Cellular signaling beyond the Wiener-Kolmogorov limit

Casey Weisenberger[1,+], **David Hathcock**[2,+], **and Michael Hinczewski**[1,*]

[1]Department of Physics, Case Western Reserve University, Cleveland, Ohio
[2]Department of Physics, Cornell University, Ithaca, New York
[*]mxh605@case.edu
[+]these authors contributed equally to this work

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 the bound being broken. To accomplish this, we develop a theoretical framework for multi-level 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.

1 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 re-interpretaton 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 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], drawing 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 closed form 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.

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 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.

2 Results

2.1 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 Fig. 1, we have an input chemical species X_0 followed by N downstream species $X_1, \ldots, X_N$. 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 $x_i = 0, 1, 2, \ldots$. Hence the state of the system can be represented by the vector $\boldsymbol{x} = (x_0, x_1, \ldots, x_N)$. Stochastic transitions between states are governed by an infinite-dimensional Markovian transition rate matrix W. The element $W_{\boldsymbol{x}',\boldsymbol{x}}$ of this matrix represents the probability per unit time to observe state $\boldsymbol{x}'$ at the next infinitesimal time step, given that the current state is $\boldsymbol{x}$. 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 $p_{\boldsymbol{x}}(t)$ of being in state $\boldsymbol{x}$ at time t evolves according to the corresponding master equation[36],

$$\frac{d}{dt}p_{\boldsymbol{x}}(t) = \sum_{\boldsymbol{x}'}\left[W_{\boldsymbol{x},\boldsymbol{x}'}p_{\boldsymbol{x}'}(t) - W_{\boldsymbol{x}',\boldsymbol{x}}p_{\boldsymbol{x}}(t)\right]. \qquad (1)$$

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

$$0 = \sum_{\boldsymbol{x}'}\left[W_{\boldsymbol{x},\boldsymbol{x}'}\mathcal{P}_{\boldsymbol{x}'} - W_{\boldsymbol{x}',\boldsymbol{x}}\mathcal{P}_{\boldsymbol{x}}\right]. \qquad (2)$$

All physical observables we consider can be expressed as averages over this stationary distribution. If $f(\boldsymbol{x})$ is

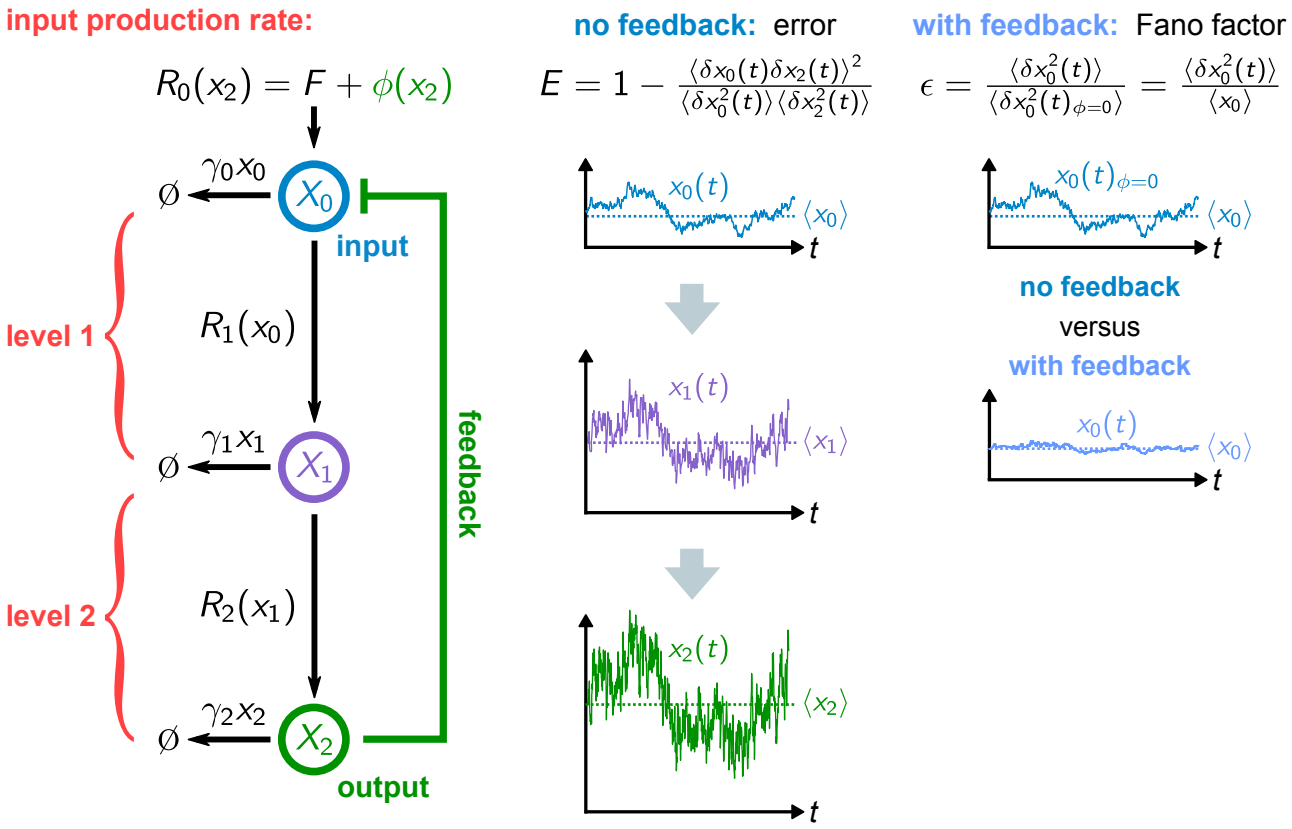

Figure 1. Overview of the N-level signaling cascade model, showing an example with $N = 2$. The signal from input species X_0 is propagated through to output species X_N, with the possibility of feedback back to the input. In the absence of feedback, the signal fidelity is measured via the error E, defined 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.

some function of state $\boldsymbol{x}$, we will use $\langle f \rangle \equiv \sum_{\boldsymbol{x}} f(\boldsymbol{x}) \mathcal{P}_{\boldsymbol{x}}$ 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 to 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],

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

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 non-exponential 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),

$$\begin{aligned} 0 = \sum_{i=0}^{N} \Big\{ &\gamma_i \left[(x_i+1)\mathcal{P}_{\boldsymbol{x}+\boldsymbol{e}_i} - x_i \mathcal{P}_{\boldsymbol{x}}\right] \\ &+ R_i(x_{i-1}) \left[\mathcal{P}_{\boldsymbol{x}-\boldsymbol{e}_i} - \mathcal{P}_{\boldsymbol{x}}\right] \Big\}. \end{aligned} \tag{5}$$

For compactness of notation, we define $x_{-1} \equiv x_N$ and introduce the $(N+1)$-dimensional unit vectors $\boldsymbol{e}_i$, where $\boldsymbol{e}_0 = (1,0,\ldots,0)$, $\boldsymbol{e}_1 = (0,1,0,\ldots,0)$, $\boldsymbol{e}_2 = (0,0,1,0,\ldots,0)$ and so on. Eq. (5) is generally analytically intractable, in the sense that we cannot usually directly solve it to find the stationary distribution $\mathcal{P}_{\boldsymbol{x}}$. 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.

2.2 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, 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 Supplementary Information (SI).

2.2.1 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,

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

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:

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

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]. The Langevin equations corresponding to Eq. (5) are:

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

where the $n_i(t)$ 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)-(7). The result is:

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

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

2.2.2 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,

$$\tilde{s}(t) = \int_{-\infty}^{\infty} dt'\, H(t-t')\left[s(t') + n(t')\right]. \tag{10}$$

The details of the functions $H(t)$ and $n(t)$, as derived from Eq. (9), are given in the SI Sec. 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 $0 \le \epsilon < \infty$. 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:

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

By construction, $E(s(t), \tilde{s}(t)) \le \epsilon(s(t), \tilde{s}(t))$ and in fact E has a restricted range: $0 \le E \le 1$. 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 $E_{\rm WK}$. Hence we have $\epsilon \geq E \geq E_{\rm WK}$ in general for linear systems, and $\epsilon = E = E_{\rm WK}$ when $H(t) = H_{\rm WK}(t)$.

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

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

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

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

This is a polynomial of degree $2N$ in λ, and hence has $2N$ roots. Because the coefficients of λ in the polynomial are real, the conjugate of any complex root must also be a root. Finally, because only even powers of λ appear in $B(\lambda)$, the negative of a root is also a root. Putting all these facts together ensures that there will always be N roots λ_j where $\mathrm{Re}(\lambda_j) > 0$. Moreover, among the set of λ_j, any complex roots come in conjugate pairs. This guarantees that the expression for $E_{\rm WK}$ in Eq. (13) is always real. Note that the choice of ordering of the roots λ_j, $j = 1, \ldots, N$ is arbitrary, since it does not affect the result. Within the linear approximation, $E_{\rm WK}$ gives a lower bound on the achievable E, and and hence an upper bound on the maximum mutual information between input and output, $\mathcal{I}_{\rm max} = -\frac{1}{2}\log_2 E_{\rm WK}$.

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:

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

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 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 $H_{\rm WK}(t)$, and hence $E = E_{\rm WK}$. (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 non-trivial; 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 $H_{\rm WK}(t)$ up to a proportionality constant. However, as shown in SI Sec. 1, for the no-feedback N-level model we can achieve $H(t) \propto H_{\rm WK}(t)$ 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 $E_{\rm WK}$ explicitly when this scenario is true. With the aid of the recursion relation in Eq. (16), we can then write $E_{\rm WK}$ in this case as:

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

where $\ell_i \equiv \gamma_{i-1}\Lambda_i/\gamma_0$ 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]:

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

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 is 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 Sec. 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 $\epsilon \geq E_{\mathrm{WK}}$.

Unlike the no-feedback cascade, where we can in principle tune the biological parameters so that $E = E_{\mathrm{WK}}$, for the linearized negative feedback system we can only asymptotically approach the bound from above, $\epsilon \to E_{\mathrm{WK}}$. 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,

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

where

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

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},

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

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

2.2.3 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)), depend 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. Achieving systems with better optimal performance can be expensive. To illustrate this, 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. 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 rougly a 10^4 increase in Λ_{eff}, assuming we are in the regime where $\Lambda_{\mathrm{eff}} \gg 1$. This extreme cost of eliminating fluctuations via negative feedback[3] has to be borne across the whole cascade:

since $\Lambda_{\rm eff}$ in Eq. (21) 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 $\Lambda_{\rm eff}$.

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 $\mathcal{I}_{\rm max} = -(1/2)\log E_{\rm WK}$ by 1 bit. In the limit of $\ell_i \gg 1$ in Eq. (18), this can be achieved for example by increasing every ℓ_i by a factor of ≈ 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 constants associated with level i as $\Lambda_i = \sigma_{i0}/(\bar{x}_{i-1}\gamma_0)$ and $\ell_i = (\gamma_{i-1}/\gamma_0)\Lambda_i = (\sigma_{i0}/\sigma_{i-1,0})(\gamma_{i-1}/\gamma_0)^2$. Hence increasing ℓ_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. Note the Λ_i parameter has a simple physical interpretation here: the average number of i molecules produced per molecule of X_{i-1} during the characteristic time interval γ_0^{-1} of input fluctuations. 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 ~ 1 to 3 bits[18–25], with most systems near the lower end of the spectrum.

2.3 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 $H_{\rm WK}(t)$ and the the corresponding error bound $E_{\rm WK}$; iv) explore if and under what conditions the system can reach optimality. But the procedure leaves open an important question: is the resulting bound $E_{\rm WK}$ 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 $E_{\rm WK}$ 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.

2.3.1 Nonlinearity in the $N = 1$ model without feedback

Ref. 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,

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

Here, $v_n(x_0;\bar{x}_0)$ is a polynomial of nth degree in x_0, which depends on $\bar{x}_0 = F/\gamma_0$ as a parameter. The functions $v_n(x_0;\bar{x}_0)$ are variants of so-called Poisson-Charlier polynomials, whose properties are described in detail in the SI Sec. 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]

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

The first few polynomials are given by:

$$\begin{aligned} &v_0(x;\bar{x}) = 1, \quad v_1(x;\bar{x}) = x - \bar{x}, \\ &v_2(x;\bar{x}) = (x-\bar{x})^2 - x. \end{aligned} \tag{26}$$

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

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

Using Eq. (26) we can write the first two coefficients as

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

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:

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

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 Fig. 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 Fig. 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 $R_1(x_0)$.

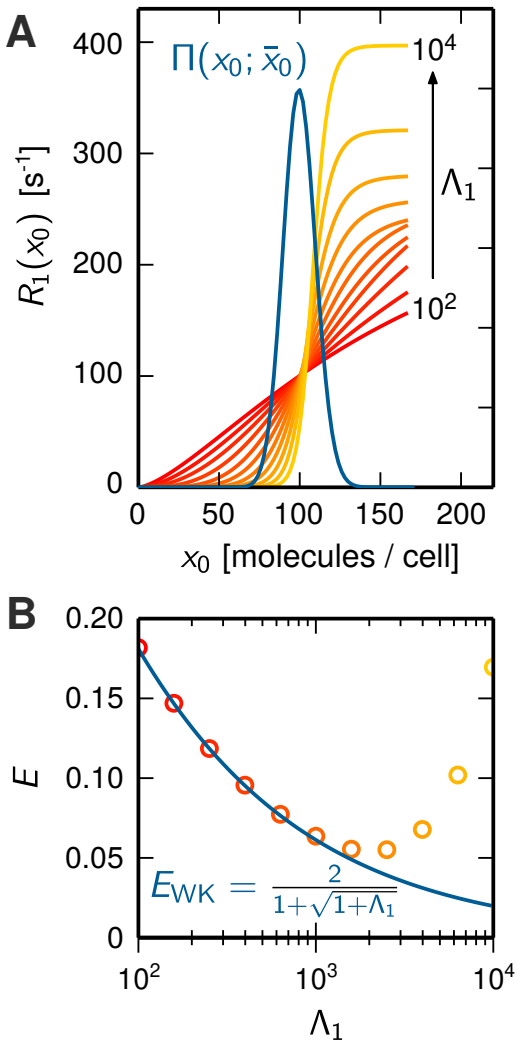

Figure 2. The case of a no-feedback $N = 1$ signaling model with a nonlinear production function $R_1(x_0)$ for parameters: $F = 1\ \mathrm{s}^{-1}$, $\gamma_0 = 0.01\ \mathrm{s}^{-1}$, $\sigma_{10} = 100\ \mathrm{s}^{-1}$. A) Examples of a variety of production functions $R_1(x_0)$, colored from red 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.

2.4 Nonlinearity in the $N = 1$ TetR negative feedback circuit

Ref. 28 studied an $N = 1$ negative feedback loop inspired by data from an experimental synthetic yeast gene circuit[45]. In this circuit, TetR messenger RNA (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 messenger RNA. The model is similar to Eq. (8) when $N = 1$,

$$\begin{aligned}\frac{d}{dt}x_0(t) &= R_0(x_1) - \gamma_0 x_0(t) + n_0(t),\\ \frac{d}{dt}x_1(t) &= R_1(x_0) - \gamma_1 x_1(t) - \Gamma(x_1(t)) + n_1(t),\end{aligned} \tag{31}$$

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

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

Note that this model has 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 $A_i, \nu_i, \theta_i, i = 1, 2$ 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 Fig. 3 show the optimization results for different Λ_1 in the range $\Lambda_1 = 2 - 10$, comparable to experimental estimates[46]. The $E_{\rm WK}$ 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

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

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

Based on 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.

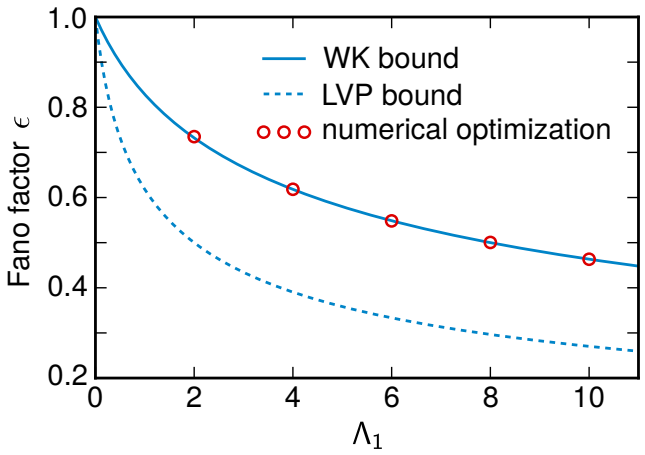

Figure 3. The Fano factor ϵ for the TetR $N = 1$ negative feedback loop of Ref. 28. Numerical optimization results are shown as circles, while the WK and LVP lower bounds (Eqs. (23) and (34) respectively) are shown as solid and dashed curves. Figure adapted from Ref. 28.

3 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 $E_{\rm WK}$. The first is from an $N = 2$ no-feedback cascade with a 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 $E \ge E_{\rm WK}$.) 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 a $N = 1$ 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).

3.1 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 $\boldsymbol{x} = (x_0, x_1, \ldots, x_N)$ that describes the full state of our system, we will define the N-dimensional truncated vector $\hat{\boldsymbol{x}} = (x_0, x_1, \ldots, x_{N-1})$ that is missing the final component x_N. In a similar way we define the truncated N-dimensional unit vectors $\hat{\boldsymbol{e}}_i$, $i = 1, \ldots, N-1$, where $\hat{\boldsymbol{e}}_0 = (1, 0, \ldots, 0)$, $\hat{\boldsymbol{e}}_1 = (0, 1, \ldots, 0)$, and so on until $\hat{\boldsymbol{e}}_{N-1} = (0, \ldots, 0, 1)$. Consider the following generating function derived from the stationary distribution $\mathcal{P}_{\mathbf{x}}$,

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

The subscript $\hat{\boldsymbol{x}}$ denotes the fact that $H_{\hat{\boldsymbol{x}}}(y)$ 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,

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

Here, the pth derivative of $H_{\hat{\boldsymbol{x}}}(y)$ with respect to y is denoted as $H^{(p)}_{\hat{\boldsymbol{x}}}(y) \equiv (d^p/dy^p) H_{\hat{\boldsymbol{x}}}(y)$. 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,

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

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 $H^{(p)}_{\hat{\boldsymbol{x}}}(1)$:

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

Here, $\sum_{\boldsymbol{x}} \equiv \sum_{x_0=0}^{\infty} \cdots \sum_{x_N=0}^{\infty}$ and $\sum_{\hat{\boldsymbol{x}}} \equiv \sum_{x_0=0}^{\infty} \cdots \sum_{x_{N-1}=0}^{\infty}$. 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

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

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 $H^{(p)}_{\hat{\boldsymbol{x}}}(1)$ 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 the SI Sec. 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,

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

Each expansion coefficient is given by the analogue of Eq. (27), averaging over a Poisson distribution:

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

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

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

Here $\Pi(\hat{\boldsymbol{x}};\hat{\bar{\boldsymbol{x}}})$ is a multi-dimensional Poisson distribution,

$$\begin{aligned}&\Pi(\hat{\boldsymbol{x}};\hat{\bar{\boldsymbol{x}}})\\ &\quad\equiv \Pi(x_0;\bar{x}_0)\Pi(x_1;\bar{x}_1)\cdots\Pi(x_{N-1};\bar{x}_{N-1}).\end{aligned} \tag{44}$$

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

$$\begin{aligned}&v_{\hat{\boldsymbol{n}}}(\hat{\boldsymbol{x}};\hat{\bar{\boldsymbol{x}}})\\ &\equiv v_{n_0}(x_0;\bar{x}_0)v_{n_1}(x_1;\bar{x}_1)\cdots v_{n_{N-1}}(x_{N-1};\bar{x}_{N-1}),\end{aligned} \tag{45}$$

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

$$J_{\hat{\boldsymbol{x}}}^{(p)} = \sum_{\hat{\boldsymbol{n}}} \mu_{\hat{\boldsymbol{n}}}^{(p)} v_{\hat{\boldsymbol{n}}}(\hat{\boldsymbol{x}};\hat{\bar{\boldsymbol{x}}}), \tag{46}$$

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

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

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

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

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

$$\begin{aligned}&\mu_{\hat{\boldsymbol{n}}}^{(p)}\\ &\quad= \frac{p\nu_{\hat{\boldsymbol{n}}}^{(p-1,N)} + \sum_{i=1}^{N-1}\left(\bar{x}_i^{-1}\nu_{\hat{\boldsymbol{n}}-\hat{\boldsymbol{e}}_i}^{(p,i)} - \gamma_i\mu_{\hat{\boldsymbol{n}}-\hat{\boldsymbol{e}}_i}^{(p)}\right)}{p\gamma_N + \sum_{i=0}^{N-1} n_i\gamma_i},\end{aligned} \tag{49}$$

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

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

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 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. But 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 $\nu_{\hat{\boldsymbol{n}}}^{(p,i)}$ that appear in Eq. (49).

3. We solve the recursive system of equations in Eq. (49) for $\mu_{\hat{\boldsymbol{0}}+\hat{\boldsymbol{e}}_0}^{(1)}$, $\mu_{\hat{\boldsymbol{0}}}^{(1)}$, and $\mu_{\hat{\boldsymbol{0}}}^{(2)}$, and 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:

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

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, $E \geq E_{\mathrm{WK}}$ always, where E_{WK} is given by Eq. (13).

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 Fig. 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}, Fig. 4 shows $\log_{10}|E_{\mathrm{min}}/E_{\mathrm{WK}} - 1|$ with the cool colored contours indicating $E_{\mathrm{min}} > E_{\mathrm{WK}}$ and warm colored contours indicating $E_{\mathrm{min}} < E_{\mathrm{WK}}$. In the purely linear case described earlier, we found that $E = E_{\mathrm{WK}}$ when the conditions from Eq. (17) are satisfied, which corresponds to $r = \sqrt{1+\Lambda_1}$, shown as as 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 Fig. 4). However the improvement relative to the WK bound is exceedingly small, roughly $\sim 0.001 - 0.01\%$ 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 $R_2(x_1) = \sigma_{21}x_1 + \sigma_{22}v_2(x_1;\bar{x}_1)$, 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 $E = E_{\mathrm{WK}}$ 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:

$$E_{\mathrm{WK}} = 1 - \frac{\rho(r-1)r^2}{(1+r)\left(1+\sqrt{1+\rho r^2}\right)^2}. \quad (54)$$

Let us focus on the regime where signaling is at least as effective as in many experimentally measured cascades[18–25], which means $I_{\mathrm{max}} \gtrsim 1$ bit or equivalently $E_{\mathrm{WK}} \lesssim 1/4$. 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 $E - E_{\mathrm{WK}}$ to second order in the perturbation parameter σ_{22},

$$E - E_{\mathrm{WK}} \approx \frac{2}{\gamma_0\rho r}\sigma_{22} + \frac{2\bar{x}_0}{\gamma_0^2\rho^2}\sigma_{22}^2. \quad (55)$$

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

$$E_{\mathrm{min}} - E_{\mathrm{WK}} \approx -\frac{1}{2\bar{x}_0 r^2}. \quad (56)$$

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 $E_{\mathrm{WK}} \sim 2/r$ for $\rho, r \gg 1$, and hence the relative magnitude $E_{\mathrm{min}}/E_{\mathrm{WK}} - 1 \approx -(4\bar{x}_0 r)^{-1}$. The negligible scale of the improvement over E_{WK} is consistent with the numerical results of Fig. 4, though the latter was calculated over a broader portion of the parameter space.

3.2 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 $N = 1$ 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 $\phi(x_1)$,

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

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.

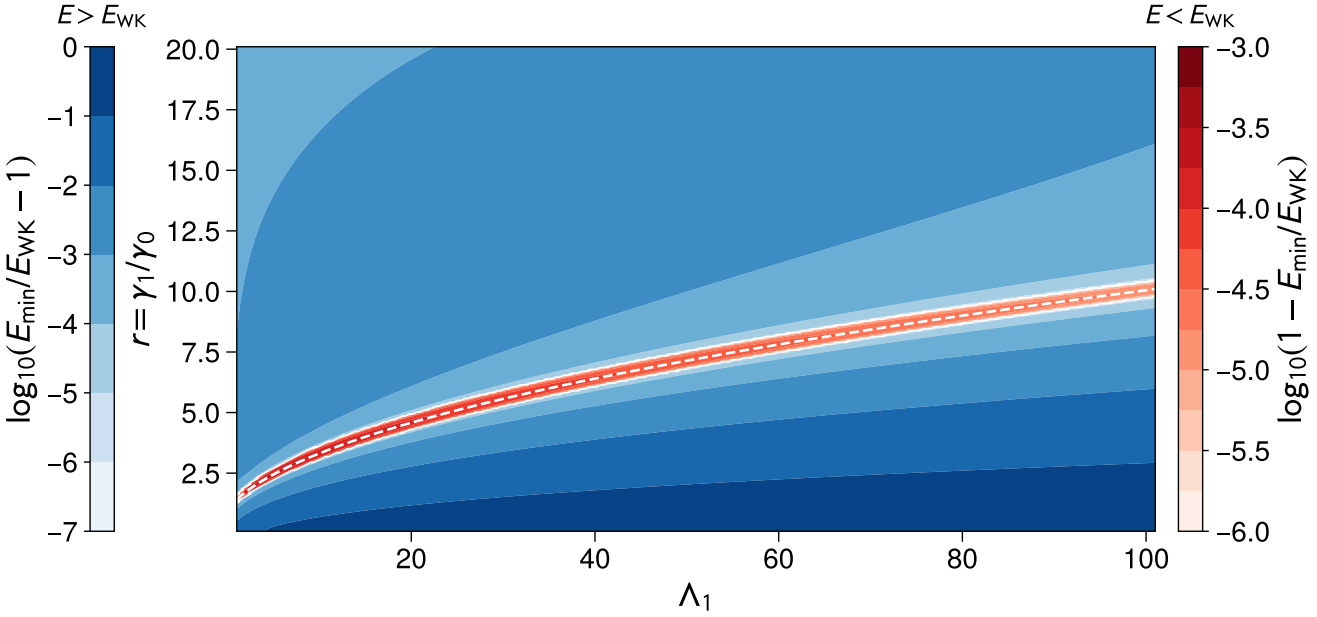

Figure 4. Contour plot of $\log_{10}|E_{\min}/E_{\mathrm{WK}} - 1|$ for the $N = 2$ cascade with linear $R_1(x_0)$ and quadratic $R_2(x_1)$. The minimum value of the error $E_{\min}$ at a given $r = \gamma_1/\gamma_0$ and Λ_1 is found by numerical minimization with respect to γ_2 and σ_{22}, with fixed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

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 $\Lambda_1 \geq 2$). Based on the 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 $\mu^{(p)}_{\hat{\boldsymbol{n}}}$ 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 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 $\mathcal{P}_{\boldsymbol{x}}$, 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 $\mathcal{P}_{\boldsymbol{x}}$ outside the allowable range of $\boldsymbol{x}$ 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 $\gamma_1 \to \infty$). Fig. 5A shows numerical results 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 Fig. 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 Fig. 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.

4 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. And 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 Fig. 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 of deriving 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

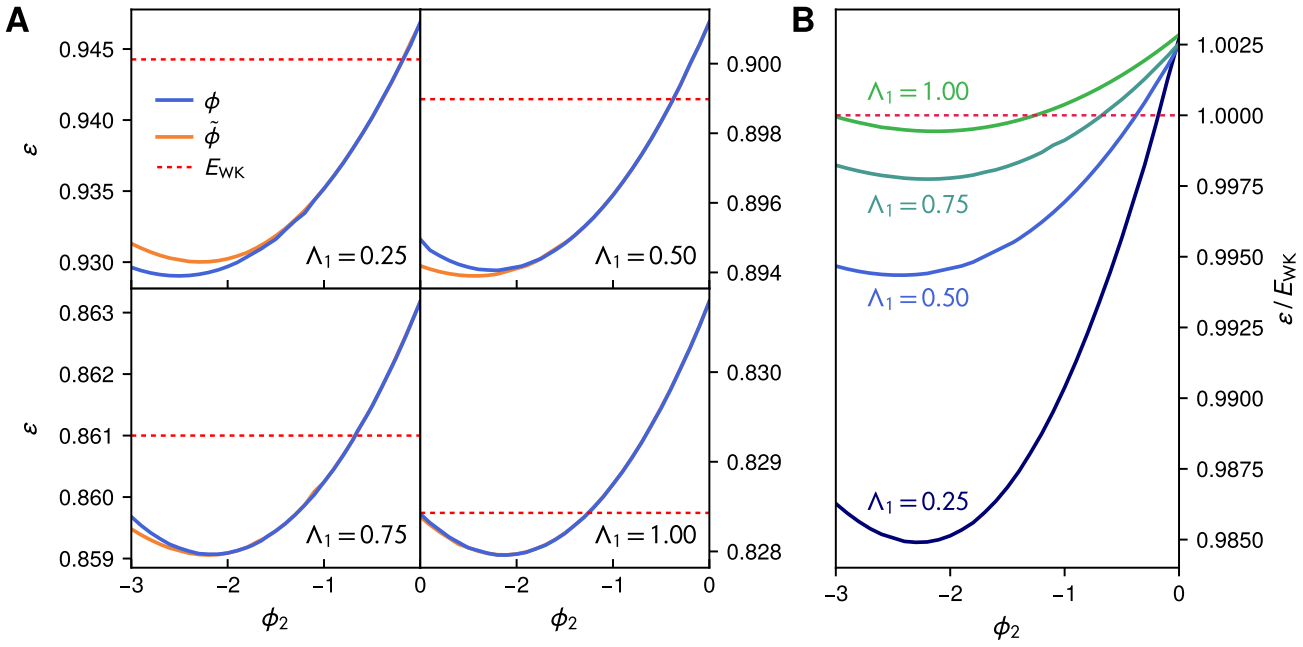

Figure 5. A) Numerically calculated Fano factor ϵ in the $N = 1$ nonlinear feedback system. The plots show ϵ versus ϕ_2 for the quadratic feedback function $\phi(x_1)$ (blue) from Eq. (57) and the monotonic alternative $\tilde{\phi}(x_1)$ (orange) from Eq. (58), using the parameters described in the 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 $\epsilon/E_{WK} = 1$.

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 between 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 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.

Acknowledgements

This article was submitted as part of the Dave Thirumalai Festschrift. M.H. would like to acknowledge Dave's absolutely formative role in his own scientific career: first as a mentor and role model during my 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.* **20**, 193 (1977).
2. von Hippel, P. H. & Berg, O. G. Facilitated target location in biological systems. *J. Biol. Chem.* **264**, 675–678 (1989).
3. Lestas, I., Vinnicombe, G. & Paulsson, J. Fundamental limits on the suppression of molecular fluctuations. *Nature* **467**, 174–178 (2010).
4. Song, Y. & Hyeon, C. Thermodynamic uncertainty relation to assess biological processes. *J. Chem. Phys.* **154**, 130901 (2021).
5. Schmiedl, T. & Seifert, U. Efficiency of molecular motors at maximum power. *EPL* **83**, 30005 (2008).
6. Hwang, W. & Hyeon, C. Energetic costs, precision, and transport efficiency of molecular motors. *J. Phys. Chem. Lett.* **9**, 513–520 (2018).
7. Lynch, M. & Marinov, G. K. The bioenergetic costs of a gene. *Proc. Natl. Acad. Sci.* **112**, 15690–15695 (2015).
8. Ilker, E. & Hinczewski, M. Modeling the growth of organisms validates a general relation between metabolic costs and natural selection. *Phys. Rev. Lett.* **122**, 238101 (2019).
9. Cheong, R., Rhee, A., Wang, C. J., Nemenman, I. & Levchenko, A. Information Transduction Capacity of Noisy Biochemical Signaling networks. *Science* **334**, 354–358 (2011).
10. Bowsher, C. G. & Swain, P. S. Environmental sensing, information transfer, and cellular decision-making. *Curr. Opin. Biotech.* **28**, 149–155 (2014).
11. Becskei, A. & Serrano, L. Engineering stability in gene networks by autoregulation. *Nature* **405**, 590–593 (2000).
12. Thattai, M. & van Oudenaarden, A. Intrinsic noise in gene regulatory networks. *Proc. Natl. Acad. Sci. USA* **98**, 8614–8619 (2001).
13. Simpson, M. L., Cox, C. D. & Sayler, G. S. Frequency domain analysis of noise in autoregulated gene circuits. *Proc. Natl. Acad. Sci. USA* **100**, 4551–4556 (2003).
14. Austin, D. W. *et al.* Gene network shaping of inherent noise spectra. *Nature* **439**, 608–611 (2006).
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* 2020.10.06.327700 (2020).
16. Thomas, P. J. & Eckford, A. W. Capacity of a simple intercellular signal transduction channel. *IEEE Trans. Inf. Theory* **62**, 7358–7382 (2016).
17. Rhee, A., Cheong, R. & Levchenko, A. The application of information theory to biochemical signaling systems. *Phys. Biol.* **9**, 045011 (2012).
18. Tkačik, G., Callan, C. G. & Bialek, W. Information flow and optimization in transcriptional regulation. *Proc. Natl. Acad. Sci.* **105**, 12265–12270 (2008).
19. Cheong, R., Rhee, A., Wang, C. J., Nemenman, I. & Levchenko, A. Information transduction capacity of noisy biochemical signaling networks. *Science* **334**, 354–358 (2011).
20. Uda, S. *et al.* Robustness and compensation of information transmission of signaling pathways. *Science* **341**, 558–561 (2013).
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.* **111**, E326–E333 (2014).
22. Selimkhanov, J. *et al.* Accurate information transmission through dynamic biochemical signaling networks. *Science* **346**, 1370–1373 (2014).
23. Potter, G. D., Byrd, T. A., Mugler, A. & Sun, B. Dynamic sampling and information encoding in biochemical networks. *Biophys. J.* **112**, 795–804 (2017).
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.* **114**, 5755–5760 (2017).
25. Keshelava, A. *et al.* High capacity in g protein-coupled receptor signaling. *Nat. Commun.* **9**, 1–8 (2018).

26. Hinczewski, M. & Thirumalai, D. Cellular signaling networks function as generalized Wiener-Kolmogorov filters to suppress noise. *Phys. Rev. X* **4** (041017, 2014).

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

28. Hinczewski, M. & Thirumalai, D. Noise control in gene regulatory networks with negative feedback. *J. Phys. Chem. B* **120**, 6166–6177 (2016).

29. Samanta, H. S., Hinczewski, M. & Thirumalai, D. Optimal information transfer in enzymatic networks: A field theoretic formulation. *Phys. Rev. E* **96**, 012406 (2017).

30. Zechner, C., Seelig, G., Rullan, M. & Khammash, M. Molecular circuits for dynamic noise filtering. *Proc. Natl. Acad. Sci. USA* **113**, 4729–4734 (2016).

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. Mat.* **5**, 3–14 (1941).

33. Bode, H. W. & Shannon, C. E. A simplified derivation of linear least square smoothing and prediction theory. *Proc. Inst. Radio. Engin.* **38**, 417–425 (1950).

34. Goldbeter, A. & Koshland, D. E. An amplified sensitivity arising from covalent modification in biological-systems. *Proc. Natl. Acad. Sci. U.S.A.* **78**, 6840–6844 (1981).

35. Martín, H., Flández, M., Nombela, C. & Molina, M. Protein phosphatases in MAPK signalling: we keep learning from yeast. *Mol. Microbiol.* **58**, 6–16 (2005).

36. van Kampen, N. G. *Stochastic processes in physics and chemistry* (Elsevier, Amsterdam, 2007).

37. Sturm, O. *et al.* The mammalian MAPK/ERK pathway exhibits properties of a negative feedback amplifier. *Sci. Signal.* **3**, *ra90*, 1–7 (2010).

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.* **2**, 16–30 (2016).

39. Gillespie, D. T. The chemical Langevin equation. *J. Chem. Phys.* **113**, 297–306 (2000).

40. Detwiler, P. B., Ramanathan, S., Sengupta, A. & Shraiman, B. I. Engineering aspects of enzymatic signal transduction: Photoreceptors in the retina. *Biophys. J.* **79**, 2801–2817 (2000).

41. Alon, U. *An Introduction to Systems Biology: Design Principles of Biological Circuits* (Chapman and Hall/CRC, 2006).

42. Mugler, A., Walczak, A. M. & Wiggins, C. H. Spectral solutions to stochastic models of gene expression with bursts and regulation. *Phys. Rev. E* **80**, 041921 (2009).

43. Walczak, A. M., Mugler, A. & Wiggins, C. H. A stochastic spectral analysis of transcriptional regulatory cascades. *Proc. Natl. Acad. Sci. USA* **106**, 6529–6534 (2009).

44. Ogura, H. Orthogonal functionals of the Poisson process. *IEEE Trans. Info. Theory* **18**, 473–481 (1972).

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. USA* **106**, 5123–5128 (2009).

46. Cai, L., Friedman, N. & Xie, X. S. Stochastic protein expression in individual cells at the single molecule level. *Nature* **440**, 358–362 (2006).

47. Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81**, 2340–2361 (1977).

48. Kuehn, C. Moment closure—a brief review. In *Control of self-organizing nonlinear systems*, 253–271 (Springer, 2016).

Supplementary Information:

Cellular signaling beyond the Wiener-Kolmogorov limit

Contents

1 Deriving the WK optimal filter results for the multi-level cascade without feedback

1.1 Mapping the system onto a noise filter

The starting point for the derivation is the system of equations in main text Eq. (9), with $\phi_1 = 0$ in the absence of feedback:

$$\begin{aligned}\frac{d}{dt}\delta x_0(t) &= -\gamma_0 \delta x_0(t) + n_0(t),\\ \frac{d}{dt}\delta x_i(t) &= -\gamma_i \delta x_i(t) + \sigma_{i1}\delta x_{i-1}(t) + n_i(t), \qquad i > 0,\end{aligned} \tag{S1}$$

where the Gaussian noise functions satisfy $\langle n_i(t) n_j(t')\rangle = 2\delta_{ij}\gamma_i \bar{x}_i \delta(t-t')$. Taking the Fourier transform of Eq. (S1), we can solve the system of equations for the fluctuation functions $\delta x_i(\omega)$ in Fourier space,

$$\begin{aligned}\delta x_0(\omega) &= \frac{n_0(\omega)}{\gamma_0 - i\omega},\\ \delta x_j(\omega) &= \frac{1}{\gamma_j - i\omega}\left(\sigma_{j1}\delta x_{j-1}(\omega) + n_j(\omega)\right), \quad j > 0,\end{aligned} \tag{S2}$$

with $f(\omega)$ denoting the Fourier transform of a function $f(t)$. Iteratively plugging the result for $\delta x_{j-1}(\omega)$ into the $\delta x_j(\omega)$ equation, starting from $j = 1$, we can solve Eq. (S2) to get the following expressions for the Fourier space input and output fluctuations:

$$\begin{aligned}\delta x_0(\omega) &= \frac{n_0(\omega)}{\gamma_0 - i\omega},\\ \delta x_N(\omega) &= \left(\prod_{j=1}^{N}\frac{\sigma_{j1}}{\gamma_j - i\omega}\right)\left[\delta x_0(\omega) + \sum_{j=1}^{N} n_j(\omega)\prod_{k=1}^{j-1}\frac{\gamma_k - i\omega}{\sigma_{j1}\sigma_{k1}}\right].\end{aligned} \tag{S3}$$

Let us compare the result for $\delta x_N(\omega)$ to the Fourier transform of main text Eq. (10), the noise filter convolution integral:

$$\tilde{s}(\omega) = H(\omega)(s(\omega)+n(\omega)). \tag{S4}$$

We can make a mapping of the system to a linear noise filter with the following choice of estimate, signal, noise, and filter function:

$$\tilde{s}(\omega) = \delta x_N(\omega), \qquad s(\omega) = \delta x_0(\omega), \qquad n(\omega) = \sum_{j=1}^{N} n_j(\omega) \prod_{k=1}^{j-1} \frac{\gamma_k - i\omega}{\sigma_{j1}\sigma_{k1}}, \qquad H(\omega) = \prod_{j=1}^{N} \frac{\sigma_{j1}}{\gamma_j - i\omega}. \tag{S5}$$

1.2 Concise overview of WK optimal filter theory

To apply WK theory to our problem, let us summarize its main results (see Ref. 1 for a more detailed review). Given a Fourier-transformed signal and noise functions $s(\omega)$ and $n(\omega)$, let us denote the corresponding power spectra $P_s(\omega)$ and $P_n(\omega)$. The spectra are defined through the relation $\langle f(\omega)f(\omega')\rangle = 2\pi P_f(\omega)\delta(\omega+\omega')$, where $f = s$ or n. For the signal corrupted by noise, $y(\omega) \equiv s(\omega)+n(\omega)$, the corresponding power spectrum is $P_y(\omega) = P_s(\omega)+P_n(\omega)$ if the noise is uncorrelated with the signal. This is indeed the case, since the Gaussian noise functions $n_j(\omega)$ in Eq. (S5) that contribute to $n(\omega)$ are uncorrelated with $n_0(\omega)$, the function that enters into the signal $\delta x_0(\omega)$ in Eq. (S3).

Once $P_s(\omega)$ and $P_s(\omega)$ are specified, one can find a corresponding optimal filter function $H_{\mathrm{WK}}(\omega)$. Optimality here means that the time-domain function $H_{\mathrm{WK}}(t)$, plugged into the convolution integral of main text Eq. (10), minimizes the error $\epsilon(s(t),\tilde{s}(t))$ between the estimate and signal defined in main text Eq. (11). In Fourier space the optimal filter takes the following form if signal and noise are uncorrelated[2]:

$$H_{\mathrm{WK}}(\omega) = \frac{1}{P_y^+(\omega)} \left\{ \frac{P_s(\omega)}{(P_y^+(\omega))^*} \right\}_+ . \tag{S6}$$

The + superscripts and subscripts denote two types of causal decompositions. For example, the function $P_y^+(\omega)$ is defined via $P_y(\omega) = |P_y^+(\omega)|^2$, where the factor $P_y^+(\omega)$ is chosen such that it has no zeros or poles in the upper half-plane. This decomposition always exists for all the physical power spectra we encounter in signaling contexts. The other decomposition, denoted by $\{G(\omega)\}_+$ for a function $G(\omega)$, can be calculated from $\{G(\omega)\}_+ \equiv \mathcal{F}[\Theta(t)\mathcal{F}^{-1}[G(\omega)]]$. Here $\mathcal{F}[f(t)]$ indicates the Fourier transform of a function $f(t)$, $\mathcal{F}^{-1}$ the inverse Fourier transform, and $\Theta(t)$ is a unit step function[3]. In practice, it is often convenient to calculate it through an alternative method: doing a partial fraction expansion of $G(\omega)$ and keeping only those terms with no poles in the upper half-plane.

To find the lower bound on ϵ, we inverse Fourier transform $H_{\mathrm{WK}}(\omega)$ back to the time domain. The minimum error E_{WK} can then be expressed compactly in the following form, which is convenient for calculations:

$$E_{\mathrm{WK}} = 1 - \frac{1}{C_s(0)} \int_0^\infty dt H_{\mathrm{WK}}(t) C_s(t), \tag{S7}$$

where $C_s(t) = \mathcal{F}^{-1}[P_s(\omega)]$ is the signal autocorrelation function, given by the inverse Fourier transform of its power spectrum.

1.3 Calculating the optimal filter function H_{WK}

Given Eqs. (S3), (S6), and the properties of the Gaussian noise functions $n_j(t)$, which in Fourier space satisfy $\langle n_i(\omega)n_j(\omega)\rangle = 4\pi\delta_{ij}\gamma_i\bar{x}_i\delta(t-t')$, the power spectra for the signal and noise can be written as:

$$P_s(\omega) = \frac{2F}{\omega^2+\gamma_0^2}, \tag{S8a}$$

$$P_n(\omega) = \frac{2F}{\gamma_0^2} \sum_{j=1}^{N} \frac{1}{\Lambda_j} \left[\prod_{k=1}^{j-1} \frac{(\omega^2+\gamma_k^2)}{\gamma_0\gamma_k\Lambda_k} \right]. \tag{S8b}$$

Here we have used the facts that $\bar{x}_0 = F/\gamma_0$, $\bar{x}_i = \sigma_{i0}/\gamma_i$ for $i > 0$, and have introduced the dimensionless constants $\Lambda_j \equiv \bar{x}_{j-1}\sigma_{j1}^2/(\sigma_{j0}\gamma_0)$. Summing $P_s(\omega)$ and $P_n(\omega)$, we can write $P_y(\omega)$ in the form:

$$P_y(\omega) = \frac{2F}{\gamma_0^2(\omega^2+\gamma_0^2)} B(i\omega), \tag{S9}$$

where $B(\lambda)$ is the polynomial from main text Eq. (14),

$$B(\lambda) = \gamma_0^2 + \sum_{j=1}^{N} \gamma_0^{2-j} \prod_{k=1}^{j} \frac{\gamma_{k-1}^2 - \lambda^2}{\gamma_{k-1}\Lambda_k}. \tag{S10}$$

As discussed in the main text, this polynomial will always have N roots λ_j, $j = 1,\ldots,N$, where $\mathrm{Re}(\lambda_j) > 0$. (The other N roots of the polynomial are just $-\lambda_j$.) Thus we can factor $B(i\omega)$ in the following way:

$$B(i\omega) = \gamma_0^2 \left(\prod_{k=1}^{N} \frac{1}{\gamma_0\gamma_{k-1}\Lambda_k} \right) \left[\prod_{j=1}^{N} (\omega + i\lambda_j) \right] \left[\prod_{j=1}^{N} (\omega - i\lambda_j) \right]. \tag{S11}$$

Since $\omega = -i\lambda_j$ for $j = 1,\ldots,N$ are all the zeros of $B(i\omega)$ in the complex lower half plane, this enables us to write down the decomposition $P_y(\omega) = P_y^+(\omega)(P_y^+(\omega))^*$ where

$$P_y^+(\omega) = \frac{\sqrt{K}}{\omega + i\gamma_0} \prod_{j=1}^{N} (\omega + i\lambda_j), \tag{S12a}$$

$$(P_y^+(\omega))^* = \frac{\sqrt{K}}{\omega - i\gamma_0} \prod_{j=1}^{N} (\omega - i\lambda_j), \tag{S12b}$$

and

$$K = 2F \prod_{k=1}^{N} \frac{1}{\gamma_0\gamma_{k-1}\Lambda_k}. \tag{S13}$$

Continuing with the calculation of $H_{\mathrm{WK}}(\omega)$, we see that:

$$\frac{P_s(\omega)}{(P_y^+(\omega))^*} = \frac{2F}{\sqrt{K}(\omega + i\gamma_0)} \prod_{j=1}^{N} \frac{1}{\omega - i\lambda_j}. \tag{S14}$$

The quantity $\left\{ \frac{P_s(\omega)}{(P_y^+(\omega))^*} \right\}_+$ is computed from taking the causal part of the partial fraction decomposition of Eq. (S14). Because the only causal pole (pole in the lower half plane) of Eq. (S14) is $-i\gamma_0$, all other terms in the decomposition are dropped, yielding:

$$\left\{ \frac{P_s(\omega)}{(P_y^+(\omega))^*} \right\}_+ = \frac{2Fi^N}{C\sqrt{K}(\omega + i\gamma_0)}, \tag{S15}$$

where $C = \prod_{j=1}^{N}(\gamma_0 + \lambda_j)$. Finally, we can divide this result by $P_y^+(\omega)$, following Eq. (S6), giving us the optimal filter:

$$\begin{aligned} H_{\mathrm{WK}}(\omega) &= \frac{2Fi^N}{CK} \prod_{j=1}^{N} \frac{1}{\omega + i\lambda_j} \\ &= \frac{2F}{CK} \prod_{j=1}^{N} \frac{i}{\omega + i\lambda_j} \\ &= \frac{2F}{CK} \prod_{j=1}^{N} \frac{1}{\lambda_j - i\omega}. \end{aligned} \tag{S16}$$

Plugging in the definitions of C and K, we can rewrite the prefactor to get the final form for the optimal filter function:

$$H_{\mathrm{WK}}(\omega) = \prod_{k=1}^{N} \frac{\gamma_0 \gamma_{k-1} \Lambda_k}{(\gamma_0 + \lambda_k)(\lambda_k - i\omega)}. \tag{S17}$$

1.4 Calculating the optimal error E_{WK}

To calculate E_{WK} from Eq. (S7), we first take the inverse Fourier transform of $H_{\mathrm{WK}}(\omega)$ from Eq. (S17), which gives a sum of exponentials in the time domain,

$$H_{\mathrm{WK}}(t) = \Theta(t) \left(\prod_{j=1}^{N} \frac{\gamma_0 \gamma_{j-1} \Lambda_j}{(\gamma_0 + \lambda_j)} \right) \left[(-1)^{N-1} \sum_{k=1}^{N} e^{-\lambda_k t} \prod_{m \neq k} \frac{1}{\lambda_k - \lambda_m} \right]. \tag{S18}$$

Using the fact that $C_s(t) = \mathcal{F}^{-1}[P_s(\omega)] = \bar{x}_0 \exp(-\gamma_0 |t|)$, we can evaluate the integral in Eq. (S7) to find

$$E_{\mathrm{WK}} = 1 - \left(\prod_{j=1}^{N} \frac{\gamma_0 \gamma_{j-1} \Lambda_j}{(\gamma_0 + \lambda_j)} \right) \left[(-1)^{N-1} \sum_{k=1}^{N} \frac{1}{\gamma_0 + \lambda_k} \prod_{m \neq k} \frac{1}{\lambda_k - \lambda_m} \right]. \tag{S19}$$

Reversing the partial fraction decomposition,

$$\begin{aligned} \prod_{k=1}^{N} \frac{1}{y + \lambda_k} &= \sum_{k=1}^{N} \frac{1}{y + \lambda_k} \prod_{m \neq k} \frac{1}{\lambda_m - \lambda_k} \\ &= (-1)^{N-1} \sum_{k=1}^{N} \frac{1}{y + \lambda_k} \prod_{m \neq k} \frac{1}{\lambda_k - \lambda_m}, \end{aligned} \tag{S20}$$

with $y = \gamma_0$, the error reduces to the value in main text Eq. (13):

$$E_{\mathrm{WK}} = 1 - \prod_{j=1}^{N} \frac{\gamma_0 \gamma_{j-1} \Lambda_j}{(\gamma_0 + \lambda_j)^2}. \tag{S21}$$

1.5 Conditions under which the system can achieve WK optimality

In order for the system to attain $E = E_{\mathrm{WK}}$, the parameters must be tuned such that $H(\omega) \propto H_{\mathrm{WK}}(\omega)$, where $H(\omega)$ and $H_{\mathrm{opt}}(\omega)$ are given by Eqs. (S5) and (S17) respectively. Comparing the two functions, we see that they are proportional to one another when $\lambda_j = \gamma_j$ for all $j = 1, \ldots, N$. Satisfying this condition actually requires a certain relationship between the different per-capita deactivation rates γ_j and the Λ_j parameters.

To see this, let us first denote $B_N(\lambda)$ as the polynomial from Eq. (S10) for a particular value of N. The explicit forms of the polynomials for the first few values of N are as follows:

$$\begin{aligned} B_1(\lambda) &= \gamma_0^2 + \frac{\gamma_0^2 - \lambda^2}{\Lambda_1}, \\ B_2(\lambda) &= \gamma_0^2 + \frac{\gamma_0^2 - \lambda^2}{\Lambda_1} + \frac{(\gamma_0^2 - \lambda^2)(\gamma_1^2 - \lambda^2)}{\gamma_0 \gamma_1 \Lambda_1 \Lambda_2}, \\ B_3(\lambda) &= \gamma_0^2 + \frac{\gamma_0^2 - \lambda^2}{\Lambda_1} + \frac{(\gamma_0^2 - \lambda^2)(\gamma_1^2 - \lambda^2)}{\gamma_0 \gamma_1 \Lambda_1 \Lambda_2} + \frac{(\gamma_0^2 - \lambda^2)(\gamma_1^2 - \lambda^2)(\gamma_2^2 - \lambda^2)}{\gamma_0^2 \gamma_1 \gamma_2 \Lambda_1 \Lambda_2 \Lambda_3}. \end{aligned} \tag{S22}$$

Consider the $N = 1$ system. There is one root λ_1 with a positive real part, and we set it to $\lambda_1 = \gamma_1$ to satisfy the condition. This requires that $B_1(\gamma_1) = 0$, which occurs when $\gamma_1 = \gamma_0 \sqrt{1 + \Lambda_1}$. Interestingly, this same value of γ_1

will also be a root for all higher polynomials $N > 1$. Because the additional terms in the higher polynomials all contain a $(\gamma_1^2 - \lambda^2)$ factor, we see that $B_N(\gamma_1) = B_1(\gamma_1) = 0$ for $N > 1$.

Thus $B_2(\lambda)$ has one root $\lambda_1 = \gamma_1 = \gamma_0\sqrt{1+\Lambda_1}$ that we have already found, and a new root $\lambda_2 = \gamma_2$ whose value we need to determine. This will be true iteratively at every higher value of N: the first $N-1$ roots $\lambda_j = \gamma_j$, $j = 1, \ldots, N-1$, will be the same roots as for $B_{N-1}(\lambda)$, and there will one new root $\lambda_N = \gamma_N$. This follows from the structure of the $B_N(\lambda)$ polynomials, where

$$B_N(\gamma_j) = B_j(\gamma_j) = 0 \qquad \text{for } N > j. \tag{S23}$$

We can find all the higher roots by induction. Let us assume that we have already found the values of $\lambda_j = \gamma_j$ for $j = 1, \ldots, N-1$ and are interested in finding $\lambda_N = \gamma_N$. The known roots allow us to completely factor $B_{N-1}(\lambda)$, and from the definition of the polynomials in Eq. (S10) that factorization has to take the form:

$$B_{N-1}(\lambda) = \gamma_0^2 \prod_{j=1}^{N-1} \frac{(\gamma_j^2 - \lambda^2)}{\gamma_0 \gamma_{j-1} \Lambda_j}. \tag{S24}$$

Note that we know the overall prefactor in the factorization above from the prefactor of the highest power $\lambda^{2(N-1)}$ in the definition of $B_{N-1}(\lambda)$. Turning to $B_N(\lambda)$, we can write this polynomial as $B_{N-1}(\lambda)$ plus an added term,

$$B_N(\lambda) = B_{N-1}(\lambda) + \frac{\gamma_0(\gamma_0^2 - \lambda^2)}{\gamma_{N-1}\Lambda_N} \prod_{j=1}^{N-1} \frac{(\gamma_j^2 - \lambda^2)}{\gamma_0 \gamma_{j-1} \Lambda_j}. \tag{S25}$$

Comparing Eq. (S25) to Eq. (S24), we see that

$$\begin{aligned} B_N(\lambda) &= B_{N-1}(\lambda) + \frac{(\gamma_0^2 - \lambda^2)}{\gamma_0 \gamma_{N-1} \Lambda_N} B_{N-1}(\lambda) \\ &= B_{N-1}(\lambda) \left[1 + \frac{(\gamma_0^2 - \lambda^2)}{\gamma_0 \gamma_{N-1} \Lambda_N} \right]. \end{aligned} \tag{S26}$$

Setting the factor in the brackets to zero allows us to find the new root $\lambda_N = \gamma_N$ in terms of the previous root γ_{N-1},

$$\gamma_N = \gamma_0 \sqrt{1 + \frac{\gamma_{N-1}}{\gamma_0} \Lambda_N}. \tag{S27}$$

Starting from the known value of $\gamma_1 = \gamma_0\sqrt{1+\Lambda_1}$, we can iteratively use Eq. (S27) to find all the higher roots. The solutions are the nested radical forms shown in main text Eq. (17),

$$\gamma_1 = \gamma_0\sqrt{1+\Lambda_1}, \qquad \gamma_2 = \gamma_0\sqrt{1+\sqrt{1+\Lambda_1}\Lambda_2}, \qquad \gamma_3 = \gamma_0\sqrt{1+\sqrt{1+\sqrt{1+\Lambda_1}\Lambda_2}\Lambda_3}, \quad \ldots. \tag{S28}$$

When these conditions are satisfied, the expression for E_{WK} simplifies to the form in main text Eq. (18),

$$E_{\mathrm{WK}} = 1 - \prod_{i=1}^{N} \frac{\ell_i}{(1+\sqrt{1+\ell_i})^2}, \tag{S29}$$

where $\ell_i = \gamma_{i-1}/\Lambda_i/\gamma_0$.

2 Deriving the WK optimal filter results for the multi-level cascade with feedback

2.1 Mapping the system onto a noise filter, finding the WK filter function and bound

The feedback derivation starts with main text Eq. (9), but with the ϕ_1 term present:

$$\begin{aligned}\frac{d}{dt}\delta x_0(t) &= -\gamma_0\delta x_0(t) - \phi_1\delta x_N(t) + n_0(t),\\ \frac{d}{dt}\delta x_i(t) &= -\gamma_i\delta x_i(t) + \sigma_{i1}\delta x_{i-1}(t) + n_i(t), \qquad i>0,\end{aligned} \tag{S30}$$

The noise filter mapping is qualitatively different from the no feedback case, taking the form of main text Eq. (19),

$$s(t) \equiv \delta x_0(t)|_{\phi=0}, \qquad \tilde{s}(t) = \delta x_0(t)|_{\phi=0} - \delta x_0(t). \tag{S31}$$

We know the $\delta x_0(t)|_{\phi=0}$ solution in Fourier space already, having calculated it in Eq. (S3),

$$s(\omega) = \delta x_0(\omega)|_{\phi_0} = \frac{n_0(\omega)}{\gamma_0 - i\omega}. \tag{S32}$$

We can manipulate the Fourier space counterpart of Eq. (S30) to relate $\tilde{s}(\omega)$ to $s(\omega)$ through a noise filter equation,

$$\tilde{s}(\omega) = H(\omega)(s(\omega) + n(\omega)), \tag{S33}$$

where

$$n(\omega) = \sum_{j=1}^{N} n_j(\omega)\prod_{k=1}^{j-1}\frac{\gamma_k - i\omega}{\sigma_{j1}\sigma_{k1}}, \qquad H(\omega) = \frac{\phi_1\prod_{j=1}^{N}\sigma_{j1}}{\prod_{j=0}^{N}(\gamma_j - i\omega) + \phi_1\prod_{j=1}^{N}\sigma_{j1}}. \tag{S34}$$

Comparing to Eq. (S5), we see that $s(\omega)$ and $n(\omega)$ in this mapping are exactly the same as in the no feedback case. Hence $P_s(\omega)$ and $P_n(\omega)$ are the same, which means the calculation of H_{WK} and E_{WK} is unchanged. The result for E_{WK} in Eq. (S21) serves as a lower bound for the error ϵ.

2.2 Conditions under which the system can achieve WK optimality

Comparing $H(\omega)$ from Eq. (S34) and $H_{\mathrm{WK}}(\omega)$ from Eq. (S17), one sees that achieving $H(\omega) = H_{\mathrm{WK}}(\omega)$, and hence $\epsilon = E_{\mathrm{WK}}$, is non-trivial. However there is one scenario where this can be approximately fulfilled. We will show that in a certain limit the N-level feedback system effectively behaves like an $N = 1$ level system with an effective Λ_1 parameter. Note that the $N = 1$ version of $P_n(\omega)$ from Eq. (S8b) looks like:

$$P_n(\omega) = \frac{2F}{\gamma_0^2\Lambda_1}. \tag{S35}$$

Let us now consider an N-level system where $\gamma_j \gg \gamma_0$ for $j > 0$. The main frequency scale in the system is set by the input signal, which has characteristic frequency γ_0, so typical frequencies ω that are relevant to the system behavior all share the property that $\omega \ll \gamma_j$ for $j > 0$. If we use this simplification in Eq. (S8b), the noise power spectrum can be approximated as:

$$P_n(\omega) \approx \frac{2F}{\gamma_0^2}\sum_{j=1}^{N}\frac{1}{\Lambda_j}\left[\prod_{k=1}^{j-1}\frac{\gamma_k}{\gamma_0\Lambda_k}\right]. \tag{S36}$$

Comparing Eq. (S35) to Eq. (S36), we note that the multi-stage noise power spectrum is approximately the same form as for an $N = 1$ system, except with Λ_1 replaced by an effective parameter Λ_{eff} given by:

$$\Lambda_{\mathrm{eff}} = \left(\sum_{j=1}^{N}\frac{1}{\Lambda_j}\left[\prod_{k=1}^{j-1}\frac{\gamma_k}{\gamma_0\Lambda_k}\right]\right)^{-1}. \tag{S37}$$

For the special case where the production functions $R_j(x_{j-1}) = \sigma_{j1} x_{j-1}$, and hence $\sigma_{j1} = \sigma_{j0}/\bar{x}_{j-1}$ for $j > 0$, the expression for Λ_{eff} simplifies to the result shown in main text Eq. (22):

$$\Lambda_{\text{eff}} = \frac{1}{F}\left[\sum_{j=1}^{N} \frac{1}{\sigma_{j0}}\right]^{-1}. \tag{S38}$$

The corresponding $N = 1$ optimal filter $H_{\text{WK}}(\omega)$ from Eq. (S17), with Λ_{eff} instead of Λ_1, can be expressed as:

$$H_{\text{WK}}(\omega) = \frac{\gamma_0(\sqrt{1+\Lambda_{\text{eff}}} - 1)}{\gamma_0\sqrt{1+\Lambda_{\text{eff}}} - i\omega}. \tag{S39}$$

Here we have used the fact that $\lambda_1 = \gamma_0\sqrt{1+\Lambda_1}$ is the root for $B_1(\lambda)$ from Eq. (S22), and substituted in Λ_{eff}.

Let us now write $H(\omega)$ from Eq. (S34) using the approximation $\omega \ll \gamma_j$ for $j > 0$,

$$H(\omega) \approx \frac{\phi_1 \prod_{j=1}^{N} \sigma_{j1}}{(\gamma_0 - i\omega)\prod_{j=1}^{N}\gamma_j + \phi_1 \prod_{j=1}^{N}\sigma_{j1}}. \tag{S40}$$

We can thus approximately have $H(\omega) \approx H_{\text{WK}}(\omega)$ from Eq. (S39) when the feedback strength is tuned to the value from main text Eq. (21),

$$\phi_1 = \gamma_0(\sqrt{1+\Lambda_{\text{eff}}} - 1)\prod_{j=1}^{N}\frac{\gamma_j}{\sigma_{j1}}, \tag{S41}$$

which then ensures that $\epsilon \approx E_{\text{WK}}$, with the latter having the $N = 1$ form,

$$E_{\text{WK}} = \frac{2}{1+\sqrt{1+\Lambda_{\text{eff}}}}. \tag{S42}$$

3 Exact error calculation in the nonlinear cascade without feedback

This section fills in the details of the calculation that transforms main text Eq. (37), a relation for the generating function $H_{\hat{\boldsymbol{x}}}(y)$ and its derivatives $H^{(p)}_{\hat{\boldsymbol{x}}}(y)$, into the recursion relation of main text Eq. (49). The ultimate goal is to use the recursion relation to find the coefficients $\mu^{(p)}_{\hat{\boldsymbol{n}}}$ in order to evaluate the exact error E given by main text Eq. (48):

$$E = 1 - \frac{\bar{x}_0\left(\mu^{(1)}_{\hat{\mathbf{0}}+\hat{\boldsymbol{e}}_0}\right)^2}{\mu^{(2)}_{\hat{\mathbf{0}}} + \mu^{(1)}_{\hat{\mathbf{0}}} - \left(\mu^{(1)}_{\hat{\mathbf{0}}}\right)^2}. \tag{S43}$$

Recall the expansions defined in the main text for all the quantities of interest:

$$\begin{aligned} R_i(x_{i-1}) &= \sum_{n=0}^{\infty} \sigma_{in} v_n(x_{i-1}; \bar{x}_{i-1}) \quad \text{for } i > 0, \\ J^{(p)}_{\hat{\boldsymbol{x}}} &= \sum_{\hat{\boldsymbol{n}}} \mu^{(p)}_{\hat{\boldsymbol{n}}} v_{\hat{\boldsymbol{n}}}(\hat{\boldsymbol{x}}; \hat{\bar{\boldsymbol{x}}}), \end{aligned} \tag{S44}$$

where

$$J^{(p)}_{\hat{\boldsymbol{x}}} = \frac{H^{(p)}_{\hat{\boldsymbol{x}}}(1)}{\Pi(\hat{\boldsymbol{x}}; \hat{\bar{\boldsymbol{x}}})}. \tag{S45}$$

Here we use the multi-dimensional versions of the Poisson distributions and Poisson-Charlier polynomials,

$$\begin{aligned}\Pi(\hat{\boldsymbol{x}};\hat{\bar{\boldsymbol{x}}}) &\equiv \Pi(x_0;\bar{x}_0)\Pi(x_1;\bar{x}_1)\cdots\Pi(x_{N-1};\bar{x}_{N-1}),\\ v_{\hat{\boldsymbol{n}}}(\hat{\boldsymbol{x}};\hat{\bar{\boldsymbol{x}}}) &\equiv v_{n_0}(x_0;\bar{x}_0)v_{n_1}(x_1;\bar{x}_1)\cdots v_{n_{N-1}}(x_{N-1};\bar{x}_{N-1}).\end{aligned} \tag{S46}$$

More details on the Poisson-Charlier polynomials can be found in the next section of the SI, which provides a brief guide to their most useful properties.

Since we know the production functions $R_i(x_{i-1})$ for our system of interest, we can easily find the coefficients σ_{in} in Eq. (S44), using main text Eq. (42). To derive the coefficients $\mu^{(p)}_{\hat{\boldsymbol{n}}}$, we start with the relation in main text Eq. (37):

$$\begin{aligned}0 = \sum_{i=0}^{N-1}\Big\{&\gamma_i\left[(x_i+1)H^{(p)}_{\hat{\boldsymbol{x}}+\hat{\boldsymbol{e}}_i}(1) - x_i H^{(p)}_{\hat{\boldsymbol{x}}}(1)\right] + R_i(x_{i-1})\left[H^{(p)}_{\hat{\boldsymbol{x}}-\hat{\boldsymbol{e}}_i}(1) - H^{(p)}_{\hat{\boldsymbol{x}}}(1)\right]\Big\}\\ &- p\gamma_N H^{(p)}_{\hat{\boldsymbol{x}}}(1) + pR_N(x_{N-1})H^{(p-1)}_{\hat{\boldsymbol{x}}}(1).\end{aligned} \tag{S47}$$

Using Eq. (S45) and the fact that Poisson distributions satisfy $(x_i+1)\Pi(x_i+1;\bar{x}_i) = \bar{x}_i\Pi(x_i;\bar{x}_i)$, we can rewrite Eq. (S47) in terms of the $J^{(p)}_{\hat{\boldsymbol{x}}}$ functions:

$$0 = \left\{\sum_{i=0}^{N-1}\gamma_i\left[\bar{x}_i J^{(p)}_{\hat{\boldsymbol{x}}+\hat{\boldsymbol{e}}_i} - x_i J^{(p)}_{\hat{\boldsymbol{x}}}\right] + R_i(x_{i-1})\left[x_i\bar{x}_i^{-1}J^{(p)}_{\hat{\boldsymbol{x}}-\hat{\boldsymbol{e}}_i} - J^{(p)}_{\hat{\boldsymbol{x}}}\right] - p\gamma_N J^{(p)}_{\hat{\boldsymbol{x}}} + pR_N(x_{N-1})J^{(p-1)}_{\hat{\boldsymbol{x}}}\right\}\Pi(\hat{\boldsymbol{x}};\hat{\bar{\boldsymbol{x}}}). \tag{S48}$$

Let us introduce one more expansion, for products of the $R_i(x_{i-1})$ and $J^{(p)}_{\hat{\boldsymbol{x}}}$ functions,

$$R_i(x_{i-1})J^{(p)}_{\hat{\boldsymbol{x}}} = \sum_{\hat{\boldsymbol{n}}}\nu^{(p,i)}_{\hat{\boldsymbol{n}}}v_{\hat{\boldsymbol{n}}}(\hat{\boldsymbol{x}};\hat{\bar{\boldsymbol{x}}}). \tag{S49}$$

Because $R_i(x_{i-1})$ and $J^{(p)}_{\hat{\boldsymbol{x}}}$ have their own individual expansions in terms of the Poisson-Charlier polynomials, defined by Eq. (S44), the coefficients $\nu^{(p,i)}_{\hat{\boldsymbol{n}}}$ are entirely determined by the coefficients σ_{in} and $\mu^{(p)}_{\hat{\boldsymbol{n}}}$ of the individual expansions. This relation, a property of the Poisson-Charlier polynomials, is explained in more detail in SI Sec. 4.5. It takes the form:

$$\nu^{(p,i)}_{\hat{\boldsymbol{n}}} = \sum_{\substack{a,b=0\\ a+b\geq n_{i-1}\\ |a-b|\leq n_{i-1}}}^{\infty}\sigma_{ia}\mu^{(p)}_{\hat{\boldsymbol{n}}+(b-n_{i-1})\hat{\boldsymbol{e}}_{i-1}}C^{ab}_{n_{i-1}}(\bar{x}_i-1), \tag{S50}$$

where $C^{mn}_k(z)$ are polynomials defined in Eqs. (S66)-(S67).

Let us define $\langle f(\hat{\boldsymbol{x}})\rangle_{\hat{\bar{\boldsymbol{x}}}} = \sum_{\hat{\boldsymbol{x}}} f(\hat{\boldsymbol{x}})\Pi(\hat{\boldsymbol{x}};\hat{\bar{\boldsymbol{x}}})$ as the average of a function $f(\hat{\boldsymbol{x}})$ with respect to $\Pi(\hat{\boldsymbol{x}};\hat{\bar{\boldsymbol{x}}})$. Using the recursion relationships for Poisson-Charlier polynomials shown in Eq. (S64), one can prove the following useful identities:

$$\begin{aligned}\left\langle v_{\hat{\boldsymbol{n}}}(\hat{\boldsymbol{x}};\hat{\bar{\boldsymbol{x}}})J^{(p)}_{\hat{\boldsymbol{x}}}\right\rangle_{\hat{\bar{\boldsymbol{x}}}} &= \zeta_{\hat{\boldsymbol{n}}}(\hat{\bar{\boldsymbol{x}}})\mu^{(p)}_{\hat{\boldsymbol{n}}},\\ \left\langle v_{\hat{\boldsymbol{n}}}(\hat{\boldsymbol{x}};\hat{\bar{\boldsymbol{x}}})R_i(x_{i-1})J^{(p)}_{\hat{\boldsymbol{x}}}\right\rangle_{\hat{\bar{\boldsymbol{x}}}} &= \zeta_{\hat{\boldsymbol{n}}}(\hat{\bar{\boldsymbol{x}}})\nu^{(p,i)}_{\hat{\boldsymbol{n}}},\\ \left\langle v_{\hat{\boldsymbol{n}}}(\hat{\boldsymbol{x}};\hat{\bar{\boldsymbol{x}}})x_iR_i(x_{i-1})J^{(p)}_{\hat{\boldsymbol{x}}-\hat{\boldsymbol{e}}_i}\right\rangle_{\hat{\bar{\boldsymbol{x}}}} &= \zeta_{\hat{\boldsymbol{n}}}(\hat{\bar{\boldsymbol{x}}})\left[\nu^{(p,i)}_{\hat{\boldsymbol{n}}-\hat{\boldsymbol{e}}_i} + \bar{x}_i\nu^{(p,i)}_{\hat{\boldsymbol{n}}}\right],\\ \left\langle v_{\hat{\boldsymbol{n}}}(\hat{\boldsymbol{x}};\hat{\bar{\boldsymbol{x}}})J^{(p)}_{\hat{\boldsymbol{x}}+\hat{\boldsymbol{e}}_i}\right\rangle_{\hat{\bar{\boldsymbol{x}}}} &= \zeta_{\hat{\boldsymbol{n}}}(\hat{\bar{\boldsymbol{x}}})\left[\mu^{(p)}_{\hat{\boldsymbol{n}}} + (n_i+1)\mu^{(p)}_{\hat{\boldsymbol{n}}+\hat{\boldsymbol{e}}_i}\right],\\ \left\langle v_{\hat{\boldsymbol{n}}}(\hat{\boldsymbol{x}};\hat{\bar{\boldsymbol{x}}})x_iJ^{(p)}_{\hat{\boldsymbol{x}}}\right\rangle_{\hat{\bar{\boldsymbol{x}}}} &= \zeta_{\hat{\boldsymbol{n}}}(\hat{\bar{\boldsymbol{x}}})\left[\mu^{(p)}_{\hat{\boldsymbol{n}}-\hat{\boldsymbol{e}}_i} + (n_i+\bar{x}_i)\mu^{(p)}_{\hat{\boldsymbol{n}}} + (n_i+1)\bar{x}_i\mu^{(p)}_{\hat{\boldsymbol{n}}+\hat{\boldsymbol{e}}_i}\right],\\ \left\langle v_{\hat{\boldsymbol{n}}}(\hat{\boldsymbol{x}};\hat{\bar{\boldsymbol{x}}})x_iJ^{(p)}_{\hat{\boldsymbol{x}}-\hat{\boldsymbol{e}}_i}\right\rangle_{\hat{\bar{\boldsymbol{x}}}} &= \zeta_{\hat{\boldsymbol{n}}}(\hat{\bar{\boldsymbol{x}}})\left[\mu^{(p)}_{\hat{\boldsymbol{n}}-\hat{\boldsymbol{e}}_i} + \bar{x}_i\mu^{(p)}_{\hat{\boldsymbol{n}}}\right],\end{aligned} \tag{S51}$$

where $\zeta_{\hat{\boldsymbol{n}}}(\hat{\tilde{\boldsymbol{x}}}) \equiv \prod_{i=0}^{N-1} n_i! \bar{x}_i^n$. By multiplying Eq. (S48) by $v_{\hat{\boldsymbol{n}}}(\hat{\boldsymbol{x}};\hat{\tilde{\boldsymbol{x}}})$ and summing over $\hat{\boldsymbol{x}}$, we can use the above averages to obtain the following relation:

$$0 = -n_0\gamma_0\mu^{(p)}_{\hat{\boldsymbol{n}}} + \sum_{i=1}^{N-1}\left(-\gamma_i\mu^{(p)}_{\hat{\boldsymbol{n}}-\hat{\boldsymbol{e}}_i} - n_i\gamma_i\mu^{(p)}_{\hat{\boldsymbol{n}}} + \bar{x}_i^{-1}\nu^{(p,i)}_{\hat{\boldsymbol{n}}-\hat{\boldsymbol{e}}_i}\right) - p\gamma_N\mu^{(p)}_{\hat{\boldsymbol{n}}} + p\nu^{(p-1,i)}_{\hat{\boldsymbol{n}}}. \tag{S52}$$

We can rearrange this obtain the recursion relation in main text Eq. (49),

$$\mu^{(p)}_{\hat{\boldsymbol{n}}} = \frac{p\nu^{(p-1,N)}_{\hat{\boldsymbol{n}}} + \sum_{i=1}^{N-1}\left(\bar{x}_i^{-1}\nu^{(p,i)}_{\hat{\boldsymbol{n}}-\hat{\boldsymbol{e}}_i} - \gamma_i\mu^{(p)}_{\hat{\boldsymbol{n}}-\hat{\boldsymbol{e}}_i}\right)}{p\gamma_N + \sum_{i=0}^{N-1} n_i\gamma_i}. \tag{S53}$$

This relation, together with $\mu^{(0)}_{\hat{\boldsymbol{0}}} = 1$ which we know from the normalization property $\sum_{\hat{\boldsymbol{x}}} H_{\hat{\boldsymbol{x}}}(1) = 1$, is sufficient for us to calculate any coefficient $\mu^{(p)}_{\hat{\boldsymbol{n}}}$ of interest.

4 Properties of the Poisson-Charlier polynomials

4.1 Definition of the polynomials

In this section, we summarize some properties of the polynomials $v_n(x;\bar{x})$ used in our analytical expansion approach for calculating moments of master equations. These are variants of Poisson-Charlier (PC) polynomials[4,5], $c_n(x;\bar{x})$, related by a trivial factor to the standard PC definition:

$$v_n(x;\bar{x}) = (-\bar{x})^n c_n(x;\bar{x}). \tag{S54}$$

The nth function $v_n(x;\bar{x})$ is a polynomial in x of degree n, depending on the parameter $\bar{x}$. It is defined as follows:

$$v_n(x;\bar{x}) = \sum_{m=0}^{n}\binom{n}{m}(-\bar{x})^m (x)_{n-m}. \tag{S55}$$

Here $(x)_k \equiv x(x-1)\cdots(x-k+1) = k!\binom{x}{k}$ is the kth falling factorial of x, with $(x)_0 \equiv 1$. The first few polynomials are given by:

$$\begin{aligned} &v_0(x;\bar{x}) = 1, \quad v_1(x;\bar{x}) = x-\bar{x}, \quad v_2(x;\bar{x}) = (x-\bar{x})^2 - x, \\ &v_3(x;\bar{x}) = (x-\bar{x})^3 - 3x(x-\bar{x}) + 2x. \end{aligned} \tag{S56}$$

These $v_n(x;\bar{x})$ appear in a variety of master equation expansion approaches, for example the spectral method of Refs. 6,7. In fact, $v_n(x;\bar{x}) = n!\langle n|x\rangle$, where $\langle n|x\rangle$ is the mixed product defined in Eq. A8 of Ref. 6 (with $\bar{x}$ substituted for the rate parameter g).

4.2 Orthogonality with respect to the Poisson distribution

One of the convenient properties of these polynomials is that they have simple averages with respect to the Poisson distribution,

$$\Pi(x;\bar{x}) = \frac{\bar{x}^x e^{-\bar{x}}}{x!}, \tag{S57}$$

where x is a non-negative integer, and $\bar{x}$ is the parameter that defines the mean of the distribution, so that $\bar{x} = \sum_{x=0}^{\infty} x\Pi(x;\bar{x})$. Let us denote the average of a function $f(x)$ with respect to the Poisson distribution $\Pi(x;\bar{x})$ in the following way:

$$\langle f(x)\rangle_{\bar{x}} \equiv \sum_{x=0}^{\infty} f(x)\Pi(x;\bar{x}). \tag{S58}$$

Then the polynomials of Eq. (S55) satisfy the following orthogonality relationship[8,9]:

$$\langle v_{n'}(x;\bar{x})v_n(x;\bar{x})\rangle_{\bar{x}} = n!\bar{x}^n\delta_{n',n}. \tag{S59}$$

Since $v_0(x;\bar{x}) = 1$, a special case of Eq. (S59) when $n' = 0$ gives an expression for the mean:

$$\langle v_n(x;\bar{x})\rangle_{\bar{x}} = \delta_{n0}. \tag{S60}$$

4.3 Using the polynomials as a basis for function expansions

The polynomials form a basis in which one can expand arbitrary functions of populations $f(x)$,

$$f(x) = \sum_{n=0}^{\infty} \alpha_n v_n(x;\bar{x}), \tag{S61}$$

for some coefficients α_n. To calculate the mth coefficient α_m, we multiply both sides of Eq. (S61) by $v_m(x;\bar{x})$ and take the average with respect to $\Pi(x;\bar{x})$:

$$\langle v_m(x)f(x)\rangle_{\bar{x}} = \sum_{n=0}^{\infty} \alpha_n \langle v_m(x;\bar{x})v_n(x;\bar{x})\rangle_{\bar{x}} = \alpha_m m!\bar{x}^m, \tag{S62}$$

where we have used the orthogonality relation Eq. (S59). Thus α_m is given by:

$$\alpha_m = \frac{\langle v_m(x;\bar{x})f(x)\rangle_{\bar{x}}}{m!\bar{x}^m} = \sum_{n=0}^{m} \frac{(-1)^{m-n}\bar{x}^{-n}}{(m-n)!} \left\langle \binom{x}{n} f(x) \right\rangle_{\bar{x}}, \tag{S63}$$

where we have plugged in the definition of $v_m(x;\bar{x})$ from Eq. (S55). For the kinds of functions we ordinarily encounter in working with master equations, the coefficients α_m rapidly decay with m, so in practice we can often form an excellent approximation by just keeping the first few ($n \leq 5$) terms in the expansion of Eq. (S61)[9].

4.4 Recursion relationships

The polynomials satisfy the following recursion relationships, as can be easily verified from their definition in Eq. (S55):

$$\begin{aligned} xv_n(x;\bar{x}) &= n\bar{x}v_{n-1}(x;\bar{x}) + (n+\bar{x})v_n(x;\bar{x}) + v_{n+1}(x;\bar{x}), \\ v_n(x+1;\bar{x}) &= nv_{n-1}(x;\bar{x}) + v_n(x;\bar{x}), \\ xv_n(x-1;\bar{x}) &= \bar{x}v_n(x;\bar{x}) + v_{n+1}(x;\bar{x}). \end{aligned} \tag{S64}$$

4.5 Expanding the product of polynomials

The final property that comes in useful in calculations is that the product of two polynomials $v_m(x;\bar{x})$ and $v_n(x;\bar{x})$ can be itself expanded in a linear combination of polynomials in the following form:

$$v_m(x;\bar{x})v_n(x;\bar{x}) = \sum_{k=|n-m|}^{n+m} v_k(x;\bar{x})C_k^{mn}(\bar{x}), \tag{S65}$$

where the coefficients $C_k^{mn}(\bar{x})$ are polynomials in $\bar{x}$ given by:

$$C_k^{mn}(\bar{x}) = \sum_{c=\max(0,n-k,m-k)}^{\left\lfloor \frac{m+n-k}{2} \right\rfloor} \Gamma_{kc}^{mn}\bar{x}^c. \tag{S66}$$

Here, the sum starts at the largest of the three values 0, $n-k$, and $m-k$, and $\lfloor z \rfloor$ denotes the largest integer less or equal to z. The quantity Γ_{kc}^{mn} is defined as:

$$\Gamma_{kc}^{mn} \equiv \frac{m!n!}{c!(c+k-m)!(c+k-n)!(m+n-k-2c)!}. \tag{S67}$$

Thus for example if one had two functions $f(x)$ and $g(x)$ with individual expansions,

$$f(x) = \sum_{n=0}^{\infty} \alpha_n v_n(x;\bar{x}), \qquad g(x) \sum_{n=0}^{\infty} \beta_n v_n(x;\bar{x}), \tag{S68}$$

then the product can be expanded as

$$f(x)g(x) = \sum_{n=0}^{\infty} \gamma_n v_n(x;\bar{x}), \tag{S69}$$

with coefficients given by

$$\gamma_n = \sum_{\substack{k,\ell \\ k+\ell \geq n \\ |k-\ell| \leq n}}^{\infty} \alpha_k \beta_\ell C_n^{k\ell}(\bar{x}). \tag{S70}$$

References

1. Hinczewski, M. & Thirumalai, D. Noise control in gene regulatory networks with negative feedback. *J. Phys. Chem. B* **120**, 6166–6177 (2016).
2. Bode, H. W. & Shannon, C. E. A simplified derivation of linear least square smoothing and prediction theory. *Proc. Inst. Radio. Engin.* **38**, 417–425 (1950).
3. Becker, N. B., Mugler, A. & ten Wolde, P. R. Optimal prediction by cellular signaling networks. *Phys. Rev. Lett.* **115** (258103, 2015).
4. Özmen, N. & Erkuş-Duman, E. On the Poisson-Charlier polynomials. *Serdica Math. J.* **41**, 457–470 (2015).
5. Roman, S. *The Umbral Calculus* (Dover, 2005).
6. Mugler, A., Walczak, A. M. & Wiggins, C. H. Spectral solutions to stochastic models of gene expression with bursts and regulation. *Phys. Rev. E* **80**, 041921 (2009).
7. Walczak, A. M., Mugler, A. & Wiggins, C. H. A stochastic spectral analysis of transcriptional regulatory cascades. *Proc. Natl. Acad. Sci. USA* **106**, 6529–6534 (2009).
8. Ogura, H. Orthogonal functionals of the Poisson process. *IEEE Trans. Info. Theory* **18**, 473–481 (1972).
9. Hinczewski, M. & Thirumalai, D. Cellular signaling networks function as generalized Wiener-Kolmogorov filters to suppress noise. *Phys. Rev. X* **4** (041017, 2014).