Supplementary Information:
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

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),

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

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 $\mathrm{Re}(\lambda_j) > 0$, and the other N roots are just $-\lambda_j$. Moreover, among the set of λ_j, any complex roots come in conjugate pairs. This guarantees that the expression for E_{WK} in main text 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. Taking all this into account, 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 $E = E_{\rm WK}$, the parameters must be tuned such that $H(\omega) \propto H_{\rm WK}(\omega)$, where $H(\omega)$ and $H_{\rm opt}(\omega)$ 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:

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

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 $H_{\hat{\mathbf{x}}}(y)$ and its derivatives $H_{\hat{\mathbf{x}}}^{(p)}(y)$, into the recursion relation of main text Eq. (49). The ultimate goal is to use the recursion relation to find the coefficients $\mu_{\hat{\mathbf{n}}}^{(p)}$ in order to evaluate the exact error E given by main text Eq. (48):

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

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

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

where

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

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

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

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 $\mu_{\hat{\boldsymbol{n}}}^{(p)}$, we start with the relation in main text Eq. (37):

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

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 $J_{\hat{\boldsymbol{x}}}^{(p)}$ functions:

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

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

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

Because $R_i(x_{i-1})$ and $J_{\hat{\boldsymbol{x}}}^{(p)}$ have their own individual expansions in terms of the Poisson-Charlier polynomials, defined by Eq. (S44), the coefficients $\nu_{\hat{\boldsymbol{n}}}^{(p,i)}$ are entirely determined by the coefficients σ_{in} and $\mu_{\hat{\boldsymbol{n}}}^{(p)}$ 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:

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

where $C_k^{mn}(z)$ 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 \nu_{\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 \nu_{\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 \nu_{\hat{\boldsymbol{n}}}(\hat{\boldsymbol{x}};\hat{\bar{\boldsymbol{x}}}) x_i R_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 \nu_{\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 \nu_{\hat{\boldsymbol{n}}}(\hat{\boldsymbol{x}};\hat{\bar{\boldsymbol{x}}}) x_i J^{(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 \nu_{\hat{\boldsymbol{n}}}(\hat{\boldsymbol{x}};\hat{\bar{\boldsymbol{x}}}) x_i 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}}-\hat{\boldsymbol{e}}_i} + \bar{x}_i\mu^{(p)}_{\hat{\boldsymbol{n}}}\right],
\end{aligned} \tag{S51}$$

where $\zeta_{\hat{\boldsymbol{n}}}(\hat{\bar{\boldsymbol{x}}}) \equiv \prod_{i=0}^{N-1} n_i!\bar{x}_i^{n}$. By multiplying Eq. (S48) by $\nu_{\hat{\boldsymbol{n}}}(\hat{\boldsymbol{x}};\hat{\bar{\boldsymbol{x}}})$ 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 $\nu_n(x;\bar{x})$ 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:

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

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

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

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:

$$\begin{aligned}
&\nu_0(x;\bar{x}) = 1, \quad \nu_1(x;\bar{x}) = x-\bar{x}, \quad \nu_2(x;\bar{x}) = (x-\bar{x})^2 - x, \\
&\nu_3(x;\bar{x}) = (x-\bar{x})^3 - 3x(x-\bar{x}) + 2x.
\end{aligned} \tag{S56}$$

These $\nu_n(x;\bar{x})$ appear in a variety of master equation expansion approaches, for example the spectral method of Refs. 6,7. In fact, $\nu_n(x;\bar{x}) = n!\langle n|x\rangle$, 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 ($n \le 5$) 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:

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

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

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

5 Additional results for nonlinear cascades without feedback

5.1 $N = 2$ cascade with nonlinear production functions at both levels

One example we considered in the main text was the $N = 2$ cascade without feedback where the first level production function $R_1(x_0) = \sigma_{11}x_0$ is linear and the second level production function $R_2(x_1) = \sigma_{21}x_1 + \sigma_{22}v_2(x_1;\bar{x}_1)$ is quadratic. In the limit $r = \gamma_1/\gamma_0 \gg 1$ and $\rho = \sigma_{20}/\sigma_{10} \gg 1$, where signaling is efficient ($E_{\text{WK}} \lesssim 1/4$), we get main text Eq. (55) for the difference $E - E_{\text{WK}}$. This is minimized in main text Eq. (56), showing a small violation of the WK bound: $E_{\min} - E_{\text{WK}} \approx -(2\bar{x}_0 r^2)^{-1}$.

Here we generalize these results to the case where both production functions are quadratic, so that $R_1(x_0) = \sigma_{11}x_0 + \sigma_{12}v_2(x_0;\bar{x}_0)$. Following the same approach as described in the main text, we find that for $r \gg 1$ and $\rho \gg 1$:

$$E - E_{\text{WK}} \approx \frac{2}{\gamma_0 \rho r}\sigma_{22} + \frac{2\bar{x}_0}{\gamma_0^2 \rho^2}\sigma_{22}^2 + \frac{4\bar{x}_0}{\gamma_0^2 r^2 \rho}\sigma_{12}\sigma_{22} + \frac{2\bar{x}_0}{\gamma_0^2 r^4}\sigma_{12}^2. \tag{S71}$$

If we keep the parameter σ_{12} fixed, we can minimize E at the following value of σ_{22}:

$$\sigma_{22} = -\frac{\rho(r\gamma_0 + 2\bar{x}_0\sigma_{12})}{2r^2\bar{x}_0}. \tag{S72}$$

The resulting minimum error value $E_{\rm min}$, generalizing main text Eq. (56) by adding a second term proportional to σ_{12}, is given by:

$$E_{\rm min} - E_{\rm WK} \approx -\frac{1}{2\bar{x}_0 r^2} - \frac{2\sigma_{12}}{\gamma_0 r^3}. \tag{S73}$$

If $\sigma_{12} > 0$, we see that the violation of the WK bound can be made larger through the additional nonlinearity at the first level. However the r^3 in the denominator keeps the σ_{12} term small relative to $E_{\rm WK} \sim 2/r$ when $r \gg 1$, so the overall magnitude of the violation generally remains tiny.

5.2 $N = 3$ cascade with a nonlinear production function only at the first level

Main text Eq. (53) shows E for an $N = 2$ system where the first level production function can be nonlinear, but the second level one is linear ($\sigma_{2n} = 0$ for $n \geq 2$). Here we generalize this to a $N = 3$ system with a nonlinear first level, but all higher levels linear ($\sigma_{2n} = \sigma_{3n} = 0$ for $n \geq 2$). The result can be expressed as:

$$E = 1 - \frac{\mathcal{N}}{\mathcal{D}}, \tag{S74}$$

where

$$\mathcal{N} = \frac{\gamma_1\gamma_2(\gamma_1+\gamma_2)\gamma_3(\gamma_1+\gamma_3)(\gamma_2+\gamma_3)\sigma_{11}^2\sigma_{21}^2\sigma_{31}^2\bar{x}_0}{(\gamma_0+\gamma_1)^2(\gamma_0+\gamma_2)^2(\gamma_0+\gamma_3)^2}, \tag{S75}$$

and

$$\begin{aligned}\mathcal{D} = \sigma_{31}^2\Bigg[\sigma_{21}^2\Bigg(&\sum_{n=1}^{\infty}\frac{n!\left((\gamma_1+\gamma_2)(\gamma_1+\gamma_3)(\gamma_2+\gamma_3)+(\gamma_1+\gamma_2+\gamma_3)\gamma_0^2 n^2+(\gamma_1+\gamma_2+\gamma_3)^2\gamma_0 n\right)\sigma_{1n}^2\bar{x}_0^n}{(\gamma_1+\gamma_0 n)(\gamma_2+\gamma_0 n)(\gamma_3+\gamma_0 n)}\\ &+(\gamma_1+\gamma_2+\gamma_3)\sigma_{10}\Bigg)+\gamma_1(\gamma_1+\gamma_2)(\gamma_1+\gamma_3)\sigma_{20}\Bigg]+\gamma_1\gamma_2(\gamma_1+\gamma_2)(\gamma_1+\gamma_3)(\gamma_2+\gamma_3)\sigma_{30}.\end{aligned} \tag{S76}$$

Qualitatively the behavior of E is similar to the $N = 1$ (main text Eq. (29)) and $N = 2$ (main text Eq. (53)) cases when only the first level is nonlinear: σ_{1n} for $n \geq 2$ contribute to the denominator $\mathcal{D}$ only through positive terms, and hence always serve to make E larger than $E_{\rm WK}$. Given this pattern for $N = 1 - 3$, it is likely that the result generalizes to cascades of any length: nonlinearity only at the first level cannot beat the WK bound.

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).