Correlation Resonance Generated by Coupled Enzymatic Processing

William H. Mather,[†‡] Natalie A. Cookson,[†§] Jeff Hasty,[†‡§] Lev S. Tsimring,[†] and Ruth J. Williams[†¶*]
[†]Biocircuits Institute, [‡]Department of Bioengineering, [§]Molecular Biology Section, Division of Biological Sciences, and [¶]Department of Mathematics, University of California, San Diego, La Jolla, California

ABSTRACT A major challenge for systems biology is to deduce the molecular interactions that underlie correlations observed between concentrations of different intracellular molecules. Although direct explanations such as coupled transcription or direct protein-protein interactions are often considered, potential indirect sources of coupling have received much less attention. Here we show how correlations can arise generically from a posttranslational coupling mechanism involving the processing of multiple protein species by a common enzyme. By observing a connection between a stochastic model and a multiclass queue, we obtain a closed form expression for the steady-state distribution of the numbers of molecules of each protein species. Upon deriving explicit analytic expressions for moments and correlations associated with this distribution, we discover a striking phenomenon that we call correlation resonance: for small dilution rate, correlations peak near the balance-point where the total rate of influx of proteins into the system is equal to the maximum processing capacity of the enzyme. Given the limited number of many important catalytic molecules, our results may lead to new insights into the origin of correlated behavior on a global scale.

INTRODUCTION

Significant correlations in the concentrations of different intracellular molecules are often observed in the context of a cellular response to an external perturbation such as stress. It is a major challenge in systems biology to deduce the molecular interactions that underlie such correlations. Intracellular interactions are frequently described by the forward propagation of a signal through specific molecular complexes, as found, for example, in MAPK cascades and feed-forward gene networks (1,2).

These forward interactions are often used as the building blocks for systems with bidirectional interactions, such as gene networks with feedback. A general and less obvious bidirectional interaction between components can arise when components share and compete for common machinery. This type of indirect interaction increases the apparent coupling between seemingly disparate components. An example of this type of coupling has recently been observed in a system where two fluorescent reporters were tagged for fast degradation by a common protease (N. A. Cookson, W. H. Mather, T. Danino, O. Mondragón-Palomino, R. J. Williams, L. S. Tsimring, and J. Hasty, unpublished). In this study, it was shown that a rate-limited interaction can lead to strong coupling between the two fluorescent reporters.

As a theoretical complement to the aforementioned experimental study, here we investigate, in detail, a stochastic model for the processing of multiple protein species by a common enzyme. Our stochastic model tracks the dynamics of the numbers of molecules of finitely many protein species subject to processing by a limited number of copies of a common enzyme. By observing a connection between this stochastic model and a multiclass queue, we obtain a closed form expression for the steady-state distribution of the numbers of molecules of each protein species. We derive explicit analytic expressions for moments and correlations of this distribution and discover a striking phenomenon that we call "correlation resonance": for small dilution rate, correlations between the numbers of molecules for different protein species peak near the balance-point, where the total rate of influx of proteins into the system is equal to the maximum processing capacity of the enzyme. The correlations at the peak become perfect when dilution becomes negligible.

Although there are various interesting studies of correlations in biochemical reaction network models, including some considering correlations in time as well as across species (see, e.g., (4,5)), we believe this is the first work to analyze correlations between different species in a stochastic model where the species are processed by a common but limited number of copies of an enzyme. Indeed, our formulas and qualitative findings concerning correlations appear to be new both for biological systems and queuing theory. Simulation results are in excellent agreement with our analytic results and suggest that our qualitative findings are robust to variations in the model that transcend our exact analysis.

To our knowledge, this is the first use of the theory of multiclass queues to investigate correlations in a genetic network model, and for this, multiclass queues seem well suited. Historically, multiclass queues have been used in business and electrical engineering to model congestion and delay arising when different classes of customers or packets "wait in line" for processing by a common limited resource. For example, in the Internet, packets from qualitatively different classes of content, e.g., voice and video, can be processed by a single server. The server has a maximum

Submitted June 26, 2010, and accepted for publication September 24, 2010.

*Correspondence: williams@stochastic.ucsd.edu

Editor: R. Dean Astumian.

0006-3495/10/11/3172/10 $2.00

doi: 10.1016/j.bpj.2010.09.057

processing speed and so even though voice messages are distinct from video, an increase in voice traffic will slow video processing. Accordingly, there is a correlation between video throughput and voice throughput.

Multiclass queues are also used in modeling manufacturing, computer, and service systems where the entities being processed can be products, tasks, or people, and are usually referred to generically as "jobs"—a term we shall adopt henceforth in describing multiclass queues. For genetic networks, molecules are analogous to "jobs", copies of the enzyme are analogous to "servers", and protein species correspond to "classes". As we illustrate in this article, the modern theory of multiclass queues can enable one to derive formulas and qualitative properties of correlations between multiple protein species processed by common enzymatic machinery. Single-class queuing models recently considered in Arazi et al. (6) and Levine and Hwa (7) in the context of genetic networks allow only a single chemical species to be processed by a given enzyme, and do not capture the kinds of correlations treated here.

The article is organized as follows. In Stochastic Model, beginning with a standard set of biochemical reactions, we define precisely our stochastic model of coupled enzymatic processing. In Main Results, we summarize the main results of the article which are derived in Queuing Connection and Steady-State Distribution, Moments and Correlations, and Generalization for Reversible Binding. These results are based on the new observation that there is a connection between our stochastic model and a multiclass queue (see Fig. 1). This connection is explained in detail in Queuing Connection and Steady-State Distribution, where we use it to obtain a closed form expression for the steady-state distribution of the numbers of molecules of each protein species in our stochastic model. In Moments and Correlations, we derive explicit formulas for moments and correlations associated with this steady-state distribution.

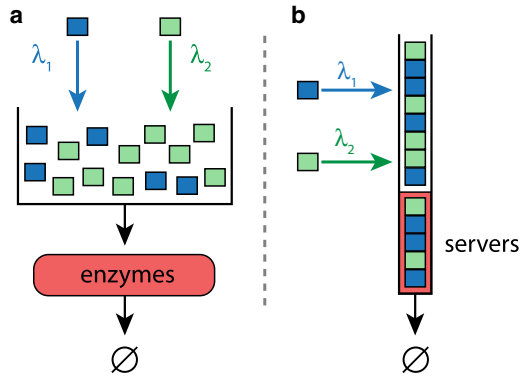

FIGURE 1 Schematic for $m = 2$ (two species of protein) for the stochastic model and an associated multiclass queue. The molecular count process in the stochastic model is equivalent in distribution to the job count process in the multiclass queue. (*a*) The stochastic model where proteins of species i are added to a volume at rate λ_i and are selected at random (without regard to type) for enzymatic processing. (*b*) A multiclass queue associated with the stochastic model where jobs are analogous to molecules and servers are analogous to copies of the enzyme. Jobs of type i arrive at rate λ_i and are inserted randomly into a queue of jobs awaiting processing; when a server becomes free, a new job is selected for processing from the head of the queue of waiting jobs.

Our observations concerning correlation resonance are based on these formulas, which are illustrated for sample parameters in Figs. 2 and 3. Generalization for Reversible Binding derives similar results to those in Moments and Correlations for a generalization of our stochastic model considered as an approximate model for the biochemical reaction system with reversible binding of the enzyme. We conclude the article with a discussion of the potential implications that correlations generated by indirect coupling may have for the behavior of genetic networks. In anticipation of potential future network applications, in W. H. Mather, J. Hasty, L. S. Tsimring, and R. J. Williams (unpublished), we generalize the queuing results used here from the situation of a single multiclass queue to networks of multiclass queues.

STOCHASTIC MODEL

For concreteness, we consider a stochastic model of enzymatic degradation of proteins as a prototype for common enzymatic processing. Other types of processing could be treated similarly, provided that products of the processing do not interact further with the other components of the model.

We consider a system with m protein species $X_1,\ldots,X_m$, which are being produced from their corresponding DNA templates $D_1,\ldots,D_m$ at constant rates $\lambda_1,\ldots,\lambda_m$, respectively. Binding of a copy of the protein X_i to an (unbound) copy of the enzyme E occurs at rate η and the complex X_iE is degraded with rate μ. We also take into account dilution of proteins at rate γ, in both their bound and unbound forms. This system is governed by the following set of biochemical reactions. For $i = 1,\ldots,m$,

$$D_i \xrightarrow{\lambda_i} X_i + D_i, \tag{1}$$

$$X_i + E \xrightarrow{\eta} X_iE, \tag{2}$$

$$X_iE \xrightarrow{\mu} E, \tag{3}$$

$$X_iE \xrightarrow{\gamma} E, \tag{4}$$

$$X_i \xrightarrow{\gamma} \varnothing. \tag{5}$$

All reaction times are exponentially distributed. We assume that the binding reaction rate η is so large that the associated binding reactions are effectively instantaneous.

More precisely, our stochastic model for this system may be described as follows. New molecules of protein species i are produced at the jump times of a Poisson process with rate $\lambda_1 > 0$, for each $i = 1,\ldots,m$. We assume there are $L > 0$ copies of the enzyme and that, when one of the L copies becomes free, it selects a protein molecule at

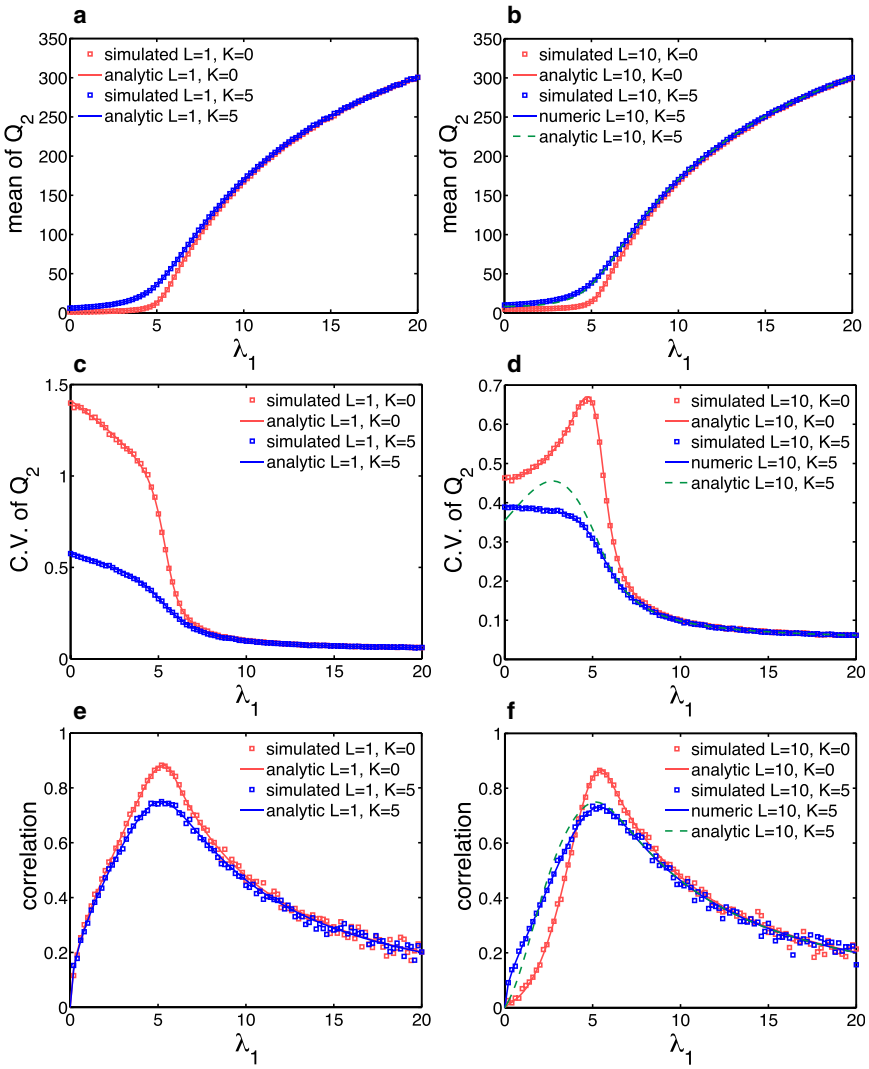

FIGURE 2 Results from stochastic simulation of the biochemical reactions compared to analytic and numerical results for two types of proteins ($m = 2$). Here "analytic" corresponds to analytic formulas obtained for the stochastic model with $\eta_+ = +\infty$ when $K = 0$, and with the Michaelis-Menten-like approximation for degradation rate when $K > 0$. The label "numeric" represents results for a refinement of the Michaelis-Menten-like approximation (when $L > 1$ and $K > 0$) where the distribution of N is computed numerically using a one-dimensional lattice method. (*a*, *c*, and *e*) Mean and coefficient of variation for Q_2, and correlation, all as functions of λ_1 for $L = 1$ (one copy of the enzyme) and $K = 0, 5$. Parameters used were $\mu = 10$, $\lambda_2 = 5$, and $\gamma = 0.01$. For simulations, we set $\eta_+ = 10^8$ when $K = 0$ and $\eta_+ = 200$, and $\eta_- = 1000$ when $K = 5$. Averages (in time) for stochastic simulations were taken from trajectories of duration 2×10^5. (*b*, *d*, and *f*) Similar to panels *a*, *c*, and *e*, with the only parameter differences being $L = 10$, $\mu = 1$, and $\eta_+ = 20$, $\eta_- = 100$ when $K = 5$.

random, binds to it instantly, and begins to degrade it. The degradation time is exponentially distributed with mean $1/\mu$ where $\mu > 0$. In addition, dilution is modeled as follows (we view this as an approximation to the discrete process of cell growth and division). Each protein molecule will remain in the system for, at most, an exponentially distributed amount of time with mean $1/\gamma$ where $\gamma > 0$. A molecule may be removed from the system by degradation before its dilution lifetime is up, or vice versa. Because the number of copies of the enzyme is often tightly regulated, we assume that dilution does not remove copies of the enzyme.

Production, degradation, and dilution are all statistically independent, production of one protein species is independent of that of any other species, and times to degrade distinct molecules as well as times until removal by dilution for distinct molecules are all independent of one another. Note that μ and γ are the same for all protein species. Arguably, a common value of γ is reasonable, as dilution is a communal effect. Although our exact analysis uses the assumptions of a common degradation rate μ and a fixed number of copies of the enzyme, simulations of the biochemical reactions for two protein species with differing degradation rates for the two protein species and a randomly varying number of copies of the enzyme (due to dilution and production) suggest that our qualitative findings regarding correlations are robust to such changes in the model (see the Supporting Material for more details).

We let $Q_i(t)$ denote the total number of molecules of protein species i that are present in the system at time t; this includes free molecules of species i as well as molecules of the species that are in the process of being degraded by enzyme. Let

$$N(t) = \sum_{i=1}^{m} Q_i(t)$$

be the total number of molecules of all protein species that are present in the system at time t. The stochastic process $\{Q_i(t), t \geq 0\}$ is the molecular count process for species i and

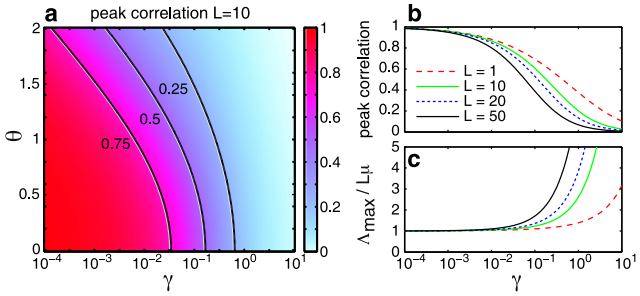

FIGURE 3 Dependence on system parameters of the peak (maximum) correlation and its location (when the correlation is considered as a function of $\Lambda = \lambda_1 + \lambda_2$ with L, μ, $\theta = \log_{10}(\lambda_1/\lambda_2)$, and γ fixed), for $m = 2$ (two species of protein). The peak correlation $r_{\max}$ and the corresponding total production rate $\Lambda_{\max}$ at the peak were found by numerical methods using the analytic formula for the correlation. (*a*) Plot of $r_{\max}$ as a function of θ and γ for $L = 10$, $\mu = 1$, and $K = 0$. Only values of peak correlation for $\theta \geq 0$ are shown because $r_{\max}(\theta,\gamma) = r_{\max}(-\theta,\gamma)$ due to symmetry. (*b*) Peak correlation as a function of γ for $\theta = 0$ and different values of L with $L\mu = 10$ and $K = 0$. (*c*) Plot of $\Lambda_{\max}/L\mu$ as a function of γ for several different values of L. Other parameters are the same as in panel *b*. The value of $\Lambda_{\max}$ is independent of θ (see the Supporting Material), though the value of the peak correlation does depend on θ. For small γ, $\Lambda_{\max} \approx L\mu$, i.e., peak correlation occurs near balance.

$$\{Q(t) = (Q_1(t), \ldots, Q_m(t)), t \geq 0\}$$

is the vector-valued count process for all species. (The letter Q here is mnemonic for quantity and the letter N is mnemonic for total number.) Because μ and γ do not depend upon i, the process $N = \{N(t), t \geq 0\}$ is a one-dimensional birth-death process and consequently is Markov. However, Q is not a Markov process. To obtain a full Markovian state descriptor for our model, we need to include some further information about the numbers of each species that are being processed at any given time. In Queuing Connection and Steady-State Distribution, we describe a multiclass queuing model whose job count process is distributionally equivalent to the molecular count process Q and give a natural Markovian state descriptor for the multiclass queue (see Fig. 1).

MAIN RESULTS

Using a connection between our stochastic model and a multiclass queue, we show that the steady-state distribution for the molecular count process, Q, can be described as follows. Conditioned on the total number of protein molecules N in the system being n, the steady-state distribution for Q is a multinomial distribution with parameters $(n; p_1,\ldots,p_m)$, where $p_i = \lambda_i/\Lambda$, $i = 1,\ldots,m$, and

$$\Lambda = \sum_{i=1}^{m} \lambda_i$$

is the total rate at which protein molecules (of all species) are being produced. In other words, in steady state, given that there is a total of n molecules in the system, the distribution of the quantities of each type is as if each of the n molecules, independently of the others, is of type i with probability p_i, the proportion of the newly produced proteins that are of type i.

Consequently, the steady-state moments for Q can be expressed in terms of the steady-state moments for N. In particular, for $i = 1,\ldots,m$, the steady-state mean and squared coefficient of variation (SCV = variance divided by the square of the mean) for Q_i are given by

$$E[Q_i] = p_i E[N], \tag{6}$$

$$SCV(Q_i) = SCV(N) - \frac{1}{E[N]} + \frac{1}{E[Q_i]}, \tag{7}$$

and the correlation between Q_i and Q_j for $j \neq i$ is given by

$$\begin{aligned} r_{ij} &= \frac{E[Q_iQ_j] - E[Q_i]E[Q_j]}{\sqrt{Var(Q_i)Var(Q_j)}} \\ &= \frac{\nu(N) - 1}{\left(\nu(N) - 1 + \frac{1}{p_i}\right)^{1/2}\left(\nu(N) - 1 + \frac{1}{p_j}\right)^{1/2}}, \end{aligned} \tag{8}$$

where $E[Y]$ denotes expectation (or mean) of a steady-state random variable Y, $Var(Y)$ denotes the variance of Y, and $\nu(N) = Var(N)/E[N]$ is the Fano factor for N.

The steady-state distribution for N (being that of a one-dimensional birth-death process) is readily seen to be given by

$$P(N = n) = c\frac{\Lambda^n}{\prod_{\ell=1}^{n}\phi(\ell)}, \quad n = 0, 1, 2, \ldots, \tag{9}$$

where

$$\phi(n) = \min(n, L)\mu + n\gamma, \quad n = 0, 1, 2, \ldots, \tag{10}$$

is the total rate at which protein molecules are being removed from the system due to the combined effects of degradation and dilution and c is the normalizing constant chosen so that the right-hand side of Eq. 9 represents a probability distribution. Explicit expressions for the steady-state moment-generating function and first and second moments for N are derived later in Moments and Correlations when there is one copy of the enzyme ($L = 1$) and in the Supporting Material when there are multiple copies ($L > 1$). These yield corresponding analytic formulas for steady-state moments and correlations for Q. In Figs. 2 and 3, we illustrate these formulas with plots for some sample parameters in the case of two protein species ($m = 2$).

In Fig. 2, sample plots are given for the mean of Q_2, coefficient of variation for Q_2, and correlation between Q_1 and Q_2, all as functions of λ_1 for $L = 1$ and $L = 10$, while the other parameters are held fixed. The plots labeled $K = 0$ correspond to the stochastic model described so far, whereas the plots for $K > 0$ correspond to a generalization considered further below. The plots show results produced by stochastic simulation of the set of biochemical reactions

as well as those obtained using our analytic formulas. Comparison of the two shows excellent agreement.

The peak in correlation apparent in Fig. 2, *e* and *f*, illustrates a general phenomenon that we observed in numerical exploration of our analytic correlation formulas: with small dilution rate ($\gamma << \mu$), the correlation has a peak near the balance-point where the total rate of influx of proteins is equal to the maximum processing capacity of the enzyme, i.e., where $\lambda_1 + \lambda_2 = L\mu$. We call this phenomenon "correlation resonance". These observations are consistent with the following simple asymptotic formula derived for $L = 1$ which shows that, in the zero dilution limit ($\gamma \to 0$), the correlation becomes perfect as the balance-point ($\rho \stackrel{\Delta}{=} \Lambda/\mu = 1$) is approached from below:

$$r_{ij} = \left(1 + \tfrac{1}{p_i}\left(\tfrac{1}{\rho} - 1\right)\right)^{-1/2}\left(1 + \tfrac{1}{p_j}\left(\tfrac{1}{\rho} - 1\right)\right)^{-1/2}, \quad i,j = 1,\ldots,m, \text{ and } j \neq i. \tag{11}$$

In Fig. 3, based on numerical analysis of our analytic formulas for two species, we illustrate the dependence of the size and location of the peak correlation on the model parameters. Fixing the number of copies of the enzyme L, the logarithm of the ratio of the production rates $\theta = \log_{10}(\lambda_1/\lambda_2)$, the degradation rate μ, and the dilution rate γ, and considering the correlation r_{12} as a function of $\Lambda = \lambda_1 + \lambda_2$, the peak correlation $r_{\max}$ and the value of $\Lambda = \Lambda_{\max}$ (where the peak occurs) were computed numerically.

The plots in Fig. 3 are representative of the results we found in numerical exploration of our formulas. Fig. 3 *a* illustrates the dependence of $r_{\max}$ on θ and γ, for $L = 10$ and $\mu = 1$. We observe that the peak correlation is large (near 1) for small dilution rate γ and nearly equal production rates ($\theta \approx 0$), and decays monotonically as either γ or $|\theta|$ is increased. For γ fixed, the maximum value of $r_{\max}$ occurs at $\theta = 0$. This maximum value is close to 1 for small γ and decays to 0 as γ is increased. Fig. 3 *b* illustrates, for $\theta = 0$ and fixed processivity $L\mu = 10$, the dependence of $r_{\max}$ on γ for several different values of L. For large and small γ, $r_{\max}$ approaches 0 and 1, respectively, independent of L. But for intermediate γ, $r_{\max}$ decays with L as expected, because increasing L reduces competition between the species for enzyme. Fig. 3 *c* illustrates the dependence of $\Lambda_{\max}$ on γ for several different values of L with $L\mu = 10$. For fixed L, μ and γ, $\Lambda_{\max}$ is independent of θ (see the Supporting Material), and for small values of γ, $\Lambda_{\max} \approx L\mu$.

The generalization of our stochastic model considered later in Generalization for Reversible Binding is intended as an approximation for the biochemical reaction system where the binding reactions in Eq. 2 are allowed to be reversible,

$$X_i + E \underset{\eta_-}{\overset{\eta_+}{\rightleftharpoons}} X_iE, \quad i = 1,\ldots,m, \tag{12}$$

and η_+ and η_- are both large. This generalization is obtained by simply replacing the degradation rate μ in our stochastic model by a Michaelis-Menten-like rate function of the form $n\mu/(K + n)$, where $K = \eta_-/\eta_+$ and n is the total number of proteins in the system. With this change in the stochastic model, the formulas in Eqs. 6–8 for the steady-state means, SCVs, and correlations of Q in terms of the steady-state moments of N still hold, but the moments of N are now different. We derive explicit formulas for these for $L = 1$ and $K > 0$ in Generalization for Reversible Binding and for $L > 1$ and $K > 0$ in the Supporting Material. Plots in Fig. 2 labeled with $K = 5$ illustrate these formulas for a two-species model.

In addition to the plots resulting from stochastic simulation of the set of biochemical reactions and from our analytic formulas, for $L = 10$ and $K = 5$, additional curves labeled "numeric" result from a refinement of the Michaelis-Menten-like approximation which gives a better fit to the simulation results. For this refinement, the degradation rate μ in the original stochastic model is replaced by a state-dependent rate $\mu k(n)$ where n is the total number of protein molecules in the system. The function $k(n)$ is the steady-state fraction of the enzymatic processors that are bound to protein molecules, computed for the fast binding/unbinding reactions given by Eq. 12, and conditioned on the total number of protein molecules in the system being fixed at n. The plot is labeled as "numeric" because even though an analytic expression is available for $k(n)$, computing the moments of N with this generalization was done numerically.

Fig. 4 shows sample trajectories illustrating the dynamic behavior of the molecular count process Q when there are 10 copies of the enzyme ($L = 10$); the other parameter values (except for λ_1) are the same as in Fig. 2, *b*, *d*, and *f*, when $K = 5$ and λ_1 is allowed to take the three values 1, 5, and 20, corresponding to underloaded, balanced, and overloaded regimes, respectively. Although our mathematical analysis does not treat dynamic behavior, the correlations exhibited here are consistent with our steady-state results and the behavior in the $\lambda_1 = 5$ case is reminiscent of process-level correlation (or state space collapse) observed for various multiclass queues near the balance-point (9,10).

In the next three sections, we give a full description of our results and explain how these results are derived.

QUEUING CONNECTION AND STEADY-STATE DISTRIBUTION

We can envisage the evolution of the stochastic model introduced in Stochastic Model as that of a multiclass queue in which molecules are analogous to jobs, copies of the enzyme are analogous to servers, and degradation corresponds to processing of a job by a server. In addition, dilution corresponds to a phenomenon called reneging, wherein jobs waiting or in service abandon the system before their processing is completed. The job count process for such a multiclass queue has the same distribution as the molecular count process Q for our stochastic model.

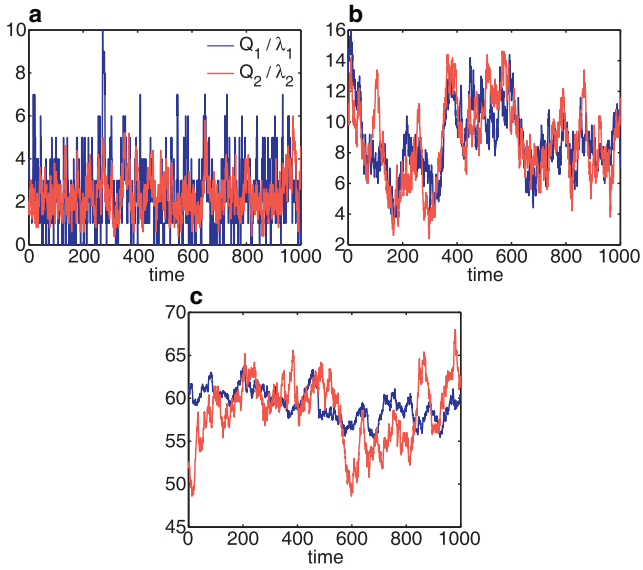

FIGURE 4 Sample trajectories for Q_1 and Q_2 resulting from stochastic simulations of the biochemical reactions in Eqs. 1, 3–5, and 12 when there are two species of protein ($m = 2$), 10 copies of the enzyme ($L = 10$), and $\lambda_2 = 5$, $\mu = 1$, $\gamma = 0.01$, $\eta_+ = 20$, and $\eta_- = 100$ (compare to Fig. 2, *b*, *d*, and *f*). Trajectories are shown in panels *a*–*c* for $\lambda_1 = 1$, $\lambda_1 = 5$, and $\lambda_1 = 20$, respectively. In each case, the last 1000 time units of a 4×10^7 time-unit run are shown. Time averages computed using the last half of each run yielded correlations of (*a*) 0.23, (*b*) 0.72, and (*c*) 0.20, respectively, which are consistent with the results displayed in Fig. 2 *f*.

We now describe the multiclass queuing model in detail. Jobs of type i arrive to a single queue according to a Poisson process with rate λ_i for $i = 1,\ldots,m$. There are L servers available for processing jobs (each server can process only one job at a time). We keep jobs in the queue while they are being processed; jobs are ordered in the queue and the jobs currently being processed are at the head of the queue. When a new job arrives to the queue, if there are fewer than L jobs in the system, the new job joins the end of the queue and immediately commences being processed by one of the free servers; on the other hand, if there are L or more jobs already in the system, then all of the servers are busy processing jobs and the new job is placed in a random position amongst those jobs in the queue that are not currently being processed. (The reason for this random placement of newly arriving jobs is explained in the next paragraph.) If n denotes the total number of jobs in the system at a given time, then when $n \leq L$, each job is being processed at an exponential rate of μ and is independently reneging at an exponential rate of γ. And when $n > L$, the first L jobs are each being processed at an exponential rate of μ, and each of the n jobs is independently reneging at an exponential rate of γ. When a job finishes being processed or reneges, the job departs from the system and the size of the queue is reduced by one, with the previous order of the remaining jobs being maintained. A newly freed server immediately selects the job that is at the head-of-the-line of waiting jobs and starts processing it. If there is no such job, the server idles until a new job is available for processing.

Random placement of new jobs amongst the queue of waiting jobs ensures that the job count process will have the same distribution as would be obtained if the server, in selecting a new job to be processed, were to choose at random from amongst the jobs waiting to be processed. This observation is the key to the fact that the job count process in this multiclass queuing model has the same distribution as the molecular count process in the stochastic model. Fig. 1 depicts the setup for the stochastic model (Fig. 1 *a*) and for the multiclass queuing model (Fig. 1 *b*) when there are two protein species/types of jobs ($m = 2$).

We have chosen to do the modeling in terms of random placement in the queue (rather than random selection to begin processing) so as to make a correspondence with the multiclass queuing model framework in Kelly (11).

To enable the reader to clearly see the correspondence with that setup, we give a few more details of this correspondence below. In Kelly's notation, when a new job arrives to the queue with n jobs in it already, if $n \geq L$, with probability $\Delta(\ell, n+1) = 1/(n+1-L)$, the new job is placed in position $\ell \in \{L+1,\ldots,n+1\}$ in the queue, and if $\ell \leq n$, jobs previously in positions $\ell, \ell+1,\ldots,n$ move to positions $\ell+1, \ell+2,\ldots,n+1$, respectively. If $n < L$, the new job is placed behind the other jobs in the queue, i.e., it is placed in position $\ell = n+1$ with probability $\Delta(n+1, n+1) = 1$.

The exponential processing and reneging are realized in the following way. Each job requires an exponentially distributed amount of service with mean 1. When there are n jobs in the queue, total service effort (from both processing and reneging) is given by $\phi(n)$ (see Eq. 10); this is the overall rate at which jobs are departing from the system when there are n jobs in the system. The proportion $\Gamma(\ell, n)$ of this effort that is directed to the job in position ℓ is $(\mu + \gamma)/\phi(n)$ for $\ell = 1,\ldots,\min(n,L)$ and is $\gamma/\phi(n)$ for $\ell = L+1,\ldots,n$ when $n > L$. A job departs from the system whenever the amount of service effort it has received is equal to its random service requirement. When a job in position $\ell \in \{1,\ldots,n\}$ departs, the jobs in positions $\ell+1,\ldots,n$ move to positions, $\ell, \ell+1,\ldots, n-1$, respectively. (Here we have used the capital Greek letters Δ,Γ instead of the lower case Greek letters used in Kelly (11) to avoid confusion with our use of the latter for other parameters.)

The m-dimensional process that tracks the number of jobs of each type that are in the queuing system (both waiting and being processed), has the same distribution as the molecular count process of the stochastic model introduced back in Stochastic Model. Accordingly, we shall use the same symbol Q to denote both processes and, as before, let

$$N = \sum_{i=1}^{m} Q_i.$$

At any time t, let $S(t)$ be the vector of length $N(t)$ where the ℓ^{th} component of this vector is the type (a member of the set

$\{1,\ldots,m\}$) of the job that is in position ℓ in the queue at time t. A vector of length zero (the empty vector) corresponds to the queue being empty. Thus, $S(t)$ is simply the list of jobs in the queue where an entry in the list specifies the type of the job in that position. The process S is a continuous time Markov chain.

It follows immediately from Theorem 3.1 and Corollary 3.4 of Kelly (11) that the steady-state distribution for S, and hence for Q and N, can be given explicitly. (Because the Markov chain S is irreducible, its steady-state distribution is the same as its stationary distribution.)

Theorem 1

The continuous time Markov chain S describing the state of the multiclass queuing model associated with our stochastic model of enzymatic processing has a unique steady-state distribution given by

$$P(S = (s_1, \ldots, s_n)) = c\prod_{\ell=1}^{n}\left(p_{s_\ell}\frac{\Lambda}{\phi(\ell)}\right), \qquad (13)$$

for $s_1, \ldots, s_n \in \{1, \ldots, m\}$, and $n = 0,1,2,\ldots$, where $p_i = \lambda_i/\Lambda$, $i = 1,\ldots,m$,

$$\Lambda = \sum\nolimits_{i=1}^{m}\lambda_i,$$

and c is the normalizing constant:

$$c = \left(\sum_{n=0}^{\infty}\frac{\Lambda^n}{\prod_{\ell=1}^{n}\phi(\ell)}\right)^{-1}. \qquad (14)$$

The associated steady-state distribution for

$$N = \sum\nolimits_{i=1}^{m}Q_i$$

is given by Eq. 9 and, conditioned on $N = n > 0$, the steady-state distribution of Q is a multinomial distribution with parameters $(n; p_1,\ldots,p_m)$, so that for any nonnegative integers $q_1,\ldots,q_m$, and

$$n = \sum\nolimits_{i=1}^{m}q_i,$$

the steady-state distribution of Q is given by

$$P(Q = (q_1, \ldots, q_m)) = P(N = n)\frac{n!}{q_1!\ldots q_m!}\prod_{i=1}^{m}p_i^{q_i}. \qquad (15)$$

A key assumption for the analysis in Kelly (11) to hold is that the service effort provided to a job does not depend on its type, just its position in the queue. The above result is a special case of well-known results on equilibrium distributions in multiclass queuing theory and is associated with the fact that the queue is quasireversible. The concept of quasireversibility is specific to queuing systems; it was first identified by Muntz (12) and independently discovered and used by Kelly (11,13,14), who also coined the term "quasireversibility".

A multiclass queue is quasireversible if the queue, when initialized with its steady-state distribution and run in reverse time, taking the departure processes as arrival processes and interchanging the parameters $\Gamma(\ell, n)$ and $\Delta(\ell, n)$, is a queue of a similar type to the original queue run in forward time. In the language of the Markov chain S, for a stationary version of S, the process run in reverse time is a Markovian state descriptor associated with a multiclass queuing model with Poisson inputs having the same parameters as the original input processes, and with the parameters $\Gamma(\ell, n)$ and $\Delta(\ell, n)$ interchanged. In particular, the multiclass queuing model has the input-output property that, in steady state, the departure processes (where a departure can occur either due to completion of processing or reneging) for each of the job types are the same as the arrival processes, i.e., independent Poisson processes with rates λ_i, $i = 1,\ldots,m$.

MOMENTS AND CORRELATIONS

Using the fact that the job count process in the multiclass queue of the previous section has the same distribution as the molecular count process for our stochastic model, by Theorem 1, the steady-state distribution for this process Q is given by Eqs. 14–15 and Eq. 9. In this section, we derive explicit analytic expressions for moments and correlations of this distribution. In particular, we obtain succinct expressions for the first two moments and for the correlations when there is one copy of the enzyme ($L = 1$) in Eqs. 22–24 below. (Formulas for $L > 1$ are derived in the Supporting Material.) We believe that our explicit formulas for the correlation and our findings concerning correlation resonance are new even to queuing theory.

In the following derivation, Q and

$$N = \sum\nolimits_{i=1}^{m}Q_i$$

are assumed to have their steady-state distributions. Due to the factorized form of Eq. 15, the moments of Q can be expressed in terms of the moments of N. In particular, the means, variances, and second moments satisfy the following for $i, j = 1,\ldots,m$,

$$E[Q_i] = p_iE[N], \qquad (16)$$

$$Var(Q_i) = p_i^2(Var(N) - E[N]) + p_iE[N], \qquad (17)$$

$$E\left[Q_i^2\right] = p_i(1 - p_i)E[N] + p_i^2E\left[N^2\right], \qquad (18)$$

$$E\left[Q_iQ_j\right] = p_ip_j\left(E\left[N^2\right] - E[N]\right) \text{ for } j \neq i, \qquad (19)$$

from which Eqs. 6–8 follow immediately. Furthermore, the moment-generating function for Q is given by

$$E\left[e^{u\cdot Q}\right] = \sum_{n=0}^{\infty}\left(\sum_{i=1}^{m}p_ie^{u_i}\right)^n P(N = n) \quad \text{for} \quad u \in \mathbb{R}^m.$$

Here N has the steady-state distribution of a one-dimensional birth-death process with constant birth rate Λ and state-dependent death rate given by ϕ; this distribution is given by Eq. 9. Below, for $L = 1$, we derive formulas for the first two moments of N and for its moment-generating function; these formulas are written succinctly in terms of confluent hypergeometric functions.

Fix $L = 1$ and let $\beta = (\mu/\gamma) + 1$, and $\delta = \Lambda/\gamma$. To obtain an expression for the moment-generating function of N, we note that

$$\prod_{\ell=1}^{n}\phi(\ell) = \gamma^n(\beta)_n,$$

where

$$(x)_n = x(x+1)(x+2)\cdots(x+n-1)$$

is the rising factorial or Pochhammer symbol. Thus, the normalizing constant c satisfies

$$c^{-1} = \sum_{n=0}^{\infty}\frac{\delta^n}{(\beta)_n} = M(1,\beta,\delta),$$

where M is the confluent hypergeometric function given by

$$M(x,y,z) = \sum_{n=0}^{\infty}\frac{(x)_n z^n}{(y)_n n!}. \tag{20}$$

This function M is sometimes also denoted by ${}_1F_1$ and is a standard function in computer packages such as MATLAB (The MathWorks, Natick, MA), Maple (Waterloo Maple, Waterloo, Ontario, Canada), and Mathematica (Wolfram Research, Champaign, IL). We then have the following formula for the moment-generating function of N:

$$E\left[e^{uN}\right] = c\sum_{n=0}^{\infty}\frac{(e^u\Lambda)^n}{\prod_{\ell=1}^{n}\phi(\ell)} = \frac{M(1,\beta,e^u\delta)}{M(1,\beta,\delta)},\ u\in\mathbb{R}. \tag{21}$$

The function M has the property that

$$\frac{d}{dz}M(x,y,z) = \frac{x}{y}M(x+1,y+1,z).$$

Using this, and differentiating the moment-generating function in Eq. 21, we obtain

$$E[N] = \frac{\delta M(2,\beta+1,\delta)}{\beta M(1,\beta,\delta)}$$

and

$$E\left[N^2\right] = E[N] + \frac{2\delta^2 M(3,\beta+2,\delta)}{\beta(\beta+1)M(1,\beta,\delta)}.$$

Substitution of these formulas into Eqs. 16 –19 and Eq. 8 yields the following key result.

Theorem 2

For our stochastic model of enzymatic processing associated with Eqs. 1–5, when there is one copy of the enzyme ($L = 1$), the steady-state means, variances, and correlations for the molecular count process Q are given by the following formulas. For $i, j = 1,\ldots,m$ and $j \neq i$,

$$E[Q_i] = \frac{p_i\delta M(2,\beta+1,\delta)}{\beta M(1,\beta,\delta)}, \tag{22}$$

$$Var(Q_i) = \frac{2p_i^2\delta^2 M(3,\beta+2,\delta)}{\beta(\beta+1)M(1,\beta,\delta)} - \left(\frac{p_i\delta M(2,\beta+1,\delta)}{\beta M(1,\beta,\delta)}\right)^2 + \frac{p_i\delta M(2,\beta+1,\delta)}{\beta M(1,\beta,\delta)}, \tag{23}$$

$$r_{ij} = \frac{h(\beta,\delta)}{\left(h(\beta,\delta)+p_i^{-1}\right)^{1/2}\left(h(\beta,\delta)+p_j^{-1}\right)^{1/2}}, \tag{24}$$

where $\beta = (\mu/\gamma) + 1$, $\delta = \Lambda/\gamma$, and

$$\Lambda = \sum_{i=1}^{m}\lambda_i,$$

$$h(\beta,\delta) = \frac{2\delta M(3,\beta+2,\delta)}{(\beta+1)M(2,\beta+1,\delta)} - \frac{\delta M(2,\beta+1,\delta)}{\beta M(1,\beta,\delta)},$$

and M is the confluent hypergeometric function given by Eq. 20. The above expression for the correlation r_{ij} is nonnegative (see the Supporting Material for a proof).

In the zero dilution limit ($\gamma \to 0$), the correlation formula simplifies. Using the asymptotic result given in (4.3.7) of Slater (15),

$$M(x,y,yw) = (1-w)^{-x}\left(1 - \frac{x(x+1)}{2y}\left(\frac{w}{1-w}\right)^2 + O\left(|y|^{-2}\right)\right), \tag{25}$$

for x fixed, w bounded, and $y \to \infty$, we obtain that in the zero dilution limit, for $\rho = \Lambda/\mu < 1$, $i, j = 1,\ldots,m$ and $j \neq i$, r_{ij} is given by Eq. 11. As $\rho \uparrow 1$ in Eq. 11, r_{ij} converges monotonically upwards to its maximum value of 1, indicating approach towards perfect correlation of the steady-state numbers of molecules of different protein species. This is reminiscent of process-level state space collapse observed for various multiclass queues in heavy traffic (9,10) and is the same limiting behavior as seen in the heavy traffic asymptotics (as $\rho \uparrow 1$) obtained by Rege and Sengupta (16) for a stationary processor sharing queue. (Processor sharing is not the same service discipline as the random-order-of-service queuing discipline considered here, but it turns out that the steady-state queue length distribution is the same for both disciplines under our model assumptions.) Despite these prior observations in the heavy traffic regime ($\rho \approx 1$), our quantitative formulas for the correlations r_{ij}, as well as our observations concerning the qualitative behavior of r_{ij} for various parameters (see Figs. 2 and 3), appear to be new.

GENERALIZATION FOR REVERSIBLE BINDING

In this section, we consider a generalization of the stochastic model which is intended as an approximate model for the biochemical reactions where the binding reactions in Eq. 2 are replaced by the reversible reactions in Eq. 12. Our generalization of the stochastic model is obtained by simply replacing the degradation rate μ by a Michaelis-Menten-like rate function of the form $n\mu/(K + n)$, where $K = \eta_-/\eta_+$ and n is the total number of proteins in the system. This is a commonly used approximation when the binding/unbinding reactions are fast (17). With this change in the stochastic model, the molecular count process still corresponds to the job count process in a quasireversible multiclass queue for which the steady-state distribution is given explicitly by Eq. 9 and Eqs. 14 and 15 with ϕ now given by

$$\phi(n) = \min(n, L)\frac{n\mu}{K + n} + n\gamma, \; n = 0, 1, 2, \ldots. \quad (26)$$

The formulas in Eqs. 16–19 and Eq. 8 for the moments and correlations of Q in terms of the moments of N still hold, but the moments of N are now different. We derive formulas for these for $L = 1$ and $K > 0$ below; derivations for $L > 1$ can be found in the Supporting Material.

Fix $L = 1$ and set $\alpha = K + 1$, $\beta = (\mu/\gamma) + \alpha$, and $\delta = \Lambda/\gamma$. Then,

$$\prod_{\ell=1}^{n}\phi(\ell) = \frac{\gamma^n n!(\beta)_n}{(\alpha)_n}$$

and the normalizing constant c satisfies $c^{-1} = M(\alpha, \beta, \delta)$, where M is the confluent hypergeometric function given by Eq. 20. In a similar manner to that in the previous section, we then have

$$E[e^{uN}] = \frac{M(\alpha, \beta, e^u\delta)}{M(\alpha, \beta, \delta)},$$
$$E[N] = \frac{\alpha\delta M(\alpha + 1, \beta + 1, \delta)}{\beta M(\alpha, \beta, \delta)},$$
$$E[N^2] = E[N] + \frac{\alpha(\alpha + 1)\delta^2 M(\alpha + 2, \beta + 2, \delta)}{\beta(\beta + 1)M(\alpha, \beta, \delta)}.$$

This yields the following generalization of the theorem of the previous section.

Theorem 3

For the generalization of our stochastic model of enzymatic processing associated with reversible binding, when there is one copy of the enzyme ($L = 1$), the steady-state means, variances, and correlations for the molecular count process Q are given by the following formulas. For $i, j = 1, \ldots, m$, $j \neq i$,

$$E[Q_i] = \frac{\alpha p_i\delta M(\alpha + 1, \beta + 1, \delta)}{\beta M(\alpha, \beta, \delta)},$$

$$Var(Q_i) = \frac{\alpha(\alpha + 1)p_i^2\delta^2 M(\alpha + 2, \beta + 2, \delta)}{\beta(\beta + 1)M(\alpha, \beta, \delta)} - \left(\frac{\alpha p_i\delta M(\alpha + 1, \beta + 1, \delta)}{\beta M(\alpha, \beta, \delta)}\right)^2 + \frac{\alpha p_i\delta M(\alpha + 1, \beta + 1, \delta)}{\beta M(\alpha, \beta, \delta)},$$

$$r_{ij} = \frac{h(\alpha, \beta, \delta)}{\left(h(\alpha, \beta, \delta) + p_i^{-1}\right)^{1/2}\left(h(\alpha, \beta, \delta) + p_j^{-1}\right)^{1/2}},$$

where $\alpha = K + 1$, $\beta = (\mu/\gamma) + \alpha$, and $\delta = \Lambda/\gamma$,

$$h(\alpha, \beta, \delta) = \frac{(\alpha + 1)\delta M(\alpha + 2, \beta + 2, \delta)}{(\beta + 1)M(\alpha + 1, \beta + 1, \delta)} - \frac{\alpha\delta M(\alpha + 1, \beta + 1, \delta)}{\beta M(\alpha, \beta, \delta)}.$$

The above expression for the correlation r_{ij} is nonnegative (see the Supporting Material for a proof).

Using the asymptotic result given in Eq. 25, we obtain in the zero dilution limit that for $\rho = \Lambda/\mu < 1$, r_{ij} is still given by the expression in Eq. 11. If we set $K = 0$ in the above formulas, we recover the formulas for the stochastic model with instantaneous irreversible binding reactions. The above formulas are illustrated in Figs. 2 and 3 for sample parameters.

DISCUSSION

In intracellular signaling and metabolic systems, enzymes often interact with multiple substrates, which may represent different pathways or branches of the same pathway (18,19). As we showed in this article, this resource sharing provides an indirect but strong mechanism for correlation among concentrations of different substrates and may lead to apparent coupling among the pathways. This result is somewhat counterintuitive, because, generally, it is assumed that correlations arise from direct interactions. By exploiting a connection between enzymatic processing and multiclass queuing theory, we were able to obtain closed-form expressions for the steady-state joint distributions of multiple substrates and to derive explicit analytic formulas for the correlations and other moments of the distributions.

We demonstrated that, for a small dilution rate, the correlation reaches a peak near balance, where the total production rate of the substrates equals the maximal processivity of the common enzyme. This transition is also characterized by a sharp increase in the mean substrate concentrations. For much smaller production rates, the queue is short or absent, and competition for the enzyme is weak, resulting in small correlations. For production rates much larger than the processivity of the enzyme, the queue is very long, the production is mainly balanced by dilution,

and correlations due to common enzymatic processing become small again. Our explicit formulas for correlations and our findings concerning correlation resonance appear to be new both in the context of biological systems and in queuing theory.

In deriving the results for the distributions and their moments, we exploited the quasireversible nature of the corresponding multiclass queuing model (11). Our theoretical results obtained here are corroborated by our experimental findings in N. A. Cookson, W. H. Mather, T. Danino, O. Mondragón-Palomino, R. J. Williams, L. S. Tsimring, and J. Hasty (unpublished). Using a synthetic two-gene circuit in *Escherichia coli* and multicolor fluorescent flow cytometry and microscopy, we demonstrated that modulating the transcription rate of one fluorescent protein affects the concentration of another if they both are endowed with the same LAA tag, and hence targeted for degradation by the common enzymatic degradation machine ClpXP.

We believe that substrate correlations due to coupled enzymatic processing play an important role in many cellular processes and should be incorporated in future computational models. This form of posttranslational regulation may explain some of the existing connections in wiring diagrams of proteomic and metabolic networks or introduce new links. Furthermore, this effect may provide a significant contribution to the apparent extrinsic noise in gene expression which is usually determined by cross-correlations of protein concentrations (20,21). This effect also has to be taken into consideration in the design of synthetic gene circuits which often rely on enzymatic processing of network components for increasing the speed of their turnover (22,23).

SUPPORTING MATERIAL

Further details, a figure, and a reference are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(10)01207-5.

R.J.W. thanks Frank Kelly for conversations concerning quasireversibility in multiclass queues and Estelle Basor for help in identifying a connection with the confluent hypergeometric function.

This work was supported in part by the National Institute of General Medical Sciences of the National Institutes of Health (grants Nos. GM079333 and GM089976) and the Department of Energy Computational Science Graduate Fellowship Program (grant No. DE-FG02-97ER25308 to N.A.C. and W.H.M.). L.S.T. was supported, in part, by Office of Naval Research, Multidisciplinary University Research Initiative grant No. N00014-07-0741. Research of R.J.W. was supported, in part, by National Science Foundation grants Nos. DMS-0825686 and DMS-0906535.

REFERENCES

1. Jordan, J. D., E. M. Landau, and R. Iyengar. 2000. Signaling networks: the origins of cellular multitasking. *Cell.* 103:193–200.
2. Alon, U. 2006. An Introduction to Systems Biology: Design Principles of Biological Circuits. CRC Press, Boca Raton, FL.
3. Reference deleted in proof.
4. Arkin, A., P. Shen, and J. Ross. 1997. A test case of correlation metric construction of a reaction pathway from measurements. *Science.* 277:1275–1279.
5. Heuett, W. J., and H. Qian. 2006. Grand canonical Markov model: a stochastic theory for open nonequilibrium biochemical networks. *J. Chem. Phys.* 124:044110.
6. Arazi, A., E. Ben-Jacob, and U. Yechiali. 2004. Bridging genetic networks and queueing theory. *Physica A.* 332:585–616.
7. Levine, E., and T. Hwa. 2007. Stochastic fluctuations in metabolic pathways. *Proc. Natl. Acad. Sci. USA.* 104:9224–9229.
8. Reference deleted in proof.
9. Bramson, M. 1998. State space collapse with application to heavy traffic limits for multiclass queueing networks. *Queue. Sys. Theory Appl.* 30:89–148.
10. Williams, R. J. 1998. Diffusion approximations for open multiclass queueing networks: sufficient conditions involving state space collapse. *Queue. Sys. Theory Appl.* 30:27–88.
11. Kelly, F. P. 1979. Reversibility and Stochastic Networks. John Wiley and Sons, New York.
12. Muntz, R. R. 1972. Poisson Departure Processes and Queueing Networks. IBM Research Report No. RC 4145. IBM Thomas J. Watson Research Center, Yorktown Heights, NY.
13. Kelly, F. P. 1975. Networks of queues with customers of different types. *J. Appl. Probab.* 12:542–554.
14. Kelly, F. P. 1976. Networks of queues. *Adv. Appl. Probab.* 8:416–432.
15. Slater, L. J. 1960. Confluent Hypergeometric Functions. Cambridge University Press, Cambridge, UK.
16. Rege, K. M., and B. Sengupta. 1996. Queue-length distribution for the discriminatory processor-sharing queue. *Oper. Res.* 44:653–657.
17. Keener, J., and J. Sneyd. 1998. Mathematical Physiology. Springer-Verlag, New York.
18. Elf, J., J. Paulsson, …, M. Ehrenberg. 2003. Near-critical phenomena in intracellular metabolite pools. *Biophys. J.* 84:154–170.
19. Ptashne, M. 2003. Regulated recruitment and cooperativity in the design of biological regulatory systems. *Philos. Trans. Math. Phys. Eng. Sci.* 361:1223–1234.
20. Swain, P. S., M. B. Elowitz, and E. D. Siggia. 2002. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl. Acad. Sci. USA.* 99:12795–12800.
21. Elowitz, M. B., A. J. Levine, …, P. S. Swain. 2002. Stochastic gene expression in a single cell. *Science.* 297:1183–1186.
22. Hasty, J., D. McMillen, and J. J. Collins. 2002. Engineered gene circuits. *Nature.* 420:224–230.
23. Heinemann, M., and S. Panke. 2006. Synthetic biology—putting engineering into biology. *Bioinformatics.* 22:2790–2799.