Gain control in molecular information processing: lessons from neuroscience

2012 Phys. Biol. 9 026003

(http://iopscience.iop.org/1478-3975/9/2/026003)

Phys. Biol. **9** (2012) 026003 (7pp) doi:10.1088/1478-3975/9/2/026003

Gain control in molecular information processing: lessons from neuroscience

Ilya Nemenman

Departments of Physics and Biology, Computational and Life Sciences Initiative, Emory University, Atlanta, GA 30322, USA

E-mail: ilya.nemenman@emory.edu

Received 3 August 2011
Accepted for publication 8 February 2012
Published 4 April 2012
Online at stacks.iop.org/PhysBio/9/026003

Abstract

Statistical properties of environments experienced by biological signaling systems in the real world change, which necessitates adaptive responses to achieve high fidelity information transmission. One form of such adaptive response is gain control. Here, we argue that a certain simple mechanism of gain control, understood well in the context of systems neuroscience, also works for molecular signaling. The mechanism allows us to transmit more than 1 bit (on or off) of information about the signal independent of the signal variance. It does not require additional molecular circuitry beyond that already present in many molecular systems, and in particular, it does not depend on existence of feedback loops. The mechanism provides a potential explanation for abundance of ultrasensitive response curves in biological regulatory networks.

1. Introduction

An important function of all biological systems is responding to signals from the surrounding environment. These signals (hereafter assumed to be scalars) $s(t)$ are often probabilistic, described by some probability distribution $P[s(t)]$. They have non-trivial temporal dynamics so that the probability of a certain value of the signal at a given time is dependent on its entire history.

Often the response $r(t)$ is produced from s by (possibly nonlinear and noisy) temporal filtering. For example, in a deterministic molecular circuit, we may have

$$\frac{\mathrm{d}r}{\mathrm{d}t} = f\left(s\left(t\right)\right) - kr, \tag{1}$$

where f is the response molecule production rate, which depends on the current value of the signal. Here, k is the rate of the first-order degradation of the molecule. Note that $r(t)$ depends on the entire history of $s(t')$, $t' < t$, and hence carries information about it. For more complicated, nonlinear degradation or for r-dependent production, equation (1) may be interpreted as linearization around the mean response. We point out that this equation can also describe dynamics of a continuous firing rate in neural systems, and this realization is one of the main motivations for the current paper.

The distribution of stimuli $P[s(t)]$ places severe constraints on the forms of f that can transduce the stimuli with high fidelity. To see this, for quasi-static signals (i.e. when the signal correlation time τ is large, $\tau \gg 1/k$), we use equation (1) to write the steady-state dose–response curve

$$r_{\mathrm{ss}} = f\left(s(t)\right)/k. \tag{2}$$

A typical monotonic, sigmoidal f is characterized by only a few large-scale parameters: the range $[f_{\min}, f_{\max}]$, the mid-point $s_{1/2}$ and the width of the transition region Δs (cf figure 1). If the signal mean $\mu \gg s_{1/2}$, then, for most signals, $r_{\mathrm{ss}} \approx f_{\max}/k$. Then, responses to two different signals s_1 and s_2 are indistinguishable as long as

$$\left.\frac{\mathrm{d}r_{\mathrm{ss}}(s)}{\mathrm{d}s}\right|_{s=s_1} (s_2 - s_1) < \delta r, \tag{3}$$

where δr is the precision of the response resolution. Similarly, when $\mu \ll s_{1/2}$, then $r_{\mathrm{ss}} \approx f_{\min}/k$. Thus, for reliably communicating information about the signal, f should be tuned, such that $s_{1/2} \approx \mu$. If a biological system can change its $s_{1/2}$ to follow changes in μ, this is called *adapting to the mean* of the signal, and if $s_{1/2}(\mu) = \mu$, then the adaptation is *perfect* [1, 2]. Similarly, if the quasi-static signal is taken from the distribution with $\sigma \equiv (\langle s(t)^2\rangle_t - \mu^2)^{1/2} \gg \Delta s$, then the response to most of the signals will be indistinguishable from the extrema. It will be near $\sim (r_{\max} + r_{\min})/2$ if $\sigma \ll \Delta s$. Thus, to use the full dynamic range of the response, a biological system must tune the width of the sigmoidal dose–response

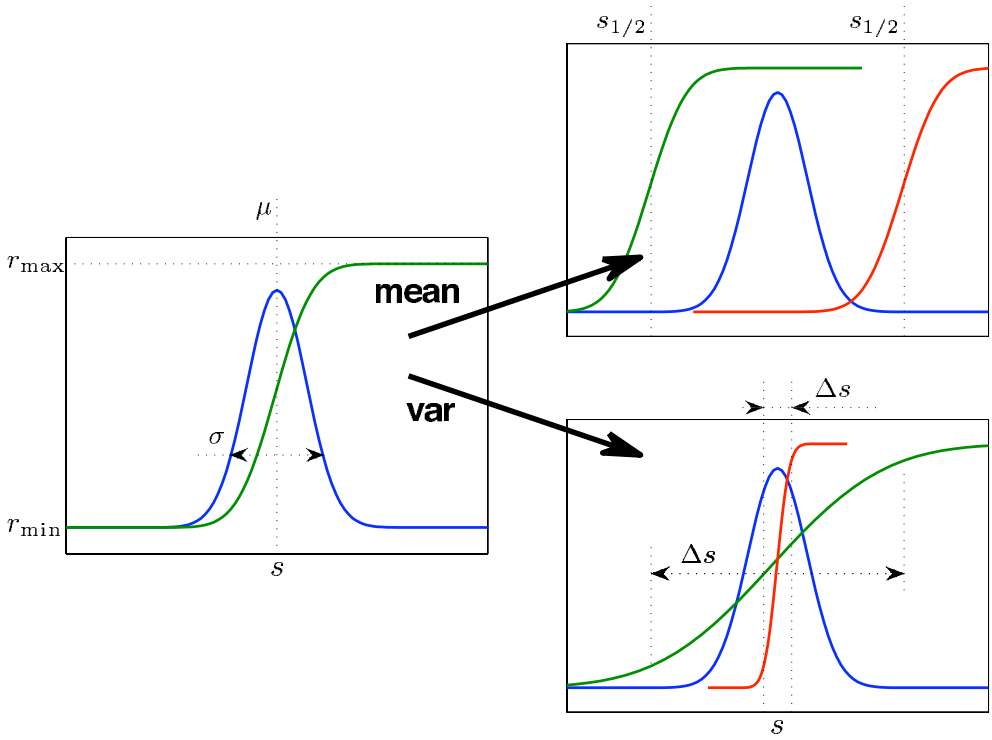

Figure 1. Parameters characterizing response to a signal. Left panel: the probability distribution of the signal, $P(s)$ (blue), and the best-matched steady-state dose–response curve r_{ss} (green). Top right: if the mid-point of the dose–response curve, $s_{1/2}$, is far away from the mean of the signal, a typical response will be extremal. Bottom right: if the width of the dose–response curve, Δs, is considerably different from the standard deviation of the signal, then the typical response is either extremal, or at its mid-point. These mismatches prevent using the entire dynamic range of the response to convey information about the signal.

curve to $\Delta s \approx \sigma$; this is called the *variance adaptation* or *gain control* [2].

Both of these adaptation behaviors can be traced to the same theoretical argument [3]: for sufficiently general conditions on the response resolution δr, the response that optimizes the fidelity of a signaling system, as measured by its information-theoretic channel capacity [4], is $r^*_{ss}(s) = \int_{-\infty}^{s} P(s')ds'$, where $P(s')$ is the probability distribution of an instantaneous signal value, obtained from $P[s(t)]$. In some situations, when the mean and the variability of the signal scale proportionally, like in fold-change detection problems [5, 6], the two adaptations are deeply intertwined. However, more generally environmental changes that lead to varying μ and σ, as well as the mechanisms of the adaptation, are distinct. Thus, it often makes sense to consider the two adaptations as separate phenomena [2].

Adaptation to the mean, sometimes also called *desensitization*, has been observed and studied in a wide variety of biological sensory systems [1, 3, 7–9], with active work persisting to date[1]. In contrast, while gain control has been investigated in neurobiology [10–12], we are not aware of its systematic analysis in molecular sensing. In this paper, we start filling in the gap. Our main contribution is the observation that a mechanism for gain control, observed in a fly motion estimation system by Borst *et al* [12], can be transferred to molecular information processing with minimal modifications. Importantly, unlike adaptation to the mean, which is implemented typically using extra feedback circuitry [1, 9, 13], the gain control mechanism we analyze requires no additional regulation. It is built-in into many molecular signaling systems. The main ingredients of the gain control mechanism in [12] is a strongly nonlinear, sigmoidal response function $f(s)$ and a realization that real-world signals are dynamic with a nontrivial temporal structure. Thus, one must move away from the steady-state response analysis, and autocorrelations within the signals will allow the response to carry more information about the signal than seems possible naïvely.

In this context, we show that, just like the neural circuits in [12], a simple biochemical circuit described in equation (1) can be made insensitive to changes in σ with no extra regulatory features. That is, for an arbitrary choice of σ, and for a wide range of other parameters, the circuit can generate an output that is informative of the input and, in particular, carries more than a single bit of information about it. For brevity, we will not review the original work on gain control in neural systems [12], but will instead develop the methodology directly in the molecular context.

[1] To illuminate the relation between the classic perfect adaptation (in *Escherichia coli* chemotaxis or elsewhere) and our terminology, we point out that we consider slowly varying chemical concentrations inputs in such experiments not as signals, but as *mean* signals. Fluctuations create additional fast signals on top of these slowly changing means. Feedback then ensures that the mean signal elicits the mean response.

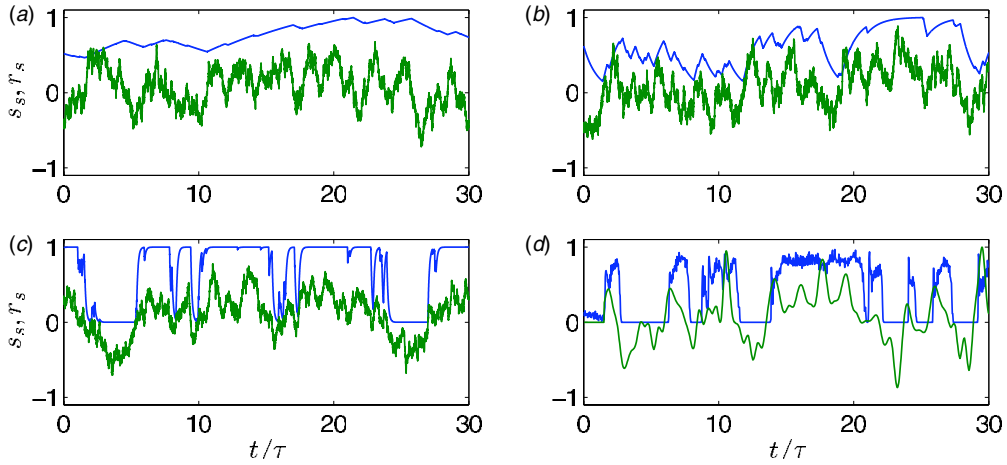

Figure 2. Examples of signals and responses for dynamics in equations (1), (4), (5). On the vertical axis, we plot normalized signals $s_s = (s - m)/\max(s)$ (green) and normalized responses $r = r/\max(r)$ (blue). On the horizontal axis, the time is rescaled by the correlation time of the signal, τ. Panels (a)–(c) have $k\tau = 0.1, 1, 10$, respectively (recall that $1/k$ is the response time of the circuit, equation (1)). In panel (d), we show, for comparison, the firing rate of a blow fly motion-sensitive neuron H1 and its driving stimulus, both rescaled to 1 (see [15] for details of this experiment). The stimulus had little power at high frequencies, but the single-exponential correlation structure held for long times. Note the similarity between the telegraph-series-like structure of the responses in panels (c) and (d). Since the H1 neuron served as a model neural system for the feedback-free gain adaptation in [12], this similarity suggests to look for gain-controlled responses in molecular signaling, equation (1), as well.

2. Results: gain control with no additional regulatory structures

Let us assume for simplicity that the signal in equation (1) has the Ornstein–Uhlenbeck dynamics with

$$\langle s(t)\rangle = \mu, \qquad \langle s(t+t')s(t)\rangle = \sigma^2 \mathrm{e}^{-t'/\tau}. \tag{4}$$

We will assume that the response has been adapted to the mean value of this signal (likely by additional feedback control circuitry, not considered here explicitly), so that the response to $s = \mu$ is half maximal. Now, we explore how insensitivity to σ can be achieved as well.

We start with a step-function approximation to the sigmoidal response production rate

$$f = f_0\theta(s-\mu) = f_0 \times \begin{cases} 0, & s < \mu, \\ 1/2, & s = \mu', \\ 1, & s > \mu, \end{cases} \tag{5}$$

where f_0 is some constant. This is a limiting case of very high Hill number dose–response curves, which have been observed in nature [14]. Figure 2 shows sample signals and responses produced by this system. Note that such f makes the system manifestly insensitive to σ. Any changes in σ will not result in changes to the response; hence, the gain is controlled *perfectly*.

Nevertheless, this choice of f is pathological, resulting in a binary steady-state response ($r_{\mathrm{ss}} = 0$ for $s < \mu$, and $r_{\mathrm{ss}} = f_0/k$ otherwise). That is, the response cannot carry more than 1 bit of information about the stimulus. However, as illustrated in figure 2, a *dynamic* response is not binary and varies over its entire dynamic range. Can this make a difference and produce a dose–response relation that is both high fidelity and insensitive to the variance of the signal?

To answer this, we first specify what we mean by the dose–response curve or the input–output relation when there is no steady-state response. For the response at a single time point t, we can write $P(r(t)|\{s(t' \leqslant t)\}) = \delta(r - r[s])$, where $\delta(\cdots)$ is the Dirac δ-function, and the functional $r[s]$ is obtained by solving equation (1). Since the signal is probabilistic, marginalizing over all but the instantaneous value of it at time $t - t'$, one obtains $P(r(t)|s(t-t'))$, the distribution of the response at time t conditional on the value of the signal at $t - t'$. Furthermore, for the distribution of the signal given by equation (4), one can numerically integrate equation (1) and evaluate the correlation $c(t') = \langle r(t)s(t-t')\rangle_t$.[2] Since equation (1) is causal, $c(t')$ has a maximum at some $t' = \Delta(\tau, k) \geqslant 0$, illustrated in figure 3. Correspondingly, in this paper, we replace the familiar notion of the dose–response curve by the delayed probabilistic input–output relation $P(r(t)|s(t-\Delta))$. This is a relatively common choice in molecular signaling [16] and in neuroscience [10].

We emphasize that, since f is a step function, $f(s) \equiv f(\alpha s)$ for any positive scalar α. Therefore, for two signals that can be mapped into each other by rescaling, $P(r(t)|s(t-\Delta)/\sigma)$ is manifestly independent of σ. In other words, the system is gain compensating *by construction*. Correspondingly, in figure 4, we plot the input–output relation for $k\tau = 10$, where s is measured in the units of σ. A smooth, probabilistic, sigmoidal response with a width of the transition region $\Delta \sim \sigma$ is clearly visible. This is because, for a step function f, the value of $r(t)$ depends not on $s(t)$, but on how long the signal has been positive prior to the current time. In its turn, this duration is correlated with s/σ, producing a probabilistic dependence between r and s/σ. The latter is manifestly invariant to variance changes.

These arguments make it clear that the fidelity of the response curve should depend on the ratio of characteristic times of the signal and the response, $k\tau$. Indeed, as shown in

[2] All simulations were performed using Matlab v 7.6 and Octave v 3.0.2 using Apple Macbook Air. The correlation time of the signal was $\tau = 300$ integration time steps, and averages were taken over 3×10^6 time steps. To change the value of $k\tau$, only k was adjusted.

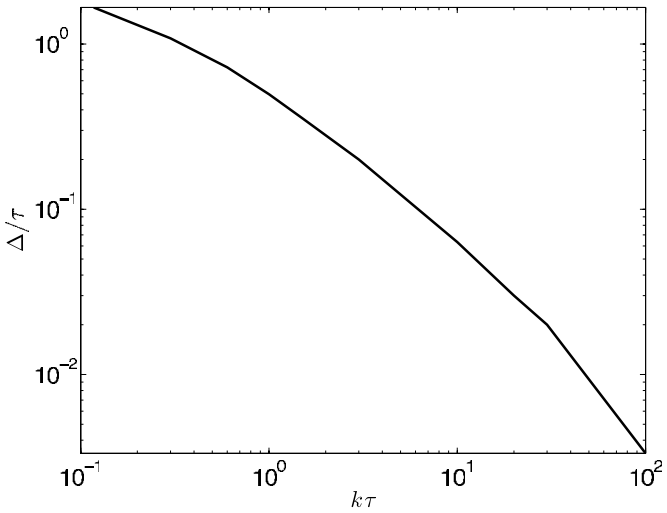

Figure 3. Dependence of the delay between the signal and the response, Δ, which achieves the maximum correlation between s and r. Here, Δ is expressed in units of the signal correlation time τ, and it is studied as a function of $k\tau$, the ratio of characteristic time scales of the signal and the response dynamics.

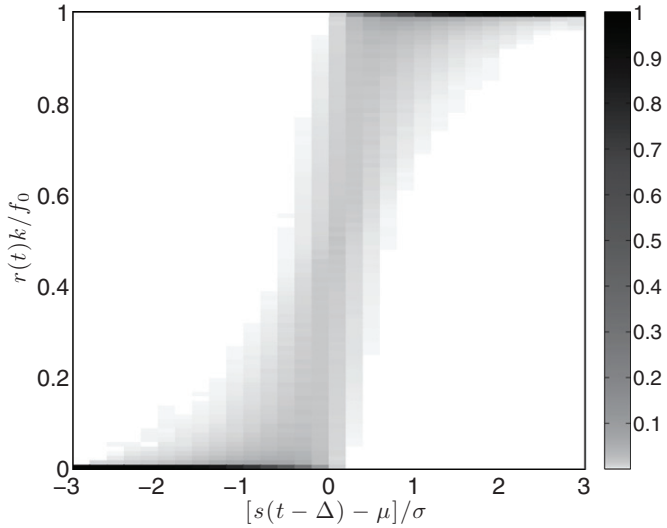

Figure 4. Conditional distribution $P(r(t)|s(t-\Delta))$ for $k\tau = 10$. The distributions depend only on s/σ, manifesting gain-compensating nature of the system. The signals are discretized into 30 values in the range of $[-3\sigma, +3\sigma]$. For each $s(t-\Delta)$, a histogram of $r(t)$ is built with 100 distinct r values. The normalized histograms are gray-scale coded as columns in the figure, with dark representing the higher conditional probability $P \sim 1$. We use a nonlinear color scale to enhance the plot.

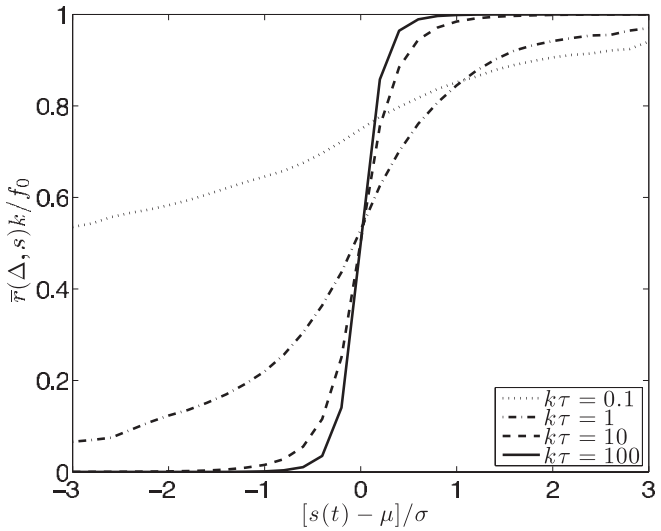

Figure 5. Mean conditional response $\bar{r}(\Delta, s)$ for different combinations of the signal and the response characteristic times, $k\tau$.

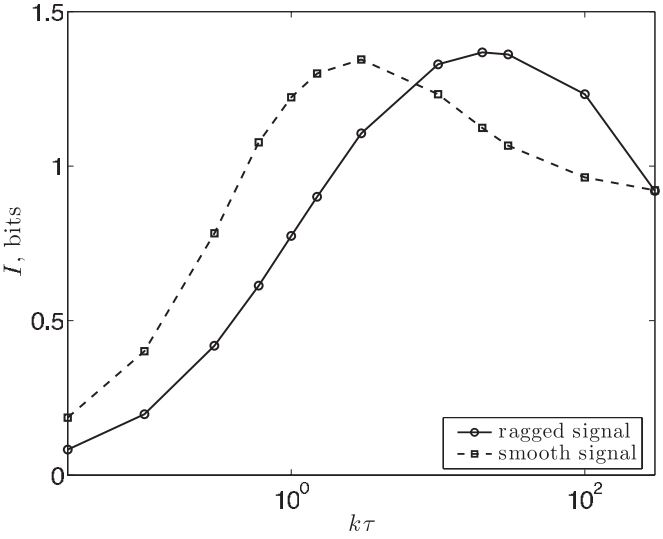

Figure 6. The signal–response mutual information at the optimal temporal delay as a function of $k\tau$. The solid line represents $I_k[r(t+\Delta), s(t)]$, the information for the Ornstein–Uhlenbeck signal, and the maximum of the information here is $I_{\max} = 1.37$ bits, achieved at $k^* \approx 20/\tau$. The dashed line stands for the same information for the smoothed signal. It is maximized at $k^* \approx 3/\tau$ with $I_{\max} = 1.35$ bits.

figure 2, for $k\tau \to 0$, the response integrates the signal over long times. It is little affected by the current value of the signal and does not span the full available dynamic range. At the other extreme of a very fast response, $k\tau \to \infty$, the system is always almost in a steady state. Then, the step nature of f is evident, and the response quickly swings between two limiting values (f_0/k and 0).

We illustrate the dependence of the response conditional distribution on the integration time in figure 5 by plotting $\bar{r}(\Delta, s) = \int \mathrm{d}r\, r(t+\Delta) P(r(t+\Delta)|s(t))$, the conditional-averaged response for different values of $k\tau$. Neither $k\tau \to 0$ nor $k\tau \to \infty$ are optimal for signal transmission. One expects existence of an optimal k^*, for which most of the dynamic range of r gets used, but the response is not completely binary. To find this optimum, we evaluate the mutual information [4] between the signal and the response at the optimal delay, $I_k[r(t+\Delta), s(t)]$, as a function of $k\tau$, cf figure 6. A broad maximum in information transmission is observed near $k^* \approx 20/\tau$, which is not too far from the quasi-static limit. However, $I_{\max} \equiv I_{k^*} = 1.37$ bits, which is substantially larger than 1. Thus, temporal correlations in the stimulus allow one to transmit 37% more information about it than the step response would suggest naïvely. This information is transmitted in a

gain-controlled manner so that changes in σ have no effect on the transmitted information amount. The value should remain above 1 bit even for non-step-like f, as long as f is sigmoidal and $\Delta s/\sigma \ll 1$.

We emphasize that the information here is *per signaling event*, i.e. per independent value of the signal. Indeed, since we consider responses that change much faster than the signals, the system is always near a steady state, and each 'new' value of the signal is encoded by an independent response value. This also make sense experimentally: measuring joint distributions of time series of stimuli and responses is very hard, and experiments often focus on information between one signal value and one response value [16]. Our analysis is relevant for interpretation of such experiments.

Effects of the signal structure. The gain insensitivity of the constructed molecular circuit model depends only weakly on details of the temporal structure of the signal. As long as there are autocorrelations, one can use them to transmit more than 1 bit about the signal in a gain-independent fashion using the strong nonlinearity of f. To verify this, we replace the Ornstein–Uhlenbeck signal, equation (4), with its low-pass filtered version, $s'(t) = 1/k \int^t \mathrm{d}t' s(t') \mathrm{e}^{-k(t-t')}$, and k is the same as in equation (1). This new signal is smoother and has less structure at high frequencies. We repeat the same analysis as above to find Δ, estimate the conditional response distribution, and then evaluate I_k, the stimulus–response information. We find that the maximum information in this case is $I_{\max} = 1.35$ bits, statistically indistinguishable from the Ornstein–Uhlenbeck case. However, the maximum is now at $k\tau \approx 3$. This is because the smooth signal changes its sign a lot less often, and smaller integration times are needed to approach the extreme values of the response.

Knowing σ in a gain-insensitive response. When gain-insensitive, the system looses information about the actual signal variance. This rarely happens in biology. For example, while we see well at different ambient light levels, we nonetheless know how bright it is outside. For the fly visual system, it was shown that variance independence of the response breaks on long time scales. The signal variance can be inferred from long-term features of the neural code [10, 17]. Correspondingly, we ask if long term observation of the response of an approximately gain-controlled molecular signaling circuit allows one to infer the signal variance σ.

To this extent, consider f as a narrow sigmoid, with the width of the crossover region $\Delta s/\sigma \ll 1$. The effect of the variance on the response is still negligible. For concreteness, we take $f = f_0[\tanh((s-\mu)/\Delta s) + 1]$. Consider now the fraction of time during which the rate of change of the response is near $\max(f)$. This requires that $r \approx 0$ (so that the degradation kr is negligible), but s is already large, $(s-\mu)/\Delta s \gg 1$. The probability of this happening depends on the signal variance and hence on the speed with which the signal crosses over the threshold region. Thus, one can estimate σ by observing a molecular circuit for a long time and counting how often the rate of change of the response is large. The probability of a large derivative will depend on the exact shape of f. However, for a signal defined by equation (4), the statistical error of any such counting estimator will scale as $\propto \sqrt{\tau/T}$. Hence, the system can be almost insensitive to σ on short time scales, but allow its determination from long observations periods $T \gg \tau$.

To verify this, we simulate the signal determined by equation (4) with $k\tau = 20$, which maximizes the signal–response mutual information. We arbitrarily choose the cutoff of 80% of the maximum possible rate of change of the response, and we calculate the mean fraction of time ϕ when the rate is above the cutoff. We further calculate the standard deviation of this fraction, σ_ϕ. We repeat this for signals with various $\Delta s/\sigma$ and for experiments of different duration, obtaining a time dependence of the Z-score for disambiguating two signals with different variances $Z = (\phi_2 - \phi_1)/\sqrt{\sigma_{\phi_1}^2 + \sigma_{\phi_2}^2}$, where the indices $1, 2$ denote the signals being disambiguated. For example, for distinguishing signals with $\Delta s/\sigma = 1/10$ and $1/20$, the data result in the fit $Z \approx 0.8(T/\tau)^{0.48\pm0.04}$. This is consistent with the square root scaling (the error bars indicate the 95% confidence interval). That is, for T/τ as little as 10, $Z > 2$, and the two signals are distinguishable. Signals with larger variances are harder to disambiguate. For example, for attempting to distinguish $\Delta s/\sigma = 1/90$ from $1/100$, $Z \approx 9.4 \times 10^{-3}(T/\tau)^{0.56\pm0.08}$, and Z crosses 2 for $T \approx 15\,000\tau$.

This long-term variance determination can be performed molecularly. For example, one can use a feed-forward incoherent loop with r as an input [18]. The loop acts as an approximate differentiator for signals that change slowly compared to its internal relaxation times [19]. The output of the loop can then activate a subsequent chemical species by a Hill-like dynamics, with the activation threshold close to the maximum of f. If this species degrades slowly, it will integrate the fraction of time when $\mathrm{d}r/\mathrm{d}t$ is above the threshold, providing the readout of the signal variance.

3. Discussion

In this work, we were able to translate the arguments of [12] to the context of simple continuous biochemical dynamics, equation (1). We have argued that, just like neural circuits, simple molecular systems can respond to signals in a gain-insensitive way without a need for explicit adaptation and feedback loops (though such loops may be needed to choose $s_{1/2}$ and k appropriately). That is, they can be sensitive only to the signal value relative to its standard deviation. To make the mechanism work, the signaling system must obey the following criteria:

- a nonlinear–linear (NL) response, i.e. a strongly nonlinear, sigmoidal response production rate f integrated (linearly) over time,
- properly matched time scales of the signal and the response dynamics.

In addition, the information about the signal variance can be recovered if

- episodes of large values of the rate of change of the response are counted over long times.

We have also argued that our results hold for a broad class of probability distributions of the signals.

Naïvely transmitted information of only 1 bit (on or off) would be possible with a step function f. However, the response in this system is a time average of a nonlinear function of the signal. This allows one to use temporal correlations in the signal to transmit more than 1 bit of information for broad classes of signals. While 1.35 bits may not seem like much more than 1, the question of whether molecular signaling systems can achieve more than 1 bit at all is still a topic of active research [16, 20, 21]. Similar use of temporal correlations has been reported to increase information transmission in other circuits, such as clocks [22]. In practice, in our case, there is a tradeoff between variance independence and high information transmission through the circuit: a wider production rate would give a higher maximal information for properly tuned signals, but then the information would drop down to zero if $\Delta s \gg \sigma$. It would be interesting to explore the optimal operational point for this tradeoff under various optimization hypotheses.

While our analysis is applicable to any molecular system, molecular or neural, that satisfies the three conditions listed above, there are specific examples where we believe it may be especially relevant. The *E. coli* chemotaxis flagellar motor has a very sharp response curve (the Hill coefficient of about 10) [14]. This system is possibly the best studied example of biological adaptation to the mean of the signal. However, the question of whether the system is insensitive to the signal variance changes has not been addressed. The ultrasensitivity of the motor suggests that it might be. Similarly, in eukaryotic signaling, push–pull enzymatic amplifiers, including MAP kinase mediated signaling pathways, are also known for their ultrasensitivity [23–25]. And yet the ability of these circuits to respond to temporally varying signals in a variance-independent way has not been explored.

We end this paper with a simple observation. While the number of biological information-processing systems is astonishing, the types of computations they perform are limited. Focusing on the computation would allow cross-fertilization between seemingly disparate fields of quantitative biology. The phenomenon studied here, lifted wholesale from neurobiology literature, is an example. Arguably, computational neuroscience has had a head start compared to computational molecular systems biology. The latter can benefit immensely by embracing well-developed results and concepts from the former.

Acknowledgments

We thank F Alexander, W Hlavacek and M Wall for useful discussions in the earlier stages of the work, participants of *The 4th International q-bio Conference* for the feedback, and F Family and the anonymous referees for commenting on the manuscript. We are grateful to R de Ruyter van Steveninck for providing the data for one of the figures. This work was supported in part by DOE under contract number DE-AC52-06NA25396 and by NIH/NCI grant number 7R01CA132629-04.

References

[1] Berg H 2003 *E. Coli in Motion* (New York: Springer)
[2] Nemenman I 2012 Information theory and adaptation *Quantitative Biology: From Molecules to Cellular Systems at press* ed M E Wall (Boca Raton, FL: CRC Press) at press (publish date 15 August 2012)
[3] Laughlin S 1981 A simple coding procedure enhances a neuron's information capacity *Z. Naturforsch.* **36**c 910
[4] Shannon C and Weaver W 1949 *The Mathematical Theory of Communication* (Urbana, IL: University of Illinois Press)
[5] Shoval O, Goentoro L, Hart Y, Mayo A, Sontag E and Alon U 2010 Fold-change detection and scalar symmetry of sensory input fields *Proc. Natl Acad. Sci. USA* **107** 15995–6000
[6] Lazova M, Ahmed T, Bellomo D, Stocker R and Shimizu T 2011 Response rescaling in bacterial chemotaxis *Proc. Natl Acad. Sci. USA* **108** 13870–5
[7] Normann R and Perlman I 1979 The effects of background illumination on the photoresponses of red and green cells *J. Physiol.* **286** 491 (Available at http://jp.physoc.org/content/286/1/491.short)
[8] MacGlashan D Jr, Lavens-Phillips S and Miura K 1998 IgE-mediated desensitization in human basophils and mast cells *Front. Biosci.* **3** 746–56 (Available at http://bioscience.org/1998/v3/d/macglash/d746-756.htm)
[9] Detwiler P, Ramanathan S, Sengupta A and Shraiman B 2000 Engineering aspects of enzymatic signal transduction: photoreceptors in the retina *Biophys. J.* **79** 2801
[10] Brenner N, Bialek W and de Ruyter van Steveninck R 2000 Adaptive rescaling optimizes information transmission *Neuron* **26** 695
[11] Gaudry K and Reinagel P 2007 Contrast adaptation in a nonadapting LGN model *J. Neurophysiol.* **98** 1287–96
[12] Borst A, Flanagin V and Sompolinsky H 2005 Adaptation without parameter change: dynamic gain control in motion detection *Proc. Natl Acad. Sci. USA* **102** 6172
[13] Ma W, Trusina A, El Samad H, Lim W and Tang C 2009 Defining network topologies that can achieve biochemical adaptation *Cell* **138** 760–73
[14] Cluzel P, Surette M and Leibler S 2000 An ultrasensitive bacterial motor revealed by monitoring signaling proteins in single cells *Science* **287** 1652
[15] Nemenman I, Lewen G, Bialek W and de Ruyter van Steveninck R 2008 Neural coding of natural stimuli: information at sub-millisecond resolution *PLoS Comput. Biol.* **4** e1000025
[16] Cheong R, Rhee A, Wang C J, Nemenman I and Levchenko A 2011 Information transduction capacity of noisy biochemical signaling networks *Science* **334** 354–8
[17] Fairhall A, Lewen G, Bialek W and de Ruyter van Steveninck R 2001 Efficiency and ambiguity in an adaptive neural code *Nature* **412** 787
[18] Mangan S and Alon U 2003 Structure and function of the feed-forward loop network motif *Proc. Natl Acad. Sci. USA* **100** 11980–5
[19] Sontag E 2010 Remarks on feedforward circuits, adaptation, and pulse memory *IET Syst. Biol.* **4** 39–51
[20] Ziv E, Nemenman I and Wiggins C 2007 Optimal signal processing in small stochastic biochemical networks *PLoS One* **2** e1077
[21] Tkacik G, Callan C and Bialek W 2008 Information capacity of genetic regulatory elements *Phys. Rev.* E **78** 011910

[22] Mugler A, Walczak A and Wiggins C 2010 Information-optimal transcriptional response to oscillatory driving *Phys. Rev. Lett.* **105** 058101

[23] Goldbeter A and Koshland D 1981 An amplified sensitivity arising from covalent modification in biological systems *Proc. Natl Acad. Sci. USA* **78** 6840–4

[24] Huang C and Ferrell J 1996 Ultrasensitivity in the mitogen-activated protein kinase cascade *Proc. Natl Acad. Sci. USA* **93** 10078–83

[25] Samoilov M, Plyasunov S and Arkin A 2005 Stochastic amplification and signaling in enzymatic futile cycles through noise-induced bistability with oscillations *Proc. Natl Acad. Sci. USA* **102** 2310–5