Optimal temporal patterns for dynamical cellular signaling

Yoshihiko Hasegawa
Department of Information and Communication Engineering, Graduate School of Information Science and Technology, The University of Tokyo, Tokyo 113-8656, Japan
E-mail: yoshihiko.hasegawa@gmail.com
Keywords: chemical thermodynamics, signal transduction, optimal control
Supplementary material for this article is available online

Abstract
Cells use temporal dynamical patterns to transmit information via signaling pathways. As optimality with respect to the environment plays a fundamental role in biological systems, organisms have evolved optimal ways to transmit information. Here, we use optimal control theory to obtain the dynamical signal patterns for the optimal transmission of information, in terms of efficiency (low energy) and reliability (low uncertainty). Adopting an activation-deactivation decoding network, we reproduce several dynamical patterns found in actual signals, such as steep, gradual, and overshooting dynamics. Notably, when minimizing the energy of the input signal, the optimal signals exhibit overshooting, which is a biphasic pattern with transient and steady phases; this pattern is prevalent in actual dynamical patterns. We also identify conditions in which these three patterns (steep, gradual, and overshooting) confer advantages. Our study shows that cellular signal transduction is governed by the principle of minimizing free energy dissipation and uncertainty; these constraints serve as selective pressures when designing dynamical signaling patterns.

1. Introduction

Cells transmit information through signal transduction and transcription networks [1, 2]. Recent studies have revealed that, along with the identity and static concentration of molecules, cells also encode information into dynamical patterns [3–9]. Examples of dynamical patterns include extracellular signal-regulated kinase (ERK), the yeast transcription factor Msn2, the transcription factor NF-κB, a protein kinase AKT, and calcium signaling. Many studies have used nonlinear and stochastic approaches to investigate the properties of dynamical cellular information processing [10–17]. Because signal transduction plays central and crucial roles in the survival of cells, the time course of dynamical patterns is expected to be highly optimized so that cells can efficiently and accurately transmit information. Although the advantages of dynamical signals over static ones have been extensively studied [8, 18], there has been little investigation into determining which dynamical signals are the best. We assume that two principles that are prevalent in many biological systems govern the optimality of signal patterns: energetic efficiency (low energy) and reliability (low uncertainty). Biological systems are often characterized by low energy consumption. For instance, neuronal systems are known to function with remarkably low energy consumption. Specifically, in neurons, information processing capability is bounded by the amount of energy consumption and it is reported that the energy consumption of the brain has limited its size ([19, 20] and references therein). Biochemical networks process information for a variety of purposes, and higher specificity, lower variation, and larger signal amplification demand more energy consumption [21]. These facts induce us to think that the energetic cost also plays important roles in information transmission of the dynamical signal transduction. A major cause of interference with reliability is molecular noise, which degrades the quality of transmitted information. Despite the stochastic nature of cellular processes, organisms have acquired several mechanisms to resist or to take advantage of noise in order to enhance biological functionalities [22]. As these two principles are of significance, the dynamical transmission of information has evolved in such a way that it optimally satisfies these principles. We divide the dynamical signal
transduction into two parts: encoding of extracellular stimuli into intracellular dynamical patterns, and decoding of the dynamical patterns into the response (in the present manuscript, the response corresponds to the concentration of output molecular species) (figure 1(a)). By viewing the dynamical signal as an input, the decoding network as the system to be controlled, and the output concentration as an output (figure 1(a)), we use optimal control theory \[23, 24\] to determine the signal dynamics that optimize energy efficiency and the reliability of transmitted information. We quantify the energetic cost by free energy dissipation when generating temporal dynamical patterns and the uncertainty by variance of output molecular species concentration. To decode the dynamical signals, we adopt an activation–deactivation network (figure 1(b)), which is a motif commonly used in biochemical networks. We optimize the dynamical patterns, i.e. the input. Decoders (decoding networks) may also coevolve to maximally reading out information from signals, but in this paper, we focus on optimizing the input. No matter how precisely the decoder is able to read dynamical signals, it is impossible to totally eliminate the uncertainty due to the inherent stochasticity. Therefore, there is a lower bound on the uncertainty and the bound is determined by input.

From our calculations, we identify three basic patterns for the signals: steep (figure 2(a)), gradual (figure 2(b)), and overshooting (figure 2(c)). We show that the steep pattern minimizes the energy, whereas the gradual pattern minimizes the uncertainty. Intriguingly, when minimizing the energy of a dynamical pattern while achieving a higher output concentration, overshooting is the optimal pattern; this pattern can often be seen in real processes. We identify the conditions in which these three patterns (steep, gradual, and overshooting) confer advantages. We note that these patterns are prevalent in signal transduction, and our calculations show that minimizing the energetic cost and uncertainty plays important roles in the evolutionary design of dynamical signaling patterns.

2. Methods

2.1. Model

Dynamical signal transduction is typically separated into two parts [7]: encoding of extracellular stimuli into intracellular dynamical patterns, and decoding of the dynamical patterns into the response (figure 1(a)); our study focuses on the latter process, and dynamical signals are optimized for readout by a particular decoder, where the optimization criteria are to minimize the energy consumed by signal generation and to minimize the uncertainty of the readout.
In cells, dynamical signals are decoded by molecular networks. We consider a molecular network consisting of \( N \) molecular species \( X_1, X_2, \ldots, X_N \) and we define \( x_i \) as the concentration of \( X_i \). The input signal is carried by a molecular species \( U \), whose concentration \( u(t) \) follows a dynamic pattern, and its onset is \( t = 0 \). In our analysis, we use optimal control theory to optimize the temporal pattern of \( u(t) \). Let the \( i_{\text{out}} \)th molecular species \( X_{i_{\text{out}}} \) be the output of the network. The network reads the information from the input \( u(t) \) (i.e., intracellular dynamical signal) and outputs the result as the concentration of \( X_{i_{\text{out}}} \) at time \( T (T > 0) \) (figure 1(a)), i.e., \( x_{i_{\text{out}}} (T) \) carries information about the input signal.

Consider the evolutionary design of the dynamical signal \( u(t) \) that attains the desired concentration of an output molecule \( X_{i_{\text{out}}} \) at \( t = T \). Although there might be many possible dynamics for \( u(t) (0 \leq t \leq T) \) that result in the desired output concentration, the most biologically preferable ones are selected. We can expect that the signals with lower energetic cost will be selected. In addition, biochemical reactions are subject to noise, due to the smallness of the cells. The noise degrades the information, and hence transmission with lower uncertainty is desirable. Considering the energy of the input and the uncertainty of the concentration of the output molecule, we wish to find a signal \( u(t) \) that minimizes a performance index \( R \), defined as

\[
R = \gamma_{i_{\text{out}}}(T) + w\Pi,
\]

where \( \gamma_i(t) = \langle (x_i(t) - \mu_i(t))^2 \rangle \) is the variance of the concentration of the \( i \)th molecule at time \( t \) (\( \mu_i(t) = \langle x_i(t) \rangle \) is the mean) which quantifies the uncertainty, \( \Pi \) is the energetic cost of the signal \( u(t) \), and \( w \) is a weight parameter in the range \( 0 < w < \infty \), which represents the importance of the energy for the performance index. As denoted, the output \( X_{i_{\text{out}}} \) has the target concentration at time \( t = T \). Therefore, the mean concentration of the output \( X_{i_{\text{out}}} \), which we denote as \( \mu_{i_{\text{out}}} (t) \), must attain the predefined target concentration \( \mu_{i_{\text{out}}}^{\text{trg}} \) (“trg” is short for “target”) at time \( t = T \), i.e.,

\[
\mu_{i_{\text{out}}}(T) = \mu_{i_{\text{out}}}^{\text{trg}}.
\]

is a boundary condition.

### 2.2. Quantification of uncertainty

We quantify the uncertainty as the variance of the output molecular species concentration \( \gamma_{i_{\text{out}}}(T) \); the derivation is shown below. The dynamics of molecular networks, which decode \( u(t) \) and output the result, can be generally captured by the following rate equation:

\[
\dot{x}_i(t) = \sum_{\ell=1}^{N_i} s_{i\ell} \psi_{\ell}(x, u),
\]

where \( x = (x_1, x_2, \ldots, x_N) \), \( \{s_{i\ell}\} = S \) is a stoichiometry matrix, \( \psi_{\ell}(x, u) \) is the reaction velocity of the \( \ell \)th reaction, and \( N_i \) is the number of reactions. Due to the smallness of the cells, chemical reactions are subject to stochasticity. We describe the noisy dynamics by the Fokker–Planck equation (FPE) [25, 26]:

\[
\frac{\partial}{\partial t} P(x; t) = -\sum_{\ell} \frac{\partial}{\partial x_\ell} \sum_{\ell} s_{i\ell} \psi_{\ell}(x, u) P(x; t) + Q \sum_{i,j} \frac{\partial^2}{\partial x_i \partial x_j} \sum_{\ell} s_{i\ell} s_{j\ell} \psi_{\ell}(x, u) P(x; t),
\]

where \( P(x; t) \) is the probability density of \( x \) at time \( t \), and \( Q \) is the noise intensity related to the volume \( V \) via \( Q = (2V)^{-1} \). Optimal control theory and related variational methods have been employed by many researchers [27–31]. Although stochastic optimal control theory has been applied in biological contexts [31], it is difficult to apply it to multivariate models. Instead, we describe the dynamics by using the time evolution of moments derived from equation (3) [32, 33]. For general nonlinear models, the naive calculation of moment equations results in an infinite hierarchy of differential equations. Because our adopted models are linear with respect to \( x \) (see equations (15) and (16)), we can obtain closed differential equations (see the supplementary material). For moments of up to the second order (mean \( \mu_i(t) \), variance \( \gamma_i(t) \), and covariance \( \gamma_{ij}(t) = \langle (x_i - \mu_i)(x_j - \mu_j) \rangle \)), we have the following moment equation with respect to \( z = (\mu_1, \ldots, \mu_N, \gamma_1, \ldots, \gamma_N, \rho_{12}, \rho_{23}, \ldots, \rho_{1N}) \):

\[
\dot{z}_i(t) = h_i(z, u),
\]

where \( h_i(z, u) \) is right-hand side of the moment equation and the dimensionality of \( z \) is \( M = N(N + 3)/2 \). With the moment equation (4), we can reduce the stochastic optimal control problem to a deterministic one.

### 2.3. Quantification of energetic cost

We next define the energetic cost of a signal (\( \Pi \) in equation (1)) as the free energy dissipated by controlling the concentration \( u(t) \) such that it follows the desired temporal dynamics. We derive the energetic cost of the signal \( U \) with a simple biochemical model after [34–36]. We assume that input molecular species \( U \) is activated from \( U \) and undergoes the following reaction:
\[ \mathcal{U} := U. \]  

Note that the total concentration \( u^{tot} = u + \bar{u} \) does not change with time, where \( u \) and \( \bar{u} \) are the concentrations of \( U \) and \( \mathcal{U} \), respectively. In general, the deactivation reaction is not the reverse of activation. Therefore, when considering the energetic cost of generating \( U \), we have to handle activation and deactivation separately. Activation is typically mediated by phosphorylation, where \( \mathcal{U} \) is activated by a kinase through the transfer of inorganic phosphate (P) to the solution. These reactions are written as

\[ \mathcal{U} + \text{ATP} \xrightarrow{\beta_d} U + \text{ADP}, \quad U \xrightarrow{\beta_d} \mathcal{U} + P, \]

where \( \beta_a, \beta_d, \alpha_a, \) and \( \beta_d \) are reaction rates. When these parameters are held constant and the system is closed, the system relaxes to an equilibrium state. Let \( c_T, c_D, \) and \( c_P \) be the concentrations of ATP, ADP, and P, respectively. In the natural cellular environment where the system is open, \( c_T, c_D, \) and \( c_P \) can be regarded as constant, due to external agents [35], which we denote as \( c_T^0, c_D^0, \) and \( c_P^0 \), respectively. Therefore the system relaxes to a nonequilibrium steady state (NESS). The steady-state concentration of \( U \) is

\[ u^{ss} = \frac{u^{tot}(\beta_d + \alpha_a)}{\beta_a + \beta_d + \alpha_a + \alpha_d}, \]

where \( \beta_a = \beta_a^0, \alpha_a = \alpha_a^0, \) and \( \beta_d = \beta_d^0. \) At the steady state, the net flux (clockwise direction in figure 2(d)) is \( J^{ss} = u^{ss} (\alpha_a - \beta_d)/(\beta_a + \beta_d + \alpha_a + \alpha_d) \). The free energy dissipated during one cycle (i.e., \( U \rightarrow U \rightarrow \mathcal{U} \) in the clockwise direction in figure 2(d)) is \( \Delta \phi = k_B T \ln(\beta_a/\beta_d/\alpha_a) \), where \( k_B \) is the Boltzmann constant and \( T \) is the temperature (see the supplementary material). Therefore the instantaneous free-energy dissipation (i.e., power) is

\[ \mathcal{P} = -J^{ss} \Delta \phi = k_B T u^{ss} (\alpha_a - \beta_d)/(\beta_a + \beta_d + \alpha_a + \alpha_d) \ln(\alpha_a/\beta_d/\beta_d). \]

Equation (8) quantifies the cost of the activation-deactivation of \( U \). Next, in order to yield the dynamics of \( U \), we assume that kinase activity is controlled by an upstream molecular species and thus \( \alpha_a \) varies temporally. We assume that the relaxation of equation (6) is very fast (i.e., \( 1/(\beta_a + \beta_d + \alpha_a) \) is very small) so that the concentration \( u \) is well approximated by \( u^{ss} \) (equation (7)) even for the case of a time-varying \( \alpha_a \). From equation (7), \( \alpha_a \) can be represented as a function of \( u \):

\[ \alpha_a = \frac{U}{u^{tot} - u} (\beta_a + \beta_d) \]

From the condition \( \alpha_a > 0 \), the minimum of \( u \) is \( u^{min} = \beta_d u^{tot}/(\beta_a + \beta_d) \). For this dynamic case, the instantaneous free energy dissipation is given as a function of \( u \) which is equation (8) along with equation (9).

The molecular species upstream from \( U \) also consumes energy; however, signal transduction generally amplifies the external stimuli, and so the concentration of the upstream molecular species is less than the concentration of \( U \)[37, 38]. Furthermore, we assume that the concentration of the molecular species downstream from \( U \) is also less than that of \( U \)[37, 38]. Indeed, in the ERK pathway, the concentration of ERK is higher than that of its downstream molecular species (that is, the energetic cost of decoding \( U \) smaller than the cost of generating it).

Therefore, we assume that the energetic cost of generating \( U \) (i.e., equation (8)) dominates the overall energetic cost. The free energy dissipated during \( 0 < t < T \) is

\[ \Pi = \int_0^T \mathcal{P} \, dt. \]

Because it is difficult to use the exact representation of \( \Pi \) in the optimal control calculation, we approximate \( \mathcal{P} \) with a simpler expression: \( \mathcal{P} = 0 \) at a zero-flux point \( u = u^d = \beta_d u^{tot}/(\beta_a + \alpha_d) \) where \( J \) vanishes. Assuming that \( u^d \) is sufficiently small, \( \mathcal{P} = 0 \) when \( u \) is very low concentration. Furthermore, \( \mathcal{P} \) increases superlinearly as \( u \) increases from the zero-flux point \( u = u^d \). Taking into account the conditions and computational feasibility, we use the approximation \( \mathcal{P} \approx \bar{\mathcal{P}} \) with \( \bar{\mathcal{P}} = ku^2, \) where \( k \) is a proportionality coefficient. Then, the free energy dissipation during the period \( 0 \leq t \leq T \) is approximated by

\[ \bar{\Pi} = \int_0^T \bar{\mathcal{P}} \, dt = \int_0^T ku^2 \, dt. \]

Figure 3 compares the exact expression of \( \mathcal{P} \) (equation (8)) with its quadratic approximation \( \bar{\mathcal{P}} \) for two settings; the exact and quadratic results are shown by solid and dashed lines, respectively. For both parameter settings, we see that the behavior of the quadratic approximation is similar to that of the exact one. The major difference between the exact and the quadratic representations is that equation (8) diverges to \( \infty \) for \( u \rightarrow u^{min} \) and \( u \rightarrow u^{tot} \). Therefore, in order for the quadratic expression to well approximate the exact energetic cost, \( u^{tot} \) should satisfy requirements in addition to the condition of \( u^d \). If \( u^{tot} \) is too small, the energy divergence at \( u = u^{tot} \) prevents the signal to have higher peaks. On the other hand, if \( u^{tot} \) is
are Lagrange multipliers that force the constraints. In equation
\( \Pi \) (dashed line), (a) \( u^m = 20.0 \),
\( \alpha_d = 5.0 \), \( \beta_a = 20.0 \), and \( \beta_d = 0.1 \) \((q = 1.9 \text{ for quadratic})\); and (b) \( u^m = 20.0 \), \( \alpha_d = 10.0 \), \( \beta_a = 10.0 \), and \( \beta_d = 0.1 \) \((q = 4.4 \text{ for quadratic})\). We use \( k_u = 1 \) and \( T = 1 \), without loss of generality.

excessively large, the exact energetic cost becomes almost linear with respect to \( u \) which makes the quadratic approximation less reliable. With this approximation, we set \( u(t) = 0 \) for \( t < 0 \) and \( u(t) \geq 0 \) for \( t > 0 \) (we define that the onset is the time when \( u(t) \) becomes positive).

2.4. Finding the optimal signaling pattern

We wish to obtain the optimal control \( u(t) \) that minimizes \( R \) of equation (1) while satisfying equation (4) and the predefined target mean concentration of equation (2) \((\mu_{\text{tar}}(T) = z_{\text{tar}}(T) = \mu_{\text{tar}}^g)\). Then, by virtue of optimal control theory \([23, 24]\), we minimize the following augmented performance index:

\[
\bar{R} = \gamma_{\text{tar}}(T) + w\bar{\Pi} + \nu(z_{\text{tar}}(T) - \mu_{\text{tar}}^g) + \sum_{i=1}^{M} \int_{0}^{T} \lambda_i(h_i(z, u) - z_i)\,dt,
\]

where \( \nu \) and \( \lambda = (\lambda_1, \lambda_2, \ldots, \lambda_M) \) are Lagrange multipliers that force the constraints. In equation (12), the exact energetic cost \( \Pi \) in \( R \) is replaced by its quadratic approximation \( \bar{\Pi} \) (we set \( q = 1 \), because the scaling of \( q \) is offset by \( w \)). Using the calculus of variations \([23, 24]\), finding an optimal signal \( u(t) \) is reduced to solving the differential equations given by equation (4) and

\[
\dot{\lambda}_i(t) = -\frac{\partial}{\partial z_i} H(z, u, \lambda),
\]

\[
0 = \frac{\partial}{\partial u} H(z, u, \lambda),
\]

where \( H(z, u, \lambda) \) is the Hamiltonian \([23, 24]\):

\[
H(z, u, \lambda) = w\bar{\Pi} + \sum_{i=1}^{M} \lambda_i h_i(z, u).
\]

We assume vanishing initial values for all moments: \( \mu(0) = 0 \), \( \gamma_i(0) = 0 \), and \( \rho_{ji}(0) = 0 \) (i.e., \( z_i(0) = 0 \) for all \( i \)). For the boundary conditions, \( \lambda_{\text{in} + \text{tar}}(T) = 1 \) and \( \lambda_i(T) = 0 \) \((i = \text{in}, \text{out}, N + \text{in} \text{out})\) are required from the optimal control theory, and \( z_{\text{tar}}(T) = \mu_{\text{tar}}^g \) for the final value of \( z \). There are boundary conditions at both \( t = 0 \) and \( t = T \); this two-point boundary value problem can be solved numerically by using general solvers (see the supplementary material).

3. Results

We consider the following activation–deactivation decoding motif (figure 1(b)): an inactive molecule \( X_1 \) is activated to become \( X_2 \), where the activation is dependent on the input molecule \( U \). The rate equation is

\[
\dot{X}_1 = \alpha_1 u(X_{\text{tot}} - X_1) - \beta_1 X_1,
\]

where \( X_{\text{tot}} \) is the total concentration \( X_{\text{tot}} = X_1 + X_2 \) which does not change with time \((X_0 \text{ is the concentration of } X_0)\), and \( \alpha_1 \) and \( \beta_1 \) are activation and deactivation rates, respectively. The activated molecule \( X_1 \) activates an output molecule \( X_2 \) (e.g., an activated transcription factor), and hence \( X_2 \),...
reports the result (i.e., \( i_{\text{out}} = 2 \)). The rate equation is \( \dot{x}_2(t) = \alpha_2 x_1(t) x_3(t) - \beta_2 x_2(t) \), where \( x_1(t) \) is the concentration of \( X_1 \). When \( x_1(t) \) is far from its saturation concentration, activation of \( X_1 \) is approximately the first-order reaction with respect to \( X_1 \), i.e., \( \dot{x}_1(t) = \alpha_1 x_1(t) - \beta_1 x_1(t) \). Then dynamics of \( X_1 \) and \( X_2 \) is represented by the following linear differential equation:

\[
\dot{x}_1(t) = \alpha_1 x_1(t) - \beta_1 x_1(t),
\]

(15)

\[
\dot{x}_2(t) = \alpha_2 x_1(t) - \beta_2 x_2(t).
\]

(16)

Because of the linearity of equations (15) and (16), the moment equation can be obtained without the truncation approximation. As denoted above, the input signal \( u(t) \) must produce the dynamics that satisfy the constraint that the target mean concentration of \( X_2 \) at time \( T \) is \( u_{\text{avg}} (i_{\text{out}} = 2 \) in equation (2)). This type of decoding motif is prevalent and can be found in various biochemical systems [13, 39–41]. We call this a two-stage model.

By incorporating intrinsic noise due to a small number of molecules, we have a corresponding FPE from equation (3) (see the supplementary material). We then calculate the moment equation from the FPE.

### 3.1. Steep pattern minimizes energetic cost, and gradual pattern minimizes uncertainty

Using the two-stage model, we calculated the optimal signal \( u(t) \). Figures 4(a)–(c) shows the optimal signal \( u(t) \) as a function of \( t \) and \( \log_{10}(w) \) for three cases: (a) \( \mu_{\text{avg}} = 1.0 \) and \( x_{t_{\text{tot}}} = 1.0 \), (b) \( \mu_{\text{avg}} = 0.2 \) and \( x_{t_{\text{tot}}} = 1.0 \), and (c) \( \mu_{\text{avg}} = 1.0 \) and \( x_{t_{\text{tot}}} = 3.0 \). Note that \( \mu_{\text{avg}} = 1.0 \) is a near-saturation value with \( x_{t_{\text{tot}}} = 1.0 \), i.e., it is close to the maximal reachable target concentration. The other parameters are shown in the caption of figure 4. The minimum value for \( w \) is determined such that \( u(t) \) satisfies \( u(t) \geq 0 \), and the maximum is \( 10^2 \) times the minimum. In all the three cases (figures 4(a)–(c)), for larger values of \( w \), we see that the optimal signals steeply increase at \( t = 0 \) and gradually decay as time elapses. As \( w \) decreases, the optimal pattern varies from steeper to more-gradual patterns. Comparing figures 4(a) and (b), we can see the effect of different target concentrations. The target concentration for figure 4(b) is lower than it is for (a) (the other parameters are the same). For figure 4(a), the decay right after \( t = 0 \) is especially rapid, and this is followed by a plateau state \((t = 0.2–0.8)\); this is a typical overshooting pattern, similar to that shown in figure 2(c). However, the optimal signal of figure 4(b) does not exhibit overshooting. We next compared figures 4(a) and (c) when all parameters are identical except for \( x_{t_{\text{tot}}} \) (for figure 4(c) is larger than it is for (a)) to see how the overshooting pattern depends on the total concentration \( x_{t_{\text{tot}}} \). When the total concentration \( x_{t_{\text{tot}}} \) is larger, the optimal signal does not overshoot. From these results, we see that the steep pattern minimizes the energy, whereas the gradual pattern minimizes the variance. Along with the condition of the steep pattern, the overshooting pattern emerges when the target concentration \( \mu_{\text{avg}} \) is relatively high and the total concentration \( x_{t_{\text{tot}}} \) is relatively low.

In figure 5(a), we compared the optimal signal (solid line), which exhibits the overshooting pattern, with a constant signal (dashed line), for the period starting at \( t = 0 \) and ending at \( t = T \); both patterns attain the same

---

1 Note that linearity is not a prerequisite for applying the moment method. For nonlinear cases, closed moment equations can be obtained by truncating higher order moments than the second.

2 When \( w \) is below the minimal values, the signal determined by the optimal control approach violates the positivity condition at an earlier time. For such values, the optimal pattern is a very low concentration at an earlier time and a peak concentration at a later time.
Because the exact energetic cost diverges to infinity, we normalized the signal as follows:

\[ \text{Normalized signal} = \frac{\text{signal}}{\text{minimum signal}} \]

We also calculated the ratio for the exact energy and uncertainty to those at the minimum weight \( w_{\text{min}} \). We set \( w = 1.0 \) for the optimal signal and \( w = 0.2 \) for the constant signal. In (a)-(d), the unspecified parameters are the same as those in figure 4(a).

3.2. Slow relaxation of the decoder causes overshooting

We considered the effects of time scale of the decoder on overshooting of the optimal signal. When increasing (decreasing) \( \beta_1 \) or \( \beta_2 \), the relaxation time of the decoder becomes shorter (longer). We first varied \( \beta_2 \) while keeping other parameters unchanged, except for \( \mu_{trg}^2 \) (the other parameters are the same as those in figure 3(a)). Since we are interested in the shape rather than the magnitude, we normalized the signal as follows:

\[ \text{Normalized signal} = \frac{\text{signal}}{\text{minimum signal}} \]

Because the exact energetic cost diverges to infinity, we truncated at \( u^\text{tr} \) (the concentration lower than \( u^\text{tr} \) is identified as \( u^\text{tr} \)).

Figure 5. Results of the two-stage model. (a) Comparison of the optimal signal (solid line) and a constant signal (dashed line), which attain the same target concentration \( \mu_{trg}^2 = 1.0 \) (\( w = 1.0 \) for the optimal signal). (b) Normalized signal \( \tilde{u}(t) \) for three values of \( \beta_2 \): \( \beta_1 = 0.2 \) and \( \mu_{trg}^2 = 1.6 \) (solid line), \( \beta_1 = 1.0 \) and \( \mu_{trg}^2 = 1.0 \) (dashed line), and \( \beta_1 = 5.0 \) and \( \mu_{trg}^2 = 0.3 \) (dotted line). We set \( w = 1.0 \). (c) Energy and uncertainty dependence on \( w \) for two cases: \( \mu_{trg}^2 = 0.2 \), \( \alpha_1 = 2.0 \), \( \alpha_2 = 0.5 \), \( \beta_1 = 1.0 \), and \( \beta_2 = 1.0 \) (parameters are the same as those in figure 4(b); solid (energy) and dotted (uncertainty) lines) and \( \mu_{trg}^2 = 0.5 \), \( \alpha_1 = 0.5 \), \( \alpha_2 = 10.0 \), \( \beta_1 = 0.1 \), and \( \beta_2 = 0.1 \) (dashed (energy) and dot-dashed (uncertainty) lines). Ranges of \( w \) are \( 1.2 \times 10^{-3} \leq w \leq 1.2 \times 10^{-1} \) for the former and \( 0.03 \leq w \leq 3.0 \) for the latter. We plot the ratio of energy and uncertainty to those at the minimum weight \( w_{\text{min}} \). Variation of the normalized signal \( \tilde{u}(t) \) in response to the target concentration \( \mu_{trg}^2 \) for \( w = 1.2 \times 10^{-3} \); \( \mu_{trg}^2 = 0.2 \) (dotted line), 0.5 (dashed line), and 1.0 (solid line). In (a)-(d), the unspecified parameters are the same as those in figure 4(a).
\[ \ddot{u}(t) = u(t)/\sqrt{\ddot{\Pi}}, \] which guarantees the unit energy \( \ddot{\Pi} = 1 \). Figure 5(b) shows the normalized signal \( \ddot{u}(t) \) for three cases: \( \beta_1 = 0.2 \) (solid line), \( \beta_2 = 1.0 \) (dashed line), and \( \beta_2 = 5.0 \) (dotted line). As shown above, the overshooting pattern emerges when \( \mu_{\text{PS}} \) is close to the saturation value. Thus, for each \( \beta_2 \) value, \( \mu_{\text{PS}} \) is set to the nearest reachable value (parameter details are shown in the caption of figure 5(b)). When \( \beta_2 = 5.0 \), the optimal signal does not overshoot, but the other two settings do. From this result, when the relaxation time of \( X_2 \) is sufficiently shorter than \( T \), the steep (overshooting) pattern does not minimize energy consumption. The same calculation was performed for \( \beta_1 \) and we found that all of the optimal signals exhibit the overshooting (see the supplementary material). This implies that the relaxation time of \( X_1 \) is not responsible for the overshooting pattern.

### 3.3. Tradeoff between energy and uncertainty

In figure 5(c), we next evaluated the dependence on the weight \( w \) of the energy \( \ddot{\Pi} \) and the uncertainty \( \gamma_{\text{unc}}(T) \) for two parameter settings. Parameters of the first setting is the same as those in figure 4(b) whose results are plotted by solid (energy) and dotted (uncertainty) lines. The second setting highlights the uncertainty variation when we change \( w \) and these results are shown by dashed (energy) and dot-dashed (uncertainty) lines (parameter details are described in the caption of figure 5(c)). In figure 5(c), we plot the ratio of the energy and uncertainty to those at the minimum weight \( w_{\text{min}} \), where \( w_{\text{min}} \) is the minimum of \( w \) for each parameter setting. For the both settings, as \( w \) increases, the uncertainty increases, and the energy decreases; there is thus a tradeoff between the energy and the uncertainty, since they cannot be minimized simultaneously. For cellular inference, the tradeoff between uncertainty and energy consumption has been confirmed by several studies [42–45]. Similarly, in biochemical clocks, it has been shown that there is a tradeoff between temporal accuracy and energy consumption [46]. These studies [43–46] calculated the entropy production, which is the energy required for maintaining a system at NESS. There is also a tradeoff between the energy cost and information coding in neural systems ([20] and references therein). We have shown that a similar relation also holds for dynamical signals.

### 3.4. Calculation with simplified model

We identified that the steep (overshooting) pattern minimizes the energy. Let us explain this mechanism with a simplified two-stage model:

\[ \dot{x}_1(t) = u(t), \quad \dot{x}_2(t) = \alpha_2 x_1 - \beta_2 x_2, \quad \] \hspace{1cm} (17)

along with a delta-function stimulus, \( u(t) = \delta(t - t_\text{s}) \) (\( t_\text{s} \) is the time of the stimulus, where \( 0 < t_\text{s} < T \)). From the equation (17), the final output concentration is

\[ x_2(T) = \frac{\alpha_2}{\beta_2} \left[ 1 - e^{-\beta_2 (t - t_\text{s})} \right]. \quad \] \hspace{1cm} (18)

When the relaxation of \( X_2 \) is very slow (\( \beta_2^{-1} \gg T \)), the output concentration is \( x_2(T) \approx \alpha_2 (T - t_\text{s}) \), which shows that the signal at an earlier time has a greater effect than at a later time; the steep pattern allows the output concentration to reach the target concentration at a lower cost. For the fast relaxation case (\( \beta_2^{-1} \ll T \)), we obtain \( x_2(T) \approx \alpha_2 \beta_2^{-1} \), which shows that \( x_2(T) \) does not depend on \( t_\text{s} \); it is not advantageous for the signal to have a peak at an earlier time. Therefore, when the relaxation time of the decoder is very fast, the steep pattern does not confer advantages for minimizing the energy consumption; this agrees with the optimal control calculation shown in figure 5(b). Next, suppose there is only a single stage required to decode the signal \( \ddot{x}_1(t) = u(t) \), where \( x_1(T) \) is the output concentration (a single-stage model). In this case, the final output concentration is \( x_1(T) \approx 1 \) and does not depend on \( t_\text{s} \), implying that the steep patterns do not minimize the energy. On the other hand, when there is a third stage \( x_3 = \alpha_3 x_2 - \beta_3 x_3 \) in addition to equations (17) (a three-stage model), we find that the final concentration \( x_3(T) \) depends on \( t_\text{s} \) unless \( \beta_3^{-1} \ll T \) (for \( \beta_2^{-1}, \beta_3^{-1} \gg T \), we have \( x_3(T) \approx \alpha_2 \alpha_3 (T - t_\text{s})^2/2 \), and for \( \beta_2^{-1}, \beta_3^{-1} \ll T \), \( x_3(T) = \alpha_2 \alpha_3 / (\beta_2 \beta_3) \)). Thus, the steep patterns generally minimize the energy consumption when the relaxation of the decoder is slow and there are more than two stages in the decoding.

The effect of the gradual pattern can be accounted for by the moment equation. The variance \( \gamma_2 \) and covariance \( \rho_{12} \) are governed by (see moment equations in the supplementary material)

\[ \dot{\gamma}_2(t) = 2(\alpha_2 \rho_{12} - \beta_2 \gamma_2) + 2Q(\alpha_2 \mu_1 + \beta_2 \mu_2), \quad \] \hspace{1cm} (19)

\[ \dot{\rho}_{12}(t) = \alpha_2 \gamma_1 - \rho_{12}(\alpha_1 u + \beta_1 + \beta_2). \quad \] \hspace{1cm} (20)

We find that the main reason for the difference between the variance \( \gamma_2(T) \) of the steep and gradual patterns is the area \( \int_0^T \rho_{12}(t) \, dt \); namely, a smaller area yields a smaller value for \( \gamma_2(T) \). From equation (20), because the decay velocity of \( \rho_{12}(t) \) depends on \( u \), a higher concentration of \( u \) at a later time results in a smaller value for \( \rho_{12}(t) \), which corresponds to the gradual pattern of \( u(t) \).
4. Discussion and conclusion

Our result provides insights into experimentally observed dynamical patterns. Reference [8] reported the dynamical pattern of ERK activity in response to different strengths of extracellular stimulus (i.e., the ligand concentration); the pattern is steep when stimulated by a strong stimulus, and it is gradual when stimulated by a weak one. This experimental observation can be accounted for by our model. We show that the optimal signal is steep for larger values of \( \mu_n^{\text{trg}} \) and gradual for smaller values (figure 5(d)). It is expected that the strong and weak ligand stimuli result in strong and weak responses, respectively, i.e., higher and lower output concentrations. Therefore, the ERK activity induced by the strong ligand stimulus may be related to \( \mu_n^{\text{trg}} = 1.0 \), and that induced by the weak one is related to \( \mu_n^{\text{trg}} = 0.2 \). When the target concentration is higher (i.e., \( \mu_n^{\text{trg}} = 1.0 \)), the magnitude of the signal is larger, and hence the effect of the energy of the signal on the objective function \( R \) (equation (1)) is greater than that of the variance \( \gamma_2(T) \). In contrast, for the smaller values of \( \mu_n^{\text{trg}} \) (i.e., \( \mu_n^{\text{trg}} = 0.2 \)), the variance \( \gamma_2(T) \) becomes the leading term because the energy of the signal is smaller. Therefore, the steep pattern is preferable when the target concentration is higher, while the gradual one is preferable when the target concentration is lower. These theoretical results qualitatively agree with the observed dynamical patterns reported in [8].

Along with conditions for the steep pattern, the overshooting dynamics minimize the energy of the input signals when the total concentration \( x_1^{104} \) is smaller and the target concentration \( \mu_n^{\text{trg}} \) is higher. Surprisingly, this behavior can be found in several dynamical patterns; for example, activities of the ERK, the IκB kinase, which regulates the transcription factor NF-κB, and the kinase AKT show this behavior [4, 39, 47–49]. These examples indicate that the pattern has biological advantages. We also note the biochemical origin of the overshoot. For example, simple incoherent feed-forward loops [1, 2, 50, 51] and activation–deactivation motifs [52] can generate such a pattern, and these motifs can indeed be found in signaling pathways. Furthermore, a strongly damped oscillation is indistinguishable from overshooting. Although NF-κB is known to exhibit damped oscillation upon stimulation, some studies [53, 54] are skeptical about the functional role of the NF-κB oscillation; that is, the NF-κB oscillation may be a by-product of inducing overshooting. Overshooting has often been observed in actual dynamical patterns, but its functional advantage has not been well understood. We have shown that this pattern produces direct benefits.

Acknowledgments

This work was supported by KAKENHI Grant No. 16K00325 from the Ministry of Education, Culture, Sports, Science and Technology.

References

[1] Alon U 2007 Nat. Rev. Genet. 8 450–61
[2] Alon U 2007 An Introduction to Systems Biology (Boca Raton, FL: CRC Press)
[3] Behar M and Hoffmann A 2010 Curr. Opin. Genet. Dev. 20 684–93
[4] Kobuta H, Noguchi R, Toyoshima Y, Ozaki Y I, Uda S, Watanabe K, Ogawa W and Kuroda S 2012 Mol. Cell 46 820–32
[5] Purvis I E, Karhoo K W, Mock C, Batchelor E, Loewer A and Lahav G 2012 Science 336 1440–4
[6] Purvis I E and Lahav G 2013 Cell 152 945–56
[7] Sonnen K F and Aulehla A 2014 Sem. Cell. Dev. Biol. 34 91–8
[8] Selimkhanov J, Taylor B, Yao J, Pilko A, Albeck J, Hoffmann A, Tsimring L and Wollman R 2014 Science 346 1370–3
[9] Lin Y, Sohn C H, Dalal C K, Cai L and Elowitz M B 2015 Nature 527 54–8
[10] Tostevin F and ten Wolde P R 2009 Phys. Rev. Lett. 102 218101
[11] Mora T and Wingreen NS 2010 Phys. Rev. Lett. 104 248101
[12] Mugler A, Walczak A M and Wiggins C H 2010 Phys. Rev. Lett. 105 058101
[13] Hansen A S and O’Shea E K 2013 Mol. Syst. Biol. 9 704
[14] Kobayashi T I 2010 Phys. Rev. Lett. 104 238104
[15] Mc Mahon S S, Lenive O, Filippi S and Stumpf M P H 2015 Science 346 1370–3
[16] Becker N B, Mugler A and ten Wolde P R 2015 Phys. Rev. Lett. 115 258103
[17] Makadia H K, Schwabe J S and Vagipalli R 2013 PLoS Comput. Biol. 11 e1004563
[18] Tostevin F, de Ronde W and ten Wolde P R 2012 Phys. Rev. Lett. 108 108104
[19] Laughlin S B 2001 Curr. Opin. Neurobiol. 11 475–80
[20] Sengupta B and Stemmler M B 2014 Proc. IEEE 102 738–50
[21] Mehta P, Lang A H and Schwab D J 2016 J. Stat. Phys. 162 1153–66
[22] McDonnell M D, Stocks N G, Pearce C E M and Abbott D 2008 Stochastic Resonance (Cambridge: Cambridge University Press)
[23] Kamien M I and Schwartz N L 2012 Dynamic Optimization: The Calculus of Variations and Optimal Control in Economics and Management (New York: Dover)
[24] Hull D G 2013 Optimal Control Theory for Applications (Berlin: Springer)
[25] Gillespie D T 2000 J. Chem. Phys. 113 297–306
[26] Klipp E, Liebermeister W, Wierling C, Kowald A, Lehrach H and Herwig R 2009 Systems Biology (Weinheim: Wiley)
[27] Forger D B and Paydarfar D 2004 J. Theor. Biol. 230 521–32
[28] Moehlis J, Shea-Brown E and Rabitz H 2006 J. Comput. Nonlinear Dyn. 1 358–67
[29] Hasegawa Y and Arita M 2014 J. R. Soc. Interface 11 20131018
[30] Hasegawa Y and Arita M 2014 Phys. Rev. Lett. 113 108101
[31] Iolov A, Ditlevsen S and Longtin A 2014 J. Neural Eng. 11 046004
[32] Rodriguez R and Tuckwell H C 1996 Phys. Rev. E 54 5855–90
[33] Tuckwell H C and Jost J 2009 Physica A 388 4115–23
[34] Qian H and Beard D A 2005 Biophys. Chem. 114 213–20
[35] Qian H 2007 Annu. Rev. Phys. Chem. 58 113–42
[36] Beard D A and Qian H 2008 Chemical Biophysics: Quantitative Analysis of Cellular Systems (Cambridge: Cambridge University Press)
[37] Aksan Kurnaz I 2004 Biotechnol. Bioeng. 88 890–900
[38] Thomson T M et al 2011 Proc. Natl Acad. Sci. USA 108 20265–70
[39] Sasagawa S, Ozaki Y i, Fujita K and Kuroda S 2005 Nat. Cell. Biol. 7 365–73
[40] Salazar C, Polit A Z and Höfer T 2008 Biophys. J. 94 1203–15
[41] Tănase-Nicola S, Warren P B and ten Wolde P R 2006 Phys. Rev. Lett. 97 068102
[42] Tu Y 2008 Proc. Natl Acad. Sci. USA 105 11737–41
[43] Mehta P and Schwab D J 2012 Proc. Natl Acad. Sci. USA 109 17978–82
[44] Lang A H, Fischer C K, Mora T and Mehta P 2014 Phys. Rev. Lett. 113 148103
[45] Barato A C, Hartich D and Seifert U 2014 New J. Phys. 16 103024
[46] Cao Y, Wang H, Ouyang Q and Tu Y 2015 Nat. Phys. 11 772–8
[47] Werner S L, Barken D and Hoffmann A 2005 Science 309 1857–61
[48] Werner S L, Kearns J D, Zadorozhnaya V, Lynch C, O’Dea E, Boldin M P, Ma A, Baltimore D and Hoffmann A 2008 Genes Dev. 22 2093–101
[49] Mathew S, Sundararaj S, Mamiya H and Banerjee I 2014 Bioinformatics 30 2334–42
[50] Mangan S and Alon U 2003 Proc. Natl Acad. Sci. USA 100 11980–5
[51] Mangan S, Itzkovitz S, Zaslaver A and Alon U 2006 J. Mol. Biol. 356 1073–81
[52] Behar M and Hoffmann A 2013 Biophys. J. 105 231–41
[53] Barken D, Wang C J, Kearns J, Cheong R, Hoffmann A and Levchenko A 2005 Science 308 52
[54] Cheong R, Hoffmann A and Levchenko A 2008 Mol. Syst. Biol. 4 192