Queueing Systems manuscript No.
(will be inserted by the editor)

Non-product form equilibrium probabilities in a class of two-station closed reentrant queueing networks

Woo-sung Kim · James R. Morrison*

Received: date / Accepted: date

Abstract While many single station queues possess explicit forms for their equilibrium probabilities, queueing networks are more problematic. Outside of the class of product form networks (e.g., Jackson, Kelly and BCMP networks), one must resort to bounds, simulation, asymptotic studies or approximations. By focusing on a class of two-station closed reentrant queueing networks under the last buffer first served (LBFS) policy, we show that *non-product form* equilibrium probabilities can be obtained. When the number of customer classes in the network is five or less, explicit solutions can be obtained. Otherwise, we require the roots of a characteristic polynomial and a matrix inversion that depend only on the network topology. The approach relies on two key points. First, under LBFS, the state space can be reduced to four dimensions independent of the number of buffers in the system. Second, there is a sense of spatial causality in the global balance equations that can then be exploited.

To our knowledge, these two-station closed reentrant queueing networks under LBFS represent the first class of queueing networks for which explicit *non-product form* equilibrium probabilities can be constructed (for five customer classes or less), the generic form of the equilibrium probabilities can

This work was supported in part by Korea Research Foundation (KRF) Grant 20110005696.

Woo-sung Kim
KAIST, 335 Gwahangno, Yuseong-gu, Daejeon 305-701, Korea
Fax: +82-42-350-3110
Tel.: +82-42-350-3167

*James R. Morrison
E-mail: james.morrison@kaist.edu, http://xs3d.kaist.edu
Corresponding author
KAIST, 335 Gwahangno, Yuseong-gu, Daejeon 305-701, Korea
Fax: +82-42-350-3110
Tel.: +82-42-350-3127

be deduced and matrix analytic approaches can be applied. As discussed via example, there may be other networks for which related observations can be exploited.

Keywords Closed queueing networks · Product form queueing networks · Closed form solution · Buffer priority policy

1 Introduction

Under reasonable conditions, the well-known product form networks of Jackson [1], Kelly [2] and BCMP [3] have equilibrium probabilities of the form

$$\pi(x_1, x_2, \ldots, x_S) = K \cdot \prod_{i=1}^{S} f_i(x_i),$$

where S is the number of stations in the network, $\pi(x_1, x_2, \ldots, x_S)$ is the steady state probability that there are x_1 customers at server 1, x_2 customers at server 2, ..., and x_S customers at server S, K is a normalizing constant and each function f_i depends only on x_i, for all i. This is the typical interpretation of "product form" for a queueing network ([4]). There has been much work on identifying queueing networks with product form, c.f., [5]-[11]. As stated in Bayer and Boucherie ([12]), "[v]irtually, all non-product form... networks do not yield explicit and exact analytical results... This is the motivation for clinging to product form models..."

A key class of networks *outside* of the product form structures are the multiclass queueing networks. Such networks arise in the modeling of semiconductor wafer fabricators ([13]) and other large scale modern systems. However, in general there are no known solutions for their equilibirum probabilities – they are not product form. To address this intractability, performance bounds (e.g., [13]-[19]), approximations (e.g., [20],[21]) and simulations (e.g., [22],[23]) have been proposed to obtain measures of system behavior. Much effort has also been devoted to developing stability conditions ([24],[25],[26]) and asymptotically optimal policies ([27],[28]). While these methods are useful, they do not provide the equilibrium probabilities.

We consider a subset of such multiclass queueing networks referred to as closed. A closed queueing network is one in which there are no external arrivals or departures (hence the "closed" qualifier) and N trapped customers circulate endlessly. As discussed in [29], such networks have been studied since 1954 ([30]). They may be used to model a network in which a new customer is released into the system every time a customer departs, so that the total number of jobs in the network is maintained at a constant level. Examples include mine haulage systems ([31]), automated material handling systems (AMHS) ([32]), pull production systems ([33]) and internet communication protocols ([34]).

We assume *exponential* service times and buffer priority policies. Because a closed network will then possess a finite state space, one can "always" numerically obtain solutions to the global balance equations. As such, these networks may be considered simpler than their open counterparts. However, as the number of trapped customers N increases, the state space increases dramatically, so that it is practically impossible to exactly obtain the equilibrium probabilities.

With the goal of completely characterizing the steady state behavior, we will focus on a class of closed reentrant two-station multiclass queueing networks operating under a last buffer first served (LBFS) policy. Outside of the abbreviated conference precursors to this work ([35, 36, 37]), there are no known multiclass queueing networks for which equilibrium probabilities can be obtained beyond the product form classes.

To address the intractability of closed two station queueing networks, Harrison and Wein [38] employed Brownian motion models for the balanced case. There, they obtained control policies conjectured as asymptotically optimal for the throughput as the number of trapped customers N approached infinity. Later, these Harrison-Wein policies (HW policies) were shown to be asymptotically optimal in an appropriate sense in [39] for the balanced case. Following work on open two station networks [40], conditions guaranteeing that a closed two station network can asymptotically achieve its bottleneck throughput under any non-idling control policy were given in [41]. In work that is perhaps closest in spirit to our study, a five-class two-station open queueing network was studied in [42].

The contributions of this work are as follows. We

- identify a limited class of closed two station reentrant queueing networks operating under an LBFS policy for which
  - independent of the number of classes, the state space can be reduced to four dimensions;
  - there is a sense of spatial causality in the global balance equations;
- demonstrate that the equilibrium probabilities are a linear combination of powers of roots of a characteristic polynomial;
- show that with five classes or less, closed form solutions for the equilibrium probabilities can be found;
- for our networks with six classes or more, obtain a general form for the equilibrium probabilities;
- study the the computational efficiency of our approach; and
- show another network that possesses some of the structural properties that may be similarly (but not as completely) exploited.

Overall, our approach is straightforward, though it is somewhat involved. It shares similarities with methods used for single queues. To our knowledge, the networks identified represent the only multiclass queueing networks for which non-product form solutions to the equilibrium probabilities have been obtained.

The remainder of the paper is organized as follows. In Section 2, the system description and performance measures are discussed. In Section 3, we derive general structural properties of our system. In Section 4, we study the case of five customer classes or less. In Section 5, we study the case of six customer classes or more to develop a general non-product form expression. In Section 6, other algorithmic approaches and computational complexity are mentioned. In Section 7, we apply the methods to the closed Lu-Kumar network. Concluding remarks are presented in Section 8. Please refer to the conference precursors of this work ([35, 36, 37]) for details omitted here.

2 Class of systems of interest

We first describe our systems and performance metrics.

2.1 System overview

We consider a class of two-station closed reentrant queueing networks. Such a network consists of two single server stations labeled σ_1 and σ_2 which provide service. There are m+n processes or customer classes, labeled P_1, P_2, P_3, $\ldots$, P_{m+n}. To each is associated an infinite capacity buffer at which customers await service. Label these b_1, ... , b_{m+n} according to their associated process. Customers receive processes P_1, P_2, P_3, $\ldots$, P_m at station σ_1. Customers receive processes P_{m+1}, P_{m+2}, $\ldots$, P_{m+n} at station σ_2. For convenience, we require $m \geq 2$ and $n \geq 2$ (cases with m or n equal to one are discussed later). Figure 1 depicts such a system. The service time for a customer in process P_i is exponentially distributed with rate μ_i, $0 < \mu_i < \infty$ (it will not thus matter whether service is preempt-resume or preempt-repeat). All service times are mutually independent. There are no exogenous arrivals and no departures; N trapped customers circulate within the network. Customers require service in sequential order from each process. That is, after receiving service from process P_i, the customer will immediately advance to buffer b_{i+1} to await service for process P_{i+1}. When i=m+n, the customer next proceeds to buffer b_1 for process P_1. Customers are processed in first-come first-served order within a particular buffer.

Since each station possesses a single server, and we assume the server must devote its entire effort to serving a single customer when one is present, there will be contention between buffers if two or more are non-empty at a station. To resolve this contention, we employ the last buffer first served (LBFS) policy. Under the LBFS policy, process P_j has higher priority than P_i if $i < j$. We assume that the policy is non-idling and preemptive. The LBFS policy is the same as the HW policy ([38]) for our network.

Let $u(t) = (\upsilon_1(t), \upsilon_2(t), ..., \upsilon_{m+n}(t))$, where $\upsilon_i(t)$ denotes the number of customers in buffer b_i at time t (including any receiving service for process P_i). Let $S = \{u : u \in Z_+^{m+n}, e^T u = N\}$, where $e = (1, 1, ..., 1)^T$, denote the non-negative integer orthant restricted to the simplex. Assume all processes are

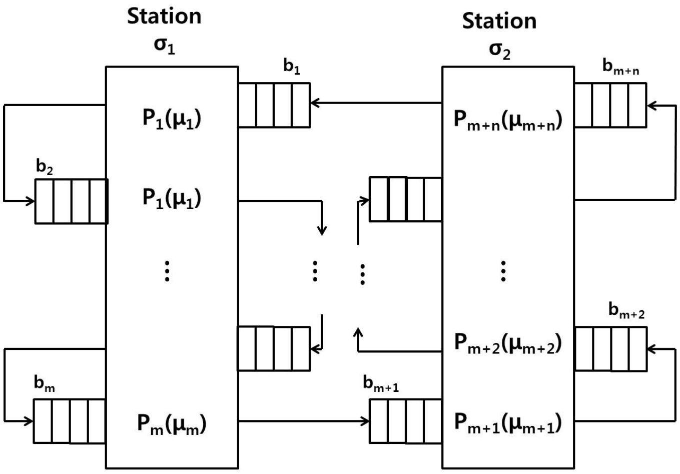

Fig. 1 A class of two station closed reentrant networks

right-continuous with left-limits. Our system is thus a finite-state, continuous-time, time-homogeneous, controlled Markov chain with state space S; u(t) denotes the state at time t. The LBFS policy depends only on the state.

We resort to uniformization ([44]). That is, we rescale time so that $\sum_{i=1}^{m+n} \mu_i$ = 1 and sample the system at all times τ_w at which either a real or a virtual service completion occurs. The sampled process $\{u(\tau_w)\}_{w=1}^{\infty}$ is a finite-state, discrete-time, time-homogeneous Markov chain. Abusing notation slightly we denote it as $\{u_w\}_{w=1}^{\infty} = \{(\upsilon_1(w), ..., \upsilon_{m+n}(w))\}_{w=1}^{\infty}$. The Markov Chain has a single communicating class because state {N,0,...,0} is reachable from every state in S. It is aperiodic because we can remain in any state for an arbitrary number of time steps (due to the virtual service completions). Its equilibrium probabilities are the same as those of the original process.

2.2 Performance metrics

Perhaps the most celebrated performance measure in a closed queueing network is throughput. Given a fixed customer population N, an initial condition detailing their locations and a scheduling policy Δ (typically assumed non-idling and non-anticipative), the random variable

$$\alpha^{\Delta}(N) = \liminf_{T\to\infty} \frac{D_{m+n}(T)}{T},$$

where $D_{m+n}(T)$ is the number of departures from process P_{m+n} in the time interval [0,T], is called *throughput*. Generally, the throughput takes a single value with probability one, increases as the number of fixed customers in the system increases, and converges to the maximum achievable throughput as N

increases. We can calculate the bottleneck throughput as

$$\alpha^* := \min_{\sigma} \frac{1}{\sum_{\{i:b_i\in\sigma\}} \frac{1}{\mu_i}}, \tag{1}$$

where $b_i \in \sigma$ indicates that buffer b_i is served by station σ. It is clear that $\alpha^{\Delta}(N) \leq \alpha^*$ a.s. This value α^* can be approached as $N \to \infty$ if the system and policy are efficient; see [41]. Under a policy such as LBFS, the throughput is well defined [47]. We may calculate it as $\mu_i \cdot Prob\{$Process P_i is in service in steady state$\}$, for any i.

3 General structural properties

Here, we derive general structural properties for our class of networks.

3.1 State space reduction

The first property follows directly from the LBFS policy.

Proposition 1. In equilibrium,

- $\sum_{i=2}^{m} v_i(w) \leq 1$ and $\sum_{i=m+2}^{m+n} v_i(w) \leq 1$, and
- The state space has an equivalent representation as the set $S' = \{(w, x, y, z) \in \mathcal{Z}_+ \times \{0, 2, \ldots, m\} \times \mathcal{Z}_+ \times \{0, 2, \ldots, n\} | w + y + I_{x\neq 0} + I_{z\neq 0} = N\}$, where $I_{x\neq 0}$ is the indicator of whether the value x is non-zero; similarly for $I_{z\neq 0}$.

We interpret a state $s = (w, x, y, z) \in S'$ as follows. At a given time instant, w and y denote the number of customers in buffers b_1 and b_{m+1}, respectively. If there is a customer in one of the buffers $b_2, \ldots, b_m$, the value x denotes the index of that buffer; otherwise, $x = 0$. If there is a customer in one of the buffers $b_{m+2}, \ldots, b_{m+n}$, the z is such that b_{m+z} is the label for that buffer; otherwise, $z = 0$. Since the system is closed, we have the simplex condition. For example, consider the state {1,2,3,4}. The first entry indicates that there is one customer in b_1. The second entry indicates that there is one customer in b_2. Similarly, there are three customers in b_{m+1}, and there is one customer in b_{m+4}. In this case we have $N = 6$.

Proof. The first follows from the LBFS policy. The second follows since it is equivalent to simply track in which buffer the customer currently in the buffers $b_2, \ldots, b_m$ resides (if there is one). Similarly for buffers $b_{m+2}, \ldots, b_{m+2}$. Residual service times are not needed with exponential service times. □

Note that, under LBFS, our reentrant network becomes a closed tandem network of two Erlang servers.

3.2 Recursive structure of the global balance equations

Employing the notation of the previous section, we endeavor to construct a convenient matrix representation for the global balance equations (GBEs). We demonstrate that they possess a recursive structure that will be amenable to standard matrix z-transform techniques in the sequel.

Since our Markov Chain model has a finite state space and single communicating class, there is a unique probability vector solution to the global balance equations ([48]); it is the vector of equilibrium probabilities. Use $\Pi_{w,x,y,z}$ to denote the equilibrium probability of state $s = (w, x, y, z)$. The transition probability from state s to $s' \neq s$ is denoted as $\lambda[s, s']$. Recalling that we have assumed $m, n \geq 2$, they are

$$\begin{aligned}
&\lambda[(w,0,y,z),(w-1,2,y,z)] = \mu_1 \cdot I_{\{w\geq 1\}},\\
&\lambda[(w,x,y,0),(w,x,y-1,2)] = \mu_{m+1} \cdot I_{\{y\geq 1\}},\\
&\lambda[(w,x,y,z),(w,x+1,y,z)] = \mu_x \cdot I_{\{2\leq x\leq m-1\}},\\
&\lambda[(w,m,y,z),(w,0,y+1,z)] = \mu_m,\\
&\lambda[(w,x,y,z),(w,x,y,z+1)] = \mu_z \cdot I_{\{2\leq z\leq n-1\}},\\
&\lambda[(w,x,y,n),(w+1,x,y,0)] = \mu_{m+n},
\end{aligned}$$

and all others are zero for $s \neq s'$. Due to the virtual service completions, $\lambda[s,s] = 1 - \sum_{s'\neq s} \lambda[s,s']$. We define $X_{i,j}(k)$ as

$$\begin{aligned}
&X_{0,0}(k) = \Pi_{N-k-1,0,k+1,0} \text{ for } k = -1, ..., N-1,\\
&X_{i,0}(k) = \Pi_{N-k-2,i,k+1,0} \text{ for } i = 2, ..., m \text{ and } k = -1, ..., N-2,\\
&X_{i,j}(k) = \Pi_{N-k-2,i,k,j} \text{ for } i = 2, ..., m, j = 2, ..., n \text{ and } k = 0, ..., N-2,\\
&X_{0,j}(k) = \Pi_{N-k-1,0,k,j} \text{ for } j = 2, ..., n \text{ and } k = 0, ..., N-1.
\end{aligned}$$

For example, $X_{0,0}(-1) = \Pi_{N,0,0,0}$.

The transition probability diagram is given in Figure 2. There, the state space is depicted as a collection of rectangles stacked on top of each other. Each rectangle in the stack represents the collection of states for which the number of customers at station σ_1 is held constant (and thereby the number of customers at station σ_2 is held constant). Within each such rectangle, the difference between the states is simply the location of the customers currently in service. Those probabilities $X_{i,j}(k)$, with *k fixed* are the equilibrium probabilities for states on the k^{th} floor of the stack of rectangles; such states have exactly N-k-1 customers at station σ_1 and k+1 customers at station σ_2. We will call those probabilities $X_{i,j}(k)$, with k fixed, as the k^{th} floor probabilities.

In the interior of the stack, the transition probabilities between states in rectangle k are the same as those between corresponding states in rectangle $k+1$. We call the probabilities $X_{i,j}(-1)$ as initial conditions. If we know the values $X_{i,j}(-1)$, we can calculate the probabilities for the far left column of states

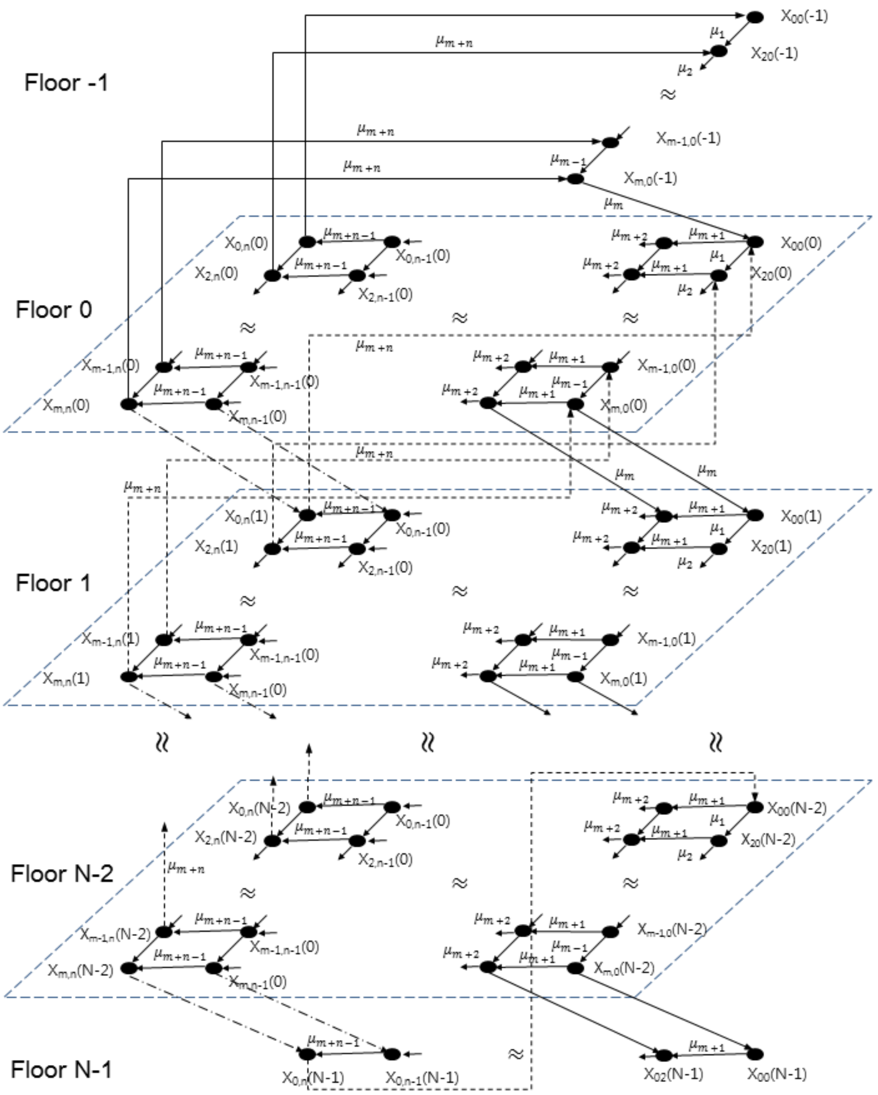

Fig. 2 Transition probability diagram

on the 0^{th} floor as

$$Y_{*n}[0] \triangleq \begin{bmatrix} X_{0,n}(0) \\ X_{2,n}(0) \\ \vdots \\ X_{m-1,n}(0) \\ X_{m,n}(0) \end{bmatrix} = \frac{1}{\mu_{m+n}} \begin{bmatrix} \mu_1 & 0 & \cdots & 0 & 0 \\ -\mu_1 & \mu_2 & \cdots & 0 & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & -\mu_{m-2} & \mu_{m-1} & 0 \\ 0 & 0 & \cdots & -\mu_{m-1} & \mu_m \end{bmatrix} \begin{bmatrix} X_{0,0}(-1) \\ X_{2,0}(-1) \\ \vdots \\ X_{m-1,0}(-1) \\ X_{m,0}(-1) \end{bmatrix}.$$

Similarly, knowing the probabilities for a column on the 0^{th} floor enables us to obtain the probabilities for the column immediately to its right. It is helpful to employ the following notation.

$$\mathrm{Y_{i*}[k]} \triangleq [\mathrm{X_{i0}(k), X_{i2}(k), X_{i3}(k),..., X_{in}(k)}]^{\mathrm{T}},$$
$$\mathrm{Y_{*j}[k]} \triangleq [\mathrm{X_{0j}(k), X_{2j}(k), X_{3j}(k),..., X_{mj}(k)}]^{\mathrm{T}},$$
$$\mathbf{Y}[k] \triangleq [Y_{*0}^T[k], Y_{*2}^T[k], ..., Y_{*n}^T[k]]^T,$$

where $Y_{i*}[k]$, $Y_{*j}[k]$ and $\mathbf{Y}[k]$ are column vectors with dimension $n \times 1$, $m \times 1$ and $mn \times 1$, respectively. (They represent rows on the k^{th} floor, columns on the k^{th} floor and the entire k^{th} floor, respectively.) We obtain the matrix form of the GBEs for the 0^{th} floor (including the left most column) as

$$\mathbf{Y}[0] = \begin{bmatrix} Y_{*0}[0] \\ Y_{*2}[0] \\ ... \\ Y_{*n-1}[0] \\ Y_{*n}[0] \end{bmatrix} = \begin{bmatrix} \Theta & R_1 & ... & \Theta & \Theta \\ \Theta & \Theta & R_2 & ... & \Theta \\ ... & ... & ... & ... & ... \\ \Theta & \Theta & ... & \Theta & R_{n-1} \\ \Theta & \Theta & ... & \Theta & \Theta \end{bmatrix} \begin{bmatrix} Y_{*0}[0] \\ Y_{*2}[0] \\ ... \\ Y_{*n-1}[0] \\ Y_{*n}[0] \end{bmatrix} + \begin{bmatrix} \Theta \\ \Theta \\ ... \\ \Theta \\ R_n \end{bmatrix} \mathbf{Y}[-1]$$

$$=: U\mathbf{Y}[0] + U_n\mathbf{Y}[-1],$$

where

$$R_i = \frac{1}{\mu_{m+i}} \begin{bmatrix} \mu_1 + \mu_{m+i+1} & 0 & \cdots & 0 & 0 \\ -\mu_1 & \mu_2 + \mu_{m+i+1} & \ddots & \vdots & 0 \\ 0 & \ddots & \ddots & 0 & \vdots \\ \vdots & 0 & -\mu_{m-2} & \mu_{m-1} + \mu_{m+i+1} & 0 \\ 0 & 0 & \cdots & -\mu_{m-1} & \mu_m + \mu_{m+i+1} \end{bmatrix}$$

for i=1,2,...,n-1,

$$R_n = \frac{1}{\mu_{m+n}} \begin{bmatrix} \mu_1 & 0 & \cdots & 0 & 0 \\ -\mu_1 & \mu_2 & \ddots & \vdots & 0 \\ 0 & \ddots & \ddots & 0 & \vdots \\ \vdots & 0 & -\mu_{m-2} & \mu_{m-1} & 0 \\ 0 & 0 & \cdots & -\mu_{m-1} & \mu_m \end{bmatrix} \text{ and } \mathbf{Y}[-1] = \begin{bmatrix} X_{0,0}(-1) \\ X_{2,0}(-1) \\ \vdots \\ X_{m-1,0}(-1) \\ X_{m,0}(-1) \end{bmatrix}.$$

All R_i matrices are $m \times m$. Thus, U is an $mn \times mn$ matrix. $\mathbf{Y}[0]$ and $\mathbf{Y}[-1]$ are $mn \times 1$ and $m \times 1$ matrices. The notation Θ indicates an $m \times m$ matrix of zeros. Letting $R = (I - U)^{-1}U_n$, we obtain a simple matrix form. The matrix $I - U$ is invertible since it is upper triangular and each row has a pivot position ([49]). We have

$$\mathbf{Y}[0] = R\mathbf{Y}[-1]. \tag{2}$$

Similarly, writing the GBEs for other floors in matrix form, we obtain

$$\mathbf{Y}[1] = S\mathbf{Y}[-1], \tag{3}$$

$$(I - A)\mathbf{Y}[k] = B\mathbf{Y}[\mathrm{k}-1] + C\mathbf{Y}[\mathrm{k}-2], \quad \text{for } k = 2, \ldots, N-2. \tag{4}$$

The definition of these matrices and the process to develop equations (2), (3) and (4) are given in Appendix A. The matrix $I - A$ is also invertible since the matrix is upper triangular and each row has a pivot position. We thus obtain the following result.

Proposition 2. The GBEs for our class of networks can be written in the recursive form (2), (3) and (4) for $k = 2, \ldots, N-2$.
Proof. The above discussion and the details of Appendix A suffice to conclude the recursive form. □

The recursive nature of the GBEs enables us to employ standard matrix z-transform techniques.

4 Networks with explicit closed form equilibrium probabilities

In this section, special cases of our networks that possess explicit solutions for their equilibrium probabilities are introduced. Note that the $m = 1$ and $n = 1$ case gives a closed Jackson network.

Theorem 1. For our networks with five customer classes or less ($m+n \leq 5$), there are explicit closed form solutions for the equilibrium probabilities.
Proof. The result is immediate for the case $m = 1$ and $n = 1$.

For the cases with $m = 1$ and $2 \leq n \leq 4$ (or equivalently, $2 \leq m \leq 4$ and $n = 1$), we denote the state space as follows. Let S'' be the set of $s'' = (x, y, z)$ that satisfies an appropriate simplex condition, where x, y and z are the number of customers in buffers b_1, b_2, and index of any customer in buffers $b_3, \ldots, b_{n+1}$, respectively. If there is no customer in buffers $b_3, \ldots, b_{n+1}$ set $z = 0$. For the cases with $m \geq 2$ and $n \geq 2$ (with $m + n \leq 5$), the state space detailed in the prequel can be employed.

Omitting the details (there is a great deal of algebra), one employs z-transform techniques on the resulting recursive form of the GBEs. In all cases, the z-transform for the equilibrium probabilities is a rational function depending on the initial condition probabilities as above. The denominator polynomial is of degree five or less. One of the roots is always $z = 1$. It is well known that the roots of any polynomial of fourth degree or less can be obtained in closed form. Thus, we obtain the remaining four roots of the denominator polynomial. The inverse z-transform can then be computed by partial fraction expansion ([50]). The initial condition probabilities are obtained from the GBEs on the final floor of the stack together with the probability normalization condition.

Refer to the conference paper precursors ([35, 36, 37]) of this work for the details and resulting explicit solutions. □

As we shall see shortly, there is a general form for the equilibrium probabilities that all of our systems possess. In the cases $m + n \leq 5$, it is possible to completely obtain them in closed form. As an example of the results that one may thus obtain, we present the following result on throughput.

Corollary 1. The throughput, $\alpha(N)$, for the case $m = 1$ and $n = 2$ is as follows. Without loss of generality, we rescale time so that $\mu_1 = 1$.

– For the unbalanced case, that is, when $\mu_1 \neq (1/\mu_2 + 1/\mu_3)^{-1}$,

$$\alpha(N) = 1 - ar_1^{N-1}\left[\tfrac{1}{\mu_3} + \tfrac{1}{\mu_2} + \tfrac{\mu_2}{\mu_3} - \tfrac{1}{\mu_3 r_1}\right]\Pi_{N,0,0}$$
$$- br_2^{N-1}\left[\tfrac{1}{\mu_3} + \tfrac{1}{\mu_2} + \tfrac{\mu_2}{\mu_3} - \tfrac{1}{\mu_3 r_1}\right]\Pi_{N,0,0}.$$

where,

$$r_1 = \tfrac{(1+\mu_2+\mu_3)+\sqrt{(1+\mu_2+\mu_3)^2-4\mu_2\mu_3}}{2\mu_2\mu_3},$$
$$r_2 = \tfrac{(1+\mu_2+\mu_3)-\sqrt{(1+\mu_2+\mu_3)^2-4\mu_2\mu_3}}{2\mu_2\mu_3},$$
$$a = \tfrac{(1+\mu_3-\mu_2)+\sqrt{(1+\mu_2+\mu_3)^2-4\mu_2\mu_3}}{2\sqrt{(1+\mu_2+\mu_3)^2-4\mu_2\mu_3}},$$
$$b = \tfrac{(1+\mu_3-\mu_2)-\sqrt{(1+\mu_2+\mu_3)^2-4\mu_2\mu_3}}{-2\sqrt{(1+\mu_2+\mu_3)^2-4\mu_2\mu_3}},$$

and

$$\Pi_{N,0,0}^{-1} =$$
$$\left(\frac{1}{\mu_3}\right) + \left(\frac{a}{1-r_1}\right)\left[1 + \left(\frac{r_1}{\mu_3}\right) + \left(\frac{r_1\mu_2}{\mu_3}\right) - \left(\frac{1}{\mu_3}\right)\right]$$
$$+ \left(\frac{b}{1-r_2}\right)\left[1 + \left(\frac{r_2}{\mu_3}\right) + \left(\frac{r_2\mu_2}{\mu_3}\right) - \left(\frac{1}{\mu_3}\right)\right]$$
$$+ \left(\frac{ar_1^N}{1-r_1}\right)\left[\frac{1}{r_1\mu_2} - 1 - \frac{1}{\mu_2} - \frac{1}{\mu_3} - \frac{\mu_2}{\mu_3} + \frac{1}{r_1\mu_3}\right]$$
$$+ \left(\frac{br_2^N}{1-r_2}\right)\left[\frac{1}{r_2\mu_2} - 1 - \frac{1}{\mu_2} - \frac{1}{\mu_3} - \frac{\mu_2}{\mu_3} + \frac{1}{r_2\mu_3}\right].$$

– For the balanced case, that is, when $\mu_1 = (1/\mu_2 + 1/\mu_3)^{-1}$,

$$[1 - \alpha(N)]^{-1} = \left(\frac{N\mu_3^2}{\mu_3^2 - \mu_3 + 1}\right) + \left(\frac{\mu_3^3 - 2\mu_3^2 + 1}{\mu_3(\mu_3^2 - \mu_3 + 1)}\right)$$
$$+ \frac{\mu_3^2}{(\mu_3^2 - \mu_3 + 1)^2}\left[1 - \frac{(\mu_3 - 1)^2}{\mu_3^3} + \frac{(\mu_3 - 1)^{N+1}}{\mu_3^{2N}}\right].$$

With some effort, the asymptotic losses may be calculated explicitly, c.f. [35]. The asymptotic loss is exactly as predicted in [38] for the balanced case. It is zero for the unbalanced case; contrary to the conjecture.

5 Networks with $m + n \geq 6$

When $m+n \geq 6$, we cannot obtain a closed form expression for the equilibrium probabilities. However, we can exploit the system structure to gain insight into their general form. We rely on the recursive structure of the GBEs to derive the z-transform of the steady state probabilities. We later invert the z-transform to obtain the equilibrium probabilities. The z-transform of equation (4), assuming it holds for all $k \geq 2$, is

$$[(\mathrm{I}-\mathrm{A})\mathrm{z}^2 - \mathrm{Bz} - \mathrm{C}]\mathrm{Y(z)} = [(\mathrm{I}-\mathrm{A})\mathrm{z}^2\mathbf{Y}[0] + \mathrm{z}(\mathrm{I}-\mathrm{A})\mathbf{Y}[1] - \mathrm{zB}\mathbf{Y}[0]]. \quad (5)$$

We can express $\mathbf{Y}[0]$ and $\mathbf{Y}[1]$ in terms of the "initial conditions" $\mathbf{Y}[-1]$ using (2) and (3). We have

$$[(\mathrm{I}-\mathrm{A})\mathrm{z}^2 - \mathrm{Bz} - \mathrm{C}]\mathrm{Y(z)} = [(\mathrm{I}-\mathrm{A})\mathrm{z}^2 R + z(I-A)S - zBR]\mathbf{Y}[-1]. \quad (6)$$

Since (4) is a causal recursion that holds only for $k = 2, ..., N-2$, the inverse of $Y(z)$ gives $\mathbf{Y}[k]$, $k = 2, ..., N-2$ (but not beyond). The point here is that we obtain a convenient $Y(z)$ by pretending the recursion of (4) holds for all $k \geq 2$, but restrict ourself to using the inverse $\mathbf{Y}[k]$ only for $k = 2, ..., N-2$. This is of course necessary since the GBE recursion (4) only holds for $k = 2, ..., N-2$.

If $[(I-A)z^2 - Bz - C]$ is invertible, we can obtain Y(z).

Lemma 1. The matrix D=$[(I-A)z^2 - Bz - C]$ is invertible for $z \neq 0$.
Proof. We can express the matrix D as

$$D = \begin{bmatrix} z^2 I & (-z^2 A_1 - zB_1) & ... & \Theta & \Theta \\ \Theta & z^2 I & ... & \Theta & \Theta \\ ... & ... & ... & ... & ... \\ \Theta & \Theta & ... & z^2 I & (-z^2 A_{n-1} - zB_{n-1}) \\ zB_n - C_1 & \Theta & ... & \Theta & z^2 I \end{bmatrix}.$$

Except for the $zB_n - C_1$ in the lower left corner, the matrix is upper triangular; every element on the diagonal is z^2. By the invertible matrix theorem in [49], if D is row equivalent to the identity matrix, then D is invertible. First, using the sub-matrix $z^2 I$ in the first m rows, we transform the sub-matrix $zB_n - C_1$ in the last m rows into Θ. This process creates non-zero entries in the sub-matrix in the last row, just to the right of the original $zB_n - C_1$. In a similar way, we can transform this last row second sub-matrix into the 0 matrix. Recursively, we can transform matrix D into an upper triangular matrix with a pivot position in each row. Thus, the matrix has mn pivot positions and is invertible. □

Multiplying both sides of equation (6) by D^{-1}, we obtain $Y(z)$.

Lemma 2. Recall that $\mathbf{Y}[k]$ is just the vector of all $X_{i,j}(k)$. The z-transform $X_{i,j}(z)$ of the signal $X_{i,j}(k)$ obeying (2), (3) and (4) for $k = 2, \ldots$ (k unbounded) may be expressed as

$$X_{i,j}(z) = C_{i,j,1}(z)X_{0,0}(-1) + \sum_{k=2}^{m} C_{i,j,k}(z)X_{k,0}(-1),$$

for $i = 2, \ldots, m$, $j = 2, \ldots, n$, where $X_{00}(-1)$ and $X_{k0}(-1)$ are the initial conditions and $C_{i,j,k}(z) = \sum\limits_{l=0}^{2mn} b_{i,j,k,l} z^l / \sum\limits_{p=0}^{2mn} a_p z^p$ are proper rational functions for appropriate constants $b_{i,j,k,l}$ and a_p.

Proof. From (6) and Cramer's rule ([49]), we can express Y(z) as

$$Y(z) = \frac{adj D}{\det D}[(\mathrm{I}-\mathrm{A})\mathrm{z}^2\mathrm{R} + \mathrm{z}(\mathrm{I}-\mathrm{A})\mathrm{S} - \mathrm{zBR}]\mathbf{Y}[-1], \tag{7}$$

where $detD$ is the determinant of the matrix D and $adjD$ is its adjoint. Since the matrix D is of dimension $mn \times mn$, and z^2 terms occur in every element on the diagonal, $detD$ is a polynomial in which the degree of z is at most $2mn$. Thus, det D can be written as:

$$\det D = \sum_{i=0}^{2mn} a_i z^i, \tag{8}$$

where a_i, $i = 0, \ldots, 2mn$, are appropriate constants. Each element in adj D contains a polynomial of degree 2mn-2 or less because each is obtained by deleting one row and column of the matrix D. Thus, the maximum degree of adj D is 2mn-2. Because the maximum degree of z of each element of $[(I-A)z^2R + z(I-A)S - zBR]$ is 2, each element of the numerator matrix of (7) also contains a polynomial with maximum degree 2mn. So Y(z) can be written as

$$Y(z) = C_z\mathbf{Y}[-1], \tag{9}$$

where

$$C_z = \begin{bmatrix} C_{0,0,1}(z) & C_{0,0,2}(z) & \ldots & C_{0,0,m-1}(z) & C_{0,0,m}(z) \\ C_{2,0,1}(z) & C_{2,0,2}(z) & \ldots & C_{2,0,m-1}(z) & C_{2,0,m}(z) \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ C_{m,0,1}(z) & C_{m,0,2}(z) & \ldots & C_{m,0,m-1}(z) & C_{m,0,m}(z) \\ C_{0,2,1}(z) & C_{0,2,2}(z) & \ldots & C_{0,2,m-1}(z) & C_{0,2,m}(z) \\ C_{2,2,1}(z) & C_{2,2,1}(z) & \ldots & C_{2,2,m-1}(z) & C_{2,2,m}(z) \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ C_{m,n,1}(z) & C_{m,n,2}(z) & \ldots & C_{m,n,m-1}(z) & C_{m,n,m}(z) \end{bmatrix}$$

and each $C_{i,j,k}(z)$ is the proper rational function $C_{i,j,k}(z) = \sum\limits_{l=0}^{2mn} b_{i,j,k,l} z^l / \sum\limits_{p=0}^{2mn} a_p z^p$, for appropriate constants $b_{i,j,k,l}$ and a_p. The result follows. □

Inverting the z-transform, and recalling that the inverse will hold for $k \leq N-2$ (since the GBE recursion of (4) holds for $k \leq N-2$), we obtain an expression for the equilibrium probabilities in terms of the m initial condition probabilities.

Lemma 3. The solution of recursion (4) is

$$X_{i,j}(k) = C_{i,j,1}(k)X_{0,0}(-1) + \sum_{l=2}^{m} C_{i,j,l}(k)X_{l,0}(-1), \tag{10}$$

for $k = 0, \ldots, N-2$,

where

$$C_{i,j,l}(k) = \sum_{u=1}^{q} \left[\left(\beta_{i,j,l,u,1} \cdot {p_{i,j,l,u}}^k\right) + \sum_{x=2}^{n_u} \frac{\prod_{a=1}^{x-1}(k+a)}{(x-1)!} \beta_{i,j,l,u,x} \cdot {p_{i,j,l,u}}^k \right] + \beta_{i,j,l,0}\delta(l). \tag{11}$$

The $\beta_{i,j,l,0}$, $\beta_{i,j,l,u}$, ... are constants, $p_{i,j,k,u}$ are the roots of the determinant of D, and n_u is the multiplicity of the root $p_{i,j,k,u}$.

Proof. Since each $C_{i,j,k}(z)$ is a rational function in z, the inverse $C_{i,j,k}(k)$ is simply as given for appropriate constants ([50]). Thus, by standard inversion of the z-transform, we have the result. □

It is a unique solution since the unilateral z-transform and its inverse have a one-to-one correspondence ([50]).

To obtain the remaining m probabilities $X_{i,0}(-1)$, $i = 1, \ldots, m$, we employ the unused balance equations on floor $N-1$. Combining these with $X_{i,j}(k)$ of Lemma 3, we have a collection of m equations in m unknowns. We thus arrive at Lemma 4.

Lemma 4. The remaining balance equations on the $N - 1^{th}$ floor and the $X_{i,j}(k)$ from Lemma 3 give

$$\begin{bmatrix} (\mu_1 + \mu_{m+1})C_{0,0}(N-2) - \mu_m C_{m,0}(N-3) - \mu_m \sum_i C_{m,i}(N-2) \\ (\mu_2 + \mu_{m+1})C_{2,0}(N-2) - \mu_1 C_{0,0}(N-2) \\ (\mu_3 + \mu_{m+1})C_{3,0}(N-2) - \mu_2 C_{2,0}(N-2) \\ (\mu_4 + \mu_{m+1})C_{4,0}(N-2) - \mu_3 C_{3,0}(N-2) \\ \vdots \\ (\mu_{m-1} + \mu_{m+1})C_{m-1,0}(N-2) - \mu_{m-2} C_{m-2,0}(N-2) \\ (\mu_m + \mu_{m+1})C_{m,0}(N-2) - \mu_{m-1} C_{m,0}(N-2) \end{bmatrix} \mathbf{Y}[-1] = \Theta, \tag{12}$$

where $C_{i,j}(k) = \left[\, C_{i,j,1}(k)\ C_{i,j,2}(k)\ \ldots\ C_{i,j,m-1}(k)\ C_{i,j,m}(k) \,\right]$.

We are now poised to provide our key result for $m + n \geq 6$.

Theorem 2. Assuming there is a unique solution to (12) and the sum of probabilities condition

$$\left[\sum_i \sum_j \sum_k C_{i,j}(k) \right] Y[-1] = 1, \tag{13}$$

then there is a unique solution to the GBEs for a network in our class, they are the equilibrium probabilities, and have the form

$$X_{i,j}(k) = C_{i,j,1}(k) X_{0,0}(-1) + \sum_{l=2}^{m} C_{i,j,l}(k) X_{l,0}(-1), \tag{14}$$

for $k = 0, \ldots, N-2$.

Proof. If there is a unique solution to (12) and (13), then Lemmas 1, 2, 3 and 4 give us a unique solution to the GBEs. Since we have an irreducible, aperiodic, discrete-time, discrete-state and time-homogeneous Markov Chain, by [48], we obtain the result. □

Thus, under the conditions of Theorem 2, there is a unique equilibrium probability distribution with the *non-product* form above.

6 Computational complexity

We study the computation required to calculate the equilibrium probabilities.

Proposition 9. Our transition matrix has the structure of a Quasi birth-death (QBD) process.

Proof. Using our state space definition from Section 3, the transition matrix T has the form

$$T = \begin{bmatrix} P_{00} & P_{01} & 0 & \cdots & 0 & 0 & 0 \\ H_2 & H_1 & H_0 & \cdots & 0 & 0 & 0 \\ 0 & H_2 & H_1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & H_1 & H_0 & 0 \\ 0 & 0 & 0 & \cdots & H_2 & H_1 & H_0 \\ 0 & 0 & 0 & \cdots & 0 & P_{N,N-1} & P_{N,N} \end{bmatrix}.$$

The details of the sub-matrices in T are given in Appendix B. This is the finite QBD process structure; c.f. [43]. □

There are many algorithms to obtain equilibrium probabilities for QBD processes; c.f., [43]. Linear level reduction and folding methods are popular. The computational complexity of various approaches is given next.

Proposition 10. The arithmetic complexity of our algorithm is $O(m^4n^4) + O(Nm^2n^2) + O(2mn(\log 2mn)^2[|\log\varepsilon| + 2mn])$, where ε is the round-off error for irrational roots.

Proof. We require $5(mn)^2$ operations to obtain the matrix D. Gaussian elimination matrix inversion requires $(mn)^3$ operations. Next, we obtain the z-transform in $5m(mn)$ operations. To invert the z-transform, we want the roots of the determinant of the matrix D. There is no technique to factorize a polynomial in the real field with polynomial complexity. Therefore, we use the algorithm in [45]. The arithmetic complexity of their algorithm is $O(2mn(\log 2mn)^2|\log\varepsilon| + (2mn)^2(\log 2mn)^2)$. After finding the roots, partial fraction expansion requires $8(mn)^4 - 7(mn)^2 + 2mn$ operations; see [46]. Evaluating the resulting expressions of (14) takes $2N(mn)^2 + (2mn-1)mn$ operations. To obtain the initial conditions (including the normalization), we invert an $m \times m$ matrix in m^3 operations. Summing these up proves the proposition. □

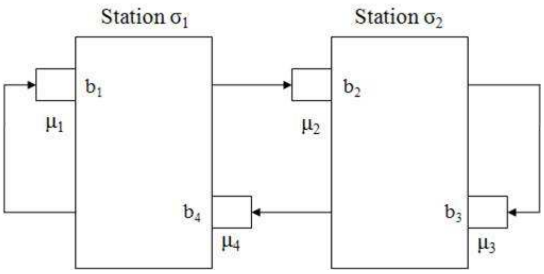

Fig. 3 The Lu-kumar network gives priority to buffers b_2 and b_4

The authors in [45] suggest that round off error $\varepsilon \leq 2^{-2mn}$ will provide adequate results. The arithmetic complexity of linear level reduction and folding methods are $2/3 N\,(mn)^3 + O(N(mn)^2)$ ([43]) and $O(m^3n^3 \log_2 N) + O(Nm^2n^2)$ ([51]), respectively. Our approach scales similarly to the folding methods as N grows large.

7 Extension to other network structures

It may be possible to identify and exploit similar properties in other networks. Here we consider the closed version of the Lu-Kumar network. It is in our class of networks with $m = n = 2$, but uses a different policy. Refer to Figure 3. For consistency with the standard labeling of this network, buffers b_1 and b_4 are served at station σ_1; buffers b_2 and b_3 are served at station σ_2. Customers require service from buffers b_1, b_2, b_3 and b_4 in order. They then return to buffer b_1. Service times are exponential with rates $\mu_1, \ldots, \mu_4$ for buffers $b_1, \ldots, b_4$, respectively. Normalize time so that $\mu_1 + \mu_2 + \mu_3 + \mu_4 = 1$. The servers give preemptive priority to buffers b_4 and b_2 at their stations. Use $s = (w_1(t), w_2(t), w_3(t), w_4(t))$, where $w_i(t)$ denotes the number of customers in buffer b_i, as the system state. Naturally, it must obey the simplex condition $w_1(t)+w_2(t)+w_3(t)+w_4(t) = N$. We will drop the dependence on time unless it is required.

It follows from a result in [40, 41] that many states are transient.

Corollary 2. At any time t in steady state, $w_2(t) \cdot w_4(t) = 0$.

The transition probability diagram is shown in Figure 4. Let Π_{w_1,w_2,w_3,w_4} denote the equilibrium probability of the state (w_1, w_2, w_3, w_4). For convenience, define $X_k(n) := \Pi_{k,0,N-k-n,n}$ and $Y_k(n) := \Pi_{N-k-n,n,k,0}$. Note that we do not have the QBD structure here.

Similar to the preceding, assume that we know the "initial condition" probabilities $X_0(0)$, $X_1(0)$, $\ldots$, $X_N(0)$. (These are also called $Y_N(0)$, $Y_{N-1}(0)$, $\ldots$, $Y_0(0)$.) Then, the probabilities on the upper left diagonal of Figure 4, $X_0(1), X_0(2), \ldots, X_0(N)$, can be obtained recursively as $X_0(k) = \mu_3^k X_0(0)/(\mu_3+$

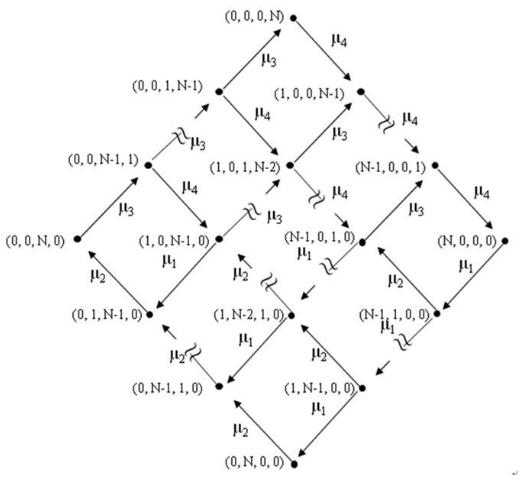

Fig. 4 The transition probability diagram

$\mu_4)^k$, for $k = 0, 1, \ldots, N$. Knowing this, we can similarly obtain the probabilities on the diagonal below it, $X_1(k)$, $k = 1, \ldots, N-1$. Repeating this process, one obtains all $X_k(n)$, $k = 0, 1, \ldots, N-1$, $n = 1, \ldots, N-k$ as an explicit function of the initial condition probabilities. Similarly, for $Y_k(n)$.

Proposition 11. As a function of the "initial condition" probabilities $X_0(0)$, $X_1(0)$, …, $X_N(0)$, we have

$$X_k(n) = \alpha^n \cdot X_k(0) + n \cdot \alpha^n \sum_{i=1}^{k} \frac{(n+2i-1)!}{i!(n+i)!} \cdot (\alpha\beta)^i \cdot X_{k-i}(0), \tag{15}$$

$$Y_k(n) = \gamma^n \cdot X_{N-k}(0) + n \cdot \gamma^n \sum_{i=1}^{k} \frac{(n+2i-1)!}{i!(n+i)!} \cdot (\gamma\varsigma)^i \cdot X_{N-k+i}(0), \tag{16}$$

where $0 \le k \le N-1$, $1 \le n \le N-k-1$, $\alpha := \mu_3/(\mu_3+\mu_4)$, $\beta := \mu_4/(\mu_3+\mu_4)$, $\gamma := \mu_1/(\mu_1+\mu_2)$ and $\varsigma := \mu_2/(\mu_1+\mu_2)$.

It remains only to determine the initial condition probabilities. Here, there is no remaining recursive structure in the GBEs that can be exploited. We must solve the $N+1$ GBEs written at the initial condition states (plus the normalization condition that probabilities sum to one) to obtain a solution. The resulting matrix has the Toeplitz structure; inversion is efficient.

Thus, rather than solving for all $(N+1)^2$ probabilities directly from the GBEs, we can focus on $N+1$ GBEs. Computation is thus significantly reduced. Explicit solutions, however, do not seem possible.

8 Concluding remarks

In this paper, we sought closed-form expressions for the steady state probabilities in a class of two-station closed reentrant queueing networks. Under LBFS, the state space for this class can be reduced to four dimensions. The matrix representation of the global balance equations possesses a recursive form. Standard z-transform techniques allowed us to obtain an explicit non-product form solution for the equilibrium probabilities with five customer classes or less. For six or more customer classes, the same procedure can be used to obtain a general form for the equilibrium probabilities. The transition probability matrix has a quasi birth-death (QBD) structure, so computation is efficient. The approach was extended to the closed Lu-Kumar network with some, but not complete, success. To our knowledge, these represent the first multiclass queueing networks found that admit an explicit non-product form solution for their equilibrium probabilities.

Appendix A. Global balance equations

Here, we obtain the matrix form of the global balance equations and introduce matrix notation omitted in the paper. Similarly to floor 0, we obtain the GBE matrix form for floor 1 as

$$\mathbf{Y}[1] = \begin{bmatrix} Y_{*0}[1] \\ Y_{*2}[1] \\ \vdots \\ Y_{*n-1}[1] \\ Y_{*n}[1] \end{bmatrix} = \begin{bmatrix} \Theta & A_1 & \Theta & \cdots & \Theta \\ \Theta & \Theta & A_2 & \ddots & \Theta \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \Theta & \Theta & \cdots & \Theta & A_{n-1} \\ \Theta & \Theta & \cdots & \Theta & \Theta \end{bmatrix} \mathbf{Y}[1] + \begin{bmatrix} \Theta & B_1 & \Theta & \cdots & \Theta \\ \Theta & \Theta & B_2 & \ddots & \Theta \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \Theta & \Theta & \cdots & \Theta & B_{n-1} \\ B_n & \Theta & \cdots & \Theta & \Theta \end{bmatrix} \mathbf{Y}[0]$$
$$+ \begin{bmatrix} \Theta \\ \Theta \\ \vdots \\ \Theta \\ C_1 \end{bmatrix} \mathbf{Y}[-1],$$

where

$$A_i = \tfrac{\mu_{m+i+1}}{\mu_{m+i}} I + \tfrac{1}{\mu_{m+i}} \begin{bmatrix} \mu_1 & 0 & \cdots & & 0 & 0 \\ -\mu_1 & \mu_2 & & & \ddots & 0 \\ 0 & & \ddots & & & \vdots \\ \vdots & \ddots & -\mu_{m-2} & & \mu_{m-1} & 0 \\ 0 & 0 & \cdots & & -\mu_{m-1} & \mu_m \end{bmatrix},$$

for $i = 1, \ldots, n-1$,

$$B_i = \tfrac{1}{\mu_{m+i}} \begin{bmatrix} 0 & 0 & \ldots & 0 & \mu_m \\ 0 & 0 & \ldots & 0 & 0 \\ \ldots & .. & \ldots & .. & \ldots \\ 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & \ldots & 0 & 0 \end{bmatrix}, \text{ for } i = 1, 2, \ldots, n-1,$$

$$B_n = \frac{\mu_{m+1}}{\mu_{m+n}} I + \frac{1}{\mu_{m+n}} \begin{bmatrix} \mu_1 & 0 & \cdots & 0 & 0 \\ -\mu_1 & \mu_2 & & \ddots & 0 \\ 0 & & \ddots & & \vdots \\ \vdots & \ddots & -\mu_{m-2} & \mu_{m-1} & 0 \\ 0 & 0 & \cdots & -\mu_{m-1} & \mu_m \end{bmatrix} \text{ and}$$

$$C_1 = \frac{1}{\mu_{m+n}} \begin{bmatrix} 0 & 0 & \dots & 0 & -\mu_m \\ 0 & 0 & \dots & 0 & 0 \\ \dots & .. & \dots & .. & \dots \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \end{bmatrix}.$$

All A_i, B_i and C_1's are $m \times m$ matrices. Note that assuming we know the initial conditions $\mathbf{Y}[-1]$, $\mathbf{Y}[1]$ can be written as $\mathbf{Y}[1] = S\mathbf{Y}[-1]$ for an appropriate matrix S. Similarly, the general equations for the GBEs can obtained as

$$\mathbf{Y}[k] = \begin{bmatrix} Y_{*0}[k] \\ Y_{*2}[k] \\ \dots \\ Y_{*n-1}[k] \\ Y_{*n}[k] \end{bmatrix} = \begin{bmatrix} \Theta & A_1 & \Theta & \dots & \Theta \\ \Theta & \Theta & A_2 & \Theta & \Theta \\ \dots & \dots & \dots & \dots & \dots \\ \Theta & \Theta & \dots & \Theta & A_{n-1} \\ \Theta & \Theta & \dots & \Theta & \Theta \end{bmatrix} \mathbf{Y}[k]$$

$$+ \begin{bmatrix} \Theta & B_1 & \dots & \Theta & \Theta \\ \Theta & \Theta & B_2 & \Theta & \Theta \\ \dots & \dots & \dots & \dots & \dots \\ \Theta & \Theta & \dots & \Theta & B_{n-1} \\ B_n & \Theta & \dots & \Theta & \Theta \end{bmatrix} \mathbf{Y}[k-1] + \begin{bmatrix} \Theta & \Theta & \dots & \Theta & \Theta \\ \Theta & \Theta & \Theta & \Theta & \Theta \\ \dots & \dots & \dots & \dots & \dots \\ \Theta & \Theta & \dots & \Theta & \Theta \\ C_1 & \Theta & \dots & \Theta & \Theta \end{bmatrix} \mathbf{Y}[k-2],$$

$2 \leq k \leq N-2$, where A_i, B_i and C_1 are as before.
Letting

$$A = \begin{bmatrix} \Theta & A_1 & \Theta & \dots & \Theta \\ \Theta & \Theta & A_2 & \Theta & \Theta \\ \dots & \dots & \dots & \dots & \dots \\ \Theta & \Theta & \dots & \Theta & A_{n-1} \\ \Theta & \Theta & \dots & \Theta & \Theta \end{bmatrix}, B = \begin{bmatrix} \Theta & B_1 & \dots & \Theta & \Theta \\ \Theta & \Theta & B_2 & \Theta & \Theta \\ \dots & \dots & \dots & \dots & \dots \\ \Theta & \Theta & \dots & \Theta & B_{n-1} \\ B_n & \Theta & \dots & \Theta & \Theta \end{bmatrix},$$

$$C = \begin{bmatrix} \Theta & \Theta & \dots & \Theta & \Theta \\ \Theta & \Theta & \dots & \Theta & \Theta \\ \dots & \dots & \dots & \dots & \dots \\ \Theta & \Theta & \dots & \Theta & \Theta \\ C_1 & \Theta & \dots & \Theta & \Theta \end{bmatrix}.$$

We obtain

$$(I - A)\mathbf{Y}[k] = B\mathbf{Y}[k-1] + C\mathbf{Y}[k-2], \text{ for } k = 2, \dots, N-1.$$

Appendix B. Matrix definitions for the transition probability matrix
Here, we give notation for the sub-matrices within the transition probability matrix. For convenience, we append virtual states to floors -1 and $N-1$, so that they too have mn states. Transitions to and from these states occur

with probability 0. We can express the transition probability matrix T as

$$T = \begin{bmatrix} P_{00} & P_{01} & 0 & \cdots & 0 & 0 & 0 \\ H_2 & H_1 & H_0 & \cdots & 0 & 0 & 0 \\ 0 & H_2 & H_1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & H_1 & H_0 & 0 \\ 0 & 0 & 0 & \cdots & H_2 & H_1 & H_0 \\ 0 & 0 & 0 & \cdots & 0 & P_{N-1,N-2} & P_{N-1,N-1} \end{bmatrix}, \text{ where}$$

$$P_{00} = \begin{bmatrix} P'_{00} & \Theta & \cdots & \Theta & \Theta \\ \Theta & \Theta & & \Theta & \Theta \\ \vdots & & \ddots & & \vdots \\ \Theta & \Theta & & \Theta & \Theta \\ \Theta & \Theta & \cdots & \Theta & \Theta \end{bmatrix} \text{ where } P'_{00} = \begin{bmatrix} 1-\mu_1 & \mu_1 & \dots & 0 & 0 \\ 0 & 1-\mu_2 & & 0 & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & & 0 & \mu_{m-1} \\ 0 & 0 & \cdots & 0 & 1-\mu_m \end{bmatrix}.$$

The matrix P'_{00} and Θs are $m \times m$ matrices.

$$P_{01} = \begin{bmatrix} P'_{01} & \Theta & \cdots & \Theta & \Theta \\ \Theta & \Theta & & \Theta & \Theta \\ \vdots & & \ddots & & \vdots \\ \Theta & \Theta & & \Theta & \Theta \\ \Theta & \Theta & \cdots & \Theta & \Theta \end{bmatrix} \text{ where } P'_{01} = \begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & & 0 & 0 \\ \mu_m & 0 & \cdots & 0 & 0 \end{bmatrix}.$$

$$H'_i = \begin{bmatrix} 1-\mu_1-\mu_{m+i} & \mu_1 & \dots & 0 & 0 \\ 0 & 1-\mu_2-\mu_{m+i} & & 0 & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & & 0 & 0 \\ 0 & 0 & \cdots & 0 & 1-\mu_m-\mu_{m+i} \end{bmatrix},$$

$$H_1 = \begin{bmatrix} H'_1 & \mu_{m+1}I & \dots & \Theta & \Theta \\ \Theta & H'_2 & & \Theta & \Theta \\ \vdots & & \ddots & & \vdots \\ \Theta & \Theta & & H'_{n-1} & \mu_{m+n-1}I \\ \Theta & \Theta & \cdots & \Theta & H'_n \end{bmatrix}, H_0 = \begin{bmatrix} P'_{01} & \Theta & \dots & \Theta & \Theta \\ \Theta & P'_{01} & & \Theta & \Theta \\ \vdots & & \ddots & & \vdots \\ \Theta & \Theta & & P'_{01} & \Theta \\ \Theta & \Theta & \cdots & \Theta & P'_{01} \end{bmatrix},$$

$$H_2 = \begin{bmatrix} \Theta & \Theta & \dots & \Theta & \Theta \\ \Theta & \Theta & & \Theta & \Theta \\ \vdots & & \ddots & & \vdots \\ \Theta & \Theta & & \Theta & \Theta \\ \mu_{m+n}I & \Theta & \cdots & \Theta & \Theta \end{bmatrix}, P_{N-1,N-1} = \begin{bmatrix} \Theta & P^1_{N-1,N-1} & \cdots & \Theta & \Theta \\ \Theta & \Theta & & \Theta & \Theta \\ \vdots & & \ddots & & \vdots \\ \Theta & \Theta & & \Theta & P^{n-1}_{N-1,N-1} \\ \Theta & \Theta & \cdots & \Theta & \Theta \end{bmatrix},$$

$$P^j_{N-1,N-1} = \begin{bmatrix} \mu_{m+j} & 0 & \dots & 0 & 0 \\ 0 & 0 & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}, \; P_{N-1,N-2} = \begin{bmatrix} \Theta & \Theta & \dots & \Theta & \Theta \\ \Theta & \Theta & & \Theta & \Theta \\ \vdots & & \ddots & & \vdots \\ \Theta & \Theta & & \Theta & \Theta \\ P'_{N-1,N-2} & \Theta & \cdots & \Theta & \Theta \end{bmatrix},$$

and $P'_{N-1,N-2} = \begin{bmatrix} \mu_{m+n} & 0 & \dots & 0 & 0 \\ 0 & 0 & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}$.

References

1. J. R. Jackson, "Jobshop-like queueing systems", *Manage. Sci.*, vol. 10, no. 1, pp. 131-142, 1963.
2. F. P. Kelly, *Reversibility and Stochastic Networks*, John Wiley & Sons, Chichester, 1979.
3. F. Baskett, K. M. Chandy, R. R. Muntz and F. G. Palacios, "Open, closed and mixed networks of queues with different classes of customers", *Journal of ACM*, vol. 22, no. 2, pp. 248-260, 1975.
4. S. Balsamo, "Product form queueing networks", In G. Haring, C. Lindemann, and M. Reiser eds., Performance Evaluation: Origins and Directions- LNCS 1769, 377-401. Springer, 2000.
5. J. M. Harrison and R. J. Williams, "Brownian models of feedforward queueing networks: Quasireversibility and product form solutions", *The Annals of Applied Probability*, vol. 2, no. 2, pp. 263-293, 1992.
6. W.H.Mather, J.Hasty, L.S. Tsimring and R.J.Williams, "Factorized time-dependent distributions for certain multiclass queueing networks and an application to enzymatic processing networks", *Queueing Syst.,* vol. 69, pp. 313-328, 2011.
7. W.Henderson and P.G.Taylor, "Product form in networks of queues with batch arrivals and batch services", *Queueing Syst.,* vol. 6, pp. 77-88, 1990.
8. J.Visschers, I. Adan, and G. Weiss, "A product form solution to a system with multi-type jobs and multi-type servers, Eurandom report 2011-002, 2011.
9. I. Adan and G.Weiss, "A loss system with skill based servers under assign to longest idle server policy",*Probability in Engineering and Informational Science*, To appear, 2011.
10. R.J. Boucherie and N.M. van Dijk, "On the arrival theorem for product form queueing networks with blocking", *Performance Evaluation*, vol. 29, pp 155-176, 1997.
11. B. Pittel, "Closed exponential networks of queues with saturation: the Jackson-type stationary distribution and its asymptotic analysis", *Math. Operat. Res.*, vol. 4, pp. 357-378, 1979.
12. N. Bayer and R.J. Boucherie, "On the structure of the space of geometric product-form models", *Probability in the Engineering and Informational Sciences*, vol. 16, pp 241-270, 2002.
13. S. Kumar and P.R. Kumar, "Performance bounds for queueing networks and scheduling policies", *IEEE Transactions on Automatic Control*, AC-39, pp. 1600-1611, 1994.
14. D. Bertsimas, I. C. Paschalidis, and J. N. Tsitsiklis, "Optimization of multiclass queueing networks: Polyhedral and nonlinear characterizations of achievable performance", *Annals of Applied Porbability*, vol. 4, pp. 43-75, 1994.
15. J. R. Morrison and P. R. Kumar, "New linear program performance bounds for queueing networks", *Journal of Optimization:Theory and Applications(Springer)*, A Volume in Honor of Professor Y.C. Ho, vol. 100, no. 3, pp. 575-597, 1999.
16. J. R. Morrison and P. R. Kumar, "New linear program performance bounds for closed queueing networks", *Discrete Event Dynamics systems*, vol. 11, pp. 291-317, 2001.

17. J. R. Morrison and P. R. Kumar, "Computational performance bounds for Markov chains with applications", *IEEE Transactions on Automatic Control*, vol. 53, no. 5, pp. 1306-1311, 2008.
18. M. C. Russell, J. Fraser, S. Rizzo and M. H. Veatch, "Comparing LP bounds for queueing networks", *IEEE Transaction on Automatic Control*, vol. 54, no. 11, pp. 2703-2707, 2009.
19. S. Rizzo and M. H. Veatch, "Performance bounds and differential cost approximations for queueing networks", Working paper, 2008.
20. W. J. Hopp and M. L. Spearman, *Factory physics: Foundations of manufacturing management*, 2nd ed. New York:McGraw-Hill, 2001.
21. J. R. Morrison and D. P. Martin, "Practical extensions to cycle time approximations for the G/G/m queue with applications", *IEEE Transactions on Automation Science and Engineering*, vol. 4, no. 4, pp. 523-532, 2007.
22. W. Whitt, "The queueing network analyzer", *Bell Syst. Tech. J*, vol 62, no.9, pp. 2279-2815 1983.
23. L. M. Wein, "Scheduling semiconductor wafer fabrication", *IEEE Transaction on Semiconductor Manufacturing*, vol. 1, no. 3, pp. 115-130, 1988.
24. P.R. Kumar and S.P. Meyn, "Stability of queueing networks and scheduling policies", *IEEE Transaction on Automatic Control*, vol. 40, no. 2, pp. 251-260, 1995.
25. J.G. Dai and G.Weiss, "Stability and instability of fluid models for reentrant lines", *Mathematics of Operations research*, vol. 21, no. 1, pp. 115-134, 1996.
26. J.J. Hasenbein, "Necessary conditions for global stability of multiclass queueing networks", *Operations Research Letters*, vol. 21, pp. 87-94, 1997.
27. J.M. Harrison and L.M. Wein, "Scheduling networks of queues: Heavy traffic analysis of a simple open network", *Queueing Systems*, vol. 5, pp. 265-280, 1989.
28. S. Kumar and P.R. Kumar, "Queueing network models in the design and analysis of semiconductor wafer fabs", *IEEE Transactions on Robotics and Automation*, vol. 17, no. 5, 2001.
29. E. Koenigsberg, "Twenty five years of cyclic queues and closed queue netoworks: A review", *Journal of Operational Research Society*, vol. 33, pp. 605-619, 1982.
30. J. Taylor and R. R. P. Jackson, "An application of the birth and death process to the provision of spare machines", *Operational research quarterly*, vol 5, pp. 96-108, 1954.
31. E. Koenigberg, "An operations research approach to mine haulage", *Min. Congr. J.* Pt 1, pp. 64-68, 1959.
32. T. M. Roeder, N. Govind and Lee W. Schruben, "A queueing network approximation of semiconductor automated material handling system: how much information do we really need?", *Proceedings of the 36th Winter simulation*, pp. 1956 -1961, 2004.
33. M. L. Spearman and M. A. Zazanis, "Push and pull production systems: Issues and comparisons", *Operations Research*, vol.40, no.3, pp. 521-532, 1992.
34. L. Kleinrock, *Queueing Systems, Computer Applications*, vol. II, John & Wiley Sons Inc, 1976.
35. J. R. Morrison and P. R. Kumar, "A counterexample to a conjecture of Harrison and Wein", *Proceedings of the 37th IEEE Conference on Decision and Control*, 1998.
36. Seunghwan Jung and James R. Morrison, "Closed form solutions for the equilibrium probability distribution in the closed Lu-Kumar network under two buffer priority policies", *Proceeding of the 8th IEEE International Conference on Control and Automation*, pp. 1488-1495, 2010.
37. Woo-sung Kim and James R. Morrison, "On equilibrium probabilities in a class of two station closed queueing network", *Proceeding of the International Conference on Control, Automation and System (ICCAS 2010)*, pp. 237-242, 2010.
38. J. M. Harrison and L. M. Wein, "Scheduling networks of queues:Heavy traffic analysis of a two-station closed network", *Operations Research*, vol. 38, no 6, pp. 1052-1064, 1990.
39. S. Kumar, "Two-server closed networks in heavy traffic: Diffusion limits and asymptotic optimality", *The Annals of Applied Probability*, vol. 10, no. 3, pp. 930-961, 2000.
40. J. G. Dai and J. H. Vande Vate, "The stability of two-station multitype fluid networks", *Operations research*, vol. 48, no. 5, pp. 721-744, 2000.
41. J. R. Morrison and P. R. Kumar, "On the guaranteed throughput and efficiency of closed reentrant lines", *Queueing Syst.,* vol. 28, pp. 33-54, 1998.

42. J. G. Dai, J. J. Hasenbein and J. H. Vande Vate, “Stability and instability of a two-station queueing network”, *The Annals of Applied Probability*, vol.14, no.1, pp. 326-377, 2004.
43. G. Latouche and V. Ramaswami, *Introduction to Matrix Analytic Methods in Stochastic Modeling*, ASA-SIAM series on Statistics and Applied Probability, 1999.
44. S. Lippman, “Applying a new device in the optimization of exponential queueing systems”, *Operations research*, vol. 23, pp. 687-710, 1975.
45. Kim M.H and Sutherland. S, “Polynomial root-finding algorithms and branched covers”, *SIAM J. Comput.,*, vol. 23, pp. 477-486, 1994.
46. J.F. Mahoney and B.D. Sivazlian, “Partial fractions expansion: a review of computational methodology and efficiency”, *Journal of Computational and Applied Mathematics*, vol. 9, pp. 247-269, 1983.
47. H. Jin, J. Ou and P. R. Kumar, “The throughput of irreducible closed markovian queueing networks: functional bounds, asymptotic loss, efficiency and the Harrison-Wein conjectures”, *Mathematics of Operations Research*, vol. 22, pp. 886-920, 1997.
48. S. Ross, *Stochastic Processes*, John & Wiley Sons, 1983.
49. D. C. Lay, *Linear algebra and its application*, second ed., Addison & Wesley, 1997.
50. J. A. Cadzow, *Discrete time systems*, Englewood Cliffs,NJ: Prentice-Hall, 1973.
51. J. Ye aand S. Q. Li, “Folding algorithm: a computational method for finite QBD processes with level-dependent transitions”, *IEEE Trans. On Comm.,*, vol. 42, no. 2, pp. 625-639, 1994.