq. (27)] we define a simple cost function, as in Ref. [3]:

$$C = X_T + N_r R_T, \quad (28)$$

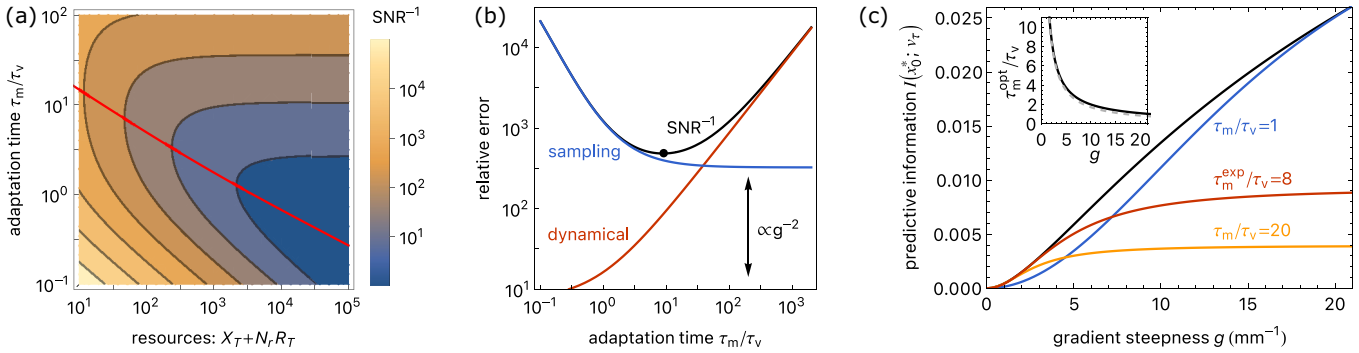

FIG. 3. The relative error is set by the resource availability, adaptation time, and gradient steepness. (a) The relative error [SNR^{-1}, Eq. (27)] as a function of the resource availability $C = X_T + N_r R_T$ and the adaptation time τ_m/τ_v. The relative error decreases monotonically with higher resource availability. The error is minimized for the optimal adaptation times indicated by the red line, which decreases with the resource availability. The ratio of readouts to receptors X_T/R_T obeys Eq. (29). (b) The relative dynamical error, sampling error, and their sum, the total relative error [SNR^{-1}, Eq. (27)], as a function of the adaptation time τ_m/τ_v. The optimal adaptation time arises from a trade-off between the sampling error, which decreases with the adaptation time, and the dynamical error, which increases with the adaptation time. The minimal total error (black dot) occurs close to the point where the sampling error saturates as a function of τ_m/τ_v. The minimal sampling error is proportional to $1/g^2$. (c) The predictive information $I(x_0^*; v_\tau) = I(\hat{p}_{\tau_r}; v_\tau) = 0.5 \log(1 + \mathrm{SNR})$, with the SNR of Eq. (27), between the current number of phosphorylated readouts $x_0^* = \overline{N}\hat{p}_{\tau_r}$ [Eq. (5)] and the future input derivative v_τ, for various adaptation times τ_m. Along the black curve, the adaptation time has been optimized; τ_m^{opt}/τ_v as a function of the gradient steepness is shown in the inset. Experiments show that for *E. coli* the adaptation time is $\tau_m^{expt}/\tau_v \approx 8$ [1,5,24], which is close to optimal for $g \lesssim 4$ mm^{-1} (red curve). Reducing the adaptation time reduces the accuracy in shallow gradients and increases it in steeper gradients (blue curve), while increasing the adaptation time reduces the accuracy in steeper gradients but does not markedly increase the accuracy in shallow gradients (yellow curve). This suggests that the system has been been optimized for sensing shallow gradients. Inset shows the optimal adaptation time τ_m^{opt}/τ_v scales inversely with the gradient steepness g, numerical result (solid black line), and analytical approximation [dashed gray line, Eq. (31)]. In panels (a) and (b) $g = 2$ mm^{-1}, in panels (b) and (c) $X_T = 10^4$ and $R_T = 8$ [1,3,35]. Other parameters are $N_r = 12$, $p = 0.3$, $f = f^{opt} = 0.5$, $\tau_r = 0.1$ s, $\tau = \tau_v = 1.16$ s, $c_0 = 100$ µm, $\sigma_{v_x} = 157$ µm^2 s^{-2} [1,3,24,35]; $\tilde{\alpha} = 2k_BT$ [24]; $K_D^I = 18$ µm and $K_D^A = 2900$ µm [19–21]. Code to reproduce this figure is available [36].

where X_T is the number of readout molecules, R_T is the number of independent receptor clusters, and N_r is the number of receptors per cluster. This cost function captures the idea that a cell must choose whether it spends its resources on making more readout molecules on the one hand, or more receptors on the other. In general, the prediction error [Eq. (27)] depends not only on the number of receptors and readout proteins, but also on the energy to drive the network [3,13,14]. These two constraints can be treated on the same footing by recognizing that the energy to synthesize the required proteins scales linearly with their copy number, with a proportionality constant set by the protein length. Here, we focus on the bound on the predictive information as set by the number of proteins, also because the effects of the energetic cost of driving the phosphorylation and methylation cycles on the optimal prediction strategy are relatively minor [3]. While the precise functional form of the constraint is somewhat arbitrary, the linear form of Eq. (28) is natural because the prediction error depends on the number of receptor and readout molecules, not on higher powers thereof.

Given a total resource availability C and a fixed number of receptors per cluster, the cell can tune the ratio of receptors to readouts. To determine what the optimal ratio is that minimizes the sampling error, we express both R_T and X_T in terms of their ratio and the total resource availability C, and we use that we can express the mean number of samples as in Eq. (6). Subsequently taking the derivative of Eq. (27) with respect to X_T/R_T and equating to zero then gives the optimal ratio,

$$\left(\frac{X_T}{R_T}\right)^{opt} = \frac{\sigma_X}{\sigma_R}\frac{p\sqrt{1+\tau_r/\tau_m}}{f(1-f)} \approx \frac{\sigma_X}{\sigma_R}\frac{p}{f(1-f)}, \tag{29}$$

where we have used that the adaptation time must be larger than the response time and thus $\sqrt{1+\tau_r/\tau_m} \approx 1$. We have further defined the noise per receptor $\sigma_R^2 \equiv \alpha p(1-p)/N_r = \tilde{\alpha} p^2(1-p)^2$ [see also Eqs. (B3), (B5), and (B19)], and the noise per readout molecule $\sigma_X^2 \equiv f(1-f)$. In terms of $\overline{N}$, using Eq. (6), we find that Eq. (29) yields an intuitive relation for optimal networks,

$$\overline{N} = \frac{\sigma_X}{\sigma_R} R_T. \tag{30}$$

This relation shows that for equal noise magnitudes per protein, the average number of samples should equal the total number of receptor clusters. This simple relation arises from the fact that the methylation noise cannot be averaged out, and a minimally redundant design is therefore one in which each receptor cluster is sampled once.

Given the optimal ratio of readouts to receptors in Eq. (29) we can compute the relative error [Eq. (27)] as a function of the total resource availability C and the adaptation time τ_m [Fig. 3(a)]. As expected, we find that the error decreases monotonically with the resource availability. More interesting is that we find a clear optimum for the adaptation time τ_m.

F. Optimal adaptation time

The optimal adaptation time τ_m, given by the red line in Fig. 3(a), arises from the antagonistic effect of the adaptation time on the sampling error and the dynamical error [Fig. 3(b)]. The sampling error decreases monotonically with the adaptation time because a longer adaptation time increases the change in the receptor activity upon the same change in the current or future signal derivative, i.e., it increases the (dynamic) gain [Eqs. (18), (19), and (27)]. However, increasing the adaptation time means that the derivative is taken over a longer time further back into the past, and this derivative will be less informative about the future derivative that the cell aims to predict: the dynamical error increases monotonically with τ_m [Eq. (27)]. The minimal total error occurs for the smallest adaptation time that is sufficiently large to lift the signal above the noise, i.e., reduce the sampling error, while minimizing the dynamical error [Fig. 3(b)].

The value of the adaptation time for which the total error is minimized depends on the resource availability C and the gradient steepness g: these parameters set the magnitude of the sampling error [Eqs. (15) and (27)]. To obtain analytical insight into the optimal adaptation time $\tau_m^{\rm opt}$, we exploit that the response time τ_r must be smaller than the adaptation time τ_m to mount a nonzero response to transient input changes. We further consider that the relevant regime for *E. coli* is likely that where gradients are shallow relative to the length of a run (see also Sec. II G). This means that the sampling error dominates over the dynamical error, although the latter is not negligible [see Eq. (27) with $\sigma_v^2 = g^2\sigma_{v_x}^2$, Eq. (15)]. To minimize the prediction error [Eq. (27)] in this regime, the adaptation time must be large relative to the signal correlation time $\tau_m \gg \tau_v$, which is set by the duration of a run. We obtain for the optimal adaptation time (see Appendix C)

$$\tau_m^{\rm opt} \approx \frac{\sqrt{2}}{\beta g \sigma_{v_x}} \sqrt{\frac{p^2}{\overline{N}} + \frac{p(1-p)}{\overline{N}_I}} \quad \text{for } \tau_m \gg \tau_v, \tau_r, \tag{31}$$

where the number of independent samples $\overline{N}_I$ is given by Eq. (23) with $R_I = R_T/\alpha$. The inset of Fig. 3(c) shows that Eq. (31) is a good approximation of the optimal adaptation time over a large range of the gradient steepness g.

Equation (31) shows how the optimal adaptation time decreases when we increase the total number of receptor samples $\overline{N} \propto X_T$ [Eq. (6)] or the number of independent receptor samples $\overline{N}_I$, which depends on both X_T and R_T [Eq. (23)]. This is because increasing these resources reduces the sampling error: decreasing the adaptation time then decreases the dynamical error more than it increases the relatively small sampling error. In Appendix C, we discuss in more detail how the optimal adaptation time varies with X_T and R_T separately. Equation (31) also shows that the optimal adaptation time decreases as the gradient steepness increases. The reason is that a steeper gradient generates a stronger signal with a larger variance [Eq. (15)], which reduces the sampling error [Eq. (27)].

G. Comparison with experiment

To check whether the uncovered design principles [Eqs. (29) and (31)] are relevant to real world biochemical networks, we evaluate the design of the *E. coli* chemotaxis network in this light.

To assess the design principle of Eq. (29), we use the definitions of σ_X and σ_R given below it. For p and f of order $1/2$ and $\tilde{\alpha} = 2$, based on experiment [24], Eq. (29) predicts an optimal number of readout molecules per receptor cluster of $X_T/R_T \approx 3$. This is in good agreement with earlier predictions [13] and the experimental data of Li and Hazelbauer [37], assuming a cluster consists of two trimers of receptor dimers and two CheA dimers [38]. With $X_T \approx 10^3$–10^4 readout molecules depending on the growth rate [37], this result, i.e., $X_T/R_T \approx 3$, suggests that the number of receptor clusters is in the range $R_T \approx 10^2$–10^3. On the other hand, fitting more recent experimental data with an MWC-based chemotaxis model as we use here suggests a much smaller number of receptor clusters of $R_T \approx 8$ [1,3,35]. However, this estimate for the number of receptor clusters was based on fitting the noise amplitude of the model [3,35] to the experimental data of Ref. [1]. Recent experiments indicate that the receptor array is poised to a critical point [39], where receptor switching becomes correlated over long distances, and it is conceivable that this small value of $R_T \approx 8$ corresponds to the small number of domains over which the receptors effectively switch in concert. More work is needed to understand whether receptor switching near a critical point can effectively be described by an MWC model and whether the design rule unveiled here [Eq. (29)], also generalizes to a receptor array near a critical point. Lastly, further study is necessary to understand whether information transmission in this system is maximized near a critical point [40].

The adaptation time of the *E. coli* chemotaxis system has repeatedly been shown to be ≈10 s, yielding $\tau_m^{\rm expt}/\tau_v \approx 8$ [1,5,24]. Given the estimated resource allocation in the effective MWC description, $X_T = 10^4$ and $R_T = 8$ with $N_r = 12$ [1,3,35], this adaptation time is close to optimal for a gradient steepness $g \lesssim 4\ \text{mm}^{-1}$ [Fig. 3(c)]. In particular, while decreasing the methylation time improves the prediction accuracy in steeper gradients, it reduces information transmission in shallower gradients. On the other hand, while increasing the methylation time beyond the measured one decreases the accuracy in steeper gradients, the improvement in shallow gradients is only very minor because the system is already very close to the fundamental bound on the predictive information as set by the resource constraint and the gradient steepness [Fig. 3(c)]. These arguments show that the methylation time of *E. coli* is indeed optimal for sensing shallow gradients with $g \lesssim 4\ \text{mm}^{-1}$. It suggests that the chemotaxis system has been optimized for navigating weak gradients. To get an idea of what this gradient steepness means we can compare it to the length of an *E. coli* cell, which is ≈1 μm. To cover a gradient length scale $g^{-1} = 1/4$ mm thus requires the cell to move at least 250 times its body length, corresponding to approximately ten runs in the same direction [41–43]. This illustrates how extremely shallow the gradients that *E. coli* can encounter likely are. Moreover, it suggests that it is most important to maximize accuracy in shallow gradients, where it is hard to distinguish signal from noise. In steeper gradients, *E. coli* would be further from the optimal design, but the total information it obtains about the signal of interest is still larger because the input fluctuations are bigger. Clearly, a network

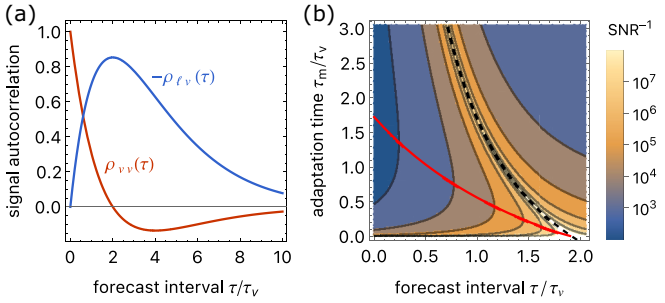

FIG. 4. Two distinct regimes appear for the chemotaxis network predicting a future derivative when the signal concentration variations are of the same order as the change in concentration during a run: $\sigma_\ell = 2\tau_v\sigma_v$ as opposed to $\sigma_\ell \gg \tau_v\sigma_v$ (Fig. 3). (a) The correlation of the current signal derivative $v(t_0)$ [red curve, Eq. (D1)] and of the current signal concentration $\ell(t_0)$ [blue curve, Eq. (D2)] with the future derivative v_τ. (b) The SNR^{-1} is shown [Eq. (D12)] as a function of the adaptation time τ_m/τ_v and the forecast interval τ/τ_v for the chemotaxis network sensing a signal with statistics shown in panel (a). The optimal adaptation time in the regime of short forecast intervals and adaptation times (red curve) was computed by numerically minimizing Eq. (D12). The error diverges where the dynamical gain is zero [Eq. (D7)], because the correlation between the network output and the future derivative vanishes (dashed black curve). Parameters other than σ_ℓ are as in Fig. 3 and the code to reproduce the figure is available [36].

that would adapt the adaptation time to the steepness of the gradient, would probably perform better over a broader range of gradient steepness. However, such a network would be (much) more complicated with a higher resource cost. It thus appears that evolution has optimized the network for sensing those gradients that are most difficult to detect.

H. Changing the signal statistics

Up to this point we have assumed that the cell navigates environments in which the nutrient concentration varies much more widely than the change in concentration over the duration of a run (Sec. II B). Indeed, we expect that this is the most natural regime for many microorganisms, and especially for *E. coli*, because it is clear that *E. coli* can indeed measure concentration changes over a wide range of background concentrations based on the dissociation constants of the receptor in its active and inactive state, which differ by two orders of magnitude [19–21]. However, to illustrate the effect that the choice of signal statistics has on our results, we here consider a scenario in which the concentration and its change during a run are of the same order. We may view this as a scenario in which the cell remains close to a constant nutrient peak.

Specifically, we consider a signal with the dynamics given in Eqs. (12) and (13), but with a concentration standard deviation $\sigma_\ell = 2\tau_v\sigma_v$, yielding $\omega_0 = (2\tau_v)^{-1}$. For these parameters, both the current concentration and the current derivative can be informative of the future derivative v_τ, depending on the forecast interval [Fig. 4(a), Eqs. (D1) and (D2)].

Previous work has shown that in this scenario, where both the current concentration value and the current concentration derivative are informative about the future derivative, the optimal network is one that bases its prediction on both the concentration and the derivative, proportional to the magnitude of their correlations with the signal of interest [3]. Our model of the chemotaxis network does not allow this because it is set up to exhibit perfect adaptation, and thus only measures the concentration change. However, in the limit of infinitely slow adaptation, $\tau_m \to \infty$, the network effectively becomes a push-pull network and measures the concentration only. Therefore, it is still interesting to investigate the optimal adaptation time that minimizes the relative error (SNR^{-1}) under these markedly different signal statistics.

To compute the SNR^{-1}, we derive the dynamic gain and the dynamical error for this signal, with $\sigma_\ell = 2\tau_v\sigma_v$ [Eqs. (D6) and (D11)]. Since we only change the signal statistics, the sampling error in estimating the receptor activity [$\sigma^{2,\mathrm{samp}}_{\hat{p}_{\tau_r}}$, Eq. (22)] and optimal resource allocation [Eq. (29)] remain unchanged. We can then compute the relative prediction error (SNR^{-1}) of the chemotaxis network under the new signal statistics in the same manner as before [via Eqs. (2) and (11)].

Figure 4(b) reveals that the prediction error now exhibits two distinct regions as a function of the adaptation time τ_m and the forecast interval τ, separated by a boundary on which the error diverges (dashed black line). These two regions reflect two distinct prediction strategies. For short forecast intervals the future derivative v_τ is more correlated with the current derivative $v(t_0)$ than with the current value $\ell(t_0)$, i.e., $\rho_{vv}(\tau) > |\rho_{\ell v}(\tau)|$ [Fig. 4(a)]. In this regime, a short adaptation time allows the network to predict the future derivative v_τ based on the current, i.e., instantaneous, derivative $v(t_0)$. As the forecast interval increases, the contribution from the dynamical error rises more strongly than that from the sampling error [Eq. (D12)], which tends to decreases the optimal adaptation time. Simultaneously however, the future derivative also starts to become more strongly correlated to the current concentration value $\ell(t_0)$ as the forecast interval increases [Fig. 4(a), blue curve]. As a result, in the regime of large forecast intervals beyond the dashed black curve in Fig. 4(b), the optimal adaptation time diverges: the network effectively becomes a push-pull network and bases its prediction on the concentration value rather than its derivative.

Figure 4(b) shows that the forecast interval at which the error diverges becomes shorter for longer adaptation times (black dashed line). The prediction error diverges when the dynamic gain [Eq. (D6)], which quantifies the covariance between the current network output and the future signal derivative, becomes zero [Eq. (D7)]. The point at which the dynamic gain becomes zero depends on the forecast interval τ via the signal correlation function $\rho_{vv}(\tau)$ [red curve Fig. 4(a), Eq. (D1)], which quantifies how much the current concentration derivative $v(t_0)$ is correlated with the future concentration derivative v_τ at the forecast interval τ. It also depends on the adaptation time τ_m, because that determines the degree to which the signal derivative taken by the network reflects the current derivative. When $\tau_m \to 0$, the system takes an instantaneous derivative; the prediction error then diverges when the forecast interval $\tau = 2\tau_v$, precisely because the current derivative is then not correlated with the future derivative at that later time τ [$\rho_{vv}(2\tau_v) = 0$, see Fig. 4(a), red curve]. For a longer adaptation time, the network takes a derivative over the signal further back into the past, i.e., a derivative centered

around a time $t \approx t_0 - \tau_m$, which, as can be inferred from $\rho_{vv}(\tau)$ in Fig. 4(a) exploiting stationarity, is uncorrelated with the future derivative at a time $\tau \approx t_0 - \tau_m + 2\tau_v$.

III. DISCUSSION

Microorganisms that navigate chemical gradients need to predict the concentration change that they will encounter. For simple input signals where the change in concentration is Markovian, the optimal way to achieve this is to measure the current time derivative of the concentration [3]. Measuring such temporal concentration changes requires perfect adaptation. Moreover, to measure the most recent concentration change, the adaptation time must be short relative to the correlation time of the input. However, building and maintaining a biochemical network costs physical resources. When the resource availability is limited, the signal is obscured by noise in the network. The only way to lift the signal above the noise in this regime, is to increase the adaptation time. This trade-off between lifting the signal above the noise, and measuring a concentration change which is informative of the future input, sets the optimal adaptation time.

The optimal adaptation time depends on the amount of resources available to maintain the network, and the magnitude of changes in the input. The latter is set by the swimming behavior of the cell and the steepness of the chemical gradient it navigates. In steeper gradients the input changes more strongly, which reduces the sampling error and increases the signal-to-noise ratio. A smaller sampling error allows for a shorter adaptation time, which mitigates the dynamical error and maximizes the overall accuracy. Therefore, the optimal adaptation time to predict the concentration change scales approximately inversely with the gradient steepness. Interestingly, simulations show that the optimal adaptation time that maximizes navigational performance also increases as the gradient becomes more shallow [44,45]. This indicates that predicting the concentration change is indeed important for successful navigation, in line with results of agent-based simulations on the interplay between prediction and navigation [46].

Our results provide a possible explanation for a puzzling observation. During chemotaxis, *E. coli* performs subsequent runs of approximately one second in different directions. Runs in the correct direction relative to the gradient are extended, and vice versa, such that the cell moves up a gradient of attractant on average. To implement this strategy, *E. coli* must predict how the concentration will change while it navigates the gradient. To this end, it seems natural to measure the change in concentration over the course of one run, i.e., over approximately one second. However, the adaptation time of *E. coli* is around ten seconds [1,5,24]. This raises the question, why would *E. coli* measure concentration changes over a timescale that is much longer than that of a run? Our work shows that the adaptation time must be this long to discern the signal from the inevitable biochemical signaling noise in shallow gradients.

Our analysis highlights that the optimal design of any sensing system depends on the signal statistics. In our previous work [3] we argued that (a) *E. coli* aims to predict the future derivative, (b) it can measure concentration changes over a range of background concentrations that is much larger than the typical concentration change during a run, (c) in this regime, the optimal signaling system for predicting the derivative is a perfectly adaptive system, as *E. coli* has, and (d) its long adaptation time of $\tau_{\rm m} \approx 10$ s is optimal for sensing shallow gradients. In fact, in this regime the predictive power of the *E. coli* chemotaxis system becomes exceedingly close to that of the optimal system. These observations together strongly suggest that *E. coli* has been optimized to sense shallow gradients, which is perhaps not so surprising since these gradients are the hardest to detect. Our previous analysis [3] also showed that, since in this regime of shallow gradients the dynamics of the concentration derivative is Markovian, the optimal design does not depend on the forecast interval. Our current analysis of a different input signal (Fig. 4) shows, however, that in general, the optimal design depends on the signal statistics and the forecast interval. What the relevant signal statistics and forecast interval are depends on the dynamics of the environment and on how the organism navigates the environment [47,48]. We leave these questions for future work.

More generally, our results provide insight into the optimal design of adaptive signaling networks. First and foremost, this improves our understanding of navigation behavior of microorganisms. But the uncovered principles might well hold more generally and shed light on other adaptive signaling networks as well, e.g., that of rod cells in the vertebrate eye [49]. Moreover, our theory facilitates the optimal design of microrobots that need to navigate environments without a map.

ACKNOWLEDGMENTS

We thank Vahe Galstyan for a careful reading of the paper. 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 (ERC) under the European Union's Horizon 2020 research and innovation program (Grant Agreement No. 885065).

APPENDIX A: THE PREDICTION ERROR

Here we derive the general expression for the prediction error $\sigma^2_{\hat{p}_{\tau_r}}$, which shows how the complete error decomposes into independent parts caused by fluctuations in the number N of samples, the error of a sampling process with a fixed number of samples and a constant input, and uninformative fluctuations from the input signal. Our starting point is the decomposition of the error in Eq. (7).

The first term of Eq. (7) is straightforward to compute, using the definition of $\hat{p}_{\tau_r}$ from Eq. (5) we obtain

$$\mathrm{Var}(\mathbb{E}[\hat{p}_{\tau_r}|N]) = \mathrm{Var}\left(\mathbb{E}\left[\frac{1}{\overline{N}}\sum_{i=1}^{N} n_i(t_i)|N\right]_{t_i,n_i}\right)_N \tag{A1}$$

$$= \frac{1}{\overline{N}^2}\mathrm{Var}(N\mathbb{E}[\langle n(t_i)\rangle]_{t_i})_N \tag{A2}$$

$$= \frac{p^2}{\overline{N}}, \tag{A3}$$

where the subscripts after the expected values and variances denote the random variables over which the expectation is taken. For instance, in Eq. (A1) the expected value is taken under a fixed number of samples N over the state $n_i \in \{0, 1\}$ of each cluster, later also denoted with angle brackets as an ensemble average, and over all sampling times t_i, which are exponentially distributed with the probability density function [13]

$$f(t_i) = \frac{1}{\tau_r} e^{-(t_0 - t_i)/\tau_r}. \tag{A4}$$

From Eq. (A2) to Eq. (A3) we use that the average number of active receptor clusters is defined as $p \equiv \mathbb{E}[\langle n(t_i)\rangle]_{t_i}$, which is constant with respect to N. The variance is subsequently taken over the Poisson distributed number N of samples, with both mean and variance $\overline{N}$. The resulting expression [Eq. (A3)] is the error that arises because the network cannot distinguish between those readout molecules that sampled an inactive cluster, and those that did not sample a cluster at all [13,15].

We decompose the second term of Eq. (7) in two steps. First, we use the definition of $\hat{p}_{\tau_r}$ [Eq. (5)] and split the self- and cross-terms in the covariance of the kinase activity:

$$\mathbb{E}[\mathrm{Var}(\hat{p}_{\tau_r}|N)] = \mathbb{E}\left[\mathrm{Var}\left(\frac{1}{\overline{N}}\sum_{i=1}^{N} n_i(t_i)|N\right)_{n_i,t_i}\right]_N \tag{A5}$$

$$= \frac{1}{\overline{N}^2}\mathbb{E}[N\mathrm{Var}(n_i(t_i)) + N(N-1)\mathrm{Cov}(n_i(t_i), n_j(t_j))]_N \tag{A6}$$

$$= \frac{p(1-p)}{\overline{N}} + \mathrm{Cov}(n_i(t_i), n_j(t_j)). \tag{A7}$$

From Eq. (A6) to Eq. (A7) we used that both the variance of each cluster and the covariance between clusters are independent of the number N of samples, and that for a Poisson distributed number of samples, N, we have $\mathbb{E}[N(N-1)] = \overline{N}^2$. To continue, the covariance between different kinases at different times can be decomposed into contributions from the receptor noise, and fluctuations in the full history of the input signal, the trajectory $\boldsymbol{s}$,

$$\mathrm{Cov}(n_i(t_i)n_j(t_j)) = \mathbb{E}[\mathrm{Cov}(n_i(t_i), n_j(t_j)|\boldsymbol{s})]_{t_i,t_j,\boldsymbol{s}} + \mathrm{Cov}(\mathbb{E}[\langle n_i(t_i)|\boldsymbol{s}\rangle]_{t_i}, \mathbb{E}[\langle n_j(t_j)|\boldsymbol{s}\rangle]_{t_j})_{\boldsymbol{s}} \tag{A8}$$

$$= \mathbb{E}[\mathrm{Cov}(n_i(t_i), n_j(t_j)|\boldsymbol{s})]_{t_i,t_j,\boldsymbol{s}} + \mathrm{Var}(\mathbb{E}[\langle n(t_i)|\boldsymbol{s}\rangle]_{t_i})_{\boldsymbol{s}}, \tag{A9}$$

where we use that $\mathbb{E}[\langle n_i(t_i)|\boldsymbol{s}\rangle]_{t_i} = \mathbb{E}[\langle n_j(t_j)|\boldsymbol{s}\rangle]_{t_j}$. The two terms on the RHS of Eq. (A9) respectively describe the covariance between clusters when the input is fixed, and the variance that is caused by input fluctuations. The first term is the receptor-level noise, which for the chemotaxis model considered in this work arises only from methylation Eqs. (B14)–(B19). The second term is the variance of the mean activity conditional on the input, which is the signal-induced variance. This signal induced variance comprises all variance caused by the input, so both the dynamical error and the variance that is informative of the signal of interest $\tilde{g}^2\sigma^2_{s_\tau}$ [Eqs. (B20)–(B25)].

Combining Eqs. (A7) and (A9) gives

$$\mathbb{E}[\mathrm{Var}(\hat{p}_{\tau_r}|N)] = \frac{p(1-p)}{\overline{N}} + \mathbb{E}[\mathrm{Cov}(n_i(t_i), n_j(t_j)|\boldsymbol{s})]_{t_i,t_j,\boldsymbol{s}} + \mathrm{Var}(\mathbb{E}[\langle n(t_i)|\boldsymbol{s}\rangle]_{t_i})_{\boldsymbol{s}}. \tag{A10}$$

Finally, the third term of Eq. (7) is the contribution of the signal of interest to the output variance:

$$\mathrm{Var}(\mathbb{E}[\hat{p}_{\tau_r}|s_\tau]) = \mathrm{Var}\left(\mathbb{E}\left[\frac{1}{\overline{N}}\sum_{i=1}^{N} n_i(t_i)|s_\tau\right]_{t_i,n_i,N}\right)_{s_\tau} \tag{A11}$$

$$= \mathrm{Var}(\mathbb{E}[\langle n(t_i)|s_\tau\rangle]_{t_i})_{s_\tau} \tag{A12}$$

$$= \tilde{g}^2\sigma^2_{s_\tau}, \tag{A13}$$

where in the last step we have used the dynamic input output relation of Eq. (1). The dynamic gain $\tilde{g}$ of the chemotaxis network is derived in Eqs. (B10)–(B13). Substituting the equalities of Eqs. (A3), (A10), and (A13) in Eq. (7) of the main text gives the complete prediction error given in Eq. (11) in the main text.

We note that this derivation deviates from that of Malaguti and Ten Wolde [15] in that Eq. (A5) includes the contributions from all signal variations, including the informative signal variations [which are then subtracted from the full variance in Eq. (7)], while in Ref. [15] the corresponding term does not contain these informative signal fluctuations. While the final result is identical, the derivation presented here is arguably easier.

APPENDIX B: THE CHEMOTAXIS NETWORK

In the *E. coli* chemotaxis network, receptors cooperatively control the activity of the kinase CheA, and the activity is adaptive due to the methylation of inactive receptors [5,22–25,34]. We here follow the widely used approach to describe the effects of receptor cooperativity and methylation on kinase activity via the Monod-Wyman-Changeux (MWC) model [1,20,21,24,26–30]. In this model, each receptor can switch between an active and inactive conformational state n and receptors are partitioned into clusters of equal size N_r. In the spirit of the MWC model, receptors within a cluster switch conformation in concert, so that each cluster is either active or inactive [26]. Furthermore, it is assumed that receptor-ligand binding and conformational switching are faster than the other timescales in the system, such that the activity state of the receptor can effectively be described by its equilibrium probability to be active, given the methylation level of the cluster m and the external ligand concentration ℓ. The probability for the receptor cluster to be active is then described by

$$a(\ell, m) \equiv \langle n|\ell, m\rangle = \{1 + \exp[\Delta F_T(\ell, m)]\}^{-1}, \tag{B1}$$

where $\Delta F_T(\ell, m) = -\Delta E_0 + N_r[\Delta F_\ell(\ell) + \Delta F_m(m)]$ is the free-energy difference between the active and inactive state, which is a function of free-energy difference arising from ligand binding and methylation:

$$\Delta F_\ell(\ell) = \ln\left(1 + \ell(t)/K_D^I\right) - \ln\left(1 + \ell(t)/K_D^A\right), \tag{B2}$$

$$\Delta F_m(m) = \tilde{\alpha}[\bar{m} - m(t)]. \tag{B3}$$

Between the two states the cluster has an altered dissociation constant, which is denoted K_D^I for the inactive state, and K_D^A for the active state. The free-energy difference due to methylation has been experimentally shown to depend approximately linearly on the methylation level [24]. We assume that inactive receptors are irreversibly methylated, and active receptors irreversibly demethylated, with zero-order ultrasensitive kinetics [21,50,51]. The methylation dynamics of a receptor cluster is then given by

$$\dot{m} = [1 - a(\ell, m)]k_R - a(\ell, m)k_B + B_m(a)\xi(t), \tag{B4}$$

with $B_m(a) = \sqrt{[1 - a(\ell, m)]k_R + a(\ell, m)k_B}$, and unit white noise $\xi(t)$. These dynamics indeed give rise to perfect adaptation, since from this equation we find that the steady-state cluster activity is given by $p \equiv \bar{a} = 1/(1 + k_B/k_R)$, thus indeed independent of the ligand concentration.

In this work we consider linear dynamics, we therefore employ a linear noise approximation [52]. The deviation of the equilibrium cluster activity from its mean $\delta a(t) = a(t) - p$ is then given by

$$\delta a(t) \equiv \langle n(t)|\delta\ell, \delta m\rangle - p = \alpha\delta m(t) - \beta\delta\ell(t), \tag{B5}$$

with $\alpha = \tilde{\alpha}N_r p(1 - p)$ and $\beta = \kappa N_r p(1 - p)$, with $\kappa = (1 + K_D^I/c_0)^{-1} - (1 + K_D^A/c_0)^{-1}$. For the methylation dynamics on one cluster we then obtain

$$\dot{\delta m} = -\delta a(t)/(\alpha\tau_m) + \eta_m(t), \tag{B6}$$

where we have introduced the adaptation time $\tau_m = [\alpha(k_R + k_B)]^{-1}$ and $\eta_m(t)$ is Gaussian white noise on a single cluster with correlation function

$$\langle \eta_{m_i}(t)\eta_{m_j}(t')\rangle = \delta_{ij}\delta(t - t')\frac{2p(1 - p)}{\alpha\tau_m} \tag{B7}$$

between the ith and jth receptor cluster, where δ_{ij} is the Kronecker delta. Combining Eqs. (B5) and (B6) yields the change in activity over time

$$\dot{\delta a} = -\delta a(t)/\tau_m - \beta v(t) + \alpha\eta_m(t), \tag{B8}$$

where we have the change in concentration over time $v(t) \equiv \dot{\delta\ell}$. Using Eq. (B8) we can also express the instantaneous activity as

$$\delta a(t) = \int_{-\infty}^{t} dt'[\alpha\eta_m(t') - \beta v(t')]e^{-(t-t')/\tau_m}. \tag{B9}$$

This expression shows that the cluster activity, when we average out the methylation noise, reflects the change in concentration weighted exponentially over the past adaptation time τ_m.

1. Dynamic gain

The dynamic gain of the network can be obtained by deriving the average response of the network to the signal of interest s_τ. In general we have the expression given in Eq. (1) for the dynamic input output relation of linear signaling networks. In our case the signal of interest is the future concentration derivative $s_\tau = v_\tau$. Using Eqs. (B5) and (B9), we find for the average conditional activity

$$\langle n(t_i)|v_\tau\rangle = \mathbb{E}[\langle n(t_i)|v_\tau, \delta\ell, \delta m\rangle]_{\delta\ell,\delta m} \tag{B10}$$

$$= p - \beta\int_{-\infty}^{t_i} dt\langle v(t)|v_\tau\rangle e^{-(t_i-t)/\tau_m} \tag{B11}$$

$$= p - e^{-(t_0+\tau-t_i)/\tau_v}\frac{\tau_m\beta v_\tau}{1 + \tau_m/\tau_v}, \tag{B12}$$

where we used that the conditional mean derivative is $\langle v(t)|v_\tau\rangle = v_\tau \exp[-(t_0 + \tau - t)/\tau_v]$, also see Eq. (14). Averaging over all sampling times, distributed as in Eq. (A4), gives

$$\mathbb{E}[\langle n(t_i)|v_\tau\rangle]_{t_i} = p - \frac{\tau_m\beta e^{-\tau/\tau_v}v_\tau}{(1 + \tau_m/\tau_v)(1 + \tau_r/\tau_v)}. \tag{B13}$$

Comparison to Eq. (1) yields the dynamic gain $\tilde{g}$ given in Eq. (18).

2. Receptor noise

The variance that is caused by receptor-level (here methylation) noise is the covariance between clusters under a fixed input trajectory, i.e., the first term of Eq. (A9). We can write this covariance in terms of the equilibrium activity as follows, using Eq. (B5) and noting that $\delta\ell(t)$ is contained in $\boldsymbol{s}$ for $t \leqslant t_0$:

$$\mathbb{E}[\mathrm{Cov}(n_i(t_i), n_j(t_j)|\boldsymbol{s})]_{t_i,t_j,\boldsymbol{s}}$$

$$= \mathbb{E}[\langle n_i(t_i)n_j(t_j)|\boldsymbol{s}, \delta m\rangle]_{t_i,t_j,\boldsymbol{s},\delta m} - p^2 \tag{B14}$$

$$= \mathbb{E}[\langle n_i(t_i)|\boldsymbol{s}, \delta m\rangle\langle n_j(t_j)|\boldsymbol{s}, \delta m\rangle - p^2]_{t_i,t_j,\boldsymbol{s},\delta m} \tag{B15}$$

$$= \mathbb{E}[\langle \delta a_i(t_i)\delta a_j(t_j)|\boldsymbol{s}\rangle]_{t_i,t_j,\boldsymbol{s},\delta m}. \tag{B16}$$

In Eq. (B14) we condition on and average over δm to make the connection between the instantaneous cluster state n_i and the cluster activity a_i [Eq. (B5)]. Then in Eq. (B15) we use the fact that when conditioned on both the signal and the methylation level, the cluster states are independent. The covariance in the cluster activity conditioned on the full past input trajectory [Eq. (B16)] depends only on the methylation noise, using Eqs. (B9) and (B7) and keeping the sampling times fixed,

$$\mathbb{E}[\langle \delta a_i(t_i)\delta a_j(t_j)|\boldsymbol{s}\rangle]_{\boldsymbol{s},\delta m} = \mathbb{E}\Bigg[\alpha^2\int_{-\infty}^{t_i} dt\int_{-\infty}^{t_j} dt'\langle \eta_{m_i}(t)\eta_{m_j} \times (t')\rangle e^{-(t_i-t)/\tau_m}e^{-(t_j-t')/\tau_m}\Bigg]_{\boldsymbol{s},\delta m} \tag{B17}$$

$$= \langle\delta_{ij}\rangle\frac{2\alpha p(1 - p)}{\tau_m}\int_{-\infty}^{t^-} dt e^{-(t^- - t)/\tau_m}e^{-(t^+ - t)/\tau_m} \tag{B18}$$

$$= \frac{\alpha p(1 - p)}{R_T}e^{-|t_i - t_j|/\tau_m}, \tag{B19}$$

where $t^+ \equiv \max(t_i, t_j)$ and $t^- \equiv \min(t_i, t_j)$, and the number of receptor clusters R_T arises as the average Kronecker delta over all clusters: $\langle\delta_{ij}\rangle = 1/R_T$. Averaging over the exponentially distributed sampling times t_i and t_j [both following Eq. (A4)], yields the receptor noise given in Eq. (21).

3. Signal-induced correlations

The covariance in the output caused by the variation in the past input signal is given by the second term of Eq. (A9).

It describes all variance in the output caused by input fluctuations, so it comprises both the dynamical error and the informative part $\tilde{g}^2\sigma^2_{s_\tau}$. We rewrite the instantaneous activity to the equilibrium activity using Eq. (B5) and considering that $\delta\ell(t)$ is contained in $\boldsymbol{s}$ for $t \leqslant t_0$:

$$\mathrm{Var}(\mathbb{E}[\langle n(t_i)|\boldsymbol{s}\rangle]_{t_i})_{\boldsymbol{s}} = \mathrm{Var}(\mathbb{E}[\langle n(t_i)|\boldsymbol{s},\delta m\rangle]_{t_i,\delta_m})_{\boldsymbol{s}} \tag{B20}$$

$$= \mathrm{Var}(p + \mathbb{E}[\langle \delta a(t_i)|\boldsymbol{s}\rangle]_{t_i,\delta_m})_{\boldsymbol{s}} \tag{B21}$$

$$= \mathrm{Var}\left(-\frac{\beta}{\tau_\mathrm{r}}\int_{-\infty}^{t_0} dt_i \int_{-\infty}^{t_i} dt\, v(t) e^{-(t_i-t)/\tau_\mathrm{m}} e^{-(t_0-t_i)/\tau_\mathrm{r}}\right), \tag{B22}$$

where in Eq. (B20) we again condition on and average over δm to make the connection between $n(t)$ and $a(t)$ [Eq. (B5)]. In Eq. (B22) we used Eq. (B9) and the sampling time distribution of Eq. (A4). Using the correlation function of the concentration derivative, Eq. (14), we continue from Eq. (B22) to obtain

$$\begin{aligned}&\mathrm{Var}\big(\mathbb{E}[\langle n(t_i)|\boldsymbol{s}\rangle]_{t_i}\big)_{\boldsymbol{s}}\\ &= \frac{\sigma_v^2\beta^2}{\tau_\mathrm{r}^2}\int_{-\infty}^{t_0} dt_i \int_{-\infty}^{t_0} dt_j\\ &\times\left(\int_{-\infty}^{t_i} dt \int_{-\infty}^{t_j} dt' e^{-|t-t'|/\tau_v} e^{-(t_i-t)/\tau_\mathrm{m}} e^{-(t_j-t')/\tau_\mathrm{m}}\right)\\ &\times e^{-(t_0-t_i)/\tau_\mathrm{r}} e^{-(t_0-t_j)/\tau_\mathrm{r}}.\end{aligned} \tag{B23}$$

First we perform the integrals over t and t', which yields

$$\mathrm{Var}(\mathbb{E}[\langle n(t_i)|\boldsymbol{s}\rangle]_{t_i})_{\boldsymbol{s}} = \frac{\sigma_v^2\beta^2/\tau_\mathrm{r}^2}{1/\tau_v^2 - 1/\tau_\mathrm{m}^2}\int_{-\infty}^{t_0} dt_i \int_{-\infty}^{t_0} dt_j$$
$$\times\left(\frac{\tau_\mathrm{m}}{\tau_v} e^{-|t_i-t_j|/\tau_\mathrm{m}} - e^{-|t_i-t_j|/\tau_v}\right) e^{-(t_0-t_i)/\tau_\mathrm{r}} e^{-(t_0-t_j)/\tau_\mathrm{r}}. \tag{B24}$$

Finally, computing the integrals over the sampling times t_i and t_j gives

$$\begin{aligned}&\mathrm{Var}\big(\mathbb{E}[\langle n(t_i)|\boldsymbol{s}\rangle]_{t_i}\big)_{\boldsymbol{s}}\\ &= \frac{\tau_\mathrm{m}^2\beta^2\sigma_v^2(1+\tau_\mathrm{r}/\tau_\mathrm{m}+\tau_\mathrm{r}/\tau_v)}{(1+\tau_\mathrm{m}/\tau_v)(1+\tau_\mathrm{r}/\tau_v)(1+\tau_\mathrm{r}/\tau_\mathrm{m})},\end{aligned} \tag{B25}$$

which is equivalent to the expression in main text Eq. (25) with the static gain of Eq. (19).

APPENDIX C: OPTIMAL ADAPTATION TIME

Here we give a comprehensive derivation of the approximate optimal adaptation time. To gain analytical insight into the optimal adaptation time we first consider that the adaptation time τ_m must be larger than the response time τ_r to yield a nonzero response to transient input changes. Subsequently taking the derivative of Eq. (27) with respect to τ_m then gives

$$\frac{\partial \mathrm{SNR}^{-1}}{\partial\tau_\mathrm{m}} = \frac{e^{2\tau/\tau_v}}{\tau_v}\left(1+\frac{\tau_\mathrm{r}}{\tau_v}\right)^2\left[1 - \frac{2(1+\tau_v/\tau_\mathrm{m})}{(\tau_\mathrm{m}\beta g\sigma_{v_x})^2}\right.$$
$$\left.\times\left(\frac{p^2}{\overline{N}} + \frac{p(1-p)}{\overline{N}_\mathrm{I}}\right)\right] \text{ for } \tau_\mathrm{m} \gg \tau_\mathrm{r}, \tag{C1}$$

where the number $\overline{N}_\mathrm{I}$ of independent samples is given by Eq. (23) with $R_\mathrm{I} = R_\mathrm{T}/\alpha$. Now considering that for *E. coli* the adaptation time is much larger than the signal correlation time gives, up to the prefactor,

$$\frac{\partial \mathrm{SNR}^{-1}}{\partial\tau_\mathrm{m}} \propto 1 - \frac{2}{(\tau_\mathrm{m}\beta g\sigma_{v_x})^2}\left(\frac{p^2}{\overline{N}} + \frac{p(1-p)}{\overline{N}_\mathrm{I}}\right), \tag{C2}$$

for $\tau_\mathrm{m} \gg \tau_\mathrm{r}, \tau_v$. Equating Eq. (C2) to zero and solving for $\tau_\mathrm{m}^\mathrm{opt}$ yields one positive solution, given in Eq. (31).

Inspection of Eq. (31) also reveals that the limiting resource of the network will set the optimal adaptation time. This becomes apparent when we express $\overline{N}$ [Eq. (6)] and $\overline{N}_\mathrm{I}$ [Eq. (23)] in terms of the number of readout molecules, X_T, and receptor clusters, R_T. Substitution in Eq. (31) then yields for the analytical approximation of the optimal adaptation time

$$\tau_\mathrm{m}^\mathrm{opt} \approx \frac{\sqrt{2}}{\beta g\sigma_{v_x}}\sqrt{\frac{p^2}{X_\mathrm{T} f(1-f)} + \frac{\alpha p(1-p)}{R_\mathrm{T}}}. \tag{C3}$$

When $R_\mathrm{T} \gg X_\mathrm{T}$ the receptor noise [final RHS term in Eq. (C3), also see Eq. (21)] becomes negligibly small, and the adaptation time is set by the number of readout molecules X_T, which set the number of receptor samples $\overline{N}$ [Eq. (6)] and thus the first two terms of the sampling error in Eq. (11). Figure 5(a) illustrates how the optimal adaptation time becomes independent of R_T when X_T is limiting. Vice versa, when $X_\mathrm{T} \gg R_\mathrm{T}$, the optimal adaptation time is set by the receptor noise [Fig. 5(b)]. The panels of Fig. 5 show that the receptors and the readout molecules limit sensing like weak links in a chain: the prediction error, and concomitantly the optimal adaptation time, is set by the limiting resource; the error cannot be lowered by increasing the other resource [13–15].

The observation that the limiting resource sets the optimal adaptation time can be understood when we consider that the optimal adaptation time arises from a trade-off between the sampling error and the dynamical error [Eq. (27)]. To minimize the dynamical error, the adaptation time τ_m must

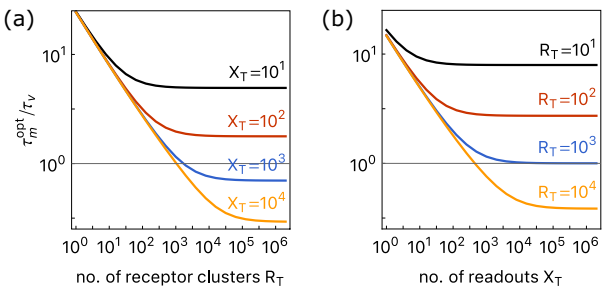

FIG. 5. The limiting resource sets the optimal adaptation time. (a) The optimal adaptation time, determined by numerically minimizing Eq. (27), as a function of the number of receptor clusters R_T, for different numbers of readout molecules X_T. (b) The optimal adaptation time as a function of the number of readout molecules X_T, for different numbers of receptor clusters R_T. The panels show that X_T and R_T limit the prediction error like weak links in a chain, as observed for a simple push-pull network without methylation feedback [13–15]. Parameters are as in Fig. 3, and the code to reproduce the figure is available [36].

be zero, such that the network output reflects the most recent signal derivative. This is analogous to the dynamical error for the push-pull network studied by Malaguti and Ten Wolde [14], which is minimized by reducing the integration time τ_r to zero such that the network output reflects the most recent signal concentration. Importantly, unlike the sampling error, the dynamical error does not depend on R_T and X_T, but only on timescales. To reduce the sampling error of the chemotaxis network, the adaptation time must increase. Again, this seems similar to the role of the integration time in the push-pull network, where an increase in the integration time can also reduce the sampling error [14]. However, in the push-pull network, increasing the integration time can reduce the sampling error because it enables the network to time-average the noise arising from receptor-ligand binding. This also means that in the push-pull network, the optimal integration time reduces to zero when $R_T \gg X_T$, because time averaging means that more receptor samples or concentration measurements are taken *per* receptor, which requires $X_T > R_T$. This is markedly different from the chemotaxis network [Fig. 5(a)], because the role of the adaptation time is fundamentally different from that of the integration time. Indeed, increasing the adaptation time reduces the sampling error because it increases the dynamical gain [Eq. (18)]. This enables the network to lift the signal above the noise.

APPENDIX D: CHANGING SIGNAL STATISTICS

To illustrate the effect of the signal statistics on the prediction error we consider a signal with the dynamics given in Eqs. (12) and (13), but with a concentration standard deviation $\sigma_\ell = 2\tau_v\sigma_v$, yielding $\omega_0 = (2\tau_v)^{-1}$. These statistics describe a critically damped harmonic oscillator. For such a signal, both the current concentration value and its derivative can be correlated to the future derivative v_τ, depending on the forecast interval:

$$\langle \delta v(t_0)\delta v(\tau)\rangle = \sigma_v^2\left(1-\frac{\tau}{2\tau_v}\right)e^{-\tau/(2\tau_v)}, \tag{D1}$$

$$\langle \delta \ell(t_0)\delta v(\tau)\rangle = -\sigma_v^2\tau e^{-\tau/(2\tau_v)}. \tag{D2}$$

These correlation functions are shown in Fig. 4(a).

The dynamic gain quantifies the covariance between the network output and the signal of interest, and therefore naturally depends on the signal statistics. To derive the dynamic gain under the new signal statistics we follow the same procedure as before, starting from Eq. (B10) and using that now $\langle v(t)|v_\tau\rangle = v_\tau[1-\tau/(2\tau_v)]\exp[-\tau/(2\tau_v)]$ [see Eq. (D1)], we obtain

$$\langle n(t_i)|v_\tau\rangle = p - \beta\int_{-\infty}^{t_i} dt\,\langle v(t)|v_\tau\rangle e^{-(t_i-t)/\tau_m} \tag{D3}$$

$$= p - e^{-(t_0+\tau-t_i)/(2\tau_v)}\frac{\tau_m\beta v_\tau}{1+\tau_m/(2\tau_v)} \times\left(1-(t_0+\tau-t_i)/(2\tau_v)-\frac{1}{1+2\tau_v/\tau_m}\right). \tag{D4}$$

Averaging over all sampling times t_i [Eq. (A4)] yields

$$\mathbb{E}[\langle n(t_i)|v_\tau\rangle]_{t_i} = p + \tilde{g}v_\tau, \tag{D5}$$

with the dynamic gain

$$\tilde{g} = \frac{-\tau_m\beta e^{-\tau/(2\tau_v)}}{[1+\tau_m/(2\tau_v)][1+\tau_r/(2\tau_v)]} \times\left(1-\frac{\tau}{2\tau_v}-\frac{1}{1+2\tau_v/\tau_m}-\frac{1}{1+2\tau_v/\tau_r}\right). \tag{D6}$$

From this expression we can see that the dynamic gain goes to zero when

$$\frac{\tau}{\tau_v} = 2\left(1-\frac{1}{1+2\tau_v/\tau_m}-\frac{1}{1+2\tau_v/\tau_r}\right). \tag{D7}$$

When this relation holds, the current network output and the future signal derivative are not correlated. When $\tau_m \to 0$ the network takes an instantaneous derivative; indeed, in this limit the dynamical gain is zero when $\tau = 2\tau_v$ (assuming $\tau_r \ll \tau_v$), because the current derivative is not correlated with the future derivative v_τ for this forecast interval τ, under these signal statistics [Eq. (D1)]. The dependence on the adaptation time arises because the adaptation time determines to what extent the current network output reflects the most recent derivative, or an exponentially weighted average of derivatives further in the past [Eq. (B9)]. Similarly, τ_r sets the timescale over which the receptor activity is averaged on the level of the readout, potentially introducing an additional delay. However, for the chemotaxis network typically $\tau_r \ll \tau_v$, such that the contribution of the final RHS term in Eq. (D7) is relatively minor.

To determine the dynamical error under the new signal statistics [Eqs. (D1) and (D2)], we first derive the total variance in the output caused by input fluctuations [second term of Eq. (A9)]. We follow the same procedure as in Eqs. (B20)–(B25), starting from Eq. (B22) and using the derivative autocorrelation of Eq. (D1) yields

$$\mathrm{Var}(\mathbb{E}[\langle n(t_i)|\boldsymbol{s}\rangle]_{t_i})_{\boldsymbol{s}} = \mathrm{Var}\left(-\frac{\beta}{\tau_r}\int_{-\infty}^{t_0}dt_i\int_{-\infty}^{t_i}dt\,v(t)e^{-t(t_i-t)/\tau_m}e^{-(t_0-t_i)/\tau_r}\right)$$

$$= \frac{\sigma_v^2\beta^2}{\tau_r^2}\int_{-\infty}^{t_0}dt_i\int_{-\infty}^{t_0}dt_j\left(\int_{-\infty}^{t_i}dt\int_{-\infty}^{t_j}dt'e^{-|t-t'|/(2\tau_v)}\left[1-\frac{|t-t'|}{2\tau_v}\right]e^{-(t_i-t)/\tau_m}e^{-(t_j-t')/\tau_m}\right) \times e^{-(t_0-t_i)/\tau_r}e^{-(t_0-t_j)/\tau_r}. \tag{D8}$$

First we perform the integrals over t and t', which yields

$$\mathrm{Var}(\mathbb{E}[\langle n(t_i)|\boldsymbol{s}\rangle]_{t_i})_{\boldsymbol{s}} = \frac{\sigma_v^2\beta^2/\tau_\mathrm{r}^2}{1/\tau_\mathrm{m}^2 - 1/(2\tau_v)^2}\int_{-\infty}^{t_0} dt_i \int_{-\infty}^{t_0} dt_j \times \left(\frac{\tau_\mathrm{m}/\tau_v}{\tau_\mathrm{m}^2/(2\tau_v)^2 - 1} e^{-|t_i-t_j|/\tau_\mathrm{m}} + e^{-|t_i-t_j|/(2\tau_v)}\left[1 - \frac{|t_i-t_j|}{2\tau_v} - \frac{2}{1-(2\tau_v)^2/\tau_\mathrm{m}^2}\right]\right) e^{-(t_0-t_i)/\tau_\mathrm{r}} e^{-(t_0-t_j)/\tau_\mathrm{r}}. \quad \text{(D9)}$$

Finally, computing the integrals over the sampling times t_i and t_j gives

$$\mathrm{Var}(\mathbb{E}[\langle n(t_i)|\boldsymbol{s}\rangle]_{t_i})_{\boldsymbol{s}} = \frac{\tau_\mathrm{m}^2\beta^2\sigma_v^2}{[1+\tau_\mathrm{m}/(2\tau_v)]^2[1+\tau_\mathrm{r}/(2\tau_v)]^2}\left(1 + \frac{\tau_\mathrm{m}\tau_\mathrm{r}}{\tau_v(\tau_\mathrm{m}+\tau_\mathrm{r})}\right), \quad \text{(D10)}$$

which describes all variance in the output caused by variations in the past input signal, i.e., it comprises both the dynamical error and the informative part $\tilde{g}^2\sigma_v^2$. Equation (D10) happens to be very similar to the variance caused by a signal that is Markovian in its derivative [Eq. (B25)], the differences are that now the denominator of the prefactor is squared and the timescale τ_v has become $2\tau_v$ in the prefactor. Using Eq. (D10) in Eq. (11) we obtain for the dynamical error

$$\sigma_{\hat{p}_{\tau_\mathrm{r}}}^{2,\mathrm{dyn}} = \frac{\sigma_v^2\tau_\mathrm{m}^2\beta^2}{[1+\tau_\mathrm{m}/(2\tau_v)]^2[1+\tau_\mathrm{r}/(2\tau_v)]^2}\left(1 + \frac{\tau_\mathrm{m}\tau_\mathrm{r}}{\tau_v(\tau_\mathrm{m}+\tau_\mathrm{r})}\right) - \tilde{g}^2\sigma_v^2. \quad \text{(D11)}$$

This dynamical error for a signal with $\sigma_\ell = 2\tau_v\sigma_v$ increases monotonically in τ_m but, in contrast with the dynamical error for a signal with $\sigma_\ell \gg \tau_v\sigma_v$ [Eq. (26)], it saturates as $\tau_\mathrm{m} \gg \tau_v$.

The absolute prediction error is $\sigma_{\hat{p}_{\tau_\mathrm{r}}}^2 = \sigma_{\hat{p}_{\tau_\mathrm{r}}}^{2,\mathrm{samp}} + \sigma_{\hat{p}_{\tau_\mathrm{r}}}^{2,\mathrm{dyn}}$, with the unchanged sampling error [Eq. (22)] and the dynamical error of Eq. (D11). Dividing the absolute prediction error by the dynamic gain [Eq. (D6)] and the signal variance σ_v^2 yields the relative error, or SNR^{-1} [see also Eq. (2)]:

$$\mathrm{SNR}^{-1} = \frac{e^{\tau/\tau_v}}{\tau_\mathrm{m}^2\beta^2\sigma_v^2}\left(1+\frac{\tau_\mathrm{m}}{2\tau_v}\right)^2\left(1+\frac{\tau_\mathrm{r}}{2\tau_v}\right)^2\left(\frac{p^2}{\overline{N}} + \frac{p(1-p)}{\overline{N}_\mathrm{I}}\right)\Bigg/\left(1 - \frac{\tau}{2\tau_v} - \frac{1}{1+2\tau_v/\tau_\mathrm{m}} - \frac{1}{1+2\tau_v/\tau_\mathrm{r}}\right)^2 + e^{\tau/\tau_v}\left(1+\frac{\tau_\mathrm{m}\tau_\mathrm{r}}{\tau_v(\tau_\mathrm{m}+\tau_\mathrm{r})}\right)\Bigg/\left(1 - \frac{\tau}{2\tau_v} - \frac{1}{1+2\tau_v/\tau_\mathrm{m}} - \frac{1}{1+2\tau_v/\tau_\mathrm{r}}\right)^2 - 1. \quad \text{(D12)}$$

The relative error is shown as a function of the forecast interval τ and the adaptation time τ_m in Fig. 4(b). When the dynamic gain vanishes [Eq. (D7)] the relative error diverges, which reflects that the network output then no longer contains information about the signal of interest.

[1] H. H. Mattingly, K. Kamino, B. B. Machta, and T. Emonet, Escherichia coli chemotaxis is information limited, Nat. Phys. **17**, 1426 (2021).

[2] H. C. Berg and E. M. Purcell, Physics of chemoreception, Biophys. J. **20**, 193 (1977).

[3] A. J. Tjalma, V. Galstyan, J. Goedhart, L. Slim, N. B. Becker, and P. R. Ten Wolde, Trade-offs between cost and information in cellular prediction, Proc. Natl. Acad. Sci. USA **120**, e2303078120 (2023).

[4] R. M. Macnab and D. E. Koshland Jr., The gradient-sensing mechanism in bacterial chemotaxis, Proc. Natl. Acad. Sci. USA **69**, 2509 (1972).

[5] J. E. Segall, S. M. Block, and H. C. Berg, Temporal comparisons in bacterial chemotaxis., Proc. Natl. Acad. Sci. USA **83**, 8987 (1986).

[6] R. Lux and W. Shi, Chemotaxis-guided movements in bacteria, Crit. Rev. Oral Biol. Med. **15**, 207 (2004).

[7] M. D. Baker, P. M. Wolanin, and J. B. Stock, Systems biology of bacterial chemotaxis, Curr. Opin. Microbiol. **9**, 187 (2006).

[8] C. V. Rao, G. D. Glekas, and G. W. Ordal, The three adaptation systems of bacillus subtilis chemotaxis, Trends Microbiol. **16**, 480 (2008).

[9] U. B. Kaupp, J. Solzin, E. Hildebrand, J. E. Brown, A. Helbig, V. Hagen, M. Beyermann, F. Pampaloni, and I. Weyand, The signal flow and motor response controlling chemotaxis of sea urchin sperm, Nat. Cell Biol. **5**, 109 (2003).

[10] B. M. Friedrich and F. Jülicher, Chemotaxis of sperm cells, Proc. Natl. Acad. Sci. USA **104**, 13256 (2007).

[11] M. Abdelgalil, Y. Aboelkassem, and H. Taha, Sea urchin sperm exploit extremum seeking control to find the egg, Phys. Rev. E **106**, L062401 (2022).

[12] S. R. Lockery, The computational worm: spatial orientation and its neuronal basis in C. elegans, Curr. Opin. Neurobiol. **21**, 782 (2011).

[13] C. C. Govern and P. R. ten Wolde, Optimal resource allocation in cellular sensing systems, Proc. Natl. Acad. Sci. USA **111**, 17486 LP (2014).

[14] G. Malaguti and P. R. ten Wolde, Theory for the optimal detection of time-varying signals in cellular sensing systems, eLife **10**, e62574 (2021).

[15] G. Malaguti and P. R. ten Wolde, Receptor time integration via discrete sampling, Phys. Rev. E **105**, 054406 (2022).

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

[17] F. Tostevin and P. R. ten Wolde, Mutual information in time-varying biochemical systems, Phys. Rev. E **81**, 061917 (2010).

[18] W. Bialek, *Biophysics: Searching for Principles*, edited by P. E. A. Inc. (Princeton University Press, Woodstock, Oxfordshire, 2012), pp. 378–381.

[19] V. Sourjik and 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. USA **99**, 12669 (2002).

[20] B. A. Mello and Y. Tu, Effects of adaptation in maintaining high sensitivity over a wide range of backgrounds for Escherichia coli chemotaxis, Biophys. J. **92**, 2329 (2007).

[21] Y. Tu, T. S. Shimizu, and H. C. Berg, Modeling the chemotactic response of Escherichia coli to time-varying stimuli, Proc. Natl. Acad. Sci. USA **105**, 14855 (2008).

[22] J. R. Maddock and L. Shapiro, Polar location of the chemoreceptor complex in the Escherichia coli cell, Science **259**, 1717 (1993).

[23] T. A. J. Duke and D. Bray, Heightened sensitivity of a lattice of membrane receptors, Proc. Natl. Acad. Sci. USA **96**, 10104 (1999).

[24] T. S. Shimizu, Y. Tu, and H. C. Berg, A modular gradient-sensing network for chemotaxis in Escherichia coli revealed by responses to time-varying stimuli, Mol. Syst. Biol. **6**, 382 (2010).

[25] J. M. Keegstra, K. Kamino, F. Anquez, M. D. Lazova, T. Emonet, and T. S. Shimizu, Phenotypic diversity and temporal variability in a bacterial signaling network revealed by single-cell FRET, eLife **6**, e27455 (2017).

[26] J. Monod, J. Wyman, and J. P. Changeux, On the nature of allosteric transitions: A plausible model, J. Mol. Biol. **12**, 88 (1965).

[27] V. Sourjik and H. C. Berg, Functional interactions between receptors in bacterial chemotaxis, Nature (London) **428**, 437 (2004).

[28] B. A. Mello and Y. Tu, An allosteric model for heterogeneous receptor complexes: Understanding bacterial chemotaxis responses to multiple stimuli, Proc. Natl. Acad. Sci. USA **102**, 17354 (2005).

[29] J. E. Keymer, R. G. Endres, M. Skoge, Y. Meir, and N. S. Wingreen, Chemosensing in Escherichia coli: Two regimes of two-state receptors, Proc. Natl. Acad. Sci. USA **103**, 1786 (2006).

[30] K. Kamino, J. M. Keegstra, J. Long, T. Emonet, and T. S. Shimizu, Adaptive tuning of cell sensory diversity without changes in gene expression, Sci. Adv. **6**, eabc1087 (2020).

[31] N. Barkai and S. Leibler, Robustness in simple biochemical networks to transfer and process information, Nature (London) **387**, 913 (1997).

[32] T. M. Yi, Y. Huang, M. I. Simon, and J. Doyle, Robust perfect adaptation in bacterial chemotaxis through integral feedback control, Proc. Natl. Acad. Sci. USA **97**, 4649 (2000).

[33] B. A. Mello and Y. Tu, Perfect and near-perfect adaptation in a model of bacterial chemotaxis, Biophys. J. **84**, 2943 (2003).

[34] J. S. Parkinson, G. L. Hazelbauer, and J. J. Falke, Signaling and sensory adaptation in Escherichia coli chemoreceptors: 2015 update, Trends Microbiol. **23**, 257 (2015).

[35] M. Reinhardt, G. Tkačik, and P. R. ten Wolde, Path weight sampling: Exact Monte Carlo computation of the mutual information between stochastic trajectories, Phys. Rev. X **13**, 041017 (2023).

[36] A. J. Tjalma, Predicting concentration changes via discrete receptor sampling (2024), doi: 10.5281/zenodo.11198862.

[37] M. Li and G. L. Hazelbauer, Cellular stoichiometry of the components of the chemotaxis signaling complex, J. Bacteriol. **186**, 3687 (2004).

[38] A. Briegel, X. Li, A. M. Bilwes, K. T. Hughes, G. J. Jensen, and B. R. Crane, Bacterial chemoreceptor arrays are hexagonally packed trimers of receptor dimers networked by rings of kinase and coupling proteins, Proc. Natl. Acad. Sci. USA **109**, 3766 (2012).

[39] J. M. Keegstra, F. Avgidis, Y. Mulla, J. S. Parkinson, and T. S. Shimizu, Near-critical tuning of cooperativity revealed by spontaneous switching in a protein signalling array, bioRxiv 2022.12.04.518992 (2022).

[40] M. Meijers, S. Ito, and P. R. ten Wolde, Behavior of information flow near criticality, Phys. Rev. E **103**, L010102 (2021).

[41] J. F. Staropoli and U. Alon, Computerized analysis of chemotaxis at different stages of bacterial growth, Biophys. J. **78**, 513 (2000).

[42] L. Turner, W. S. Ryu, and H. C. Berg, Real-time imaging of fluorescent flagellar filaments, J. Bacteriol. **182**, 2793 (2000).

[43] N. C. Darnton, L. Turner, S. Rojevsky, and H. C. Berg, On torque and tumbling in swimming escherichia coli, J. Bacteriol. **189**, 1756 (2007).

[44] N. W. Frankel, W. Pontius, Y. S. Dufour, J. Long, L. Hernandez-Nunez, and T. Emonet, Adaptability of non-genetic diversity in bacterial chemotaxis, eLife **3**, e03526 (2014).

[45] Y. S. Dufour, X. Fu, L. Hernandez-Nunez, and T. Emonet, Limits of feedback control in bacterial chemotaxis, PLoS Comput. Biol. **10**, e1003694 (2014).

[46] N. B. Becker, A. Mugler, and P. R. ten Wolde, Optimal prediction by cellular signaling networks, Phys. Rev. Lett. **115**, 258103 (2015).

[47] D. R. Brumley, F. Carrara, A. M. Hein, Y. Yawata, S. A. Levin, and R. Stocker, Bacteria push the limits of chemotactic precision to navigate dynamic chemical gradients, Proc. Natl. Acad. Sci. USA **116**, 10792 (2019).

[48] J. M. Keegstra, F. Carrara, and R. Stocker, The ecological roles of bacterial chemotaxis, Nat. Rev. Microbiol. **20**, 491 (2022).

[49] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell*, 5th ed. (Garland Science, New York, NY, 2008), pp. 917–919.

[50] T. Emonet and P. Cluzel, Relationship between cellular response and behavioral variability in bacterial chemotaxis, Proc. Natl. Acad. Sci. USA **105**, 3304 (2008).

[51] F. Tostevin and P. R. ten Wolde, Mutual information between input and output trajectories of biochemical networks, Phys. Rev. Lett. **102**, 218101 (2009).

[52] N. Van Kampen, *Stochastic Processes in Physics and Chemistry* (North Holland, Amsterdam, 1992).