THE JOURNAL OF PHYSICAL CHEMISTRY B
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCB

Article

Cellular Signaling beyond the Wiener–Kolmogorov Limit

Published as part of The Journal of Physical Chemistry virtual special issue "Dave Thirumalai Festschrift".

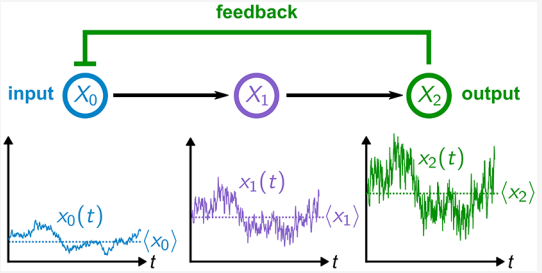
Casey Weisenberger,[§] David Hathcock,[§] and Michael Hinczewski*

Cite This: *J. Phys. Chem. B* 2021, 125, 12698–12711

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Accurate propagation of signals through stochastic biochemical networks involves significant expenditure of cellular resources. The same is true for regulatory mechanisms that suppress fluctuations in biomolecular populations. Wiener–Kolmogorov (WK) optimal noise filter theory, originally developed for engineering problems, has recently emerged as a valuable tool to estimate the maximum performance achievable in such biological systems for a given metabolic cost. However, WK theory has one assumption that potentially limits its applicability: it relies on a linear, continuum description of the reaction dynamics. Despite this, up to now no explicit test of the theory in nonlinear signaling systems with discrete molecular populations has ever seen performance beyond the WK bound. Here we report the first direct evidence of the bound being broken. To accomplish this, we develop a theoretical framework for multilevel signaling cascades, including the possibility of feedback interactions between input and output. In the absence of feedback, we introduce an analytical approach that allows us to calculate exact moments of the stationary distribution for a nonlinear system. With feedback, we rely on numerical solutions of the system's master equation. The results show WK violations in two common network motifs: a two-level signaling cascade and a negative feedback loop. However, the magnitude of the violation is biologically negligible, particularly in the parameter regime where signaling is most effective. The results demonstrate that while WK theory does not provide strict bounds, its predictions for performance limits are excellent approximations, even for nonlinear systems.



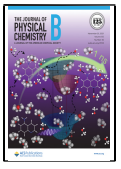
INTRODUCTION

Fundamental mathematical limits on the behavior of biochemical reaction networks[1−6] provide fascinating insights into the design space of living systems. Though these limits remain notoriously permeable compared to their analogues in physics (subject to reinterpretaton and exceptions as additional biological complexities are discovered), they still give a rough guide to what is achievable by natural selection for a given set of resources. They also raise other interesting issues:[7,8] Is selection actually strong enough to push a particular system toward optimality? When is performance sacrificed due to metabolic costs or the randomizing forces of genetic drift?

Information processing in cellular networks has been a particularly fertile ground for discussing optimality. Certain cellular processes like environmental sensing rely on accurate information transfer through intrinsically stochastic networks of reactions.[9,10] Other processes in development and regulation depend on suppressing noise through homeostatic mechanisms like negative feedback.[11−14] Either scenario, whether maintaining a certain signal fidelity or suppressing fluctuations, can be quite expensive in terms of metabolic resources,[3,15] and hence potentially can be an area where optimization is relevant.

Discussions of signaling performance limits are often framed in terms of information theory concepts like channel capacity[16,17] and complemented by direct experimental estimates.[18−25] In recent years, another tool has emerged for understanding constraints on biological signal propagation: optimal noise filter theory.[15,26−30] This draws on the classic work of Wiener and Kolmogorov (WK) in engineered communications systems.[31−33] The theory maps the behavior of a biological network onto three basic components: a signal time series, noise corrupting the signal, and a filter mechanism to remove the noise. Once the identification is made, the payoff is substantial: One can use the WK solution for the optimal noise filter function to derive the closed forms of analytical bounds on measures of signal fidelity (like mutual information) or noise suppression (like Fano factors). These bounds depend on the network's reaction rate parameters, allowing us to determine a minimum energetic price associated with a certain level of performance.[15] Finally, the theory specifies the conditions under which optimality can be realized in a particular network.

Received: September 7, 2021
Revised: October 25, 2021
Published: November 10, 2021



https://doi.org/10.1021/acs.jpcb.1c07894
J. Phys. Chem. B 2021, 125, 12698–12711

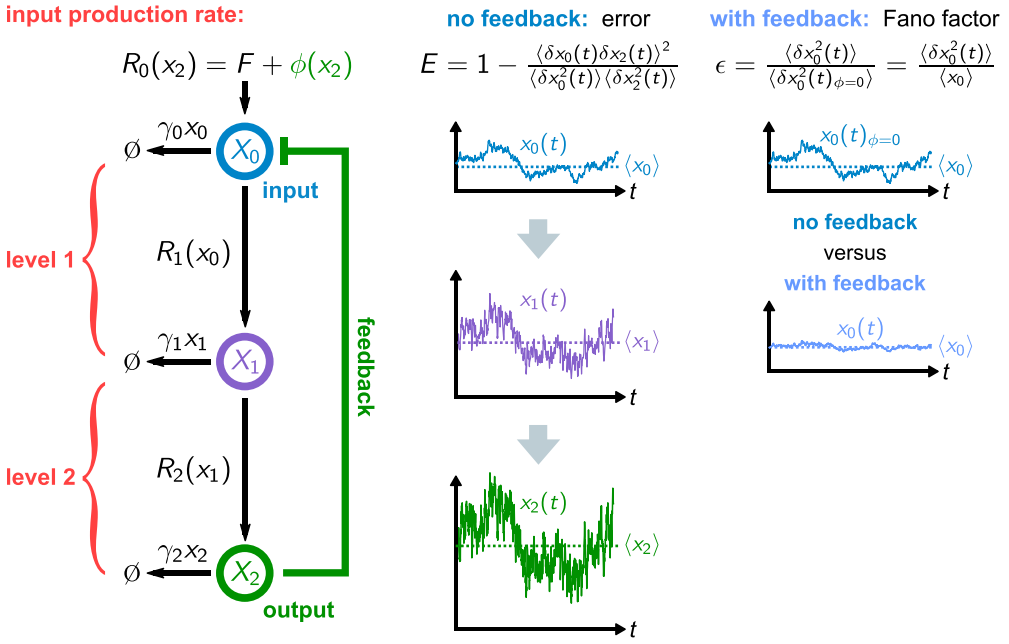


Figure 1. Overview of the N-level signaling cascade model, showing an example with $N = 2$. The signal from input species X_0 is propagated through to output species X_N, with the possibility of feedback back to the input. In the absence of feedback, the signal fidelity is measured via the error E, defined in terms of correlations between the input fluctuations $\delta x_0(t) = x_0(t) - \langle x_0 \rangle$ and output fluctuations $\delta x_N(t) = x_N(t) - \langle x_N \rangle$. In the linearized system, the error is related to the input−output mutual information $\mathcal{I}$ through $\mathcal{I} = -\frac{1}{2}\log_2 E$. For the system with feedback, the quantity we focus on is ϵ, the ratio of input variance $\langle \delta x_0^2(t) \rangle$ with feedback to the variance $\langle \delta x_0^2(t)_{\phi=0} \rangle$ without. This is also equal to the Fano factor $\langle \delta x_0^2(t) \rangle / \langle x_0 \rangle$, which measures the effectiveness of feedback in suppressing input fluctuations.

To date, however, there has been one major caveat: The WK theory relies on a continuum description of the molecular populations in the network and assumes all reaction rates are linearly dependent on the differences of these population numbers from their mean values. While this may be a good approximation in certain cases (i.e., large populations, with small fluctuations relative to the mean), it certainly raises doubts about the universal validity of the bounds derived from the theory. Biology is rife with nonlinearities, for example, so-called ultrasensitive, switch-like rate functions[34] in signaling cascades. Could these nonlinear effects allow a system to substantially outperform a WK bound derived using linear assumptions? Curiously, every earlier attempt to answer this question for specific systems[26,28] (summarized below) has yielded the same answer: The WK bound seemed to hold rigorously even when nonlinearities and discrete populations were taken into account.

The current work shows that this is not the full story. We have found for the first time two biological examples that can be explicitly proven to violate their WK bounds: a two-level signaling cascade and a negative feedback loop. To demonstrate this, we start by describing a general theoretical framework for signaling cascades with arbitrary numbers of intermediate species (levels), with the possibility of feedback interactions between the input and output species. We show how to calculate WK bounds based on the linearized, continuum version of this system, generalizing earlier WK results for single-level systems. In order to check the validity of the WK bound, we introduce an analytical approach for calculating exact moments of the discrete stationary probability distribution of molecular populations, starting from the underlying master equation. Our method works for arbitrarily long cascades in the absence of feedback. It allows us to find cases in a nonlinear two-level signaling cascade where the WK bound holds, as well as cases where it is violated. A similar picture emerges in a nonlinear single-level system with negative feedback, but here we use an alternative numerical approach to tackle the master equation. Remarkably, for the cases where nonlinearity helps beat the WK bound, the magnitude of the violation is tiny, typically fractions of a percent. We observe a trend that as the signaling efficiency increases, improving the biological function of the system, the size of the violation decreases or vanishes. This makes the WK value an excellent estimate for the actual performance limit in the biologically relevant parameter regime. Thus, while the results show that the WK theory does not rigorously bound the behavior of nonlinear signaling systems, they also put the theory on a more solid foundation for practical applications.

RESULTS AND DISCUSSION

Signaling Network. We begin by defining a general model of an N-level cellular signaling cascade. Each specific system we consider in our analysis will be a special case of this model. As shown schematically in Figure 1, we have an input chemical species X_0 followed by N downstream species X_1, ..., X_N. For example, if this was a model of a mitogen-activated kinase (MAPK) cascade,[35] the input X_0 would be an activated kinase, which activates another kinase via phosphorylation (X_1), which in turn leads to a sequence of downstream activations until we reach the final activated kinase X_N. The copy number of species X_i is denoted by $x_i = 0, 1, 2,...$ Hence, the state of the system can be represented by the vector $\mathbf{x} = (x_0, x_1, ..., x_N)$. Stochastic transitions between states are governed by an infinite-dimensional Markovian transition rate matrix $\mathbf{W}$. The element $W_{\mathbf{x}',\mathbf{x}}$ of this matrix represents the probability per unit time to observe state $\mathbf{x}'$ at the next infinitesimal time step, given that

https://doi.org/10.1021/acs.jpcb.1c07894
J. Phys. Chem. B 2021, 125, 12698−12711

the current state is $\mathbf{x}$. The values of these elements will depend on the rates of the chemical reactions that are possible in our signaling network, as described below. The probability $p_{\mathbf{x}}(t)$ of being in state $\mathbf{x}$ at time t evolves according to the corresponding master equation:[36]

$$\frac{\mathrm{d}}{\mathrm{d}t}p_{\mathbf{x}}(t) = \sum_{\mathbf{x}'} [W_{\mathbf{x},\mathbf{x}'}p_{\mathbf{x}'}(t) - W_{\mathbf{x}',\mathbf{x}}p_{\mathbf{x}}(t)] \tag{1}$$

The first term on the right represents the gain of probability in state $\mathbf{x}$ due to transitions out of all other states $\mathbf{x}'$ into $\mathbf{x}$, and the second term represents the loss due to transitions out of $\mathbf{x}$ into all other states. We will focus on systems where W is time-independent and the system reaches a unique stationary distribution $\mathcal{P}_{\mathbf{x}}$. The latter satisfies eq 1 with the left-hand side set to zero:

$$0 = \sum_{\mathbf{x}'} [W_{\mathbf{x},\mathbf{x}'}\mathcal{P}_{\mathbf{x}'} - W_{\mathbf{x}',\mathbf{x}}\mathcal{P}_{\mathbf{x}}] \tag{2}$$

All physical observables we consider can be expressed as averages over this stationary distribution. If $f(\mathbf{x})$ is some function of state $\mathbf{x}$, we will use $\langle f\rangle \equiv \sum_{\mathbf{x}} f(\mathbf{x})\mathcal{P}_{\mathbf{x}}$ to denote the associated stationary average.

The detailed form of eq 2 for our cascade requires specifying all the possible chemical reactions in our network. We start with species X_0, which is produced with some rate $R_0(x_N) \geq 0$. We treat "production" as occurring with a single effective rate, encompassing all the substeps involved in activation of X_0 from some inactive form (not explicitly included in the model). The functional form of the rate $R_0(x_N)$ can be decomposed into two parts:

$$R_0(x_N) = F + \phi(x_N) \tag{3}$$

Here, F represents a constant baseline activation rate and $\phi(x_N)$ the perturbation of that rate due to feedback from the final downstream species X_N. $\phi(x_N)$ is a potentially nonlinear function, with $\phi'(x_N) < 0$ corresponding to negative feedback (production of X_0 inhibited by increases in x_N), and $\phi'(x_N) > 0$ corresponding to positive feedback (production of X_0 enhanced by increases in x_N). In the absence of feedback, $\phi(x_N) = 0$. The possibility of feedback from the last species to an upstream one has analogues in biological systems like the ERK MAPK pathway.[37] Of course, there may be feedback to multiple upstream species (as is the case for ERK), but here we only consider one feedback interaction as a starting point for modeling.

In a similar spirit, the baseline rate F is a constant for simplicity, representing the net effect of processes leading to the activation of X_0 that are not explicitly part of the model. There are also deactivation processes for X_0 (i.e., the action of phosphatases) which we model by an overall deactivation rate $\gamma_0 x_0$ proportional to the current population. We denote the constant γ_0 as the per-capita deactivation rate. For the case of no-feedback, the marginal stationary probability of the input X_0 is a Poisson distribution $\Pi(x_0; \bar{x}_0)$[38]

$$\Pi(x_0; \bar{x}_0) = \frac{(\bar{x}_0)^{x_0}\mathrm{e}^{-\bar{x}_0}}{(x_0)!} \tag{4}$$

where $\bar{x}_0 \equiv F/\gamma_0$, which in this case is equal to the mean and variance: $\langle x_0\rangle = \langle(\delta x_0)^2\rangle = \bar{x}_0$, with $\delta x_0 \equiv x_0 - \langle x_0\rangle$. Dynamically, the input signal has exponentially decaying autocorrelations, with characteristic time γ_0^{-1}. More complex types of input (for example with time-dependent $F(t)$ or nonexponential autocorrelations) can also be considered in generalizations of the model.[26,27] For our system, once feedback is turned on, the input distribution is no longer simply described by eq 4 and in general will not have a closed form analytical solution.

For $i > 0$, the production function for the ith species X_i is $R_i(x_{i-1}) \geq 0$, depending on the population x_{i-1} directly upstream. The deactivation rate at the ith level is $\gamma_i x_i$. We allow the R_i functions to be arbitrary and, hence, possibly nonlinear. Putting everything together, we can now write out the explicit form of eq 2:

$$0 = \sum_{i=0}^{N} \{\gamma_i[(x_i + 1)\mathcal{P}_{\mathbf{x}+\mathbf{e}_i} - x_i\mathcal{P}_{\mathbf{x}}] + R_i(x_{i-1})[\mathcal{P}_{\mathbf{x}-\mathbf{e}_i} - \mathcal{P}_{\mathbf{x}}]\} \tag{5}$$

For compactness of notation, we define $x_{-1} \equiv x_N$ and introduce the $(N+1)$-dimensional unit vectors $\mathbf{e}_i$, where $\mathbf{e}_0 = (1, 0, ..., 0)$, $\mathbf{e}_1 = (0, 1, 0, ..., 0)$, $\mathbf{e}_2 = (0, 0, 1, 0, ..., 0)$, and so on. Equation 5 is generally analytically intractable, in the sense that we cannot usually directly solve it to find the stationary distribution $\mathcal{P}_{\mathbf{x}}$. Despite this limitation, we can still make progress on understanding signaling behavior in the cascade via alternative approaches. Linearization of the production functions, described in the next section, is one such approach. Crucially, this approximation facilitates deriving bounds on signaling fidelity via the WK filter formalism. Later on, we will also introduce exact analytical as well as numerical methods for tackling certain cases of eq 5, to explore the validity of the WK bounds in the presence of nonlinearities.

WK Filter Formalism. In this section, we provide a brief overview of linearizing our signaling model and mapping it to a WK filter, generalizing the approach developed in refs 26, 28, and 38. This mapping allows us to derive bounds on various measures of signaling fidelity, which we know are valid at least within the linear approximation. The aim here is to summarize the bounds that we will later try to beat by introducing nonlinearities. Additional details of the WK approach can be found in the review of ref 38, which presents three special cases of our model: the $N = 1$ and $N = 2$ cascades without feedback, and the $N = 1$ system with feedback. The WK bound for the general N-level cascade, with and without feedback, is presented here for the first time, with the complete analytical derivation shown in the Supporting Information (SI).

Linearization. If we consider the limit where the mean copy numbers of all the chemical species in the cascade are large, we can approximately treat each population x_i as a continuous variable. If the magnitude of fluctuations in the stationary state is small relative to the mean, we can also approximate all the production functions to linear order around their mean values

$$\begin{aligned} R_0(x_N) &\approx F - \phi_1(x_N - \bar{x}_N) \\ R_i(x_{i-1}) &\approx \sigma_{i0} + \sigma_{i1}(x_{i-1} - \bar{x}_{i-1}) \qquad i > 0 \end{aligned} \tag{6}$$

with some coefficients ϕ_1, σ_{i0}, and σ_{i1}. Here, we have absorbed the zeroth-order Taylor coefficient of $\phi(x_N)$ around $\bar{x}_N$ into F. Note that the sign convention for the first-order coefficient ϕ_1 means that $\phi_1 > 0$ corresponds to negative feedback. The stationary averages in the linearized case are

https://doi.org/10.1021/acs.jpcb.1c07894
J. Phys. Chem. B 2021, 125, 12698−12711

$$\bar{x}_0 = \frac{F}{\gamma_0} \quad \bar{x}_i = \frac{\sigma_{i0}}{\gamma_i} \quad \text{for } i > 0 \tag{7}$$

We will use bar notation like $\bar{x}_i$ to exclusively denote the linearized stationary mean values. Brackets like $\langle x_i \rangle$ will always denote the true mean, whether the system is linear (in which case $\langle x_i \rangle = \bar{x}_i$) or not.

One advantage of linearization is the ability to express dynamics in an analytically tractable form, using the chemical Langevin approximation.[39] This approximation applies to systems with large copy numbers of chemical species, so that the populations $x_i(t)$ can be approximated as continuous functions. Another assumption involves the number of reactions of each possible type that occur in a microscopic time step, which ordinarily obeys Poisson statistics. In the chemical Langevin approach, these numbers are assumed to be sufficiently large that the Poisson distributions can be approximated by Gaussian ones. This leads to dynamical equations which have the form of deterministic rate equations with added Gaussian noise. We will later directly test the validity of these assumptions for signaling systems when we compare the WK bounds derived from the Langevin approach to our exact results based on eq 5. The Langevin equations corresponding to eq 5 are

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}x_0(t) &= R_0(x_N) - \gamma_0 x_0(t) + n_0(t) \\ \frac{\mathrm{d}}{\mathrm{d}t}x_i(t) &= R_i(x_{i-1}) - \gamma_i x_i(t) + n_i(t) \qquad i > 0 \end{aligned} \tag{8}$$

where the $n_i(t)$ components are Gaussian noise functions with correlations $\langle n_i(t) n_j(t') \rangle = 2\delta_{ij}\gamma_i \bar{x}_i \delta(t - t')$. We can rewrite eq 8 in terms of deviations from the mean, $\delta x_i(t) \equiv x_i(t) - \langle x_i \rangle$, plugging in eqs 6 and 7. The result is

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\delta x_0(t) &= -\gamma_0 \delta x_0(t) - \phi_1 \delta x_N(t) + n_0(t) \\ \frac{\mathrm{d}}{\mathrm{d}t}\delta x_i(t) &= -\gamma_i \delta x_i(t) + \sigma_{i1}\delta x_{i-1}(t) + n_i(t) \qquad i > 0 \end{aligned} \tag{9}$$

As shown in the SI Section 1, this system of stochastic differential equations can be solved using a Fourier transform.

Finding Bounds on Signal Fidelity by Mapping the System onto a Noise Filter. The linear chemical Langevin approach also allows us to map the system onto a classic noise filter problem from signal processing theory. We describe two versions of this mapping here, the first for the system without feedback, and the second with feedback.

1. No-Feedback System. Imagine we are interested in understanding correlations between two dynamical quantities in our system, as a measure of how accurately signals are transduced through the cascade. The choice of these two quantities, one of which we will label the "true signal" $s(t)$ and the other the "estimated signal" $\tilde{s}(t)$ within the filter formalism, depends on the biological question we would like to ask. For the cascade without feedback ($\phi_1 = 0$), a natural question is how well the output X_N reflects the input X_0. The function of the cascade can be to output an amplified version of the input,[40] but there is inevitably corruption of the signal as it is transduced from level to level due to the stochastic nature of the biochemical reactions in the network. If we assign $s(t) \equiv \delta x_0(t)$ and $\tilde{s}(t) \equiv \delta x_N(t)$, it turns out that because of the linearity of the dynamical system in eq 9 the two are related through a convolution:

$$\tilde{s}(t) = \int_{-\infty}^{\infty} \mathrm{d}t' H(t - t')[s(t') + n(t')] \tag{10}$$

The details of the functions $H(t)$ and $n(t)$, as derived from eq 9, are given in the SI Section 1. One can interpret eq 10 as a linear noise filter: a signal $s(t)$ corrupted with additive noise $n(t)$ (a function which depends on the Langevin noise terms $n_i(t)$) is convolved with a filter function $H(t)$ to yield an estimate $\tilde{s}(t)$. The filter function, which encodes the effects of the entire cascade, obeys an important physical constraint: $H(t) = 0$ for all $t < 0$. This enforces causality, since it ensures that the current value of $\tilde{s}(t)$ (the output in our case) only depends on the past history of the input plus noise, $s(t') + n(t')$ for $t' < t$.

The traditional version of filter optimization[31−33] is searching among all possible causal filter functions $H(t)$ for the one that minimizes the relative mean squared error between the signal and estimate:

$$\epsilon(s(t), \tilde{s}(t)) = \frac{\langle (\tilde{s}(t) - s(t))^2 \rangle}{\langle s^2(t) \rangle} \tag{11}$$

Since the averages are taken in a stationary state, ϵ is time-independent and can have values in the range $0 \leq \epsilon < \infty$. For the case of a biological cascade, however, where $s(t)$ and $\tilde{s}(t)$ are the times series of input and output fluctuations $\delta x_0(t)$ and $\delta x_N(t)$, respectively, we expect that the output may be an amplified version of the input. Hence, a better measure of fidelity may be a version of eq 11 that is independent of the scale differences between signal and estimate. To define this scale-free error, note that the optimization search over all allowable $H(t)$ necessarily involves searching over all constant prefactors A that might multiply a filter function $H(t)$. Using $AH(t)$ as the filter function instead of $H(t)$ is equivalent to switching from $\tilde{s}(t)$ to $A\tilde{s}(t)$, as can be seen from eq 10. If we were to look at the error $\epsilon(s(t), A\tilde{s}(t))$ for a given $\tilde{s}(t)$ and $s(t)$, we can readily find the value of A that minimizes this error, which is given by $A = \langle \tilde{s}(t)s(t) \rangle / \langle \tilde{s}^2(t) \rangle$. Plugging this value in, we can define a scale-free error E as follows:

$$E(s(t), \tilde{s}(t)) = \min_A \epsilon(s(t), A\tilde{s}(t)) = 1 - \frac{\langle \tilde{s}(t)s(t) \rangle^2}{\langle s^2(t) \rangle \langle \tilde{s}^2(t) \rangle} \tag{12}$$

By construction, $E(s(t), \tilde{s}(t)) \leq \epsilon(s(t), \tilde{s}(t))$ and in fact E have a restricted range: $0 \leq E \leq 1$. The independence of E from the relative scale of the output versus the input makes it an attractive measure of the fidelity of information transmission through the cascade. In fact, within the linear chemical Langevin approximation, one can show that $E = 2^{-2\mathcal{I}}$, where $\mathcal{I}$ is the instantaneous mutual information in bits between $s(t)$ and $\tilde{s}(t)$.[38] Thus, E will be the main measure of signal fidelity we focus on when we discuss the no-feedback cacade.

For a given $s(t)$ and $n(t)$, we denote the causal filter function $H(t)$ that minimizes $\epsilon(s(t), \tilde{s}(t))$ as the Wiener−Kolmogorov (WK) optimal filter $H_{\mathrm{WK}}(t)$. Because this optimization includes exploring over all possible prefactors of $H(t)$, the same WK filter function simultaneously minimizes $\epsilon(s(t), \tilde{s}(t))$ and $E(s(t), \tilde{s}(t))$, and the minima of the two error types coincide. We will denote this minimum as E_{WK}.

https://doi.org/10.1021/acs.jpcb.1c07894
J. Phys. Chem. B 2021, 125, 12698−12711

Hence, we have $\epsilon \geq E \geq E_{WK}$ in general for linear systems, and $\epsilon = E = E_{WK}$ when $H(t) = H_{WK}(t)$.

The procedure for calculating $H_{WK}(t)$ for a specific system, and then finding the optimal error bound E_{WK}, is based on analytical manipulation of the power spectra associated with $s(t)$ and $n(t)$.[33,38] We illustrate the details in SI Section 1, applying the method to our cascade model. This yields the following value for E_{WK} for an N-level cascade without feedback:

$$E_{WK} = 1 - \prod_{j=1}^{N} \frac{\gamma_0 \gamma_{j-1} \Lambda_j}{(\gamma_0 + \lambda_j)^2} \tag{13}$$

Here, $\Lambda_j \equiv \bar{x}_{j-1}\sigma_{j1}^2/(\sigma_{j0}\gamma_0)$ is a dimensionless parameter associated with the jth level, and $\lambda = \lambda_j$ is the jth root with a positive real part ($\mathrm{Re}(\lambda_j) > 0$) of the following polynomial $B(\lambda)$:

$$B(\lambda) = \gamma_0^2 + \sum_{j=1}^{N} \gamma_0^{2-j} \prod_{k=1}^{j} \frac{\gamma_{k-1}^2 - \lambda^2}{\gamma_{k-1}\Lambda_k} \tag{14}$$

Various properties of the roots of this polynomial are summarized in SI Section 1.3. Within the linear approximation, E_{WK} gives a lower bound on the achievable E and, hence, an upper bound on the maximum mutual information between input and output, $\mathcal{I}_{max} = -\frac{1}{2}\log_2 E_{WK}$.

Special cases of eq 13 recover earlier results. For $N = 1$, we find the single root $\lambda_1 = \gamma_0\sqrt{1 + \Lambda_1}$, and we can rewrite eq 13 in a simple form

$$E_{WK} = \frac{2}{1 + \sqrt{1 + \Lambda_1}} \text{ for } N = 1 \tag{15}$$

which is the result found in ref 26. Similarly, the $N = 2$ version, with more complicated but still analytically tractable roots λ_1 and λ_2, was derived in ref 38. For the case of general N, the roots λ_j can be found numerically. However, there is one scenario where we know closed form expressions for all the λ_j values for any N. This turns out to be the case where the biological parameters of the cascade are tuned such that filter function $H(t)$ in eq 10 is proportional to $H_{WK}(t)$, and hence, $E = E_{WK}$. (We do not need strict equality of the filter functions, because the resulting value of E is independent of an overall constant in front of $H(t)$.) That this is even possible is itself nontrivial; generally, when we vary biological parameters in a system mapped onto a noise filter, we allow $H(t)$ to explore a certain subspace of all possible filter functions. It is not guaranteed that any $H(t)$ in that subspace will coincide with $H_{WK}(t)$ up to a proportionality constant. However, as shown in SI Section 1 for the no-feedback N-level model, we can achieve $H(t) \propto H_{WK}(t)$ when the following conditions are met:

$$\gamma_j = \gamma_0\sqrt{1 + \frac{\gamma_{j-1}}{\gamma_0}\Lambda_j} \qquad j = 1, \ldots, N \tag{16}$$

These can be solved recursively to give nested radical forms

$$\begin{aligned} \gamma_1 &= \gamma_0\sqrt{1 + \Lambda_1} \\ \gamma_2 &= \gamma_0\sqrt{1 + \sqrt{1 + \Lambda_1}\Lambda_2} \\ \gamma_3 &= \gamma_0\sqrt{1 + \sqrt{1 + \sqrt{1 + \Lambda_1}\Lambda_2}\Lambda_3} \end{aligned} \tag{17}$$

and so on. When these conditions are satisfied, the roots λ_j have straightforward analytical forms, namely, $\lambda_j = \gamma_j$ for all j. Hence, we can substitute the values in eq 17 for λ_j in eq 13 to get E_{WK} explicitly when this scenario is true. With the aid of the recursion relation in eq 16, we can then write E_{WK} in this case as

$$E_{WK} = 1 - \prod_{i=1}^{N} \frac{l_i}{(1 + \sqrt{1 + l_i})^2} \tag{18}$$

where $l_i \equiv \gamma_{i-1}\Lambda_i/\gamma_0$ are dimensionless positive constants. The simple form of the bound in eq 18 makes it useful for analyzing the energetic cost of increasing signal fidelity in a cascade. The biological implications of this bound are discussed later on.

2. System with Feedback. The case with feedback uses a qualitatively different, and more abstract, mapping of the system onto a noise filter. Here the true and estimated signals are identified with the following quantities:[28,38]

$$s(t) \equiv \delta x_0(t)|_{\phi=0} \qquad \tilde{s}(t) = \delta x_0(t)|_{\phi=0} - \delta x_0(t) \tag{19}$$

The subscript in $\delta x_0(t)|_{\phi=0}$ denotes that this $\delta x_0(t)$ is obtained by solving eq 9 with the feedback turned off, $\phi(x_N) = 0$ or equivalently $\phi_1 = 0$. The $\delta x_0(t)$ without the subscript represents the solution with the feedback present. With this mapping, the error ϵ from eq 11 can be written as

$$\epsilon(s(t), \tilde{s}(t)) = \frac{\langle(\tilde{s}(t) - s(t))^2\rangle}{\langle s^2(t)\rangle} = \frac{\langle \delta x_0^2(t)\rangle}{\langle \delta x_0^2(t)|_{\phi=0}\rangle} \tag{20}$$

The underlying motivation is that negative feedback can serve as a homeostasis mechanism, dampening fluctuations $\delta x_0(t)$ in the X_0 species that are the direct target of the feedback. Achieving a small ϵ, by making $\tilde{s}(t)$ as close as possible to $s(t)$, translates to an efficient suppression of X_0 fluctuations (relative to their undamped magnitude in the absence of feedback). Note that in this case ϵ, rather than the scale-free error E, is the quantity used to specify system performance. Despite this difference, the problem is still a question of accurate information propagation through the cascade, because we need $\delta x_N(t)$ to encode a faithful representation of the input fluctuations in order to be able to effectively suppress them via negative feedback. Since the X_0 fluctuations in the no-feedback system follow the Poisson distribution of eq 4, the denominator in eq 20 is given by $\langle \delta x_0^2(t)|_{\phi=0}\rangle = \bar{x}_0$. Thus, $\epsilon = \langle \delta x_0^2(t)\rangle/\bar{x}_0$, which is also known as the Fano factor (ratio of variance to the mean), a standard measure for the size of fluctuations. Poisson distributions have Fano factors $\epsilon = 1$, but negative feedback in optimal cases can reduce ϵ to values much smaller than 1.

The close connections between the no-feedback and feedback analysis are apparent when we consider the analogue of the convolution in eq 10 for the feedback case. It turns out that $s(t)$ and $n(t)$ have the same functional forms as in the no-feedback case, but the filter function $H(t)$ is different (details in SI Section 2). Because the WK bound depends only on the power spectra of $s(t)$ and $n(t)$, the result for the bound E_{WK} is exactly the same as eq 13, with roots λ_j specified by eq 14. The interpretation of eq 13 in this case is as a lower bound for the error in eq 20, namely, $\epsilon \geq E_{WK}$.

Unlike the no-feedback cascade, where we can in principle tune the biological parameters so that $E = E_{WK}$, for the

https://doi.org/10.1021/acs.jpcb.1c07894
J. Phys. Chem. B 2021, 125, 12698−12711

linearized negative feedback system we can only asymptotically approach the bound from above, $\epsilon \to E_{WK}$. This limit is easiest to describe in the case where the production functions at each level are directly proportional to the upstream species, $R_i(x_{i-1}) \propto x_{i-1}$ for $i > 0$. In terms of eq 6, this corresponds to setting $\sigma_{i1} = \sigma_{i0}/\bar{x}_{i-1}$, so that $R_i(x_{i-1}) = \sigma_{i1}x_{i-1}$. The following two conditions are then needed to approach WK optimality: (i) The levels in the cascade have fast deactivation rates relative to the inverse autocorrelation time of the input, $\gamma_i \gg \gamma_0$ for $i > 0$; (ii) the coefficient of the negative feedback function is tuned to the value

$$\phi_1 = \gamma_0(\sqrt{1 + \Lambda_{eff}} - 1)\prod_{i=1}^{N}\frac{\gamma_i}{\sigma_{i1}} \tag{21}$$

where

$$\Lambda_{eff} = \frac{1}{F}\left[\sum_{i=1}^{N}\frac{1}{\sigma_{i0}}\right]^{-1} \tag{22}$$

In this limit, ϵ approaches E_{WK}, with eq 13 evaluating to the same form as eq 15, except for Λ replaced by Λ_{eff}

$$E_{WK} = \frac{2}{1 + \sqrt{1 + \Lambda_{eff}}} \tag{23}$$

The $N = 1$ special case of this result, where $\Lambda_{eff} = \Lambda_1 = \sigma_{i0}/F$, was the focus of ref 28.

Optimal Bounds and Metabolic Costs. The general E_{WK} bound in eq 13, and its corresponding values in various special cases (eqs 15, 18, 23), depends on the production rate parameters σ_{i0}, σ_{i1} and the per-capita deactivation rates γ_i at each state i. These processes have associated metabolic costs. If production involves activation of a substrate via phosphorylation, the cell has to maintain a sufficient population of inactive substrate and also consumes ATP during phosphorylation. Similarly, deactivation requires maintaining a population of phosphatases. Directly relating metabolic costs (i.e., the rate of ATP consumption) in these models to the thermodynamic cost (entropy production rate) is not always straightforward, since the models are coarse-grained and do not account for the dissipation involved in reaction substeps that are not explicitly described. However, one can roughly estimate the entropy production rate at level i of the cascade by multiplying the mean production rate σ_{i0} by $n\Delta\mu/T$, where $\Delta\mu$ is the chemical potential energy of ATP hydrolysis (typically $\approx$21−29 k_BT in biological systems[15]), T is the temperature, and n is the number of ATP molecules consumed in an activation reaction. Thus, higher production rates of active species entail larger thermodynamic costs.

The parameter Λ_{eff} indirectly reflects these costs, since from eq 22 it is proportional to the harmonic mean of all the per-capita production rates σ_{i0}, $i = 1, ..., N$, relative to the baseline activation rate F of the input species X_0. Increasing the production rate for any level will increase Λ_{eff}. The $\Lambda_i \equiv \bar{x}_{i-1}\sigma_{i1}^2/(\sigma_{i0}\gamma_0)$ parameters have a slightly different interpretation: If we consider production functions of the form $R_i(x_{i-1}) = \sigma_{i1}x_{i-1}$, with $\sigma_{i1} = \sigma_{i0}/\bar{x}_{i-1}$, as described in the previous section, then Λ_i simplifies to $\Lambda_i = \sigma_{i0}/(\bar{x}_{i-1}\gamma_0)$. Thus, in this case Λ_i is the average number of X_i molecules produced per molecule of X_{i-1} during the characteristic time interval γ_0^{-1} of input fluctuations.

These interpretations of the parameters allow us to understand the expense involved in achieving systems with better optimal performance. For example, since ϵ from eq 20 is given in terms of relative variance, a 10-fold decrease in the standard deviation of fluctuations would require a 100-fold decrease in ϵ. To decrease the optimal E_{WK} from eq 23 by a factor of 100, one would need roughly a 10^4 increase in Λ_{eff}, assuming we are in the regime where $\Lambda_{eff} \gg 1$. This extreme cost of eliminating fluctuations via negative feedback[3] has to be borne across the whole cascade: Since Λ_{eff} in eq 22 is potentially bottlenecked by one σ_{i0} much smaller than the others, the mean production rates for all the levels must be hiked up in order to increase Λ_{eff}.

An analogous story emerges when we analyze the same system without negative feedback. The relevant measure here is the scale-free error E between the time series of input and output populations, or equivalently the mutual information $\mathcal{I}$. Imagine we would like to increase the mutual information upper bound $\mathcal{I}_{max} = -(1/2)\log E_{WK}$ by 1 bit. In the limit of $I_i \gg 1$ in eq 18, this can be achieved for example by increasing every I_i by a factor of $\approx$16, regardless of N. Given $\sigma_{i1} = \sigma_{i0}/\bar{x}_{i-1}$ and $\bar{x}_{i-1} = \sigma_{i-1,0}/\gamma_{i-1}$, we can evaluate the dimensionless constant I_i associated with level i as

$$I_i = (\gamma_{i-1}/\gamma_0)\Lambda_i = (\sigma_{i0}/\sigma_{i-1,0})(\gamma_{i-1}/\gamma_0)^2$$

Hence, increasing I_i requires either increasing the relative mean production between the ith level and its predecessor, or the per-capita deactivation rate of the latter (if $i > 0$), or some combination of both. The massive cost of achieving multiple bits of mutual information between input and output in a biological signaling cascade is consistent with the narrow range of experimentally measured $\mathcal{I}$ values, spanning $\sim$1 to 3 bits,[18−25] with most systems near the lower end of the spectrum.

Nonlinearity in $N = 1$ Signaling Models: Earlier Attempts to Go beyond the WK Limit. The linearized noise filter approach described above provides a general recipe for deriving bounds on signaling: (i) start with a linear chemical Langevin description of the system; (ii) identify signal and estimate time series that are based on observables of interest, and are related via convolution in terms of some system-specific filter function $H(t)$; (iii) derive the optimal filter function $H_{WK}(t)$ and the corresponding error bound E_{WK}; and (iv) explore if and under what conditions the system can reach optimality. However, the procedure leaves open an important question: Is the resulting bound E_{WK} a useful approximation describing the system's performance limits, or can biology potentially harness nonlinearity to enhance performance significantly beyond the WK bound? We know that nonlinear, Hill-like functional relationships are a regular feature of biological signaling,[41] manifested in some cases as an extreme switch-like input−output relation known as ultrasensitivity.[34] Is E_{WK} still relevant in these scenarios? This section summarizes previous efforts to answer this question (all for the $N = 1$ case), setting the stage for our main calculations.

Nonlinearity in the $N = 1$ Model without Feedback. Reference 26 derived an exact solution for the no-feedback $N = 1$ system with an arbitrary production function $R_1(x_0)$ and discrete populations. The input signal remains the same as in the linear case, governed by production rate $R_0(x_N) = F$ and deactivation rate γ_0. The starting point is expanding $R_1(x_0)$ in terms of a series of polynomials

https://doi.org/10.1021/acs.jpcb.1c07894
J. Phys. Chem. B 2021, 125, 12698−12711

$$R_1(x_0) = \sum_{n=0}^{\infty} \sigma_{1n} \nu_n(x_0; \bar{x}_0) \tag{24}$$

Here, $\nu_n(x_0; \bar{x}_0)$ is a polynomial of nth degree in x_0, which depends on $\bar{x}_0 = F/\gamma_0$ as a parameter. The functions $\nu_n(x_0; \bar{x}_0)$ are variants of so-called Poisson−Charlier polynomials, whose properties are described in detail in SI Section 4. Similar expansions have found utility in spectral solutions of master equations.[42,43] The most important characteristic of these polynomials is that they are orthogonal with respect to averages over the Poisson distribution $\Pi(x_0; \bar{x}_0)$ defined in eq 4. If we denote $\langle f(x)\rangle_{\bar{x}} \equiv \sum_{x=0}^{\infty} f(x)\Pi(x; \bar{x})$, the average of a function $f(x)$ with respect to a Poisson distribution $\Pi(x; \bar{x})$, then[26,44]

$$\langle \nu_{n'}(x; \bar{x})\nu_n(x; \bar{x})\rangle_{\bar{x}} = n!\bar{x}^n \delta_{n',n} \tag{25}$$

The first few polynomials are given by

$$\nu_0(x; \bar{x}) = 1 \qquad \nu_1(x; \bar{x}) = x - \bar{x}$$
$$\nu_2(x; \bar{x}) = (x - \bar{x})^2 - x \tag{26}$$

Equation 25 allows the coefficients σ_{1n} from eq 24 to be evaluated in terms of moments with respect to $\Pi(x_0; \bar{x}_0)$

$$\sigma_{1n} = \frac{\langle \nu_n(x_0; \bar{x}_0) R_1(x_0)\rangle_{\bar{x}_0}}{\bar{x}_0^n n!} \tag{27}$$

Using eq 26, we can write the first two coefficients as

$$\sigma_{10} = \langle R_1(x_0)\rangle_{\bar{x}_0} \quad \sigma_{11} = \bar{x}_0^{-1}\langle (x_0 - \bar{x}_0) R_1(x_0)\rangle_{\bar{x}_0} \tag{28}$$

They have a simple physical interpretation: σ_{10} is the mean production rate, and σ_{11} is a measure of how steep the production changes with x near $\bar{x}$, and they are exactly the same as the coefficients in the linear expansion of eq 6. If $\sigma_{1n} \neq 0$ for any $n \geq 2$, then the production function $R_1(x_0)$ is nonlinear.

The exact expression for the error E derived in ref 26 takes the form

$$E = 1 - \frac{\bar{x}_0 \gamma_1 \sigma_{11}^2}{(\gamma_0 + \gamma_1)^2}\left[\sigma_{10} + \sum_{n=1}^{\infty} \frac{\sigma_{1n}^2 n! \bar{x}_0^n}{\gamma_1 + n\gamma_0}\right]^{-1} \tag{29}$$

The nonlinear σ_{1n} coefficients for $n \geq 2$ contribute to the expression in the brackets as σ_{1n}^2 multiplying a positive factor, and hence if nonzero always act to increase the error regardless of their sign. It turns out that eq 29 is bounded from below by the $N = 1$ WK limit from eq 15

$$E \geq E_{\mathrm{WK}} = \frac{2}{1 + \sqrt{1 + \Lambda_1}} \tag{30}$$

with $\Lambda_1 = \bar{x}_0 \sigma_{11}^2/(\sigma_{10}\gamma_0)$. The WK limit is achieved when $\gamma_1 = \gamma_0\sqrt{1 + \Lambda_1}$, just as predicted by eq 17, and when $R_1(x_0)$ has the optimal linear form, $R_1^{\mathrm{opt}}(x_0) = \sigma_{10} + \sigma_{11}(x_0 - \bar{x}_0)$, with all $\sigma_{1n} = 0$ for $n \geq 2$.

Increasing the slope of $R_1(x_0)$ at $x_0 = \bar{x}_0$ will increase σ_{11} and hence Λ_1, progressively decreasing the E_{WK} limit. This can be seen in Figure 2, which illustrates different production functions and the corresponding error values. Eventually, the slope will become so steep that it is impossible to have a purely linear function $R_1(x_0)$ with that value of σ_{11}. This is because σ_{11} must always be smaller than $\sigma_{10}/\bar{x}_0$ to have a linear production function that is everywhere non-negative, $R_1(x_0) > 0$ for all $x_0 \geq 0$. σ_{11} can be arbitrarily large for very steep, sigmoidal production functions $R_1(x_0)$, but in this case, the error will be significantly larger than E_{WK} due to the contributions from the nonlinear coefficients σ_{1n}, $n \geq 2$. We see this for the largest values of Λ_1 in Figure 2B, with the added error due to nonlinearity overwhelming the benefit from large Λ_1. In summary, for the $N = 1$ no-feedback model, there is no way to beat the WK limit, regardless of the choice of $R_1(x_0)$.

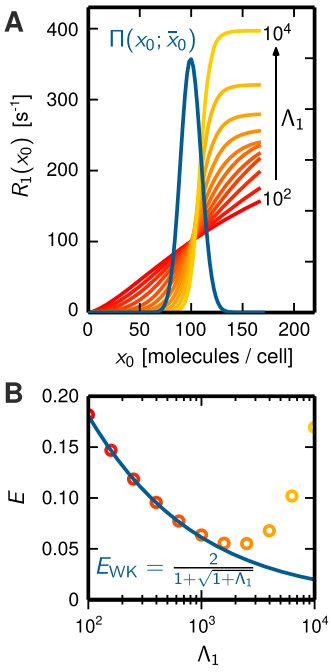


Figure 2. Case of a no-feedback $N = 1$ signaling model with a nonlinear production function $R_1(x_0)$ for parameters $F = 1$ s^{-1}, $\gamma_0 = 0.01$ s^{-1}, $\sigma_{10} = 100$ s^{-1}. (A) Examples of a variety of production functions $R_1(x_0)$, colored from red to yellow based on their steepness at $\bar{x}_0$ and, hence, the size of the corresponding parameter Λ_1. Superimposed in black is the marginal distribution $\Pi(x_0; \bar{x}_0)$ of the input species X_0. (B) For the production functions shown in panel A, the corresponding exact error E from eq 29 (circles) as a function of Λ_1. The WK bound E_{WK} from eq 30 is shown in blue for comparison. Adapted from ref 26 under the Creative Commons Attribution 3.0 License.

Nonlinearity in the $N = 1$ TetR Negative Feedback Circuit. Reference 28 studied an $N = 1$ negative feedback loop inspired by data from an experimental synthetic yeast gene circuit.[45] In this circuit, TetR mRNA (the X_0 species) leads to the production of TetR protein (the X_1 species), while the protein in turn binds to the promoter of the TetR gene, inhibiting the production of the mRNA. The model is similar to eq 8 when $N = 1$

$$\frac{\mathrm{d}}{\mathrm{d}t}x_0(t) = R_0(x_1) - \gamma_0 x_0(t) + n_0(t)$$
$$\frac{\mathrm{d}}{\mathrm{d}t}x_1(t) = R_1(x_0) - \gamma_1 x_1(t) - \Gamma(x_1(t)) + n_1(t) \tag{31}$$

with a linear production function $R_1(x_0) = \sigma_{11}x_0$, but with a sigmoidal Hill function form for the feedback $R_0(x_1)$

https://doi.org/10.1021/acs.jpcb.1c07894
J. Phys. Chem. B 2021, 125, 12698−12711

$$R_0(x_1) = \frac{A_1\theta_1^{\nu_1}}{\theta_1^{\nu_1} + x_1^{\nu_1}} \tag{32}$$

Note that this model has an additional contribution to the deactivation of the output, a function $\Gamma(x_1)$ that is also sigmoidal:

$$\Gamma(x_1) = \frac{A_2\theta_2^{\nu_2}}{\theta_2^{\nu_2} + x_1^{\nu_2}} \tag{33}$$

The parameters A_i, ν_i, θ_i, $i = 1, 2$ are all non-negative and determine the shape of the two Hill functions, which are common phenomenological expressions for regulatory interactions in biology.[41]

Using numerical methods to solve the corresponding master equation (similar to those described below), one can carry out a parameter search to solve the following optimization problem: With the constraints of fixed γ_0, γ_1, σ_{11}, $\bar{x}_0$, $\bar{x}_1$ (and hence also fixed $\Lambda_1 = \sigma_{11}/\gamma_0$), one can vary the Hill function parameters to find the smallest possible ϵ. The circles in Figure 3 show the optimization results for different Λ_1 in the range Λ_1 = 2−10, comparable to experimental estimates.[46] The E_{WK} bound from eq 23 is shown for comparison as a solid curve. The dashed curve shows an exact bound for the system, derived by Lestas, Vinnicombe, and Paulsson (LVP)[3] using information theory, which applies when $R_1(x_0)$ is linear and where the negative feedback from X_1 back to X_0 can occur via any function (linear or nonlinear). This exact bound is given by

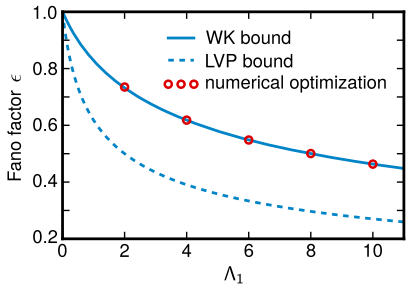


Figure 3. Fano factor ϵ for the TetR $N = 1$ negative feedback loop of ref 28. Numerical optimization results are shown as circles, while the WK and LVP lower bounds (eqs 23 and 34, respectively) are shown as solid and dashed curves. Reproduced with permission from ref 28. Copyright 2016 American Chemical Society.

$$E_{LVP} = \frac{2}{1 + \sqrt{1 + 4\Lambda_1}} \tag{34}$$

Note that $E_{LVP} \leq E_{WK}$, which opens the possibility that a nonlinear system could fall somewhere between the two curves, beating E_{WK}. However, as we see in Figure 3, the optimal numerical results were only able to approach from above, and never outperformed, the WK limit.

On the basis of the nonlinear $N = 1$ results with and without feedback described above, one could plausibly imagine that WK theory somehow provides universal bounds. Despite the fact that the WK limit was derived for linear systems, it surprisingly gives a rigorous bound for the nonlinear no-feedback model, and a numerical search to beat the limit proved fruitless in the feedback case. However, as we demonstrate in the next section, such a conclusion is premature.

BEATING THE WK LIMIT

To explore the validity of the WK limit more broadly, we need to be able to obtain precise error results in a wider range of nonlinear signaling systems. In this section, we provide two lines of evidence that demonstrate for the first time error values below E_{WK}. The first is from an $N = 2$ no-feedback cascade with linear $R_1(x_0)$ and quadratic $R_2(x_1)$ production functions. (We will prove that the related case, where $R_1(x_0)$ is nonlinear but $R_2(x_1)$ linear, always gives $E \geq E_{WK}$.) These results use an exact expression for E that is valid for any $N > 1$ no-feedback system, based on a recursion relation derived from the master equation of eq 5 that can be evaluated numerically to arbitrary precision. The second line of evidence is from an $N = 1$ negative feedback loop with linear production function $R_1(x_0)$ and a feedback function $\phi(x_1)$ that includes a quadratic contribution. Here, a solvable recursion relation for E is not possible, so we use a numerical solution of eq 5.

Exact Calculation of Error in the Nonlinear, Discrete $N > 1$ Model without Feedback. In order to understand the behavior of more complex nonlinear signaling cascades, we need to generalize the exact $N = 1$ no-feedback error expression from eq 29 to systems with $N > 1$. To start, let us introduce some convenient notation to deal with multiple-level systems. Along with the $(N + 1)$-dimensional vector $\mathbf{x} = (x_0, x_1, ..., x_N)$ that describes the full state of our system, we will define the N-dimensional truncated vector

$$\hat{\mathbf{x}} = (x_0, x_1, ..., x_{N-1})$$

that is missing the final component x_N. In a similar way, we define the truncated N-dimensional unit vectors $\hat{\mathbf{e}}_i$, $i = 1, ..., N - 1$, where $\hat{\mathbf{e}}_0 = (1, 0, ..., 0)$, $\hat{\mathbf{e}}_1 = (0, 1, ..., 0)$, and so on until $\hat{\mathbf{e}}_{N-1} = (0, ..., 0, 1)$. Consider the following generating function derived from the stationary distribution $\mathcal{P}_{\mathbf{x}}$

$$H_{\hat{\mathbf{x}}}(y) = \sum_{x_N=0}^{\infty} y^{x_N}\mathcal{P}_{\mathbf{x}} \tag{35}$$

The subscript $\hat{\mathbf{x}}$ denotes the fact that $H_{\hat{\mathbf{x}}}(y)$ depends on all components x_0 through x_{N-1}, but x_N has been eliminated through the sum. If one carries out the sum over x_N on both sides of eq 5, one can rewrite the master equation entirely in terms of generating functions

$$\begin{aligned} 0 = \sum_{i=0}^{N-1} &\{\gamma_i[(x_i + 1)H_{\hat{\mathbf{x}}+\hat{\mathbf{e}}_i}(y) - x_i H_{\hat{\mathbf{x}}}(y)] \\ &+ R_i(x_{i-1})[H_{\hat{\mathbf{x}}-\hat{\mathbf{e}}_i}(y) - H_{\hat{\mathbf{x}}}(y)]\} + \gamma_N(1 - y)H^{(1)}_{\hat{\mathbf{x}}}(y) \\ &+ R_N(x_{N-1})(y - 1)H_{\hat{\mathbf{x}}}(y). \end{aligned} \tag{36}$$

Here, the pth derivative of $H_{\hat{\mathbf{x}}}(y)$ with respect to y is denoted as $H^{(p)}_{\hat{\mathbf{x}}}(y) \equiv (d^p/dy^p)H_{\hat{\mathbf{x}}}(y)$. If we take p derivatives with respect to y of both sides of eq 36, and then set $y = 1$, we get the following relation

$$\begin{aligned} 0 = \sum_{i=0}^{N-1} &\{\gamma_i[(x_i + 1)H^{(p)}_{\hat{\mathbf{x}}+\hat{\mathbf{e}}_i}(1) - x_i H^{(p)}_{\hat{\mathbf{x}}}(1)] \\ &+ R_i(x_{i-1})[H^{(p)}_{\hat{\mathbf{x}}-\hat{\mathbf{e}}_i}(1) - H^{(p)}_{\hat{\mathbf{x}}}(1)]\} - p\gamma_N H^{(p)}_{\hat{\mathbf{x}}}(1) \\ &+ pR_N(x_{N-1})H^{(p-1)}_{\hat{\mathbf{x}}}(1) \end{aligned} \tag{37}$$

https://doi.org/10.1021/acs.jpcb.1c07894
J. Phys. Chem. B 2021, 125, 12698−12711

The above relation turns out to be the main one we need to evaluate the scale-free error E. To see this, let us rewrite the expression for E from eq 12 with the noise filter mapping $\tilde{s}(t) \equiv \delta x_N(t)$, $s(t) \equiv \delta x_0(t)$:

$$\begin{aligned} E &= 1 - \frac{\langle \tilde{s}(t)s(t)\rangle^2}{\langle s^2(t)\rangle\langle \tilde{s}^2(t)\rangle} \\ &= \frac{\langle \delta x_N(t)\delta x_0(t)\rangle^2}{\langle (\delta x_0(t))^2\rangle\langle (\delta x_N(t))^2\rangle} \\ &= \frac{(\langle x_N(t)x_0(t)\rangle - \langle x_N(t)\rangle\langle x_0(t)\rangle)^2}{(\langle x_0^2(t)\rangle - \langle x_0(t)\rangle^2)(\langle x_N^2(t)\rangle - \langle x_N(t)\rangle^2)} \end{aligned} \tag{38}$$

All the moments on the right-hand side of eq 38 with respect to the stationary distribution that involve $x_N(t)$ can in fact be expressed in terms of $H_{\hat{\mathbf{x}}}^{(p)}(1)$:

$$\begin{aligned} \langle x_N(t)\rangle &= \sum_{\mathbf{x}} x_N \mathcal{P}_{\mathbf{x}} = \sum_{\hat{\mathbf{x}}} H_{\hat{\mathbf{x}}}^{(1)}(1) \\ \langle x_N(t)x_0(t)\rangle &= \sum_{\mathbf{x}} x_N x_0 \mathcal{P}_{\mathbf{x}} = \sum_{\hat{\mathbf{x}}} x_0 H_{\hat{\mathbf{x}}}^{(1)}(1) \\ \langle x_N^2(t)\rangle - \langle x_N(t)\rangle &= \sum_{\mathbf{x}} x_N(x_N - 1)\mathcal{P}_{\mathbf{x}} = \sum_{\hat{\mathbf{x}}} H_{\hat{\mathbf{x}}}^{(2)}(1) \end{aligned} \tag{39}$$

Here, $\sum_{\mathbf{x}} \equiv \sum_{x_0=0}^{\infty} \cdots \sum_{x_N=0}^{\infty}$ and $\sum_{\hat{\mathbf{x}}} \equiv \sum_{x_0=0}^{\infty} \cdots \sum_{x_{N-1}=0}^{\infty}$. The remaining moments in eq 38, those that involve only $x_0(t)$, are known from the fact that the marginal distribution of the input X_0 is just the Poisson distribution $\Pi(x_0; \bar{x}_0)$ of eq 4. This yields

$$\langle x_0(t)\rangle = \bar{x}_0 \qquad \langle x_0^2(t)\rangle = \bar{x}_0^2 + \bar{x}_0 \tag{40}$$

Recall that the barred notation denotes the linearized stationary averages defined in eq 7. Thus, the approach to finding E is as follows: (i) Use eq 37 to derive properties of $H_{\hat{\mathbf{x}}}^{(p)}(1)$ that allow us to evaluate the moments in eq 39; (ii) together with eq 40, we can then plug the moment results into eq 38 to derive an expression for E. Here we will summarize the final result, with the full details of the derivation shown in SI Section 3.

To facilitate the solution, we expand the production function in terms of Poisson−Charlier polynomials, just as in eq 24 for the $N = 1$ case

$$R_i(x_{i-1}) = \sum_{n=0}^{\infty} \sigma_{in} \nu_n(x_{i-1}; \bar{x}_{i-1}) \text{ for } i > 0 \tag{41}$$

Each expansion coefficient is given by the analogue of eq 27, averaging over a Poisson distribution

$$\sigma_{in} = \frac{\langle \nu_n(x_{i-1}; \bar{x}_{i-1}) R_i(x_{i-1})\rangle_{\bar{x}_{i-1}}}{\bar{x}_{i-1}^n n!} \tag{42}$$

To tackle $H_{\hat{\mathbf{x}}}^{(p)}(1)$, we will define new functions $J_{\hat{\mathbf{x}}}^{(p)}$ through the relation

$$H_{\hat{\mathbf{x}}}^{(p)}(1) \equiv J_{\hat{\mathbf{x}}}^{(p)} \Pi(\hat{\mathbf{x}}; \hat{\bar{\mathbf{x}}}) \tag{43}$$

Here, $\Pi(\hat{\mathbf{x}}; \hat{\bar{\mathbf{x}}})$ is a multidimensional Poisson distribution

$$\Pi(\hat{\mathbf{x}}; \hat{\bar{\mathbf{x}}}) \equiv \Pi(x_0; \bar{x}_0)\Pi(x_1; \bar{x}_1) \cdots \Pi(x_{N-1}; \bar{x}_{N-1}) \tag{44}$$

Thus, any $J_{\hat{\mathbf{x}}}^{(p)} \neq 1$ represents the deviation of $H_{\hat{\mathbf{x}}}^{(p)}(1)$ from a simple multidimensional Poisson distribution. Similarly, we can define a multidimensional version of the Poisson−Charlier polynomials

$$\nu_{\hat{\mathbf{n}}}(\hat{\mathbf{x}}; \hat{\bar{\mathbf{x}}}) \equiv \nu_{n_0}(x_0; \bar{x}_0)\nu_{n_1}(x_1; \bar{x}_1) \cdots \nu_{n_{N-1}}(x_{N-1}; \bar{x}_{N-1}) \tag{45}$$

where $\hat{\mathbf{n}} = (n_0, n_1, \ldots, n_{N-1})$ is an N-dimensional vector of integers $n_i \geq 0$. Let us expand $J_{\hat{\mathbf{x}}}^{(p)}$ in terms of these polynomials

$$J_{\hat{\mathbf{x}}}^{(p)} = \sum_{\hat{\mathbf{n}}} \mu_{\hat{\mathbf{n}}}^{(p)} \nu_{\hat{\mathbf{n}}}(\hat{\mathbf{x}}; \hat{\bar{\mathbf{x}}}) \tag{46}$$

defining expansion coefficients $\mu_{\hat{\mathbf{n}}}^{(p)}$. It turns out the moments in eq 39 are all just linear combinations of the $\mu_{\hat{\mathbf{n}}}^{(p)}$, which follows from the properties of the Poisson−Charlier polynomials averaged with respect to Poisson distributions:

$$\begin{aligned} \langle x_N(t)\rangle &= \mu_{\hat{0}}^{(1)} \\ \langle x_N(t)x_0(t)\rangle &= \bar{x}_0(\mu_{\hat{0}}^{(1)} + \mu_{\hat{0}+\hat{\mathbf{e}}_0}^{(1)}) \\ \langle x_N^2(t)\rangle - \langle x_N(t)\rangle &= \mu_{\hat{0}}^{(2)} \end{aligned} \tag{47}$$

Here $\hat{0} \equiv (0, 0, \ldots, 0)$ is the N-dimensional zero vector. Plugging this into eq 38 gives

$$E = 1 - \frac{\bar{x}_0(\mu_{\hat{0}+\hat{\mathbf{e}}_0}^{(1)})^2}{\mu_{\hat{0}}^{(2)} + \mu_{\hat{0}}^{(1)} - (\mu_{\hat{0}}^{(1)})^2} \tag{48}$$

The final piece of the solution is converting eq 37 into a recursion relation for the coefficients $\mu_{\hat{\mathbf{n}}}^{(p)}$:

$$\mu_{\hat{\mathbf{n}}}^{(p)} = \frac{p\nu_{\hat{\mathbf{n}}}^{(p-1,N)} + \sum_{i=1}^{N-1} (\bar{x}_i^{-1}\nu_{\hat{\mathbf{n}}-\hat{\mathbf{e}}_i}^{(p,i)} - \gamma_i \mu_{\hat{\mathbf{n}}-\hat{\mathbf{e}}_i}^{(p)})}{p\gamma_N + \sum_{i=0}^{N-1} n_i \gamma_i}, \tag{49}$$

with $\mu_{\hat{0}}^{(0)} = 1$. The coefficients $\nu_{\hat{\mathbf{n}}}^{(p,i)}$ are given by the following expansion in terms of σ_{in} and $\mu_{\hat{\mathbf{n}}}^{(p)}$:

$$\nu_{\hat{\mathbf{n}}}^{(p,i)} = \sum_{\substack{a,b=0 \\ a+b \geq n_{i-1} \\ |a-b| \leq n_{i-1}}}^{\infty} \sigma_{ia} \mu_{\hat{\mathbf{n}}+(b-n_{i-1})\hat{\mathbf{e}}_{i-1}}^{(p)} C_{n_{i-1}}^{ab}(\bar{x}_{i-1}) \tag{50}$$

and $C_k^{mn}(z)$ are polynomials in z given by

$$C_k^{mn}(z) = \sum_{c=\max(0,n-k,m-k)}^{\lfloor \frac{m+n-k}{2} \rfloor} \Gamma_{kc}^{mn} z^c \tag{51}$$

with

$$\Gamma_{kc}^{mn} = \frac{m!n!}{c!(c+k-m)!(c+k-n)!(m+n-k-2c)!} \tag{52}$$

Here the sum starts at the largest of the three values 0, $n-k$, or $m-k$, and $\lfloor w \rfloor$ denotes the largest integer less than or equal to w.

The general procedure for calculating E works as follows:

1. For a given set of production functions $R_i(x_{i-1})$, we calculate the expansion coefficients σ_{in} using eq 42. If

https://doi.org/10.1021/acs.jpcb.1c07894
J. Phys. Chem. B 2021, 125, 12698−12711

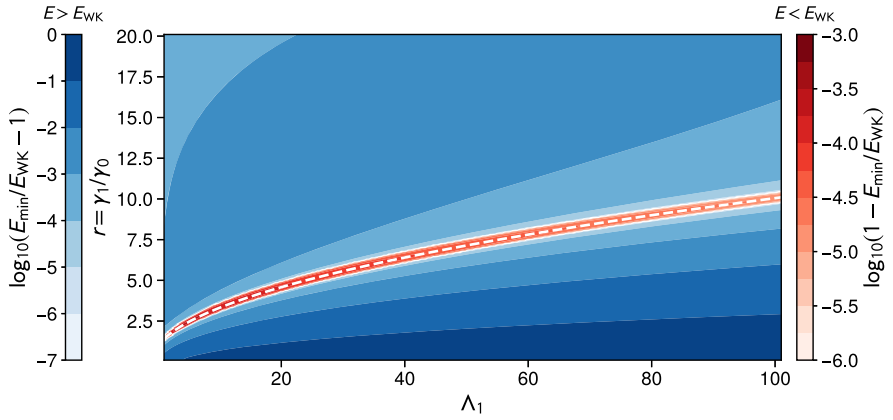


Figure 4. Contour plot of $\log_{10}|E_{min}/E_{WK} - 1|$ for the $N = 2$ cascade with linear $R_1(x_0)$ and quadratic $R_2(x_1)$. The minimum value of the error E_{min} at a given $r = \gamma_1/\gamma_0$ and Λ_1 is found by numerical minimization with respect to γ_2 and σ_{22}, with fixed $\Lambda_2 = 5$. Cool colors denote regions where $E_{min} > E_{WK}$ and warm colors where $E_{min} < E_{WK}$. The dashed white curve corresponds to $r = \sqrt{1 + \Lambda_1}$.

necessary, we truncate the expansion above some order M, setting $\sigma_{in} = 0$ for $n > M$. In practice, because of the rapid convergence of eq 41 for x_{i-1} near $\bar{x}_{i-1}$, choosing $M = 3$ or 4 is sufficient. However, we can increase the cutoff M to get whatever numerical precision we desire.

2. We plug the resulting σ_{in} into eq 50, and this in turn defines the $\nu_{\hat{\mathbf{n}}}^{(p,i)}$ components that appear in eq 49.
3. We solve the recursive system of equations in eq 49 for $\mu_{\hat{0}+\hat{\mathbf{e}}_0}^{(1)}$, $\mu_{\hat{0}}^{(1)}$, and $\mu_{\hat{0}}^{(2)}$, and we use these to find E from eq 48.

Though complex in appearance, the procedure is easy to implement as a numerical algorithm, and with any finite cutoff M is guaranteed to yield a value for E. As we increase M, we generally quickly converge to the exact E for the system.

In some cases, the entire procedure can be carried out analytically to give exact closed form expressions for E. When $N = 1$, we recover the result in eq 29, as expected. Another example is the $N = 2$ system where the first level production function $R_1(x_0)$ is arbitrary, but the second level function $R_2(x_1)$ is linear (and hence $\sigma_{2n} = 0$ for $n \geq 2$). Here E is given by

$$E = 1 - \frac{\bar{x}_0\gamma_1\gamma_2(\gamma_1 + \gamma_2)\sigma_{11}^2\sigma_{21}^2}{(\gamma_0 + \gamma_1)^2(\gamma_0 + \gamma_2)^2} \times \left[\gamma_1(\gamma_1 + \gamma_2)\sigma_{20} + \sigma_{21}^2\left(\sigma_{10} + \sum_{n=1}^{\infty} \frac{\sigma_{1n}^2 n!\bar{x}_0^n(n\gamma_0 + \gamma_1 + \gamma_2)}{(n\gamma_0 + \gamma_1)(n\gamma_0 + \gamma_2)}\right)\right]^{-1} \tag{53}$$

Just as in eq 29, any nonlinear contributions to $R_1(x_0)$ always increase E, since the coefficients σ_{1n} for $n \geq 2$ only appear in the brackets in eq 53 as σ_{1n}^2 multiplying positive factors. In this scenario, $E \geq E_{WK}$ always, where E_{WK} is given by eq 13. As shown in SI Section 5, a similar story holds for $N = 3$ when only the first level production function is allowed to be nonlinear. It is likely that nonlinearity only at the first level cannot violate the WK bound for any N.

The simplest case where we are able to observe a violation of the E_{WK} limit is for $N = 2$ when $R_1(x_0)$ is linear and the $R_2(x_1)$ is quadratic: $\sigma_{1n} = 0$ for $n \geq 2$ and $\sigma_{2m} = 0$ for $m \geq 3$. The resulting analytical expression for E is complicated, but we can investigate its optimal behavior numerically. In Figure 4, we conducted a numerical minimization of E with respect to γ_2 and the quadratic coefficient σ_{22} for various combinations of $r \equiv \gamma_1/\gamma_0$ and Λ_1, keeping Λ_2 fixed. If we denote this minimum value E_{min}, Figure 4 shows $\log_{10}|E_{min}/E_{WK} - 1|$ with the cool colored contours indicating $E_{min} > E_{WK}$ and warm colored contours indicating $E_{min} < E_{WK}$. In the purely linear case described earlier, we found that $E = E_{WK}$ when the conditions from eq 17 are satisfied, which corresponds to $r = \sqrt{1 + \Lambda_1}$, shown as a dashed white curve in the figure. With the addition of the quadratic term in $R_2(x_1)$, the region near that curve now supports solutions that beat the WK limit (the warm colored band in Figure 4). However, the improvement relative to the WK bound is exceedingly small, roughly 0.001−0.01% better.

To understand the small size of the improvement, let us look at a subset of the parameter space that is analytically tractable. Set the linear portions of the production functions to be directly proportional to the upstream population, $R_1(x_0) = \sigma_{11}x_0$ and $R_2(x_1) = \sigma_{21}x_1 + \sigma_{22}\nu_2(x_1; \bar{x}_1)$, which means $\sigma_{n1} = \sigma_{n0}/\bar{x}_{n-1}$ for $n = 1, 2$. Furthermore, imagine that the conditions of eq 17 are fulfilled for γ_1 and γ_2, which means $E = E_{WK}$ when the quadratic perturbation $\sigma_{22} = 0$. In this case $\Lambda_1 = r^2 - 1$, with $r \equiv \gamma_1/\gamma_0 > 1$, and $\Lambda_2 = \rho r$, with $\rho \equiv \sigma_{20}/\sigma_{10}$. E_{WK} from eq 18 can then be written as

$$E_{WK} = 1 - \frac{\rho(r-1)r^2}{(1+r)(1+\sqrt{1+\rho r^2})^2} \tag{54}$$

Let us focus on the regime where signaling is at least as effective as in many experimentally measured cascades,[18−25] which means $I_{max} \gtrsim 1$ bit or equivalently $E_{WK} \lesssim 1/4$. This generally requires $r \gg 1$ and $\rho \gg 1$. In this limit, the complicated full expression for E simplifies, and we can expand the difference $E - E_{WK}$ to second-order in the perturbation parameter σ_{22}

$$E - E_{WK} \approx \frac{2}{\gamma_0\rho r}\sigma_{22} + \frac{2\bar{x}_0}{\gamma_0^2\rho^2}\sigma_{22}^2 \tag{55}$$

There is a minimum $E = E_{min}$ at $\sigma_{22} = -\gamma_0\rho/(2\bar{x}_0 r)$, with

$$E_{min} - E_{WK} \approx -\frac{1}{2\bar{x}_0 r^2} \tag{56}$$

https://doi.org/10.1021/acs.jpcb.1c07894
J. Phys. Chem. B 2021, 125, 12698−12711

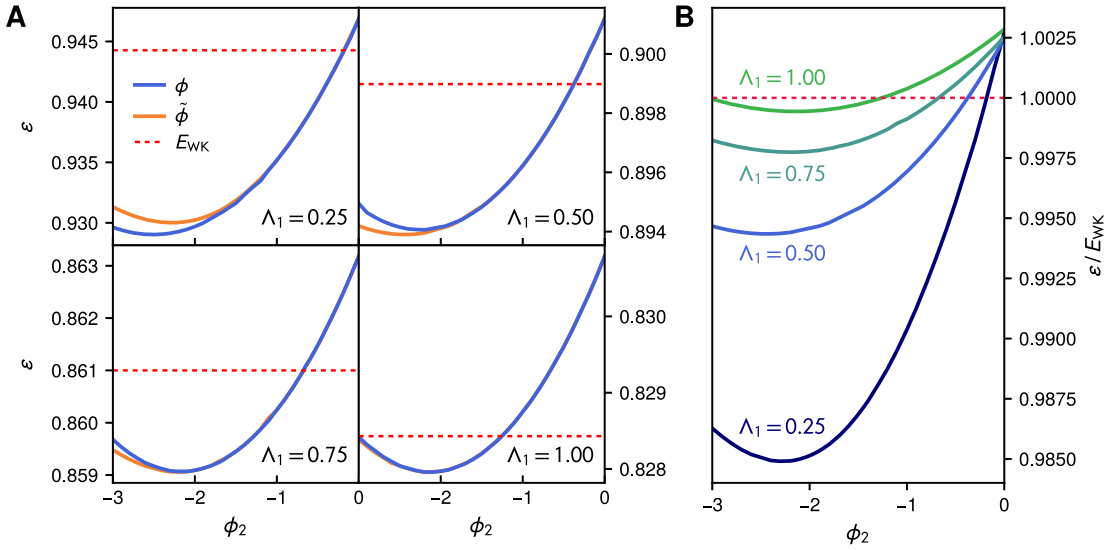


Figure 5. (A) Numerically calculated Fano factor ϵ in the $N = 1$ nonlinear feedback system. The plots show ϵ versus ϕ_2 for the quadratic feedback function $\phi(x_1)$ (blue) from eq 57 and the monotonic alternative $\tilde{\phi}(x_1)$ (orange) from eq 58, using the parameters described in the text. The WK bound E_{WK} in shown as a dashed red line. The subgraphs depict cases with four different values of Λ_1 between 0.25 and 1. (B) The Fano factor results from panel A, using the feedback function $\tilde{\phi}(x_1)$, but normalized with respect to E_{WK}. The dashed red line is $\epsilon/E_{WK} = 1$.

Though we can violate the E_{WK} bound, the size of the violation becomes small for $r \gg 1$. From eq 54, we know that $E_{WK} \sim 2/r$ for ρ, $r \gg 1$, and hence, the relative magnitude $E_{min}/E_{WK} - 1 \approx -(4\bar{x}_0 r)^{-1}$. The negligible scale of the improvement over E_{WK} is consistent with the numerical results of Figure 4, though the latter was calculated over a broader portion of the parameter space. In SI Section 5, we also considered the more complex $N = 2$ case where the production functions at both levels are quadratic. This additional nonlinearity can increase the magnitude of the violation, but the benefit is generally small: Equation 56 becomes amended with a correction term $\propto r^{-3}$.

Revisiting Nonlinearity in the $N = 1$ Model with Feedback. The violation of the E_{WK} bound in the no-feedback case raises the question of whether similar results are possible in the presence of feedback. We return to the $N = 1$ system used for the TetR model above, but with several simplifications: (i) We do not include the additional nonlinear degradation term $\Gamma(x_1)$; (ii) rather than a Hill function for $R_0(x_1)$, we use eq 3, with a quadratic form for the feedback function $\phi(x_1)$

$$\phi(x_1) = -\phi_1(x_1 - \bar{x}_1) + \phi_2(x_1 - \bar{x}_1)^2 \tag{57}$$

Note that the motivation here is not to create a quadratic approximation of a Hill function. Since ref 28 did not find any violation of E_{WK} in extensive numerical tests with Hill feedback, we surmised that higher-order nonlinear terms (ϕ_n for $n > 2$ if the Hill function was written as full Taylor series around $\bar{x}_1$) might have acted to keep E above E_{WK}, similar to what we found in eqs 29 and 53. Thus, we decided to use an alternative, simpler form for the feedback, keeping only the lowest-order nonlinear term (the quadratic one). This allows us to focus on the effect of a simple nonlinear perturbation, parametrized by ϕ_2, and see whether we can now achieve $E < E_{WK}$.

Depending on the values of ϕ_1 and ϕ_2, there could be a range of x_1 where $R_0(x_1)$ in eq 3 becomes negative, which is unphysical. In our numerical calculations, we thus always use $\max(R_0(x_1), 0)$ as the feedback function. However, for the parameters we explored, the range of x_1 where the sign switch in $R_0(x_1)$ occurs is far outside the typical range of stationary state x_1 fluctuations, so the precise details of the cutoff have a negligible influence on the results. A final important difference from the TetR model is that we will also investigate the regime of smaller Λ_1 (the numerics in the earlier study were confined to $\Lambda_1 \geq 2$). On the basis of intuition from the no-feedback case, we guess that any violation of the E_{WK} bound might become very small for large Λ_1 and, hence, difficult to detect numerically.

Though eq 57 has a simple form that is convenient for parameter exploration, it has one feature that makes it somewhat unrealistic from a biological perspective. For $\phi_1 > 0$, $\phi_2 < 0$ (the case that will be of interest to us below), the slope $d\phi(x_1)/dx_1$ becomes positive for

$$x_1 < x_1^* = \bar{x}_1 + \phi_1/(2\phi_2)$$

corresponding to positive feedback for smaller x_1 populations. Since we would like to concentrate on systems with negative feedback, we also define an alternative feedback function $\tilde{\phi}(x_1)$ that avoids this issue by being constant for $x_1 \leq x_1^*$ and monotonically decreasing for $x_1 > x_1^*$:

$$\tilde{\phi}(x_1) = \begin{cases} \phi(x_1) & x_1 > x_1^* \\ \phi(x_1^*) & x_1 \leq x_1^* \end{cases} \tag{58}$$

As we will see below, it turns out that both $\phi(y)$ and $\tilde{\phi}(y)$ give qualitatively similar results.

The Poisson−Charlier expansion approach of the previous example can also be applied to a general N-level feedback system, yielding a set of coupled linear equations for the coefficients $\mu_{\mathbf{n}}^{(p)}$ analogous to eq 49. However, because of the feedback interaction between x_N and x_0, these equations are no longer particularly useful: lower-order coefficients depend on higher-order ones in an infinite hierarchy of equations that has

https://doi.org/10.1021/acs.jpcb.1c07894
J. Phys. Chem. B 2021, 125, 12698−12711

no closure for any nonlinear $\phi(x_1)$. We thus turn to an alternative approach: solving the master equation, eq 5, for the 2D stationary probability $\mathcal{P}_{\mathbf{x}}$, where $\boldsymbol{x} = (x_0, x_1)$. Since x_0 and x_1 can be any non-negative integer, eq 5 is an infinite linear system of equations. To make it amenable to a fast numerical solution, we truncate the range of allowable (x_0, x_1) to be within six standard deviations of $\bar{x}_0$ and $\bar{x}_1$. We estimate the standard deviations from the linear case ($\phi_2 = 0$), where closed form expressions are available in terms of the system parameters. The actual standard deviations in the presence of nonzero ϕ_2 for the parameter range we considered were not perturbed significantly, so this estimation procedure worked well. Similarly, $\bar{x}_0$ and $\bar{x}_1$ were good estimates for the actual $\langle x_0 \rangle$ and $\langle x_1 \rangle$, because the mean of the distribution shifts only a small amount with ϕ_2. The window established by this procedure had a typical width of around ~100 for x_1 and ~700 for x_0 for parameters in the range described below. In eq 5, all $\mathcal{P}_{\mathbf{x}}$ outside the allowable range of $\mathbf{x}$ were set to zero. This means that eq 5 becomes a finite system of linear equations that can be solved efficiently using sparse matrix methods. Once the stationary distribution is known numerically, one can then easily calculate the error ϵ from eq 20 by finding the marginal distribution of x_0 and calculating its first and second moments. We checked for convergence and boundary effects by redoing the solution using window widths that were different than six standard deviations and verified that the results were unchanged up to the desired precision ($<10^{-4}$ for the calculation of ϵ). For select parameter sets, we also validated the moments of the stationary distribution against kinetic Monte Carlo simulations,[47] though for the latter achieving high precision is difficult because of the computational time required.

We used the following parameter values (all in units of s^{-1}): $\gamma_0 = 2$, $\gamma_1 = 200$, $\sigma_{10} = 8000$, $\sigma_{11} = 2$. The value of F was varied to allow for a range of possible

$$\Lambda_1 = \bar{x}_1\sigma_{11}^2/(\gamma_0\sigma_{10}) = F\sigma_{11}^2/(\gamma_0^2\sigma_{10})$$

The value of ϕ_1 was set to the optimality condition from eq 21

$$\phi_1 = \frac{\gamma_0\gamma_1}{\sigma_{11}}(\sqrt{1+\Lambda_1} - 1) \tag{59}$$

where we have used the fact that $\Lambda_{\text{eff}} = \Lambda_1$ for $N = 1$. This guarantees that, in the linear feedback case of $\phi_2 = 0$, the system should be close to the WK limit (up to correction factors due to finite γ_1, since technically the WK limit is only approached in the feedback case when $\gamma_1 \to \infty$). Figure 5A shows numerical results for the Fano factor ϵ as a function of ϕ_2 for different values of Λ_1 between 0.25 and 1. In all cases for linear feedback ($\phi_2 = 0$), we see that $\epsilon > E_{\text{WK}}$, where E_{WK} is given by eq 23. The fact that ϵ is above E_{WK} for the linear system is due to the fact that γ_1 is finite. For the case $\phi_2 > 0$ (not shown in the graphs), the error increases, while for $\phi_2 < 0$ we see that the error decreases, until it dips below the E_{WK} line before increasing again. The choice of feedback function, $\phi(x_1)$ or $\tilde{\phi}(x_1)$, does not make a significant difference. Interestingly, the violation of the WK bound is quite small, just as in the no-feedback case, as we can see more clearly in Figure 5B, where the ratio ϵ/E_{WK} is plotted, in this case using the $\tilde{\phi}(x_1)$ function. The largest dip we observed is still only about 1.5% below E_{WK}. Moreover, in order to see any violation at all we had to look at small $\Lambda_1 \leq 1$. In this regime, E_{WK} is quite large, just below the Poissonian Fano factor value of 1. Hence, the fluctuations are only slightly reduced by the feedback. Once Λ_1 becomes larger, in the more biologically relevant regime where negative feedback is effective at suppressing fluctuations, we found it impossible to observe any violations of E_{WK}. This could possibly explain the lack of any evidence of violations in the earlier TetR study[28] (see Figure 3), where only $\Lambda_1 \geq 2$ was considered. Though the figures show results for only one set of parameter values, other sets we tried produced qualitatively similar results: The nonlinear case beat the WK limit for small Λ_1, but it was always by a small amount.

CONCLUSIONS

Using a combination of analytical and numerical approaches, we have been able to show that the Wiener−Kolmogorov optimal error E_{WK} is not a universal lower bound for biological signaling cascades, both with and without feedback. However, far from undermining the usefulness of the WK theory, our results actually strengthen its practical value as a general purpose approximation to estimate performance limits in signaling systems. In some cases, for example, the $N = 1$ or $N = 2$ no-feedback systems with nonlinear production in the first level, the E_{WK} bound continues to hold rigorously despite nonlinearity. Also, in all cases where the bound is broken, the extent of the violation is negligible and decreases or vanishes in the regime where the system is effective at its respective task (either propagating the upstream signal with high fidelity or suppressing fluctuations). Further study is needed to see if the performance gain beyond the E_{WK} bound can be made substantial, for example, by combining the effects of nonlinearity from multiple levels in the cascade. However, additional nonlinearity is not necessarily beneficial: In eqs 29 and 53, and as depicted in Figure 2, each higher-order nonlinear contribution pushes us further away from the E_{WK} limit.

Thus, for practical purposes, the WK approach remains an excellent way to derive biological bounds that remain meaningful even when the underlying assumptions of the theory (like linearity) no longer strictly hold. Equally importantly, the theory allows one to ascertain under what conditions one can actually achieve this kind of optimality. In all the signaling systems investigated so far, E_{WK} is either directly attainable or can be asymptotically approached by tuning parameters. This is in contrast to a rigorous bound like E_{LVP} from eq 34, which holds for arbitrarily complex feedback mechanisms in a system with linear production. However, it has overestimated the optimal capabilities of all the feedback networks we have investigated: None of our systems ever gets close to E_{LVP}. A recent example of the versatility of the WK theory is the study of kinase-phosphatase signaling networks in ref 15. A simple analytical WK bound, derived from a linearized $N = 1$ network, explains a previously unknown optimal relationship among signal fidelity, bandwidth, and minimum ATP consumption. It holds across a vast biological parameter space deduced from bioinformatic databases and remains valid even when all the microscopic, nonlinear reaction details of the system are taken into account. The robustness of the WK bound, highlighted in the results of the current study, help us understand the theory's success in such contexts.

Beyond future applications of WK theory to other specific systems, and possible experimental validation, there is still work to be done in developing the analytical techniques (like

https://doi.org/10.1021/acs.jpcb.1c07894
J. Phys. Chem. B 2021, 125, 12698−12711

the Poisson−Charlier expansion) which we used for the no-feedback cascade. Exact results in nonlinear systems are relatively rare and hence valuable in themselves, and also as benchmarks for a variety of simpler approximations like the WK theory. The expansion method we described is currently limited by cases where the recursive system of equations does not close (i.e., in the presence of feedback, and more generally in biochemical networks with loops). Carefully tailored moment closure approaches[48] might provide a way forward and broaden the applicability of the method to systems with different types of feedback and other more complex network motifs.

Finally, it will be interesting to explore the relationship between WK theory and other intrinsic bounds satisfied by thermodynamic systems. The recently developed thermodynamic uncertainty relation (TUR),[4,49] for example, provides a bound on the variance of current-like observables in a wide range of systems, expressed in terms of energy dissipation. A case like our negative feedback cascade, where decreasing the variance of the X_0 species requires a significant expenditure of resources, might be possible to reformulate as a TUR problem. Does WK optimal signaling saturate the TUR bound? A deeper understanding of these kinds of interconnections will help us better delineate the fundamental limits of biological processes.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcb.1c07894.

Detailed derivations of our analytical results (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Michael Hinczewski − *Department of Physics, Case Western Reserve University, Cleveland, Ohio 44106, United States;* orcid.org/0000-0003-2837-7697; Email: mxh605@case.edu

Authors

Casey Weisenberger − *Department of Physics, Case Western Reserve University, Cleveland, Ohio 44106, United States*

David Hathcock − *Department of Physics, Cornell University, Ithaca, New York 14853, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcb.1c07894

Author Contributions

§C.W. and D.H. contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

C.W. and M.H. would like to acknowledge support from National Science Foundation MCB/BIO grant 1651560. D.H. is supported by an NSF Graduate Research Fellowship, grant DGE-1650441. M.H. would like to acknowledge Dave Thirumalai's absolutely formative role in his own scientific career: first as a mentor and role model during postdoctoral studies, and continuing to this day as a collaborator and friend. Discussions with Dave shaped my view of what it means to be a biophysical theorist (and so much else), and I happily see his influence live on in my own scientific mentoring of students. The work described in the current article is a coda to a line of research that first began under Dave's auspices in 2013, when we developed the Wiener−Kolmogorov approach for biological signaling systems. The open question raised by our first article on the topic (can the WK bound ever be beaten via nonlinearity?) is here answered in the affirmative, but instead of weakening the applicability of WK theory, the surprising insignificance of nonlinear enhancements actually strengthens the case for it.

■ REFERENCES

(1) Berg, H. C.; Purcell, E. M. Physics of chemoreception. *Biophys. J.* **1977**, *20*, 193.

(2) von Hippel, P. H.; Berg, O. G. Facilitated target location in biological systems. *J. Biol. Chem.* **1989**, *264*, 675−678.

(3) Lestas, I.; Vinnicombe, G.; Paulsson, J. Fundamental Limits on the Suppression of Molecular Fluctuations. *Nature* **2010**, *467*, 174−178.

(4) Song, Y.; Hyeon, C. Thermodynamic uncertainty relation to assess biological processes. *J. Chem. Phys.* **2021**, *154*, 130901.

(5) Schmiedl, T.; Seifert, U. Efficiency of molecular motors at maximum power. *EPL* **2008**, *83*, 30005.

(6) Hwang, W.; Hyeon, C. Energetic costs, precision, and transport efficiency of molecular motors. *J. Phys. Chem. Lett.* **2018**, *9*, 513−520.

(7) Lynch, M.; Marinov, G. K. The bioenergetic costs of a gene. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 15690−15695.

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

(9) Cheong, R.; Rhee, A.; Wang, C. J.; Nemenman, I.; Levchenko, A. Information Transduction Capacity of Noisy Biochemical Signaling Networks. *Science* **2011**, *334*, 354−358.

(10) Bowsher, C. G.; Swain, P. S. Environmental sensing, information transfer, and cellular decision-making. *Curr. Opin. Biotechnol.* **2014**, *28*, 149−155.

(11) Becskei, A.; Serrano, L. Engineering Stability in Gene Networks by Autoregulation. *Nature* **2000**, *405*, 590−593.

(12) Thattai, M.; van Oudenaarden, A. Intrinsic Noise in Gene Regulatory Networks. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 8614−8619.

(13) Simpson, M. L.; Cox, C. D.; Sayler, G. S. Frequency Domain Analysis of Noise in Autoregulated Gene Circuits. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 4551−4556.

(14) Austin, D. W.; Allen, M. S.; McCollum, J. M.; Dar, R. D.; Wilgus, J. R.; Sayler, G. S.; Samatova, N. F.; Cox, C. D.; Simpson, M. L. Gene Network Shaping of Inherent Noise Spectra. *Nature* **2006**, *439*, 608−611.

(15) Wang, T.-L.; Kuznets-Speck, B.; Broderick, J.; Hinczewski, M. The price of a bit: energetic costs and the evolution of cellular signaling. *bioRxiv preprint* **2020**, DOI: 10.1101/2020.10.06.327700.

(16) Thomas, P. J.; Eckford, A. W. Capacity of a simple intercellular signal transduction channel. *IEEE Trans. Inf. Theory* **2016**, *62*, 7358−7382.

(17) Rhee, A.; Cheong, R.; Levchenko, A. The application of information theory to biochemical signaling systems. *Phys. Biol.* **2012**, *9*, 045011.

(18) Tkačik, G.; Callan, C. G.; Bialek, W. Information flow and optimization in transcriptional regulation. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 12265−12270.

(19) Cheong, R.; Rhee, A.; Wang, C. J.; Nemenman, I.; Levchenko, A. Information Transduction Capacity of Noisy Biochemical Signaling Networks. *Science* **2011**, *334*, 354−358.

(20) Uda, S.; Saito, T. H.; Kudo, T.; Kokaji, T.; Tsuchiya, T.; Kubota, H.; Komori, Y.; Ozaki, Y.-i.; Kuroda, S. Robustness and compensation of information transmission of signaling pathways. *Science* **2013**, *341*, 558−561.

(21) Voliotis, M.; Perrett, R. M.; McWilliams, C.; McArdle, C. A.; Bowsher, C. G. Information transfer by leaky, heterogeneous, protein

kinase signaling systems. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, E326−E333.
(22) Selimkhanov, J.; Taylor, B.; Yao, J.; Pilko, A.; Albeck, J.; Hoffmann, A.; Tsimring, L.; Wollman, R. Accurate information transmission through dynamic biochemical signaling networks. *Science* **2014**, *346*, 1370−1373.
(23) Potter, G. D.; Byrd, T. A.; Mugler, A.; Sun, B. Dynamic sampling and information encoding in biochemical networks. *Biophys. J.* **2017**, *112*, 795−804.
(24) Suderman, R.; Bachman, J. A.; Smith, A.; Sorger, P. K.; Deeds, E. J. Fundamental trade-offs between information flow in single cells and cellular populations. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 5755−5760.
(25) Keshelava, A.; Solis, G. P.; Hersch, M.; Koval, A.; Kryuchkov, M.; Bergmann, S.; Katanaev, V. L. High capacity in G protein-coupled receptor signaling. *Nat. Commun.* **2018**, *9*, 876.
(26) Hinczewski, M.; Thirumalai, D. Cellular Signaling Networks Function as Generalized Wiener-Kolmogorov Filters to Suppress Noise. *Phys. Rev. X* **2014**, *4*, 041017.
(27) Becker, N. B.; Mugler, A.; ten Wolde, P. R. Optimal Prediction by Cellular Signaling Networks. *Phys. Rev. Lett.* **2015**, *115*, 258103.
(28) Hinczewski, M.; Thirumalai, D. Noise control in gene regulatory networks with negative feedback. *J. Phys. Chem. B* **2016**, *120*, 6166−6177.
(29) Samanta, H. S.; Hinczewski, M.; Thirumalai, D. Optimal information transfer in enzymatic networks: A field theoretic formulation. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **2017**, *96*, 012406.
(30) Zechner, C.; Seelig, G.; Rullan, M.; Khammash, M. Molecular circuits for dynamic noise filtering. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 4729−4734.
(31) Wiener, N. *Extrapolation, Interpolation and Smoothing of Stationary Times Series*; Wiley: New York, 1949.
(32) Kolmogorov, A. N. Interpolation and Extrapolation of Stationary Random Sequences. *Izv. Akad. Nauk SSSR., Ser. Mater.* **1941**, *5*, 3−14.
(33) Bode, H. W.; Shannon, C. E. A Simplified Derivation of Linear Least Square Smoothing and Prediction Theory. *Proc. IRE* **1950**, *38*, 417−425.
(34) Goldbeter, A.; Koshland, D. E. An amplified sensitivity arising from covalent modification in biological-systems. *Proc. Natl. Acad. Sci. U. S. A.* **1981**, *78*, 6840−6844.
(35) Martín, H.; Flández, M.; Nombela, C.; Molina, M. Protein phosphatases in MAPK signalling: we keep learning from yeast. *Mol. Microbiol.* **2005**, *58*, 6−16.
(36) van Kampen, N. G. *Stochastic Processes in Physics and Chemistry*; Elsevier: Amsterdam, 2007.
(37) Sturm, O.; Orton, R.; Grindlay, J.; Birtwistle, M.; Vyshemirsky, V.; Gilbert, D.; Calder, M.; Pitt, A.; Kholodenko, B.; Kolch, W. The mammalian MAPK/ERK pathway exhibits properties of a negative feedback amplifier. *Sci. Signaling* **2010**, *3* (153), ra90.
(38) Hathcock, D.; Sheehy, J.; Weisenberger, C.; Ilker, E.; Hinczewski, M. Noise filtering and prediction in biological signaling networks. *IEEE Trans. Mol. Biol. Multi-Scale Commun.* **2016**, *2*, 16−30.
(39) Gillespie, D. T. The Chemical Langevin Equation. *J. Chem. Phys.* **2000**, *113*, 297−306.
(40) Detwiler, P. B.; Ramanathan, S.; Sengupta, A.; Shraiman, B. I. Engineering aspects of enzymatic signal transduction: Photoreceptors in the retina. *Biophys. J.* **2000**, *79*, 2801−2817.
(41) Alon, U. *An Introduction to Systems Biology: Design Principles of Biological Circuits*; CRC Press, 2020.
(42) Mugler, A.; Walczak, A. M.; Wiggins, C. H. Spectral solutions to stochastic models of gene expression with bursts and regulation. *Phys. Rev. E* **2009**, *80*, 041921.
(43) Walczak, A. M.; Mugler, A.; Wiggins, C. H. A stochastic spectral analysis of transcriptional regulatory cascades. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 6529−6534.
(44) Ogura, H. Orthogonal functionals of the Poisson process. *IEEE Trans. Inf. Theory* **1972**, *18*, 473−481.
(45) Nevozhay, D.; Adams, R. M.; Murphy, K. F.; Josic, K.; Balazsi, G. Negative Autoregulation Linearizes the Dose-Response and Suppresses the Heterogeneity of Gene Expression. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 5123−5128.
(46) Cai, L.; Friedman, N.; Xie, X. S. Stochastic Protein Expression in Individual Cells at the Single Molecule Level. *Nature* **2006**, *440*, 358−362.
(47) Gillespie, D. T. Exact Stochastic Simulation of Coupled Chemical Reactions. *J. Phys. Chem.* **1977**, *81*, 2340−2361.
(48) Kuehn, C. *Control of Self-Organizing Nonlinear Systems*; Springer, 2016; pp 253−271.
(49) Horowitz, J. M.; Gingrich, T. R. Thermodynamic uncertainty relations constrain non-equilibrium fluctuations. *Nat. Phys.* **2020**, *16*, 15−20.

https://doi.org/10.1021/acs.jpcb.1c07894
J. Phys. Chem. B 2021, 125, 12698−12711