A stochastic spectral analysis of transcriptional regulatory cascades

Aleksandra M. Walczak[a,1], Andrew Mugler[b], and Chris H. Wiggins[c]

[a]Princeton Center for Theoretical Science, Princeton University, Princeton, NJ 08544; and Departments of [b]Physics and [c]Applied Physics and Applied Mathematics, Center for Computational Biology and Bioinformatics, Columbia University, New York, NY 10027

Edited by Peter G. Wolynes, University of California at San Diego, La Jolla, CA, and approved February 18, 2009 (received for review November 25, 2008)

The past decade has seen great advances in our understanding of the role of noise in gene regulation and the physical limits to signaling in biological networks. Here, we introduce the spectral method for computation of the joint probability distribution over all species in a biological network. The spectral method exploits the natural eigenfunctions of the master equation of birth–death processes to solve for the joint distribution of modules within the network, which then inform each other and facilitate calculation of the entire joint distribution. We illustrate the method on a ubiquitous case in nature: linear regulatory cascades. The efficiency of the method makes possible numerical optimization of the input and regulatory parameters, revealing design properties of, e.g., the most informative cascades. We find, for threshold regulation, that a cascade of strong regulations converts a unimodal input to a bimodal output, that multimodal inputs are no more informative than bimodal inputs, and that a chain of up-regulations outperforms a chain of down-regulations. We anticipate that this numerical approach may be useful for modeling noise in a variety of small network topologies in biology.

gene regulation | noise | signaling | biological network | information theory

Transcriptional regulatory networks are composed of genes and proteins, which are often present in small numbers in the cell (1, 2), rendering deterministic models poor descriptions of the counts of protein molecules observed experimentally (3–9). Probabilistic approaches have proven necessary to account fully for the variability of molecule numbers within a homogenous population of cells. A full stochastic description of even a small regulatory network proves quite challenging. Many efforts have been made to refine simulation approaches (10–14), which are mainly based on the varying step Monte Carlo or "Gillespie" method (15, 16). Yet expanding full molecular simulations to larger systems and scanning parameter space is computationally expensive. However, the interaction of many protein and gene types makes analytical methods hard to implement. A wide class of approximations to the master equation, which describes the evolution of the probability distribution, focuses on limits of large concentrations or small switches (17–19). Approximations based on timescale separation of the steps of small signaling cascades have been successfully used to calculate escape properties (20, 21). Variational techniques have also been used to calculate approximate distributions (22).

In this article we introduce a method for calculating the steady-state distributions of chemical reactants (code is available at http://specmark.sourceforge.net). The procedure, which we call the *spectral method*, relies on exploiting the natural basis of a simpler problem from the same class. The full problem is then solved numerically as an expansion in this basis, reducing the master equation to a set of linear algebraic equations. We break up the problem into two parts: a preprocessing step, which can be solved algorithmically; and the parameter-specific step of obtaining the actual probability distributions. The spectral method allows for huge computational gains with respect to simulations.

We illustrate the spectral method for the case of regulatory cascades: downstream genes responding to concentrations of transcription factors produced by upstream genes that are linked to external cues. Cascades play an important role in a diversity of cellular processes (23–25), from decision making in development (26) to quorum sensing among cells (27). We take a coarse-grained approach, modeling each step of a cascade with a general regulatory function that depends on the copy number of the reactant at the previous step (Fig. 1). Although the method as implemented describes arbitrary regulation functions, we optimize the information transmission in the case of the most biologically simple regulation function: a discontinuous threshold, in which a species is created at a high or low rate depending on the copy count of the species directly upstream. In the next sections, we outline the spectral method and present in detail our findings regarding signaling cascades.

Method

We calculate the steady-state joint distribution for L chemical species in a cascade (Fig. 1). The approach we take involves two key observations: the master equation, being linear,* benefits from solution in terms of its eigenfunctions; and the behavior of a given species should depend only weakly on distant nodes given the proximal nodes.

The second of these observations can be illustrated succinctly by considering a three-gene cascade in which the first gene may be eliminated by marginalization. For three species obeying $s \xrightarrow{q_s} n \xrightarrow{q_n} m$ as in Fig. 1, we have the linear master equation

$$\begin{aligned}\dot{p}_{snm} = \tilde{\rho}[&g p_{(s-1)nm} - g p_{snm} + (s+1)p_{(s+1)nm} - s p_{snm}] \\ &+ q_s p_{s(n-1)m} - q_s p_{snm} + (n+1)p_{s(n+1)m} - n p_{snm} \\ &+ \rho[q_n p_{sn(m-1)} - q_n p_{snm} + (m+1)p_{sn(m+1)} - m p_{snm}]. \quad [1]\end{aligned}$$

Here, time is rescaled by the second gene's degradation rate, so that each gene's creation rate (g, q_s, or q_n) is normalized by its respective degradation rate; $\tilde{\rho}$ and ρ are the ratios of the first and third gene's degradation rate to the second's, respectively.

To integrate out the first species, we sum over s. We then introduce g_n, the effective regulation of n, by

$$\sum_s q_s p_{nms} = p_{nm} \sum_s q_s p_{s|nm} \approx p_{nm} \sum_s q_s p_{s|n} \equiv g_n p_{nm}. \quad [2]$$

Here, we have made the Markovian approximation that s is conditionally independent of m given n. Generally speaking, the probability distribution depends on all steps of the cascade. However, since there are no loops in the cascades we consider here, we assume in Eq. **2** that at steady state each species is not affected by species two or more steps away in the cascade. The validity

Author contributions: A.M.W., A.M., and C.H.W. designed research, performed research, contributed new reagents/analytic tools, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: awalczak@princeton.edu.

*Although chemical kinetic rates can be nonlinear in the coordinate (e.g., through q_s or q_n in Eq. **1**), the master equation itself is a linear equation in the unknown p.

This article contains supporting information online at www.pnas.org/cgi/content/full/0811999106/DCSupplemental.

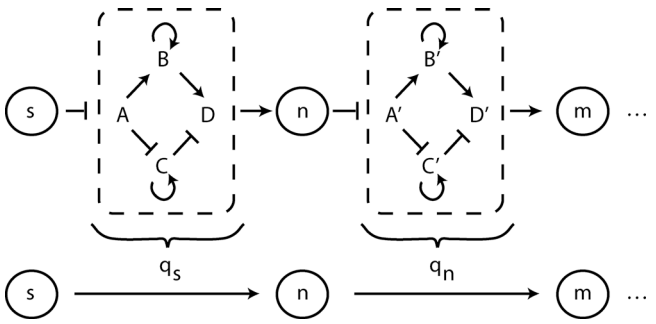

Fig. 1. A schematic representation of a general signaling cascade. Interactions between species of interest may include intermediate processes; we take a coarse-grained approach, condensing these intermediate processes into a single effective regulatory function. For example, the regulatory function q_n describes the creation rate of a species with copy count m as a function of the copy count n of the previous species.

of the Markovian approximation is tested by using both a non-Markovian tensor implementation of the spectral method and a stochastic simulation using the Gillespie algorithm (16), as discussed in supporting information (SI) *Appendix*. We find that the approximation produces accurate results for all but the most strongly discontinuous regulation functions; even in these cases qualitative features such as modality of the output distribution and locations of the modes are preserved. Armed with the Markovian approximation the equation for the remaining two species simplifies to

$$\dot{p}_{nm} = g_{n-1}p_{(n-1)m} - g_n p_{nm} + (n+1)p_{(n+1)m} - np_{nm} + \rho[q_n p_{n(m-1)} - q_n p_{nm} + (m+1)p_{n(m+1)} - mp_{nm}]. \quad [3]$$

This procedure can be extended indefinitely for a cascade of arbitrary length L, in which modules consisting of pairs of adjacent species are each described by two-dimensional master equations.

The distribution for the first two species is obtained by summing over all other species, which gives an equation of the same form as Eq. **3** but with $g_n = g =$ constant. If instead the input distribution is an arbitrary p_n, the distribution for the first two species is still described by Eq. **3**, but with g_n calculated recursively from p_n via $g_n = (-np_n + g_{n-1}p_{n-1} + (n+1)p_{n+1})/p_n$ with $g_0 = p_1/p_0$ to initialize.[†] Describing the start of a cascade (with arbitrary input distribution) and describing subsequent steps, both amount to solving Eq. **3** with g_n given by either the recursive equation or Eq. **2**, respectively.

We solve Eq. **3** by defining the generating function (28) $G(x,y) = \sum_{n,m} p_{nm}x^n y^m$ over complex variables x and y.[‡] It will prove more convenient to write the generating function in a state space as $|G\rangle = \sum_{n,m} p_{nm}|n,m\rangle$,[§] with inverse transform $p_{nm} = \langle n,m|G\rangle$, where the states $|i\rangle$ and $\langle i|$, for $i \in \{n,m\}$, along with the inner product $\langle i|i'\rangle = \delta_{ii'}$, define the protein number basis. With these definitions, Eq. **3** at steady state becomes $0 = \hat{H}|G\rangle$, where

$$\hat{H} = \hat{b}_n^+\hat{b}_n^-(n) + \rho\hat{b}_m^+\hat{b}_m^-(n). \quad [4]$$

Here, we have introduced raising and lowering operators in protein space (22, 29–32) obeying $\hat{b}_i^+|i\rangle = |i+1\rangle - |i\rangle$ for $i \in \{n,m\}$, $\hat{b}_n^-(n)|n\rangle = n|n-1\rangle - \hat{g}_n|n\rangle$ and $\hat{b}_m^-(n)|n,m\rangle = m|n,m-1\rangle - \hat{q}_n|n,m\rangle$,[¶] and the regulation functions have become operators obeying $\hat{g}_n|n\rangle = g_n|n\rangle$ and $\hat{q}_n|n\rangle = q_n|n\rangle$.

Were the operators $\hat{b}_n^-(n)$ and $\hat{b}_m^-(n)$ not n-dependent, $\hat{H}$ would be easily diagonalizable. In fact, this corresponds to the uncoupled case, in which there is no regulation, and both upstream and downstream genes undergo independent birth–death processes with Poisson steady-state distributions. We exploit this fact by working with the respective deviations of g_n and q_n from some constant creation rates $\bar{g}$ and $\bar{q}$. Then $\hat{H}$ can be partitioned as $\hat{H} = \hat{H}_0 + \hat{H}_1$, where

$$\hat{H}_0 = \hat{b}_n^+\bar{b}_n^- + \rho\hat{b}_m^+\bar{b}_m^- \quad [5]$$

$$\hat{H}_1 = \hat{b}_n^+\hat{\Gamma}_n + \rho\hat{b}_m^+\hat{\Delta}_n, \quad [6]$$

and we define operators $\bar{b}_n^-|n\rangle = n|n-1\rangle - \bar{g}|n\rangle$, $\bar{b}^-|m\rangle = m|m-1\rangle - \bar{q}|m\rangle$, $\hat{\Gamma}_n = \bar{g} - \hat{g}_n$, and $\hat{\Delta}_n = \bar{q} - \hat{q}_n$. $\hat{\Gamma}_n$ and $\hat{\Delta}_n$ capture the respective deviations of g_n and q_n from $\bar{g}$ and $\bar{q}$, and $\hat{H}_0$ is diagonal in the eigenbases $|j\rangle$ and $|k\rangle$ of the uncoupled birth–death processes at rates $\bar{g}$ and $\bar{q}$, respectively[‖]; specifically $\hat{H}_0|j,k\rangle = (j+\rho k)|j,k\rangle$. Projecting Eq. **4** onto the eigenbasis yields the linear equation of motion

$$(j+\rho k)G^{jk} + \sum_{j'}\Gamma_{j-1,j'}G^{j'k} + \rho\sum_{j'}\Delta_{jj'}G^{j',k-1} = 0, \quad [7]$$

where $G^{jk} = \langle j,k|G\rangle$, $\Gamma_{jj'} = \sum_n(\bar{g}-g_n)\langle j|n\rangle\langle n|j'\rangle$, and $\Delta_{jj'} = \sum_n(\bar{q}-q_n)\langle j|n\rangle\langle n|j'\rangle$. Eq. **7** exploits the subdiagonal nature of the k-dependence; it is initialized by using $G^{j0} = \sum_n p_n\langle j|n\rangle$, then solved exactly by matrix inversion for each subsequent k. The joint distribution is retrieved via the inverse transform as

$$p_{nm} = \sum_{jk}\langle n|j\rangle G^{jk}\langle m|k\rangle. \quad [8]$$

One computational advantage is that the overlap integrals $\langle n|j\rangle$ and $\langle j|n\rangle$ need only be evaluated explicitly for $\langle n|j=0\rangle = e^{-\bar{g}}\bar{g}^n/n!$ and $\langle j|n=0\rangle = (-\bar{g})^j/j!$; all other values can be obtained recursively by using the selection rules $\langle n|j+1\rangle = \langle n-1|j\rangle - \langle n|j\rangle$ and $\langle j|n+1\rangle = \langle j-1|n\rangle + \langle j|n\rangle$.[**] The same holds for $\langle m|k\rangle$, taking $n \to m, j \to k$, and $\bar{g} \to \bar{q}$. Note that once $\bar{g}$ and $\bar{q}$ have been chosen, the calculation can be separated into a preprocessing step, in which the matrices $\langle n|j\rangle$, $\langle j|n\rangle$, and $\langle m|k\rangle$ are calculated (and potentially reused at subsequent steps of the cascade or for subsequent steps in an optimization), and the actual step of calculating G^{jk} via Eq. **7**. The choices of $\bar{g}$ and $\bar{q}$ can affect the numerical stability of the method.

By exploiting the basis of the uncoupled system, we have reduced Eq. **3** to a set of simple linear algebraic equations. Eq. **7**, which dramatically speeds up the calculation without sacrificing accuracy (cf. *Results* and *SI Appendix*). More precisely, we have turned an $N^2 \times N^2$ matrix solve (where N is a cutoff in copy count) into K length-J vector solves (where J and K are cutoffs in eigenmodes j and k, respectively). The method is applicable for any input function g_n and regulation function q_n. Solutions using other bases and further generalizations to systems with feedback are outside the scope of the current work. One must take particular care in applying a Markov approximation when there is feedback,

[†] The recursion can equivalently be performed in the reverse direction, with $g_N = 0$, $g_{N-1} = Np_N/p_{N-1}$, and $g_{n-1} = (g_n p_n + np_n - (n+1)p_{n+1})/p_{n-1}$, where N is a cutoff in n. In several test cases we found that reverse recursion is more numerically stable than forward recursion at large N.

[‡] Setting $x = e^{ik_1}$ and $y = e^{ik_2}$ makes clear that the generating function is simply the Fourier transform.

[§] $G(x,y)$ can be recovered by projecting onto position space $\langle x,y|$, with $\langle x|n\rangle = x^n$ and $\langle y|m\rangle = y^m$.

[¶] The adjoint operations are $\langle i|\hat{b}_i^+ = \langle i-1| - \langle i|$ for $i \in \{n,m\}$, $\langle n|\hat{b}_n^-(n) = (n+1)\langle n+1| - g_n\langle n|$, and $\langle n,m|\hat{b}_m^-(n) = (m+1)\langle n,m+1| - q_n\langle n,m|$.

[‖] In position space the eigenfunctions are $\langle x|j\rangle = (x-1)^j e^{\bar{g}(x-1)}$ and $\langle y|k\rangle = (y-1)^k e^{\bar{q}(y-1)}$. The operators $\hat{b}^+$ and $\bar{b}^-$ raise and lower in eigenspace: $\hat{b}_n^+|j\rangle = |j+1\rangle$ and $\bar{b}_n^-|j\rangle = j|j-1\rangle$ (or $\langle j|\hat{b}_n^+ = \langle j-1|$ and $\langle j|\bar{b}_n^- = (j+1)\langle j+1|$), and similarly for $n \to m$ and $j \to k$.

[**] The selection rules are derived by starting with $\langle n|\hat{b}_n^+|j\rangle$ or $\langle j|\hat{b}_n^+|n\rangle$ and allowing $\hat{b}_n^+$ to act both to the left and to the right. Alternatively, one may use $\bar{b}_n^-$, obtaining $\langle n+1|j\rangle = (j\langle n|j-1\rangle + \bar{g}\langle n|j\rangle)/(n+1)$ and $\langle j+1|n\rangle = (n\langle j|n-1\rangle - \bar{g}\langle j|n\rangle)/(j+1)$, initialized with $\langle n=0|j\rangle = (-1)^j e^{-\bar{g}}$ and $\langle j=0|n\rangle = 1$. We find the latter relations yield smoother distributions p_{nm} for large cutoffs N and J.

since species in a feedback loop will be nonnegligibly dependent on all other species in the loop if the loop is small enough.

Results

The Spectral Method Is Fast and Accurate. To demonstrate the accuracy and computational efficiency of the spectral method (all numerical procedures in this article are implemented in MATLAB), we compare it with both an iterative numerical solution of Eq. **3** and a stochastic simulation using the Gillespie algorithm (16) for a cascade of length $L = 2$ with a Poisson input ($g_n = g =$ constant) and the discontinuous threshold regulation function

$$q_n = \begin{cases} q_- & \text{for } n \le n_0 \\ q_+ & \text{for } n > n_0. \end{cases} \quad [9]$$

The spectral method achieves an agreement up to machine precision with the iterative method in ~0.01 s, which is ~1,000 times faster than the iterative method's run time and $\sim 10^8$ faster than the run time necessary for the stochastic simulation to achieve the same accuracy; see *SI Appendix* for details. The huge gain in computational efficiency over both the iterative method and the stochastic simulation makes the spectral method extremely useful, particularly for optimization problems, in which the probability function must be evaluated multiple times. In the next sections we exploit this feature to optimize information transmission in signaling cascades.

Information Processing in Signaling Cascades. Linear signaling cascades are a ubiquitous feature of biological networks, used to transmit relevant information from one part of a cellular system to another (23–27). Information processing in a cascade is quantified by the mutual information (33), which measures in bits how much information about an input signal is transmitted to the output signal in a noisy process. For a cascade of length L, the mutual information between an input species (with copy number n_1) and an output species (with copy number n_L)[††] is $I = \sum_{n_1,n_L} p(n_1,n_L)\log_2[p(n_1,n_L)/p(n_1)p(n_L)]$. In this study we define the capacity I^* as the maximum mutual information over either regulatory parameters, the input distribution, or both. Depending on the signal-to-noise ratio, a high-capacity cascade functions either where the input signal is strongest or where the transmission process is least noisy (34, 35).

We first consider a cascade of length L in which the regulation function q_n is a simple threshold (Eq. **9**) with fixed parameters that are identical for each cascade step. It is worth noting that while a threshold-regulated creation rate represents the simplest choice biologically, it is the most taxing choice computationally: as the discontinuity $\Delta = |q_+ - q_-|$ increases, we find both that (*i*) a larger cutoff K in eigenmodes is required for a desired accuracy, and (*ii*) the accuracy of the Markovian approximation decreases (*SI Appendix*). The results herein therefore constitute a stringent numerical challenge for the spectral method.

We take the input $p(n_1)$ to be a Poisson distribution (i.e., $g_n = g =$ constant). In the extreme cases, when the threshold is infinite or zero, the output is a Poisson distribution centered at q_- or q_+, respectively. Similarly, when the input median is below or above threshold, the output mean should be biased toward q_- or q_+, respectively. For example, in Fig. 2*A*, $\langle n_1\rangle < n_0$, and the output distribution is shifted toward q_-. This effect is amplified at each step of the cascade, such that $\langle n_L\rangle \to q_-$ for large L. Similarly, $\langle n_L\rangle \to q_+$ for large L when $\langle n_1\rangle > n_0$ (Fig. 2*C*). When $\langle n_1\rangle \sim n_0$ (Fig. 2*B*), the output is balanced between q_- and q_+;

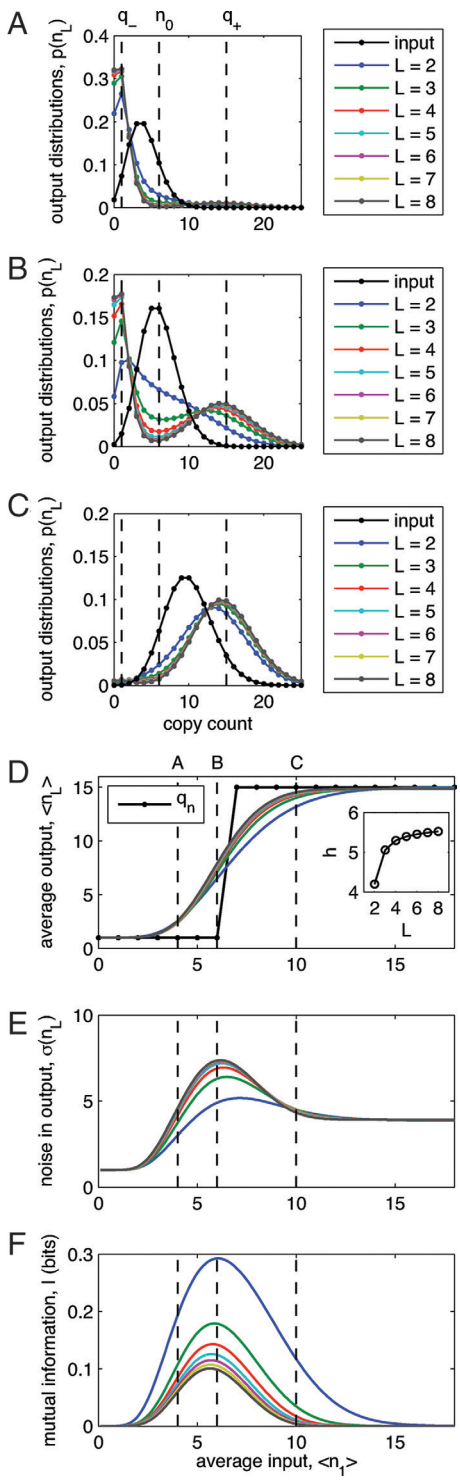

Fig. 2. Transfer functions and noise in a signaling cascade. (*A–C*) Plots of input distribution $p(n_1)$ (black) and output distributions $p(n_L)$ (colors; see legend) for various cascade lengths L. Input distribution is a Poisson centered at $\langle n_1\rangle = 4$ (*A*), $\langle n_1\rangle = n_0 = 6$ (*B*), or $\langle n_1\rangle = 10$ (*C*). The regulation function q_n for all steps is a threshold (Eq. **9**) shown in *D* (black line with dots), with parameters q_-, q_+, and n_0 overlaid as dashed lines in *A–C*. The degradation rate ratio is $\rho = 1$, and Eq. **3** is solved by using the spectral method with $\bar{q} = 10$ and $\bar{g} = \langle g_n\rangle$ for each step in the cascade. (*D–F*) Transfer functions (average output $\langle n_L\rangle$) (*D*), noise (standard deviation of the output $\sigma(n_L)$) (*E*), and mutual information I (*F*) as functions of average input $\langle n_1\rangle$ for various cascade lengths L (colors as in *A–C*). As in *A–C*, input is Poisson at every $\langle n_1\rangle$; dashed lines correspond to the specific $\langle n_1\rangle$ values in *A*, *B*, and *C*. (*D Inset*) Cooperativity h as a function of L.

[††]Although the spectral method only involves the calculation of joint distributions between adjacent species in the cascade, the input-output distribution can be obtained by using $p(n_1,n_\ell) = \sum_{n_{\ell-1}} p(n_1,n_{\ell-1})p(n_{\ell-1},n_\ell)/p(n_{\ell-1})$, initialized with $\ell = 3$ and run up to $\ell = L$. This assumes $p(n_1|n_{\ell-1},n_\ell) = p(n_1|n_{\ell-1})$, which at worst (at $\ell = 3$) is equivalent to the Markovian approximation, Eq. **2**.

if the discontinuity $\Delta = |q_+ - q_-|$ is sufficiently large, the output is bimodal, as discussed in more detail in the next section.

Mutual information I decreases monotonically with L for all $\langle n_1 \rangle$ (Fig. 2*F*), as required by the data processing inequality (36) (i.e., one cannot learn more information from the output of an $(L+1)$-gene cascade than one could from an L-gene cascade with identical regulation, only less). I is maximal for $\langle n_1 \rangle \sim n_0$ which makes intuitive sense, as it corresponds to the input taking advantage of both rates q_- and q_+ roughly equally in producing the output. A simple calculation quantifies this intuition. Approximating the steady-state distribution for the moment as a strict switch conditional on n_0 (i.e., $p(n_L|n_1) = p_-(n_L)$ if $n_1 \leq n_0$ and $p(n_L|n_1) = p_+(n_L)$ if $n_1 > n_0$ for some distributions $p_\pm(n_L)$), it follows from the definition of I that

$$I_{\text{switch}} = S - \sum_{\pm} \sum_{n_L} p_\pm(n_L) \log_2 \left[1 + \frac{\pi_\mp p_\mp(n_L)}{\pi_\pm p_\pm(n_L)} \right], \qquad [10]$$

where $S = -\sum_\pm \pi_\pm \log_2 \pi_\pm$, $\pi_- = \sum_{n_1 \leq n_0} p(n_1)$ and $\pi_+ = \sum_{n_1 > n_0} p(n_1) = 1 - \pi_-$. If there is little overlap between $p_-(n_L)$ and $p_+(n_L)$, then $I_{\text{switch}} \sim S$, which is maximal when $\pi_- = \pi_+$, i.e., when the median of the input distribution $p(n_1)$ lies at the threshold n_0. Additionally, since the maximal value of S (and I_{switch}, since the summand of the second term in Eq. **10** is always nonnegative) is 1 bit, this calculation also suggests that the capacity of threshold regulation (in the limit of strict switch-like behavior) is limited to 1 bit. Again, this result is intuitive, as the cascade is only passing on the binary information of whether particles in the input distribution are below or above the threshold.

In an experimental setup one might have access to only the mean response (or "transfer function"), or the variance in response across cells (or "noise"), of a signaling cascade to its input. Since our method yields the full distributions, such summary statistics are readily computed. Despite the sharpness of threshold regulation, the transfer functions are quite smooth even at $L = 2$ (Fig. 2*D*). The effect of the intrinsic noise is to smooth out a sharp discontinuity in creation rates, producing a continuous mean response. The transfer functions shown are least-squares fit to Hill functions of the form $\langle n_L|n_1 \rangle = \alpha_- + (\alpha_+ - \alpha_-)(n_1)^h/[(n_1)^h + (k_d)^h]$. As one would expect, for all L, best fit values of α_-, α_+ are near the rates q_- and q_+, respectively, and best fit k_d values are near the threshold n_0. As L increases, the transfer function sharpens, and cooperativity h increases (Fig. 2*D Inset*), due to the amplified migration of the output to either q_- and q_+ in longer cascades (as in Fig. 2 *A* and *C*).

The strength of the noise increases with L (Fig. 2*E*), consistent with the reduction in I with L (Fig. 2*F*), and the noise is peaked at the threshold. The "switch approximation" (Eq. **10**) illustrates the gain in information when the median of the input coincides with the threshold; the "small noise approximation" (34, 35), however, illustrates the loss in information when the peak of the input coincides with the peak of the noise. The trade-off between these two trends thwarts information transmission with unimodal input distributions (e.g., those used in Fig. 2) and suggests an input distribution with two or more modes should be able to transduce more information. Such distributions are the subject of study below, and, in related work (35, 37) are shown to be the optimal strategy and to be observed in biology for a regulatory system in which peak noise and threshold coincide.

Bimodal Output from a Unimodal Input. A striking feature of Fig. 2*B* is that the unimodal input is converted to a bimodal output for cascades of length $L = 3$ or longer. Bimodality can arise from a system with two genes whose proteins repress each other or from a single gene whose proteins activate its own expression. Here, we demonstrate that cascades with sufficiently strong regulation constitute an information-optimal mechanism for a cell to achieve bimodality.

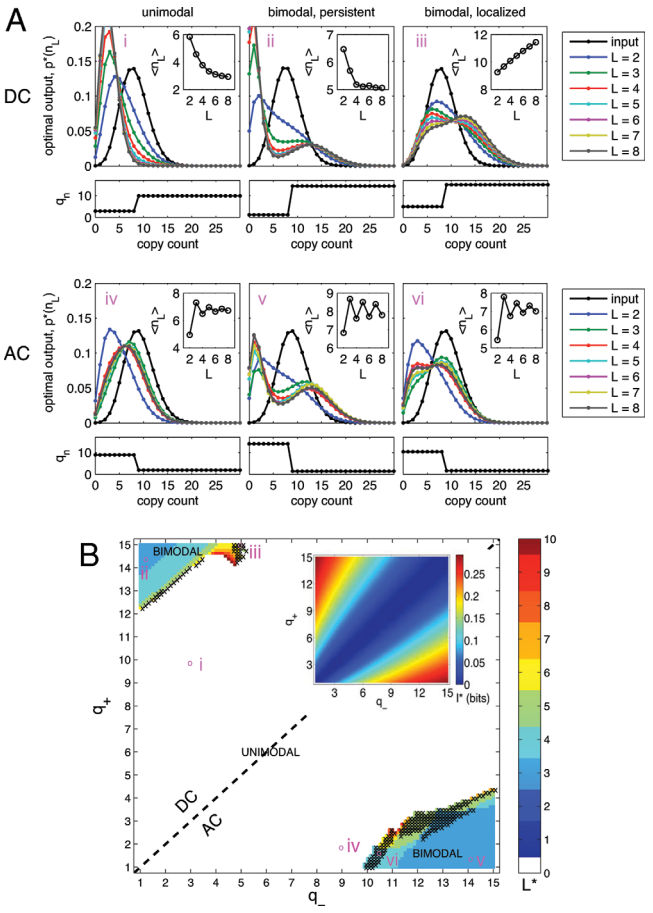

Fig. 3. Optimal output modality in cascades with unimodal input. (*A*) Plots of optimal input distribution $p^*(n_1)$ (black) and optimal output distributions $p^*(n_L)$ (colors; see legend) for different cascade lengths L (optimal input distributions are qualitatively the same; only that for $L = 2$ is shown), corresponding to regulation functions q_n (identical for each step) plotted underneath ($\rho = 1$, and solutions used $\bar{q} = 10$ and $\bar{g} = \langle g_n \rangle$). Mutual information I is optimized as a function of the mean g of the Poisson input distribution. Magenta numbers on plots correspond to magenta points in *B*. *Insets* show plots of average output $\langle n_L \rangle$ vs. cascade length L. In the first column of *A*, the output is always unimodal; in the second column, the output is bimodal for cascade lengths $L \geq L^*$ for some L^* ("persistent" bimodality); in the third column, the output is bimodal for a range of L values, then unimodal once more for large L ("localized" bimodality). The first row shows "DC" cascades, in which each step is up-regulating, and the second row shows "AC" cascades, in which each step is down-regulating. (*B*) Phase diagram of optimal output modality as a function of q_- and q_+ ($n_0 = 8$). White is unimodal, and color is bimodal, with color corresponding to L^*. Distinction between persistent (no 'x') and localized ('x') bimodality is shown up to $L = 10$. Dashed line separates DC cascades from AC cascades. (*Inset*) Capacity I^* in bits as a function of q_- and q_+ for the same data, with $L = 2$.

Recall that Fig. 2*B* corresponds to the case where the input distribution is optimally matched with the regulation function, i.e., the bimodal output represents optimal information transmission. By optimizing over the mean g of a Poisson input distribution, we find that the most informative output distribution in a cascade with unimodal input can be unimodal or bimodal, depending on regulatory parameters and the length of the cascade. Fig. 3*A* shows examples of regulation functions which produce output distributions that are unimodal, bimodal for cascades as long or longer than some L^* (which we term "persistent" bimodality), and bimodal for short cascades but unimodal at both initial and final nodes for longer cascades (which we term "localized" bimodality).

Bimodality is found both in cascades in which each step is down-regulating, which we call "AC" cascades, and in those in which

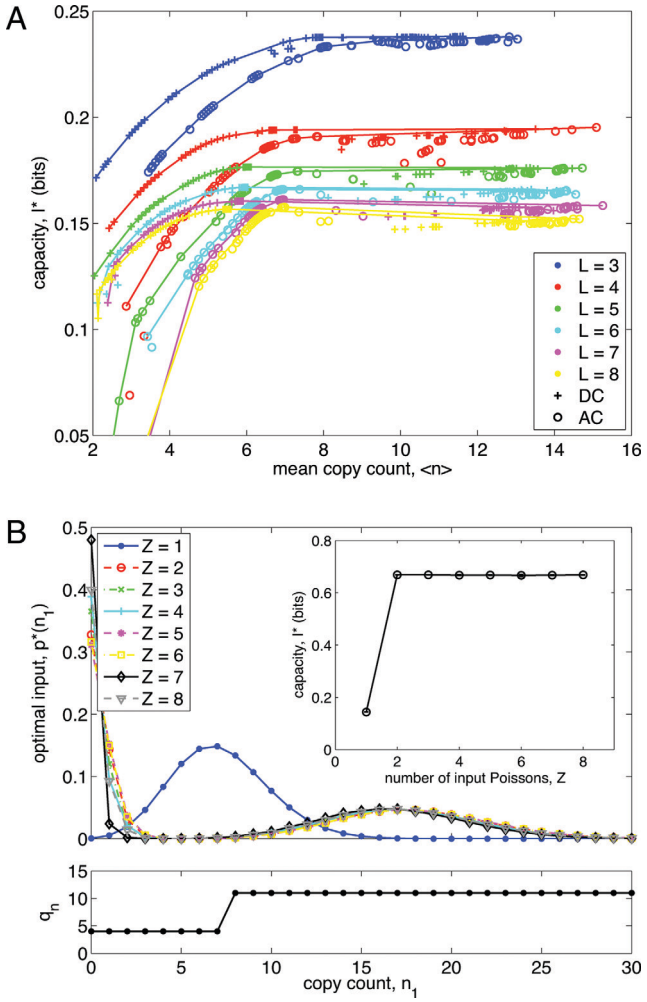

Fig. 4. Design principles of information-optimal cascades. (*A*) Capacities of AC and DC cascades. Capacity I^* vs. average copy count $\langle n\rangle$ (over all species) for AC (circles) and DC (plus signs) cascades of different lengths L (color), with Poisson input distributions. Results were obtained by optimizing the objective function $\mathcal{L}$ over all parameters (q_-, q_+, n_0, and g) for $\lambda = 1\times10^{-6} - 3\times10^{-2}$ (n_0 is constrained to be an integer, the regulation function is the same for every step, $\rho = 1$, and solutions used $\bar{q} = 10$ and $\bar{g} = \langle g_n\rangle$). Lines show convex hulls. (*B*) Optimal input distributions with different numbers Z of Poisson distributions. The cascade length is $L = 2$, the degradation rate ratio is $\rho = 1$, and the regulation function q_n is plotted; solutions used $\bar{q} = 10$ and $\bar{g} = \langle g_n\rangle$. The objective function $\mathcal{L}$ is optimized with $\lambda = 10^{-4}$ over all input parameters g_i and π_i. (*Inset*) Capacity I^* vs. Z, averaged at each Z over 7 optimizations with different initial conditions.

each step is up-regulating, which we call "DC" cascades. In DC cascades, as seen in Fig. 3*A i* (*Inset*)–*iii* (*Inset*) the average output either monotonically decreases or monotonically increases with L. In the former case, since $q_- < n_0$, the probability that the output is below the threshold given that the input is below threshold is large. Such successive regulations drive the probability of being below the threshold toward 1, successively decreasing $\langle n\rangle$ at each step in the cascade. In the latter case, since $q_+ > n_0$, the same picture holds, and $\langle n\rangle$ monotonically increases with L. Whether the monotonically increasing or decreasing behavior is the more informative is determined by the relationship among q_+, q_-, and n_0. In AC cascades, an analogous picture holds but with alternation: $\pi_- < \pi_+$ for the even-numbered links (Eq. **10**), and the AC condition $q_- > q_+$ leads to $\pi_+ < \pi_-$ for the odd-numbered links, as illustrated in Fig. 3*A iv* (*Inset*)–*vi* (*Inset*). These behaviors motivate the names "AC" and "DC," analogous to alternating and direct current flow. Performance of AC and DC cascades is compared in more detail in the next section.

Fig. 3*B* shows a phase diagram of optimal output modality as a function of the rates q_- and q_+: bimodality is found at high values of the discontinuity $\Delta = |q_+ - q_-|$ (specifically, for $\Delta \gtrsim 11$ in AC circuits and $\Delta \gtrsim 12$ in DC circuits when $n_0 = 8$). Intuitively, since the weight of the output is distributed between q_- and q_+ for long cascades, increasing their separation spreads the weights apart and creates bimodal distributions. Furthermore, as Δ increases, the bimodality becomes more robust: it goes from localized to persistent, and its onset occurs at a smaller cascade length L^*. Fig. 3*B Inset* shows that capacity I^* also increases with Δ; cascades with bimodal output therefore have higher capacities than those with unimodal output. As Δ increases, the information transmission properties of a regulatory cascade are better approximated by simple switch-like regulation (see Eq. **10**). In short, summarizing the input distribution by π_+ and π_- is a more informative summary of the distribution as the regulation becomes more discontinuous.

Channel Capacity in AC/DC Cascades. Our setup provides a way to ask quantitatively whether a cascade with up-regulating steps (AC) can transmit information with more or less fidelity than a cascade with up-regulating steps (DC). Since a cell must expend time and energy to make proteins, a fair comparison between cascade types can only be made when the species involved in each type are present in equal copy number. As in ref. 38, we introduce the objective function $\mathcal{L} = I - \lambda\langle n\rangle$, where I is mutual information and $\langle n\rangle = \sum_{\ell=1}^{L}\langle n_\ell\rangle/L$ is an average copy count over all species in the cascade. Here, λ represents the metabolic cost of making proteins, and optimizing L for different values of λ allows a comparison of AC and DC capacities I^* at similar values of $\langle n\rangle$.

For both AC and DC cascades, I^* increases with $\langle n\rangle$ as more proteins are made available to encode the signal,[‡‡] and I^* decreases[§§] with L at all $\langle n\rangle$ (Fig. 4*A*). Both AC and DC capacities converge to an L-dependent asymptotic value at high copy count, but DC cascades attain higher capacities per output protein than AC cascades. The difference is most pronounced at low copy count ($\langle n\rangle \lesssim 7$), and more pronounced still for longer cascades. The difference is easily explained: AC and DC cascades of the same length with the same discontinuity $\Delta = |q_+ - q_-|$ have the same capacity but have different mean numbers of proteins. Recall from Fig. 3*B* that large Δ leads to high-capacity, bimodal solutions. The difference between AC and DC cascades is in the placement of their optimal distributions for a given Δ. We observe that optimal AC cascades tend to exhibit $\langle n\rangle \gtrsim n_0$, whereas optimal DC cascades tend to exhibit $\langle n\rangle \lesssim n_0$. Ultimately, this allows DC cascades to achieve the same capacity for the same regulation parameters (Fig. 3*B Inset*), but use fewer proteins. These results suggest that DC cascades transmit with more information per protein than AC cascades when protein production is costly.

A Multimodal Input to a Threshold Is No More Informative than a Bimodal Input. If the first species is governed by more than a simple birth–death process, the input to a cascade will not be a simple Poisson distribution. To investigate the role of input multimodality in information transmission, we consider inputs defined by a

[‡‡] There is a slight decrease in I^* with $\langle n\rangle$ beginning near $\langle n\rangle \approx 8$ that is more pronounced at higher L. This is likely due to the decrease in accuracy of the Markovian approximation with increasing Δ (*see SI Appendix*), since large $\langle n\rangle$ requires large Δ. Calculations with the full joint distribution (via stochastic simulation) at $L = 3$ and 4 give qualitatively similar results, but with I^* increasing monotonically with $\langle n\rangle$.

[§§] The decrease of I^* with L is consistent with, but not a direct consequence of, the data processing inequality (36), as each $p^*(n_1,\ldots,n_L)$ results from a separate optimization for each subsequent choice of L.

mixture of Poisson distributions, $p(n_1) = \sum_{i=1}^{Z} \pi_i e^{-g_i} g_i^{n_1}/n_1!$, (with $\sum_{i=1}^{Z} \pi_i = 1$). As before, we expect information to increase with copy number, and we use the objective function $\mathcal{L}$ when optimizing over the input distribution $p(n_1)$.

All optimal input distributions with $Z \geq 2$ are bimodal, with one mode on either side of the threshold (Fig. 4*B*). When Z is 3 or more, either all but two π_i values are driven by the optimization to 0, or all the g_i values with nonzero weights are driven to one of two unique values. The two modes are roughly equally weighted (i.e., $\pi_1 \approx \pi_2 \approx 0.5$), consistent once more with our calculation that the median of the optimal input distribution falls roughly at the threshold (Eq. **10**). As an additional verification, a plot of capacity I^* vs. Z in Fig. 4*B Inset* reveals that I^* remains constant for $Z = 2$ and beyond. A threshold regulation function presents a binary choice, and the optimal input is a bimodal distribution that equally utilizes both sides. Last, we point out that the capacities in Fig. 4*B Inset* are <1 bit. Even with a bimodal input distribution, a short cascade ($L = 2$), and strong regulation functions (i.e., large discontinuities Δ), we do not find capacities >1 bit. This is consistent with our calculation under the approximation of threshold regulation as a switch (Eq. **10**) and with the intuition that a threshold represents a binary decision.

Conclusions

We have introduced a method, the *spectral* method, that exploits the linear algebraic structure of the master equation, and expands the full problem in terms of its natural eigenfunctions. We have illustrated our method by probing the optimal transmission properties of signaling cascades with threshold regulation. We have shown that sufficiently long cascades with sufficiently strong regulation functions optimally convert a unimodal input to a bimodal output. A bimodal input is optimal for information transmission across a threshold, and a multimodal input offers no further processing power. Sustained bimodality of the output distribution requires large discontinuities Δ between the production rates below and above the threshold. The value of Δ controls the maximum information transmitted by a cascade with threshold regulation in a similar way for cascades of up-regulations (DC) and cascades of down-regulations (AC), but a DC cascade outperforms an AC cascade by using fewer average copies of its species. We emphasize that the application of the spectral method to signaling cascades represents only a beginning. Variations on the natural bases in which to expand, and extensions of the method to other small network topologies, should be the subject of future studies. More generally, however, we anticipate that the method will prove useful in the direct solution of a large class of master equations describing a wide variety of biological systems.

ACKNOWLEDGMENTS. We thank Jake Hofman and Melanie Lee for their help in this project. A.M.W. was supported by the Princeton Center for Theoretical Science Fellowship and Columbia's Professional Schools Diversity Short-Term Visiting Fellowship. C.W. was supported by National Institutes of Health Grants 5PN2EY016586-03 and 1U54CA121852-01A1; C.W. and A.M. were supported by National Science Foundation Grant ECS-0332479; A.M. was supported by National Science Foundation Grant DGE-0742450.

1. Hooshangi S, Thiberge S, Weiss R (2005) Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proc Natl Acad Sci USA* 102:3581–3586.
2. Thattai M, van Oudenaarden A (2002) Attenuation of noise in ultrasensitive signaling cascades. *Biophys J* 82:2943–2950.
3. Paulsson J, Berg OG, Ehrenberg M (2000) Stochastic focusing: Fluctuation-enhanced sensitivity of intracellular regulation. *Proc Natl Acad Sci USA* 97:7148–7153.
4. Elowitz MB, Levine AJ, Siggia ED, Swain PS (2002) Stochastic gene expression in a single cell. *Science* 297:1183–1186.
5. Ozbudak EM, Thattai M, Kurtser I, Grossman AD, van Oudenaarden A (2002) Regulation of noise in the expression of a single gene. *Nat Genet* 31:69–73.
6. Swain PS, Elowitz MB, Siggia ED (2002) Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc Natl Acad Sci USA* 99:12795–12800.
7. Acar M, Becskei A, van Oudenaarden A (2005) Enhancement of cellular memory by reducing stochastic transitions. *Nature* 435:228–232.
8. Pedraza JM, van Oudenaarden A (2005) Noise propagation in gene networks. *Science* 307:1965–1969.
9. Thattai M, van Oudenaarden A (2001) Intrinsic noise in gene regulatory networks. *Proc Natl Acad Sci USA* 98:8614–8619.
10. van Zon JS, Morelli MJ, Tănase-Nicola S, ten Wolde PR (2006) Diffusion of transcription factors can drastically enhance the noise in gene expression. *Biophys J* 91:4350–4367.
11. van Zon JS, ten Wolde PR (2005) Green's-function reaction dynamics: A particle-based approach for simulating biochemical networks in time and space. *J Chem Phys* 123:234910.
12. Allen RJ, Warren PB, ten Wolde PR (2005) Sampling rare switching events in biochemical networks. *Phys Rev Lett* 94:18104.
13. Macnamara S, Burrage K, Sidje RB (2007) Multiscale modeling of chemical kinetics via the master equation. *Multiscale Model Simul* 6:1146–1168.
14. Ushikubo T, Inoue W, Yoda M, Sasai M (2006) Testing the transition state theory in stochastic dynamics of a genetic switch. *Chem Phys Lett* 430:139–143.
15. Bortz AB, Kalos MH, Lebowitz JL (1975) A new algorithm for Monte Carlo simulation of Ising spin systems. *J Comput Phys* 17:10–18.
16. Gillespie DT (1977) Exact stochastic simulation of coupled chemical reactions. *J Phys Chem* 81:2340–2361.
17. Paulsson J (2004) Summing up the noise in gene networks. *Nature* 427:415–418.
18. Tănase-Nicola S, Warren PB, ten Wolde PR (2006) Signal detection, modularity, and the correlation between extrinsic and intrinsic noise in biochemical networks. *Phys Rev Lett* 97:68102.
19. Walczak AM, Onuchic JN, Wolynes PG (2005) Absolute rate theories of epigenetic stability. *Proc Natl Acad Sci USA* 102:18926–18931.
20. Lan Y, Papoian GA (2006) The interplay between discrete noise and nonlinear chemical kinetics in a signal amplification cascade. *J Chem Phys* 125: 154901.
21. Lan Y, Papoian GA (2007) Stochastic resonant signaling in enzyme cascades. *Phys Rev Lett* 98:228301.
22. Lan Y, Wolynes PG, Papoian GA (2006) A variational approach to the stochastic aspects of cellular signal transduction. *J Chem Phys* 125:124106.
23. Gomperts BD, Krame, IM, Tantham PER (2002) *Signal Transduction* (Academic, San Diego).
24. Ting AY, Endy D (2002) Decoding NF-kappaB signaling. *Science* 298:1189–1190.
25. Detwiler PB, Ramanathan S, Sengupta A, Shraiman BI (2000) Engineering aspects of enzymatic signal transduction: photoreceptors in the retina. *Biophys J* 79:2801–2817.
26. Bolouri H, Davidson EH (2003) Transcriptional regulatory cascades in development: Initial rates, not steady state, determine network kinetics. *Proc Natl Acad Sci USA* 100:9371–9376.
27. Bassler BL (1999) How bacteria talk to each other: Regulation of gene expression by quorum sensing. *Curr Opin Microbiol* 2:582–587.
28. van Kampen NG (1992) *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam).
29. Mattis DC, Glasser ML (1998) The uses of quantum field theory in diffusion-limited reactions. *Rev Mod Phys* 70:979–1001.
30. Zel'Dovich YB, Ovchinnikov AA (1978) The mass action law and the kinetics of chemical reactions with allowance for thermodynamic fluctuations of the density. *Sov JETP* 47:829.
31. Doi M (1976) Second quantization representation for classical many-particle system. *J Phys A Math Gen* 9:1465–1477.
32. Sasai M, Wolynes PG (2003) Stochastic gene expression as a many-body problem. *Proc Natl Acad Sci USA* 100:2374–2379.
33. Shannon CE (1949) Communication in the presence of noise. *Proc IRE* 37:10–21.
34. Tkačik G, Callan CG, Bialek W (2008) Information capacity of genetic regulatory elements. *Phys Rev E* 78:11910.
35. Tkacik G, Callan CG, Bialek W (2008) Information flow and optimization in transcriptional regulation. *Proc Natl Acad Sci USA* 105:12265–12270.
36. Cover TM, Thomas JA (1991) *Elements of Information Theory* (Wiley, New York).
37. Gregor T, Tank DW, Wieschaus EF, Bialek W (2007) Probing the limits to positional information. *Cell* 130:153–164.
38. Ziv E, Nemenman I, Wiggins CH (2007) Optimal signal processing in small stochastic biochemical networks. *PLoS ONE* 2:e1077.