Factorized time-dependent distributions for certain multiclass queueing networks and an application to enzymatic processing networks

W.H. Mather · J. Hasty · L.S. Tsimring · R.J. Williams

Abstract Motivated by applications in biological systems, we show for certain multiclass queueing networks that time-dependent distributions for the multiclass queue-lengths can have a factorized form which reduces the problem of computing such distributions to a similar problem for related single-class queueing networks. We give an example of the application of this result to an enzymatic processing network.

W.H. Mather · J. Hasty · L.S. Tsimring · R.J. Williams
BioCircuits Institute, University of California, San Diego, USA

W.H. Mather · J. Hasty
Department of Bioengineering, University of California, San Diego, USA

J. Hasty
Molecular Biology Section, Division of Biological Sciences, University of California, San Diego, USA

R.J. Williams (✉)
Department of Mathematics, University of California, 9500 Gilman Drive, San Diego, La Jolla, CA 92093-0112, USA
e-mail: williams@math.ucsd.edu

1 Introduction

Recent advances in single-cell experimental techniques in molecular biology have provided strong impetus for the development and analysis of dynamic stochastic models for intracellular reaction networks. Of particular importance is the need to allow for the fact that some molecular species, such as transcripts, transcription factors, or housekeeping enzymes, may only be present in small numbers. Frequently, Markov chain models accompanied by discrete-event simulation [12] are used for such stochastic modeling. However, the computational burden grows rapidly with the complexity of the systems being modeled. Accordingly, there is considerable interest in methods that reduce the dimensionality of such models. Some authors have explored a multiscaling approach, for example, see [3] for recent work on a model of intracellular viral kinetics. We have been investigating an alternative approach using dimension reduction results for multiclass queueing networks.

In recent work [17], we used results from Kelly [16] on the factorized form of the stationary distribution for a certain multiclass queue to study the steady-state behavior of a stochastic model for coupled enzymatic processing of multiple protein species. In this application, protein species corresponded to job classes, copies of an enzyme were analogous to servers, and dilution corresponded to reneging. We developed formulas for correlations between the numbers of molecules of different protein species in steady-state in terms of the steady-state Fano factor (variance divided by the mean) for the total number of molecules in the system. Exploration of these formulae for sample parameters showed that the correlations could be tuned by adjusting the overall exogenous arrival rate. In particular, we observed that for small dilution rate, the correlations peaked near the balance point where the total rate of influx of new proteins into the system was equal to the maximum processing capacity of the enzymatic processors. In the current paper, motivated by the desire to consider time-dependent and network analogues of the stochastic model studied in [17], we extend the single-station multiclass queueing result to certain networks of multiclass queues and to time-dependent as well as stationary distributions. In particular, for certain multiclass queueing networks, we show that time-dependent (and stationary) distributions for the queue compositions can have a factorized form which reduces the problem of computing these distributions to a similar problem for related single-class queueing networks. We think of this as a kind of state space collapse [6, 21] at the level of distributions. We illustrate the use of this result with an example of an enzymatic processing network and explore tuning of correlations there, by adjusting an internal processing rate parameter.

Our paper is organized as follows. We define our multiclass queueing network model precisely in Sect. 2. A Markovian state descriptor, S, for this model tracks the ordered list of jobs of multiple types waiting for service at each station. Under our model assumptions, the derived process, N, that tracks the total number of jobs at each station is Markov, being a multidimensional birth–death process. The forward Kolmogorov equations for the evolution of the transition probabilities of S and N are given in Sect. 3. Our main result is stated and proved in Sect. 4. This result is codified in Theorem 4.1, which states that if the initial distribution of S has a certain factorized form, then this form is preserved in the time-dependent distributions of S. This reduces the problem of computing such distributions to the equivalent problem for the

simpler process N. In particular, this implies that if a stationary distribution for N is known, then a stationary distribution for S follows immediately. We deduce from our main result that correlations between numbers of jobs of different types at a given station can be computed in terms of the Fano factor for the total number of jobs at the station. In Sect. 5, we apply our results to an example of an enzymatic processing network. To simplify our presentation, we first develop our main results when the arrival rates are constant, and the service effort, effort allocations, and routing probabilities at a station depend only on the current number of jobs at the station. We then observe (see Remark 1 following Theorem 4.1) that the same derivations extend immediately to the case where these parameters can depend on time and on the numbers of jobs at each of the stations in the network, provided that the arrival rates at each station still satisfy a key proportionality relation (1). The case of time-dependent rates is of particular interest for applications to biological systems involving signalling. We plan to pursue such applications involving time-dependent internal processing rates in future work. The relationship between the multiclass queue-length and waiting time processes is a further topic that we have not addressed, but which is worthy of investigation, especially for queues with random order of service (see [5] and references therein for a start on this topic).

2 Multiclass queueing network model

The multiclass queueing networks considered here have many similarities and a few important differences from the models considered in Sect. 3.1 of Kelly's book [16] on Reversibility and Stochastic Networks. To help the reader appreciate the commonalities and differences, we give a description of these here. A full detailed description of the multiclass queueing networks that we consider is given further below. In terms of similarities, in our model jobs will remain of one type throughout their entire sojourn in the network, all jobs at a given station form an ordered queue, exogenous arrivals to the queue occur according to independent Poisson arrival processes (one for each type), and these arrivals are inserted into the queue in such a way that the insertion position does not depend on job type, and service effort provided to a job will be blind to its type although this effort could depend on the position of the job in the queue. In terms of differences, in contrast to the deterministic routing depending on type in [16], we allow routing to be probabilistic, and our routing probabilities can depend on the state of the queue that a job is leaving from, although the probabilities must be blind to type. Thus, our model can incorporate reneging of jobs from a queue where the reneging rate and other routing probabilities can depend on the total number of jobs in the queue, but not on the composition in terms of types. An important assumption is that we allow arrivals of any type at any station; however, at each station with nonzero arrivals, the ratio of the arrival rates for different types is fixed and the same for all stations. This is a restrictive assumption in general but is not an additional restriction if arrivals only occur at one station.

We consider a multiclass queueing network model with the following structure. The network has finitely many stations labeled by $k = 1, \ldots, K$. Jobs enter the network from outside, receive processing at a succession of stations in the network, and

eventually depart the network. We assume that there are finitely many types of job labeled by $i = 1, \ldots, I$. A job remains of the same type during its entire sojourn through the network. At each station there is a single ordered queue of jobs to be processed by the station; this queue typically contains multiple types of job. Jobs are kept in the queue while they are being processed. For station k, jobs of type i arrive to the station from outside the network according to a Poisson process with rate λ_i^k. Each time a job visits a station, it requires an exponentially distributed amount of "service" with a mean of one. When a job arrives to station k (whether from inside or outside the network) and there are n jobs in the queue there already, with probability $\delta^k(\ell, n+1)$, the new job is placed in position ℓ in the updated 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. Here $\delta^k(\ell, n+1) \in [0, 1]$ for $\ell = 1, \ldots, n+1$, and $\sum_{\ell=1}^{n+1} \delta^k(\ell, n+1) = 1$. When there are n jobs in the queue at station k, total "service effort" is supplied to jobs at station k at the rate $\phi^k(n)$, where $\phi^k(n) \geq 0$. This is the overall rate at which jobs are departing from station k when there are n jobs there. A proportion $\gamma^k(\ell, n)$ of this effort is directed to the job in position ℓ, where $\gamma^k(\ell, n) \in [0, 1]$ for each $\ell \in \{1, \ldots, n\}$, and $\sum_{\ell=1}^{n} \gamma^k(\ell, n) = 1$. A job departs from the queue at station k whenever the amount of service effort it has received is equal to its random exponentially distributed service requirement. When a job in position $\ell \in \{1, \ldots, n\}$ at station k departs and there were n jobs at the station just before the departure, the job goes to station l with probability $\beta_l^k(\ell, n)$, $l = 1, \ldots, K$, and it departs the system with probability $\beta_0^k(\ell, n)$. Here $\beta_l^k(\ell, n) \in [0, 1]$ for $l = 0, 1, \ldots, K$, $\beta_k^k = 0$, and $\sum_{l=0}^{K} \beta_l^k(\ell, n) = 1$. (Here, as in [16], with the assumption that $\beta_k^k = 0$, we do not allow immediate self-feedback to a station.) When the job departs from position ℓ in the queue, the jobs in positions $\ell+1, \ldots, n$ move to positions $\ell, \ell+1, \ldots, n-1$, respectively. We assume that the interarrival times for all exogenous arrivals and all service times are mutually independent.

An important assumption in our setup is that the parameters $\delta^k, \phi^k, \gamma^k, \beta^k$ do not depend on the type of job being handled, although they can depend on the total number of jobs in the queue at station k and $\delta^k, \gamma^k, \beta^k$ can also depend on the position of the job in the queue. A single-station model (in which case β^k is not needed) with this *type-blind* property is sometimes referred to as a multiclass homogeneous queue of Kelly type [7]. As far as routing is concerned, in contrast to the situation in [16] which has deterministic routing of jobs depending on type, our routing parameter β^k does not depend on type but can be state-dependent by varying with the total number of jobs in the queue at station k or even with the position in the queue of the job being routed. This allows, for example, for the possibility that a job may leave a station either due to completion of its processing or due to reneging and that its routing can depend on which mechanism caused the job to leave. An important additional assumption for our analysis to hold is that there are strictly positive probabilities p_i, $i = 1, \ldots, I$, satisfying $\sum_{i=1}^{I} p_i = 1$ such that for each $k = 1, \ldots, K$, $i = 1, \ldots, I$,

$$\lambda_i^k = p_i \Lambda^k, \tag{1}$$

where $\Lambda^k = \sum_{i=1}^{I} \lambda_i^k$ is the total exogenous arrival rate for station k, and this is assumed to be strictly positive for at least one k. In other words, we assume that

there are nonzero arrivals to at least one station and the proportion of the arrivals to a given station that are due to each type is the same for all stations. If the exogenous arrivals occur only at a single station k and $\lambda_i^k > 0$ for all i, then the assumption is automatically satisfied; this situation arises, for example, from a reentrant line in manufacturing applications.

As an illustration, we describe how FIFO stations that have finitely many servers processing jobs, independent reneging of jobs, and probabilistic routing of processed jobs, can be accommodated in this modeling framework. Suppose that for each k, station k has L_k servers available for processing jobs in FIFO order with i.i.d. exponential processing times having mean $1/\mu_k$, that independently each job at the station (whether being processed or awaiting processing) can renege, where the reneging times are i.i.d. exponential random variables with mean $1/\kappa_k$, and that, after completion of processing by a server at station k, a job goes to station $l \neq k$ with probability p_{kl}, or if a job at station k reneges before it is processed by a server, it leaves the system. This situation can be modeled by setting $\delta^k(n+1, n+1) = 1$ (for FIFO order), $\phi^k(n) = \min(n, L_k)\mu_k + n\kappa_k$ (combining the service effects of processing and reneging), $\gamma^k(\ell, n) = (\mu_k + \kappa_k)/\phi^k(n)$ for $\ell = 1, \ldots, \min(n, L_k)$ and $\gamma^k(\ell, n) = \kappa_k/\phi^k(n)$ for $\ell = L_k + 1, \ldots, n$ if $n > L_k$, and for $l \neq k$ and $l \in \{1, \ldots, K\}$, $\beta_l^k(\ell, n) = p_{kl}\mu_k/(\mu_k + \kappa_k)$ for $\ell = 1, \ldots, \min(n, L_k)$ and $\beta_l^k(\ell, n) = 0$ for $\ell = L_k + 1, \ldots, n$ if $n > L_k$.

For $k = 1, \ldots, K$ and $t \geq 0$, let $N^k(t)$ denote the total number of jobs (of all types) that are at station k at time t, and let $S^k(t)$ denote the vector of length $N^k(t)$ where the ℓth component of this vector is the type (a member of the set $\{1, \ldots, I\}$) of the job that is in position ℓ in the queue at station k at time t. (A vector of length zero corresponds to an empty queue.) Let $N = (N^1, \ldots, N^K)$ and $S = (S^1, \ldots, S^K)$. Then both N and S are continuous-time Markov chains. The process N is a K-dimensional birth–death process. We let $\mathcal{N}$ and $\mathcal{S}$ denote the state spaces for N and S, respectively, and we let $\mathbf{n} = (n^1, \ldots, n^K)$ and $\mathbf{s} = (s^1, \ldots, s^K)$ denote generic values in these two spaces.

3 Kolmogorov forward equations

Henceforth, we assume that the multidimensional birth–death process N does not explode in finite time. Since there are only finitely many states of S associated with each state of N and N is assumed not to explode, it follows that S does not explode in finite time.

Let π^o be an initial probability distribution for S, and let ν^o be the associated initial probability distribution for N. Then,

$$\pi_t(\mathbf{s}) = P\big(S(t) = \mathbf{s}\big), \quad \mathbf{s} \in \mathcal{S}, \tag{2}$$

specifies the distribution for S at time t, and

$$\nu_t(\mathbf{n}) = P\big(N(t) = \mathbf{n}\big), \quad \mathbf{n} \in \mathcal{N}, \tag{3}$$

specifies the associated distribution for N at time t.

Then, since S does not explode in finite time, π is the unique probability distribution-valued function satisfying the following Kolmogorov forward equation (see, for example, Chap. 2.2 of [2]). For a Markov chain associated with a biochemical reaction network, this equation is referred to as the chemical master equation,

$$\frac{\partial \pi_t(\mathbf{s})}{\partial t} = \sum_{\tilde{\mathbf{s}} \in \mathcal{S}} \pi_t(\tilde{\mathbf{s}}) \Gamma(\tilde{\mathbf{s}}, \mathbf{s}) \quad \text{for all } \mathbf{s} \in \mathcal{S}, \ t > 0, \tag{4}$$

with the initial condition $\pi_0(\mathbf{s}) = \pi^o(\mathbf{s})$ for all $\mathbf{s} \in \mathcal{S}$, where Γ is the infinitesimal generator for S given by

$$\Gamma(\tilde{\mathbf{s}}, \mathbf{s}) = \begin{cases} \lambda^k_{s^k_\ell} \delta^k(\ell, n^k) & \text{if } \tilde{\mathbf{s}} = \mathbf{s}^{-,k,\ell} \text{ for some} \\ & \quad \ell \in \{1, \ldots, n^k\}, n^k = |s^k| > 0, \\ & \quad \text{and } k \in \{1, \ldots, K\}, \\[2mm] \phi^k(n^k + 1)\gamma^k(\ell, n^k + 1) & \\ \quad \times \beta^k_l(\ell, n^k + 1)\delta^l(m, n^l) & \text{if } \tilde{\mathbf{s}} = \mathbf{s}^{t,k,\ell,l,m} \text{ for some} \\ & \quad \ell \in \{1, \ldots, n^k + 1\}, \\ & \quad m \in \{1, \ldots, n^l\}, n^l = |s^l| > 0, \\ & \quad k, l \in \{1, \ldots, K\}, k \neq l, \\[2mm] \phi^k(n^k + 1)\gamma^k(\ell, n^k + 1) & \\ \quad \times \beta^k_0(\ell, n^k + 1) & \text{if } \tilde{\mathbf{s}} = \mathbf{s}^{+,k,\ell,i} \text{ for some} \\ & \quad i \in \{1, \ldots, I\}, \\ & \quad \ell \in \{1, \ldots, n^k + 1\}, \text{ and} \\ & \quad k \in \{1, \ldots, K\}, \\[2mm] 0 & \text{for all other } \tilde{\mathbf{s}} \neq \mathbf{s}, \\[2mm] -\sum_{\mathbf{u} \neq \mathbf{s}} \Gamma(\mathbf{s}, \mathbf{u}) & \text{if } \tilde{\mathbf{s}} = \mathbf{s}. \end{cases}$$

When $n^k = |s^k| > 0$, $\mathbf{s}^{-,k,\ell}$ denotes the modification of $\mathbf{s}$ obtained by deleting the entry in s^k in position ℓ. Then a transition from $\tilde{\mathbf{s}} = \mathbf{s}^{-,k,\ell}$ to $\mathbf{s}$ corresponds to an exogenous arrival of a type s^k_ℓ job to station k which is inserted in position ℓ. When $n^l = |s^l| > 0$, $\mathbf{s}^{t,k,\ell,l,m}$ denotes the modification of $\mathbf{s}$ obtained by deleting the entry in s^l in position m and inserting an entry of type s^l_m in position ℓ in s^k (and moving the elements of s^k previously in positions $\ell, \ldots, n^k$ up by one). Then a transition from $\tilde{\mathbf{s}} = \mathbf{s}^{t,k,\ell,l,m}$ to $\mathbf{s}$ corresponds to a departure from position ℓ in the queue at station k, which is then inserted into position m in the queue at station l. The quantity $\mathbf{s}^{+,k,\ell,i}$ denotes the modification of $\mathbf{s}$ obtained by inserting a job of type i in position ℓ in s^k (and moving the elements of s^k previously in positions $\ell, \ldots, n^k$ up by one). A transition from $\tilde{\mathbf{s}} = \mathbf{s}^{+,k,\ell,i}$ to $\mathbf{s}$ corresponds to a departure of a job of type i from position ℓ in the queue at station k, where that departing job exits the network. The off-diagonal entries in Γ correspond to the infinitesimal rates for transitions between states corresponding to exogenous arrivals, transitions within the network due to departures from one station going to another station, and departures from the network. As usual, the row sums of Γ are zero and so one can determine the (negative) diagonal terms from the values of the off-diagonal terms.

Similarly, by the nonexplosion condition that we assumed on N, ν is the unique probability distribution-valued function satisfying the Kolmogorov forward equation [2]:

$$\frac{\partial \nu_t(\mathbf{n})}{\partial t} = \sum_{\tilde{\mathbf{n}} \in \mathcal{N}} \nu_t(\tilde{\mathbf{n}}) \Delta(\tilde{\mathbf{n}}, \mathbf{n}) \quad \text{for all } \mathbf{n} \in \mathcal{N}, \ t > 0, \tag{5}$$

with the initial condition $\nu_0(\mathbf{n}) = \nu^o(\mathbf{n})$ for all $\mathbf{n} \in \mathcal{N}$, where Δ is the infinitesimal generator for N given by

$$\Delta(\tilde{\mathbf{n}}, \mathbf{n}) = \begin{cases} \Lambda^k & \text{if } \tilde{\mathbf{n}} = \mathbf{n} - \mathbf{e}^k \text{ and } n^k > 0, \\ \phi^k(n^k+1)\tilde{\beta}_l^k(n^k+1) & \text{if } \tilde{\mathbf{n}} = \mathbf{n} + \mathbf{e}^k - \mathbf{e}^l \text{ and } n^l > 0, l \neq k, \\ \phi^k(n^k+1)\tilde{\beta}_0^k(n^k+1) & \text{if } \tilde{\mathbf{n}} = \mathbf{n} + \mathbf{e}^k, \\ 0 & \text{for all other } \tilde{\mathbf{n}} \neq \mathbf{n}, \\ -\sum_{\mathbf{m} \neq \mathbf{n}} \Delta(\mathbf{n}, \mathbf{m}) & \text{if } \tilde{\mathbf{n}} = \mathbf{n}, \end{cases} \tag{6}$$

where $\mathbf{e}^k$ is the unit vector in the direction of the kth coordinate axis, and

$$\tilde{\beta}_l^k(n^k+1) = \sum_{\ell=1}^{n^k+1} \gamma^k(\ell, n^k+1)\beta_l^k(\ell, n^k+1), \quad l = 0, 1, \ldots, K, \tag{7}$$

is the probability that, when there are $n^k + 1$ jobs at station k, the next departure from station k will go to station l. Here going to station 0 corresponds to exiting the system.

4 Factorized time-dependent distributions

We now show that if the initial probability distribution for S has the factorized form

$$\pi^o(\mathbf{s}) = \nu^o(\mathbf{n}) \prod_{k=1}^{K} \prod_{\ell=1}^{n^k} p_{s_\ell^k}, \quad \mathbf{s} \in \mathcal{S}, \tag{8}$$

where ν^o is a fixed (but arbitrary) initial probability distribution for N and $n^k = |s^k|$, $k = 1, \ldots, K$, then the distribution of S at time t has a similar factorized form. Recall that $\{p_i, \ i = 1, \ldots, I\}$ are the fixed exogenous arrival proportions for the various types of jobs. (Here a product over an empty index set is defined to be equal to one.) Our factorization result is the following. A proof is given in the Appendix.

Theorem 4.1 *Suppose that the continuous-time Markov chain S is initialized with the distribution given by* (8) *and let $\{\nu_t(\mathbf{n}), \mathbf{n} \in \mathcal{N}, t \geq 0\}$ specify the distribution of N as a function of time when its initial distribution is given by $\{\nu^o(\mathbf{n}), \mathbf{n} \in \mathcal{N}\}$.*

Then the distribution for S at time t is given by

$$\pi_t(\mathbf{s}) = \nu_t(\mathbf{n}) \prod_{k=1}^{K} \prod_{\ell=1}^{n^k} p_{s_\ell^k}, \quad \mathbf{s} \in \mathcal{S}, \ t \geq 0, \tag{9}$$

where $n^k = |s^k|$, $k = 1, \ldots, K$. In particular, this result applies if the system starts empty. Furthermore, if N is irreducible and positive recurrent with steady-state distribution v^*, then the steady-state distribution for S is given by (9) with v^* in place of v_t.

Remarks

1. The astute reader will observe that our proof of the factorized form for the time-dependent distributions (9) readily extends to situations where for $k = 1, \ldots, K$, $\delta^k, \phi^k, \gamma^k, \beta^k$ are allowed to depend on time and on $\mathbf{n}$ (rather than just on n^k), and the arrival rates λ_i^k, Λ^k are also allowed to be functions of time and $\mathbf{n}$, provided that the critical proportionality relation (1) is maintained. To simplify our presentation, we did not include these extensions from the outset. Also, although the dimension reduction from S to N is still dramatic, with these more complicated dependencies, it will generally be more difficult to compute the distributions associated with the derivative process N.

2. For the case of stationary distributions and $K = 1$, the above result is a special case of Theorem 3.1 in [16] on stationary distributions for quasi-reversible queues. However, we could not find our general network result in the literature. This is perhaps because much of the literature on exact solutions for multiclass queueing networks has focused on product form stationary distributions with little attention being paid to exact simplifications for time-dependent distributions or to situations where the single-class network is not of product form. We note however that in a different context and for a particular model, in Exercise 7.1.10 of his book [16], Kelly noted that a stochastic population model has a factorized time-dependent distribution when the system starts empty.

In words, we may describe the result (9) as follows: conditioned on $N(t)$, the distribution of $S(t)$ is as if, for each of the positions in each queue, the job in that position is of type i with probability p_i, independent of the types of the jobs in other positions or queues. We now describe some consequences of this result for the queue-length processes.

For all $k = 1, \ldots, K$, $i = 1, \ldots, I$, and $t \geq 0$, let

$$Q_i^k(t) = \sum_{\ell=1}^{n^k} 1_{\{S_\ell^k = i\}},$$

the number of jobs of type i that are at station k at time t. Let $\mathcal{Q}$ denote the state space for the process $Q = (Q_i^k(t), \ k = 1, \ldots, K, \ i = 1, \ldots, I; \ t \geq 0)$. Because of the general insertion and removal mechanisms we have allowed for the queues, Q will typically not be a Markov process. However, we can still compute time-dependent (or stationary) distributions for it, provided that the assumptions of Theorem 4.1 are satisfied and such distributions are known for the process N. In the following, for a random variable Y, $E[Y]$ denotes expectation (or mean) of Y, $Var(Y)$ denotes the variance of Y, $SCV(Y) = Var(Y)/(E[Y])^2$ denotes the squared coefficient of variation for Y, and $F(Y) = Var(Y)/E[Y]$ is the Fano factor for Y.

Corollary 4.1 *Suppose that the assumptions of Theorem 4.1 hold. Then for all $t \geq 0$ and $\mathbf{q} = (q_i^k, \ k = 1, \ldots, K, i = 1, \ldots, I) \in \mathcal{Q}$, we have*

$$P\big(Q(t) = \mathbf{q}\big) = P\big(N(t) = \mathbf{n}\big) \prod_{k=1}^{K} \left(\frac{n^k!}{q_1^k! \cdots q_I^k!} \prod_{i=1}^{I} p_i^{q_i^k} \right), \tag{10}$$

where $n^k = \sum_{i=1}^{I} q_i^k, \ k = 1, \ldots, K$. In particular, for all $t \geq 0$, $k = 1, \ldots, K$, and $i, j \in \{1, \ldots, I\}$,

$$E\big[Q_i^k(t)\big] = p_i E\big[N^k(t)\big], \tag{11}$$

$$SCV\big(Q_i^k(t)\big) = SCV\big(N^k(t)\big) - \frac{1}{E[N^k(t)]} + \frac{1}{E[Q_i^k(t)]}, \tag{12}$$

and the correlation between $Q_i^k(t)$ and $Q_j^k(t)$ for $j \neq i$ is given by

$$\begin{aligned}
r_{ij}^k(t) &= \frac{E[Q_i^k(t) Q_j^k(t)] - E[Q_i^k(t)] E[Q_j^k(t)]}{\sqrt{Var(Q_i^k(t)) Var(Q_j^k(t))}} \\
&= \frac{F(N^k(t)) - 1}{(F(N^k(t)) - 1 + \frac{1}{p_i})^{1/2} (F(N^k(t)) - 1 + \frac{1}{p_j})^{1/2}},
\end{aligned} \tag{13}$$

where $F(N^k(t))$ is the Fano factor for $N^k(t)$.

Proof The expression for the distribution of $Q(t)$ follows directly from that for $S(t)$ by aggregation. The formulas for the moments and correlations associated with Q^k follow from simple computations using the multinomial distribution. We leave the details to the reader. $\qquad \square$

In [17], we considered a stochastic model of coupled enzymatic processing and connected it to a single-station multiclass queue that is homogeneous of Kelly type. For this queue, there are $L \geq 1$ copies of an enzymatic processor that each work at an exponential rate of μ; jobs of type i arrive according to a Poisson process of rate λ_i for $i = 1, \ldots, I$; upon arrival, jobs either enter processing immediately if there is a free processor or they are inserted randomly into the queue of jobs awaiting processing; and jobs leave the system either by completion of enzymatic processing or by reneging (which occurs at a fixed constant exponential rate for each job in the system). The fact that the total number of jobs in this system is a one-dimensional birth–death process enabled us to obtain explicit formulae for the associated steady-state moments and correlations for the numbers of each type in terms of confluent and generalized hypergeometric functions. For sample parameters, we explored these formulas numerically and observed for small reneging rate that the correlations peaked near the balance point where the total rate of influx of new jobs into the system is equal to the maximum processing capacity of the enzymatic processors. This is consistent with the following formula that we obtained for the steady-state correlation in
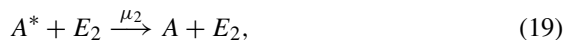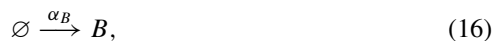
the limit when the reneging rate tends to zero:

$$r_{ij} = \left(1 + \frac{1}{p_i}\left(\frac{1}{\rho} - 1\right)\right)^{-1/2} \left(1 + \frac{1}{p_j}\left(\frac{1}{\rho} - 1\right)\right)^{-1/2}, \quad i \neq j \in \{1, \ldots, I\}, \quad (14)$$

for $\rho = \Lambda/(L\mu) < 1$ and $\Lambda = \sum_{i=1}^{I} \lambda_i$. In this limit, the correlation tends to a peak of one as the traffic intensity parameter ρ approaches the (heavy traffic) balance point $\rho = 1$. This formula and our numerical observations concerning correlations developed in [17] seem to be new. In the next section, we give an example from systems/synthetic biology to illustrate the use of Corollary 4.1 in a network context, and we explore correlation properties there.

5 Example: enzymatic processing network

We consider a system containing two nearly identical but distinguishable protein types that are acted upon by common enzymes (for background, see [1, 13, 14]). Proteins of each type are created in an unphosphorylated state, labeled A and B for the different types. These proteins can be converted by an enzyme E_1 into their phosphorylated states, labeled A^* and B^*, respectively. Additionally, A^* and B^* can be converted back into A and B, respectively, by an enzyme E_2. Proteins of either type, in either state, can be removed from the system by degradation. The system is governed by the following set of biochemical reactions:

$$\varnothing \xrightarrow{\alpha_A} A, \quad (15)$$

$$\varnothing \xrightarrow{\alpha_B} B, \quad (16)$$

$$A + E_1 \xrightarrow{\mu_1} A^* + E_1, \quad (17)$$

$$B + E_1 \xrightarrow{\mu_1} B^* + E_1, \quad (18)$$

$$A^* + E_2 \xrightarrow{\mu_2} A + E_2, \quad (19)$$

$$B^* + E_2 \xrightarrow{\mu_2} B + E_2, \quad (20)$$

$$A, B, A^*, B^* \xrightarrow{\kappa} \varnothing. \quad (21)$$

Assuming that all reaction times are exponentially distributed, the behavior of this stochastic system may be described in more detail as follows. New molecules of protein in the unphosphorylated state A, respectively B, are produced at the jump times of a Poisson process with rate α_A, respectively α_B. Assuming that there are $L_1 > 0$ copies of the enzyme E_1, when one of the copies becomes free, it selects a protein molecule at random from the available pool of unphosphorylated protein molecules and begins to process the molecule into its phosphorylated state. The time to complete the phosphorylation is exponentially distributed with mean $1/\mu_1$ where $\mu_1 > 0$. Similarly, assuming that there are $L_2 > 0$ copies of the enzyme E_2, when one of the copies becomes free, it selects a protein molecule at random from the available pool of phosphorylated protein molecules and begins to process the molecule into its
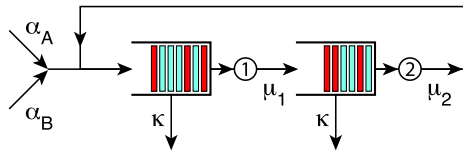
unphosphorylated state. The time to complete this processing is exponentially distributed with mean $1/\mu_2$ where $\mu_2 > 0$. Each of the protein molecules will eventually be removed from the system by degradation where the degradation rate is $\kappa > 0$, i.e., independent of all other considerations, each protein molecule remains in the system for an exponentially distributed amount of time with mean $1/\kappa$.

Reaction systems similar to those in (15)–(21) have been explored in the context of ultrasensitivity of signal propagation in cellular networks [4, 11, 13, 20]. Though such networks are typically explored in the case of a single protein type, natural biological pathways frequently involve diverging pathways with multiple proteins targeted by a common enzyme, for example, as is the case with the common action of the protein kinase Ste11 on the mating and high osmolarity signaling pathways in yeast [19]. The assumption of type-blindness concerning activity of the enzymes provides a useful baseline for how correlations may arise when enzymes treat multiple target proteins identically, as compared to preferentially. Alternatively, type-blindness may arise when nearly identical homologues of a target protein exist in a network, for example, after a gene duplication event [15].

For the above stochastic model, the process that tracks the number of molecules of each type in each state has the same distribution as the queue-length process Q for a multiclass queueing network of the kind described in Sect. 2. A depiction of the multiclass network is given in Fig. 1. For this network, following an observation of Kelly [16] (see Exercise 3.1.3), we use random insertion of arriving jobs into a queue of jobs awaiting enzymatic processing in place of the random selection of jobs for processing in the above stochastic model. (This replacement does not affect the distribution of the queue-length processes since random insertion of new jobs into the queue of jobs awaiting processing ensures that the waiting jobs are in random order, and so choosing the next job for processing from the head of this line yields a distributionally equivalent queue-length process to that obtained by selecting a job at random from the set of waiting jobs.) More specifically, the multiclass queueing model is described as follows. It has two types of jobs corresponding to the two types of proteins. Jobs of the two types arrive at station 1 from outside the system at rates $\lambda_1^1 = \alpha_A$ and $\lambda_2^1 = \alpha_B$, respectively. When a job arrives at station 1 (from inside or outside the system) and there are n^1 jobs there already, the job is inserted into position ℓ in the queue with probability $\delta^1(\ell, n^1 + 1) = 1/(n^1 + 1 - L_1)$ for $\ell \in \{L_1 + 1, \ldots, n^1 + 1\}$ if $n^1 \geq L_1$, or in position $n^1 + 1$ with probability $\delta^1(n^1 + 1, n^1 + 1) = 1$ if $n^1 < L_1$. (This captures the random insertion in the queue of jobs awaiting enzymatic processing—if there is a free enzymatic processor when the job arrives, the job immediately enters processing and joins the end of the "queue.") When there are n^1 jobs at station 1, total service effort there (from processing and reneging) is given by $\phi^1(n^1) = \min(n^1, L_1)\mu_1 + n^1\kappa$. The proportion of this effort that is directed to the job in position ℓ in the queue at station 1 is given by $\gamma^1(\ell, n^1) =$
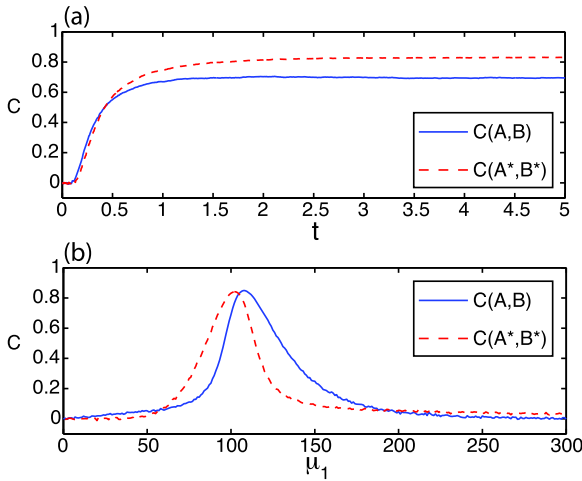
Fig. 2 (**a**) Time-dependent behavior of correlations for the network in Fig. 1 with sample parameters $\alpha_A = \alpha_B = 50$, $\mu_1 = 100$, $\mu_2 = 100$, $L_1 = L_2 = 10$, and $\kappa = 1$. Plotted are $C(A, B)$ (the correlation between A and B at station 1) and $C(A^*, B^*)$ (the correlation between A^* and B^* at station 2) as functions of time starting from an empty system. (**b**) Approximate steady-state results for the parameters in (**a**) but with varying μ_1. Plotted are the correlations evaluated at $t = 20$. In both (**a**) and (**b**), correlations were computed using (13) where the Fano factors for $N^1(t)$, $N^2(t)$ were approximated using Monte Carlo simulation involving 32 000 realizations of the pair of processes (N^1, N^2)

$(\mu_1 + \kappa)/\phi(n^1)$ for $\ell = 1, \ldots, \min(n^1, L_1)$ and $\gamma^1(\ell, n^1) = \kappa/\phi(n^1)$ for $\ell = L_1 + 1, \ldots, n^1$ and $n^1 > L_1$. Jobs at station 1 that complete enzymatic processing there are routed next to station two, whereas jobs that leave the station due to reneging exit the system. Thus, $\beta_2^1(\ell, n^1) = \mu_1/(\mu_1 + \kappa)$ for $\ell = 1, \ldots, \min(n^1, L_1)$, $\beta_2^1(\ell, n^1) = 0$ for $\ell = L_1 + 1, \ldots, n^1$ and $n^1 > L_1$, and $\beta_0^1(\ell, n^1) = 1 - \beta_2^1(\ell, n^1)$ for $\ell = 1, \ldots, n^1$. Similarly, at station two, the insertion probabilities, service effort, effort proportions, and routing parameters are given by the same expressions as for station 1, but with the indices 1 and 2 interchanged.

Assuming that the system is initialized in the empty state, the hypotheses of Theorem 4.1 are satisfied and Corollary 4.1 applies. This implies that, at any given time, conditioned on the total number of molecules at each station, the distribution of types at each station has a binomial distribution. This reduces the problem of computing the time-dependent distributions for the numbers of molecules of each type in each state to the corresponding problem for a single-class network with just one type of protein (e.g., containing A and A^* only, with production rate $\alpha = \alpha_A + \alpha_B$). Due to the state-dependent service rate associated with the degradation/reneging, such a single-class network is not of product form, and an analytic expression for either stationary or time-dependent distributions is not known. However, one can use Monte Carlo simulation to approximate these distributions. Thus, by using the factorization result to reduce to the single-class situation, the computational overhead is considerably reduced from what it would be for a simulation of the original multiclass network.

We illustrate the use of this semi-analytic/semi-computational approach for our example with sample parameters in Fig. 2. We obtained approximate values of the Fano

factors for $N^1(t)$ and $N^2(t)$ by using Monte Carlo approximations based on discrete-event simulations of a single-class network, implemented with a GPU-accelerated variant of the Gillespie algorithm [10, 12] using the pseudorandom number generator "ran2" in [18]. We then applied (13) to extract approximate values of time-dependent correlations between the numbers of each type at each station. In Fig. 2, for convenience of interpretation, we use $C(A, B)$ to denote the correlation between numbers of molecules of A and B, and $C(A^*, B^*)$ for the correlation between numbers of molecules of A^* and B^*. (These quantities are denoted by r_{12}^k, $k = 1, 2$, in (13).)

We now make several observations about the results displayed in Fig. 2. For the time-dependent behavior depicted in Fig. 2(a), the correlation between types at each station rapidly increases from zero to an apparent steady-state value. Initially, there are few jobs in the system, and the stations tend to operate like infinite server queues with little correlation between the numbers of each type. However, as time goes forward, the numbers in the system approach a steady-state, each of the stations becomes more heavily loaded, and jobs of different types awaiting service become substantially correlated in their queue-lengths. In Fig. 2(b), we plot the correlations at time $t = 20$ as functions of μ_1. Based on the time-dependent behavior in Fig. 2(a), it is reasonable to hypothesize that distributions at $t = 20$ are close to steady-state. We note that the correlation coefficients have nonmonotonic behavior as the rate μ_1 is varied. Specifically, we observe a peak in the correlation near the point where $\mu_1 = \mu_2$, i.e., where both "stations" would be bottlenecks in the closed network obtained by eliminating production and degradation [8, 9]. This indicates an ability to tune correlations of the two types of protein by adjusting the internal enzymatic processing rate μ_1.

Appendix

Proof of Theorem 4.1 Suppose that $\pi_t(\mathbf{s})$ is given by (9) for all $\mathbf{s} \in \mathcal{S}$ and $t \geq 0$. To verify the first claim of the theorem, it suffices [2] to verify that this expression satisfies the forward Kolmogorov equation (4). For notational brevity, we let

$$\psi(\mathbf{s}) = \prod_{k=1}^{K} \prod_{\ell=1}^{n^k} p_{s_\ell^k}, \quad \mathbf{s} \in \mathcal{S}. \tag{22}$$

Fix $\mathbf{s} \in \mathcal{S}$ and $t > 0$. On substituting the expression from (9) into the right-hand member of (4), we obtain

$$\sum_{\tilde{\mathbf{s}} \in \mathcal{S}} \pi_t(\tilde{\mathbf{s}}) \Gamma(\tilde{\mathbf{s}}, \mathbf{s})$$

$$= \sum_{k=1}^{K} \sum_{\ell=1}^{n^k} \pi_t(\mathbf{s}^{-,k,\ell}) \lambda_{s_\ell^k}^k \delta^k(\ell, n^k)$$

$$+ \sum_{k=1}^{K} \sum_{l \neq k} \sum_{\ell=1}^{n^k+1} \sum_{m=1}^{n^l} \pi_t\big(\mathbf{s}^{t,k,\ell,l,m}\big) \phi^k\big(n^k+1\big) \gamma^k\big(\ell, n^k+1\big) \beta_l^k\big(\ell, n^k+1\big)$$

$$\times \, \delta^l\big(m, n^l\big)$$

$$+ \sum_{k=1}^{K} \sum_{\ell=1}^{n^k+1} \sum_{i=1}^{I} \pi_t\big(\mathbf{s}^{+,k,\ell,i}\big) \phi^k\big(n^k+1\big) \gamma^k\big(\ell, n^k+1\big) \beta_0^k\big(\ell, n^k+1\big)$$

$$- \pi_t(\mathbf{s}) \sum_{k=1}^{K} \sum_{\ell=1}^{n^k+1} \sum_{i=1}^{I} \lambda_i^k \delta^k\big(\ell, n^k+1\big)$$

$$- \pi_t(\mathbf{s}) \sum_{k=1}^{K} \sum_{l \neq k} \sum_{\ell=1}^{n^k} \sum_{m=1}^{n^l+1} \phi^k\big(n^k\big) \gamma^k\big(\ell, n^k\big) \beta_l^k\big(\ell, n^k\big) \delta^l\big(m, n^l+1\big)$$

$$- \pi_t(\mathbf{s}) \sum_{k=1}^{K} \sum_{\ell=1}^{n^k} \phi^k\big(n^k\big) \gamma^k\big(\ell, n^k\big) \beta_0^k\big(\ell, n^k\big)$$

$$= \sum_{k=1}^{K} \sum_{\ell=1}^{n^k} \nu_t\big(\mathbf{n} - \mathbf{e}^k\big) \frac{\psi(\mathbf{s})}{p_{s_\ell^k}} \lambda_{s_\ell^k}^k \delta^k\big(\ell, n^k\big)$$

$$+ \sum_{k=1}^{K} \sum_{l \neq k} \sum_{\ell=1}^{n^k+1} \sum_{m=1}^{n^l} \nu_t\big(\mathbf{n} + \mathbf{e}^k - \mathbf{e}^l\big) \psi(\mathbf{s}) \phi^k\big(n^k+1\big) \gamma^k\big(\ell, n^k+1\big) \beta_l^k\big(\ell, n^k+1\big)$$

$$\times \, \delta^l\big(m, n^l\big)$$

$$+ \sum_{k=1}^{K} \sum_{\ell=1}^{n^k+1} \sum_{i=1}^{I} \nu_t\big(\mathbf{n} + \mathbf{e}^k\big) \psi(\mathbf{s}) p_i \phi^k\big(n^k+1\big) \gamma^k\big(\ell, n^k+1\big) \beta_0^k\big(\ell, n^k+1\big)$$

$$- \nu_t(\mathbf{n}) \psi(\mathbf{s}) \sum_{k=1}^{K} \sum_{\ell=1}^{n^k+1} \sum_{i=1}^{I} \lambda_i^k \delta^k\big(\ell, n^k+1\big)$$

$$- \nu_t(\mathbf{n}) \psi(\mathbf{s}) \sum_{k=1}^{K} \sum_{l \neq k} \sum_{\ell=1}^{n^k} \sum_{m=1}^{n^l+1} \phi^k\big(n^k\big) \gamma^k\big(\ell, n^k\big) \beta_l^k\big(\ell, n^k\big) \delta^l\big(m, n^l+1\big)$$

$$- \nu_t(\mathbf{n}) \psi(\mathbf{s}) \sum_{k=1}^{K} \sum_{\ell=1}^{n^k} \phi^k\big(n^k\big) \gamma^k\big(\ell, n^k\big) \beta_0^k\big(\ell, n^k\big)$$

$$= \sum_{k=1}^{K} 1_{\{n^k > 0\}} \nu_t\big(\mathbf{n} - \mathbf{e}^k\big) \psi(\mathbf{s}) \Lambda^k$$

$$+ \sum_{k=1}^{K} \sum_{l \neq k} \nu_t \left(\mathbf{n} + \mathbf{e}^k - \mathbf{e}^l\right) \psi(\mathbf{s}) \phi^k \left(n^k + 1\right) \tilde{\beta}_l^k \left(n^k + 1\right) 1_{\{n^l > 0\}}$$

$$+ \sum_{k=1}^{K} \nu_t \left(\mathbf{n} + \mathbf{e}^k\right) \psi(\mathbf{s}) \phi^k \left(n^k + 1\right) \tilde{\beta}_0^k \left(n^k + 1\right)$$

$$- \nu_t(\mathbf{n}) \psi(\mathbf{s}) \sum_{k=1}^{K} \Lambda^k - \nu_t(\mathbf{n}) \psi(\mathbf{s}) \sum_{k=1}^{K} 1_{\{n^k > 0\}} \phi^k \left(n^k\right)$$

$$= \sum_{\tilde{\mathbf{n}} \in \mathcal{N}} \nu_t(\tilde{\mathbf{n}}) \Delta(\tilde{\mathbf{n}}, \mathbf{n}) \psi(\mathbf{s})$$

$$= \frac{\partial \nu_t(\mathbf{n})}{\partial t} \psi(\mathbf{s})$$

$$= \frac{\partial \pi_t(\mathbf{s})}{\partial t},$$

and so the expression in (9) satisfies (4). Here we used (1), the facts that $\sum_{\ell=1}^{n^k} \delta^k(\ell, n^k) = 1_{\{n^k > 0\}}$, $\sum_{\ell=1}^{n^k} \gamma^k(\ell, n^k) = 1_{\{n^k > 0\}}$, $\sum_{i=1}^{I} p_i = 1$, and $\Lambda^k = \sum_{i=1}^{I} \lambda_i^k$, the definition of $\tilde{\beta}_l^k$ from (7), and the fact that $\sum_{l=0}^{K} \tilde{\beta}_l^k = 1$. Here we use the conventions that a product over an empty set has the value one and a sum over an empty set has the value zero.

The second claim of the theorem follows immediately since the distribution concentrated on the empty state has the form of (8). Finally, if ν^* is a stationary distribution for N, then the expression in (9) with ν^* in place of ν_t will be a stationary distribution for S. If furthermore N is irreducible and positive recurrent, one can verify that the same will be true for S, and then the stationary distribution will be the steady-state distribution. $\qquad \square$

References

1. Alon, U.: An Introduction to Systems Biology: Design Principles of Biological Circuits. Chapman and Hall/CRC, Baton Rouge (2007)
2. Anderson, W.J.: Continuous-time Markov Chains. Springer, New York (1991)
3. Ball, K., Kurtz, T.G., Popovic, L., Rempala, G.: Asymptotic analysis of multiscale approximations to reaction networks. Ann. Appl. Probab. **16**, 1925–1961 (2006)
4. Berg, O.G., Paulsson, J., Ehrenberg, M.: Fluctuations and quality of control in biological cells: Zero-order ultrasensitivity reinvestigated. Biophys. J. **79**, 1228–1236 (2000)
5. Borst, S.C., Boxma, O.J., Morrison, J.A., Queija, R.N.: The equivalence between processor sharing and service in random order. Oper. Res. Lett. **31**, 254–262 (2003)
6. Bramson, M.: State space collapse with application to heavy traffic limits for multiclass queueing networks. Queueing Syst. **30**, 89–148 (1998)
7. Bramson, M.: Stability of queueing networks. In: École d'Été de Probabilités de Saint-Flour XXXVI-2006. Lecture Notes in Mathematics, vol. 1950. Springer, Berlin (2008)
8. Chen, H., Mandelbaum, A.: Discrete flow networks—bottleneck analysis and fluid approximations. Math. Oper. Res. **16**, 408–446 (1991)
9. Chen, H., Mandelbaum, A.: Stochastic discrete flow networks—diffusion approximations and bottlenecks. Ann. Probab. **19**, 1463–1519 (1991)

10. Dematté, L., Prandi, D.: GPU computing for systems biology. Brief. Bioinform. **11**, 323–333 (2010)
11. Elf, J., Paulsson, J., Berg, O.G., Ehrenberg, M.: Near-critical phenomena in intracellular metabolite pools. Biophys. J. **84**, 154–170 (2003)
12. Gillespie, D.T.: Exact stochastic simulation of coupled chemical-reactions. J. Phys. Chem. **81**, 2340–2361 (1977)
13. Goldbeter, A., Koshland, D.E.: An amplified sensitivity arising from covalent modification in biological-systems. Proc. Natl. Acad. Sci. USA Biol. Sci. **78**, 6840–6844 (1981)
14. Hasty, J., McMillen, D., Collins, J.J.: Engineered gene circuits. Nature **420**, 224–230 (2002)
15. Hughes, A.L.: Gene duplication and the origin of novel proteins. Proc. Natl. Acad. Sci. USA **102**, 8791–8792 (2005)
16. Kelly, F.P.: Reversibility and Stochastic Networks. Wiley, New York (1979)
17. Mather, W.H., Cookson, N.A., Hasty, J., Tsimring, L.S., Williams, R.J.: Correlation resonance generated by coupled enzymatic processing. Biophys. J. **99**, 3172–3181 (2010)
18. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Numerical Recipes in C++: The Art of Scientific Computing. Cambridge University Press, Cambridge (2002)
19. Saito, H., Tatebayashi, K.: Regulation of the osmoregulatory HOG MAPK cascade in yeast. Journal of Biochemistry **136**, 267–272 (2004)
20. Wang, X., Hao, N., Dohlman, H.G., Elston, T.C.: Bistability, stochasticity, and oscillations in the mitogen-activated protein kinase cascade. Biophys. J. **90**, 1961–1978 (2006)
21. Williams, R.J.: Diffusion approximations for open multiclass queueing networks: Sufficient conditions involving state space collapse. Queueing Syst. **30**, 27–88 (1998)