QUEUES, STORES, AND TABLEAUX

MOEZ DRAIEF * ** AND

JEAN MAIRESSE,* *** *Université Paris 7*

NEIL O'CONNELL,**** *The University of Warwick*

Abstract

Consider the single-server queue with an infinite buffer and a first-in–first-out discipline, either of type M/M/1 or Geom/Geom/1. Denote by $\mathcal{A}$ the arrival process and by s the services. Assume the stability condition to be satisfied. Denote by $\mathcal{D}$ the departure process in equilibrium and by r the time spent by the customers at the very back of the queue. We prove that $(\mathcal{D}, r)$ has the same law as $(\mathcal{A}, s)$, which is an extension of the classical Burke theorem. In fact, r can be viewed as the sequence of departures from a dual *storage* model. This duality between the two models also appears when studying the transient behaviour of a tandem by means of the Robinson–Schensted–Knuth algorithm: the first and last rows of the resulting semistandard Young tableau are respectively the last instant of departure from the queue and the total number of departures from the store.

Keywords: Single-server queue; storage model; Burke's theorem; noncolliding random walk; tandem of queues; Robinson–Schensted–Knuth algorithm

2000 Mathematics Subject Classification: Primary 60K25; 60K35
Secondary 60C05; 82B41

1. Introduction

The main purpose of this paper is to clarify the interplay between two models of queueing theory. The first model is the classical single-server queue with an infinite buffer and a first-in–first-out (FIFO) discipline. The second, less common but very natural, model can be described as a queue operating in slotted time with batch arrivals and services. It was studied, for instance, in [4]. The literature on discrete-time queues is less complete than is its continuous-time counterpart. The interest in discrete-time queues is, however, gaining ground, due to their relevance in ATM-like communication systems [6], [36]. In this paper, to clearly distinguish between the two models, we choose to describe the second one as a *storage model*, using different terminology.

We first prove that the two models are linked in a very strong way. We set up an abstract model with an ordered pair of input variables $(\mathcal{A}, s)$ and an ordered pair of output variables $(\mathcal{D}, r) = \Phi(\mathcal{A}, s) = (\Phi_1(\mathcal{A}, s), \Phi_2(\mathcal{A}, s))$. On the one hand, the queueing model corresponds to $\mathcal{D} = \Phi_1(\mathcal{A}, s)$, with $\mathcal{A}$ and s being the arrivals and services, respectively, and $\mathcal{D}$ the

departures. On the other hand, the storage model corresponds to $r = \Phi_2(\mathcal{A}, s)$, with s and $\mathcal{A}$ being the supplies (arrivals) and requests (services), respectively, and r the departures. The interpretation of either r for the queue or $\mathcal{D}$ for the store is much less natural.

We then assume that the random variables driving the dynamics are either exponentially or geometrically distributed, and we consider the models in equilibrium (under the stability condition). In this situation, it is well known that a Burke-type theorem holds: the departures and the arrivals have the same law [7], [30], [4]. This can be considered as one of the cornerstones of queueing theory; see, for instance, the books [5], [15], and [31] for discussions and related material. Here we prove a 'joint' version of the result, namely that $(\mathcal{D}, r)$ has the same law as $(\mathcal{A}, s)$. This joint Burke theorem is new, although it is similar in spirit to the results proved in [25] and [18] for a variant: queues with *unused services*. Also, in the geometric case, we use an original method of proof based on the reversibility of a symmetrized version of the workload process (instead of the queue length process). As in [25] and [18], the joint Burke theorem can be used to obtain a representation theorem for the joint law of two independent random walks with exponential or geometric jumps conditioned on never colliding.

A second facet of the duality between queues and stores appears when considering the Robinson–Schensted–Knuth (RSK) algorithm. The RSK algorithm is an important object in representation and symmetric functions theory, and it has been known for some time that it has connections with queueing theory, although these connections have remained somewhat mysterious until now. In this paper, we elucidate these connections completely via the above duality. The results we obtain, apart from being interesting in their own right, can be used as a starting point for the application of symmetric functions theory (and the associated analysis) to obtain interchangeability as well as asymptotic results for models that are constructed by putting the above queues and stores in series. These results are also applicable in the context of the corresponding interacting particle systems, where there is much interest in obtaining exactly solvable microscopic models for Burger-type equations.

Let us now detail the results. Consider K tandem queues (or stores): the departures from a queue (or a store) are the arrivals (or supplies) at the next one. Initially, the network is empty but for an infinite number of customers in the first queue (or an infinite supply at the first store).

The particle systems associated with the above tandems (of queues or stores) are described in Section 5.2. One model is original and worth putting forward: the *bus stop* model, which is the zero-range model associated with stores in tandem.

Here we assume only that the variables driving the dynamics are $\mathbb{N}$-valued. Building on ideas developed in [3] and [22], we can study the transient evolution as follows. Consider the family of random variables (RVs) $(u(i, j), 1 \leq i \leq N, 1 \leq j \leq K)$, where $u(i, j)$ is the service of the ith customer in queue j, or the request at time slot i at store $K + 1 - j$. Apply the RSK algorithm [17], [32], [34] to this family and let P be the resulting semistandard Young tableau (here we do not consider the recording tableau Q). Let $\lambda_1 \geq \cdots \geq \lambda_K \geq 0$ be the lengths of the successive rows of P. Classically [17], we have $\lambda_1 = \max_{\pi \in \Pi} \sum_{(i, j) \in \pi} u(i, j)$, where Π is the set of paths from $(1, 1)$ to (N, K) in $\mathbb{N}^2$ that are increasing and consist of adjacent points. Moreover, it is well known in queueing theory [11], [21], [37], [35] that $\max_{\pi \in \Pi} \sum_{(i, j) \in \pi} u(i, j)$ is the instant of departure of customer N from queue K, which we denote by D. Combining the two results gives the accepted result that $\lambda_1 = D$. Here we complete the picture by proving that $\lambda_K = R$, where R is the total number of departures from the last store in the tandem up to time slot N. Again, this identity is proved in two steps by showing that $\lambda_K = \min_{\pi \in \widetilde{\Pi}} \sum_{(i, j) \in \pi} u(i, j) = R$, where $\widetilde{\Pi}$ is a different set of paths in the lattice. To summarize, we obtain on the same Young tableau the total departures for the two tandem models.

A short version of the paper, with proofs omitted, appears in the proceedings of the conference *Discrete Random Walks 2003* [10].

2. Notation

We work on a probability space $(\Omega, \mathcal{F}, \mathrm{P})$. The indicator function of an event $E \in \mathcal{F}$ is denoted by $\mathbf{1}_E$. We use the symbol '$\overset{\mathrm{D}}{=}$' to denote the equality in distribution of random variables. Depending on the context, $|A|$ is the cardinality of the set A or the length of the word A. We set $\mathbb{N}^* = \mathbb{N} \setminus \{0\}$, $\mathbb{R}_+^* = \mathbb{R}_+ \setminus \{0\}$, and $x^+ = x \vee 0 = \max(x, 0)$. We use the convention that $\sum_{i=j}^k u_i = 0$ when $j > k$.

Below, a *point process* is a stochastic simple point process on $\mathbb{R}$ with an infinite number of positive and negative points. We identify a point process $\mathcal{A}$ with the random ordered sequence of its points, i.e. $\mathcal{A} = (A_n)_{n \in \mathbb{Z}}$ with $A_n < A_{n+1}$ for all n. Observe that the numbering of the points is defined up to a translation in the indices. For any interval I, we define the *counting random variable* $\mathcal{A}(I) = \sum_{n \in \mathbb{Z}} \mathbf{1}_{\{A_n \in I\}}$.

A *marked point process* is a couple $(\mathcal{A}, c) = (A_n, c_n)_{n \in \mathbb{Z}}$ in which $\mathcal{A} = (A_n)_{n \in \mathbb{Z}}$ is a point process and $c = (c_n)_{n \in \mathbb{Z}}$ is a sequence of RVs valued in some state space. We say that the *mark* c_n is *associated* with the point A_n of the point process. For more details on point processes, see [9].

Given a point process $\mathcal{A} = (A_n)_{n \in \mathbb{Z}}$, the *reversed* point process $\mathcal{R}(\mathcal{A})$ is the point process obtained by reversing the direction of time, i.e. $\mathcal{R}(\mathcal{A}) = (-A_{-n})_{n \in \mathbb{Z}}$. Given a marked point process $(\mathcal{A}, c) = (A_n, c_n)_{n \in \mathbb{Z}}$, the reversed marked point process is $\mathcal{R}(\mathcal{A}, c) = (-A_{-n}, c_{-n})_{n \in \mathbb{Z}}$.

Given a càdlàg, i.e. right-continuous and left-limited, random process $Y = (Y(t))_{t \in \mathbb{R}}$ valued in $\mathbb{R}$, we define the reversed process $\mathcal{R} \circ Y = (\mathcal{R} \circ Y(t))_{t \in \mathbb{R}}$ as the càdlàg modification of the process $(Y(-t))_{t \in \mathbb{R}}$. Denote by $\mathcal{N}_+(Y)$ and $\mathcal{N}_-(Y)$ the point processes (with possibly finite numbers of points) corresponding respectively to the positive and negative jumps of Y; that is, for any interval I of $\mathbb{R}$,

$$\mathcal{N}_+(Y)(I) = \int_I \mathbf{1}_{\{Y(u) > Y(u-)\}} \, \mathrm{d}u, \qquad \mathcal{N}_-(Y)(I) = \int_I \mathbf{1}_{\{Y(u) < Y(u-)\}} \, \mathrm{d}u.$$

3. The model

Let $\mathcal{A} = (A_n)_{n \in \mathbb{Z}}$ be a point process and assume that $A_0 \le 0 < A_1$. We define the sequence of $\mathbb{R}_+^*$-valued RVs $a = (a_n)_{n \in \mathbb{Z}}$ by $a_n = A_{n+1} - A_n$. Let $s = (s_n)_{n \in \mathbb{Z}}$ be another sequence of $\mathbb{R}_+^*$-valued RVs. The marked point process $(\mathcal{A}, s)$ is the *input* of the model.

Define the sequence of RVs $\mathcal{D} = (D_n)_{n \in \mathbb{Z}}$ by

$$D_n = \sup_{k \le n} \left[A_k + \sum_{i=k}^n s_i \right]. \tag{3.1}$$

A priori, the D_n are valued in $\mathbb{R} \cup \{\infty\}$. Assume from now on that $(\mathcal{A}, s)$ is such that the D_n are almost surely finite. They satisfy the recursive equations

$$D_{n+1} = \max(D_n, A_{n+1}) + s_{n+1}. \tag{3.2}$$

Set $d_n = D_{n+1} - D_n$ and define an additional sequence of $\mathbb{R}_+^*$-valued RVs $r = (r_n)_{n \in \mathbb{Z}}$ by

$$r_n = \min(D_n, A_{n+1}) - A_n. \tag{3.3}$$

The marked point process $(\mathcal{D}, r)$ is the *output* of the model. By summing (3.2) and (3.3), we obtain the following interesting relation between the input and output variables:

$$r_n + d_n = a_n + s_{n+1}. \tag{3.4}$$

For the purposes of our future analysis, it is convenient to define the following auxiliary variables. Let $w = (w_n)_{n \in \mathbb{Z}}$ be the sequence of $\mathbb{R}_+^*$-valued RVs defined by

$$w_n = D_n - s_n - A_n = \sup_{k \leq n-1} \left[\sum_{i=k}^{n-1} (s_i - a_i) \right]^+. \tag{3.5}$$

These RVs satisfy the recursive equations

$$w_{n+1} = [w_n + s_n - a_n]^+. \tag{3.6}$$

Using the variables w_n, we can give alternative definitions of D_n and r_n:

$$D_n = \left[w_l + A_l + \sum_{i=l}^n s_i \right] \vee \max_{l < k \leq n} \left[A_k + \sum_{i=k}^n s_i \right] \quad \text{for all } l \leq n,$$

$$r_n = \min(w_n + s_n, a_n) = s_n + w_n - w_{n+1}. \tag{3.7}$$

Finally, we define the $\mathbb{N}$-valued càdlàg random process $Q = (Q(t))_{t \in \mathbb{R}}$ by

$$Q(t) = \sum_{n \in \mathbb{Z}} \mathbf{1}_{\{A_n \leq t < D_n\}}. \tag{3.8}$$

Lemma 3.1. *We have $\mathcal{N}_+(Q) = \mathcal{A}$ and $\mathcal{N}_-(Q) = \mathcal{D}$. Furthermore, $D_{-Q(0)} \leq 0 < D_{-Q(0)+1}$.*

Proof. Recall that $Q(0) = \sum_{n \in \mathbb{Z}} \mathbf{1}_{\{A_n \leq 0 < D_n\}}$ and that $A_0 \leq 0 < A_1$. Hence, $Q(0) = |k|$, where $k = \sup\{n \in \mathbb{Z}_- : D_n \leq 0\}$. The result follows.

We now interpret the variables defined above in two different contexts: those of a queueing model and a storage model.

3.1. The single-server queue

A bi-infinite string of customers is served at a queueing facility with a single server. Each customer is characterized by an instant of arrival in the queue and a service demand. Customers are served upon arrival in the queue and in their order of arrival. Since there is a single server, a customer may have to wait in a buffer before the beginning of its service. In Kendall's classical nomenclature, our model is a $\cdot / \cdot / 1 / \infty / \text{FIFO}$ queue.

The customers are numbered, using integers, according to their order of arrival in the queue (customer 1 being the first to arrive strictly after instant 0). Let A_n be the instant of arrival of customer n and s_n its service time. Then the variables defined in (3.1), (3.3), (3.5), and (3.8) have the following interpretations:

- D_n is the instant of departure of customer n from the queue, after completion of its service;

- w_n is the waiting time of customer n in the buffer between its arrival and the beginning of its service;
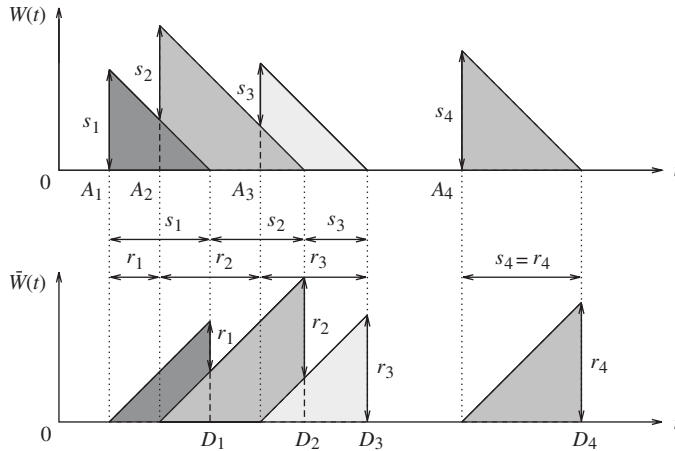
FIGURE 1: The dual variables $(s_n)_n$ and $(r_n)_n$.

- $Q(t)$ is the number of customers in the queue at instant t (either in the buffer or in service) and $Q = (Q(t))_t$ is called the *queue length process*;

- r_n is the time spent by customer n at the very back of the queue.

The variables $(r_n)_n$ are less often considered in queueing theory than the others, although they have been; see [27]. They should be viewed as being dual to the services $(s_n)_n$, as illustrated in Figure 1. In the upper part of the figure, we have represented the *workload process* $(W(t))_t$, where $W(t)$ is the waiting time of a virtual customer arriving at instant t. The dual of the workload process, $(\bar{W}(t))_t$, is shown in the lower part. (See (4.1), below, for the formal definition of $(W(t))_t$.)

3.2. The storage model

Some product P is supplied, sold, and stocked in a store in the following way. Events occur at integer-valued epochs, called *slots*. At each slot, an amount of P is supplied and an amount of P is demanded by potential buyers. The rule is to meet all the demand, if possible. The demand of a given slot that is not met is lost. The supply of a given slot that is not sold is not lost but is stocked for future use.

Let s_n be the amount of P supplied at slot $n + 1$, and let a_n be the amount of P asked for at the same slot. The variables in (3.5) and (3.3) can be interpreted in this context:

- w_n is the level of the stock at the end of slot n and evolves according to (3.6);

- r_n is the demand met at slot $n + 1$ (see (3.7)), that is, the amount of P departing at slot $n + 1$.

The variables $(D_n)_n$ and the process Q do not have natural interpretations in this model.

The evolution of the store is summarized in Figure 2. Our indexing notation (e.g. the use of s_n, a_n, and r_n at time slot $n + 1$) may seem unusual. This system was chosen for clarity of notation in Section 5.

It is important to remark that, while the equations describing the *single-server queue* and the *storage model* are exactly the same, the same variables do not all make sense in each model. The important variables are the ones corresponding to the departures from the system. The
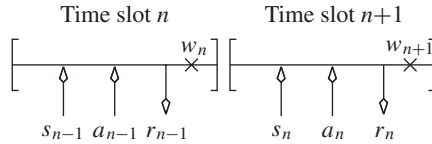
FIGURE 2: The storage model.

departures are coded in the variables $(D_n)_n$ for the single-server queue and in the variables $(r_n)_n$ for the storage model. On the other hand, the interpretations of the variables $(r_n)_n$ in the single-server queue or the variables $(D_n)_n$ in the storage model are not so immediate. To summarize, the output variables $(D_n)_n$ and $(r_n)_n$ are equally relevant, but in different modelling contexts.

Observe that it would be possible to describe the above storage model in 'queueing' terminology (as a queue with slotted time, batch arrivals $(s_n)_n$, and batch services $(a_n)_n$); this is the description used in [4], for instance. We have deliberately avoided it to clearly separate the models of Sections 3.1 and 3.2 and therefore minimize the possibility of confusion.

4. Equilibrium behaviour around Burke's output theorem

We now consider the exponential or geometric version of our model in equilibrium. We prove a Burke-type result: $(\mathcal{D}, r) \overset{\mathrm{D}}{=} (\mathcal{A}, s)$. The relevant result for the sequence of departures in one of the two models will then follow by suppressing one of the two variables. In the exponential case, our proof uses the queue length process Q, following [30]. Our method of proof in the geometric case is original and based on the zigzag process, a symmetrized version of the workload process. A discussion of the literature and of the advances made here is carried out in Section 4.3.

4.1. Output theorem in the exponential case

Let $\mathcal{A}$ be a homogeneous Poisson process of intensity $\lambda \in \mathbb{R}_+^*$. We set $\mathcal{A} = (A_n)_{n \in \mathbb{Z}}$ with $A_0 < 0 < A_1$. Recall that $a_n = A_{n+1} - A_n$. Then $(a_n)_{n \geq 1}$ is a sequence of independent, identically distributed (i.i.d.) RVs with an exponential distribution of parameter λ. Let $s = (s_n)_{n \in \mathbb{Z}}$ be a sequence of i.i.d. random variables, independent of $\mathcal{A}$, with an exponential distribution of parameter μ. We assume that $\lambda < \mu$. Consider the marked point process $(\mathcal{A}, s)$ as being the input of the model of Section 3. The sequence $(w_n)_n$ is a random walk valued in $\mathbb{R}_+$ with a barrier at 0. Under the *stability* condition $\lambda < \mu$, this random walk has a negative drift, implying that the random variables D_n defined in (3.1) are almost surely finite.

We now present our first theorem. Figure 2 provides a visual illustration of the result. The principle of the proof is the same as those of Theorems 3 and 4 of [30]. However, the adaptation is not completely immediate. Some care is necessary in dealing with the indices, to make sure that the variable r_n is the mark of the point D_n.

Theorem 4.1. *In the exponential case, the marked point process $(\mathcal{D}, r)$ has the same law as the marked point process $(\mathcal{A}, s)$.*

Proof. Knowing Q, one can recover the input of the model: $(A_n, s_n)_n = \varphi(Q)$. We are going to prove that

$$\varphi \circ \mathcal{R}(Q) = (-D_{-n+Q(0)+1}, r_{-n+Q(0)+1})_n = \mathcal{R}(\mathcal{D}, r).$$

Let $\mathcal{T}_n$ be the operator mapping a point process to its nth point, with the convention that $\mathcal{T}_0(\cdot) \leq 0 < \mathcal{T}_1(\cdot)$. For instance, $A_n = \mathcal{T}_n(\mathcal{A}) = \mathcal{T}_n \circ \mathcal{N}_+(Q)$. Taking into account the shift in the indices and using Lemma 3.1, we find that $\mathcal{T}_n(\mathcal{D}) = \mathcal{T}_n \circ \mathcal{N}_-(Q) = D_{n-Q(0)}$. We thus have

$$s_n = D_n - \max(A_n, D_{n-1})$$
$$= \mathcal{T}_{n+Q(0)} \circ \mathcal{N}_-(Q) - \max(\mathcal{T}_n \circ \mathcal{N}_+(Q), \mathcal{T}_{n+Q(0)-1} \circ \mathcal{N}_-(Q)).$$

Now define an operator $\mathcal{S}_n$ such that $s_n = \mathcal{S}_n(Q)$, and let us apply the operators $\mathcal{T}_n \circ \mathcal{N}_+$ and $\mathcal{S}_n$ to the reversed process $\mathcal{R}(Q)$. We first have

$$\mathcal{T}_n \circ \mathcal{N}_+ \circ \mathcal{R}(Q) = -D_{-n-Q(0)+1}$$

and, furthermore, $\mathcal{T}_{n+\mathcal{R}(Q)(0)-1} \circ \mathcal{N}_- \circ \mathcal{R}(Q) = -A_{-n-Q(0)+2}$. Thus,

$$\mathcal{S}_n \circ \mathcal{R}(Q) = -A_{-n-Q(0)+1} - \max(-D_{-n-Q(0)+1}, -A_{-n-Q(0)+2})$$
$$= -A_{-n+Q(0)+1} + \min(D_{-n+Q(0)+1}, A_{-n+Q(0)+2})$$
$$= r_{-n+Q(0)+1}.$$

Hence, we find that $\varphi \circ \mathcal{R}(Q) = \mathcal{R}(\mathcal{D}, r)$.

The process Q is a stationary birth-and-death process, and hence reversible: $\mathcal{R}(Q) \overset{\mathrm{D}}{=} Q$. This implies that $(\mathcal{A}, s) = \varphi(Q) \overset{\mathrm{D}}{=} \varphi \circ \mathcal{R}(Q) = \mathcal{R}(\mathcal{D}, r)$. Therefore, $\mathcal{R}(\mathcal{D}, r)$ is a homogeneous Poisson process marked with an i.i.d. sequence of RVs. This means that $\mathcal{R}(\mathcal{D}, r) \overset{\mathrm{D}}{=} (\mathcal{D}, r)$, which concludes the proof.

Corollary 4.1. *In the queueing model, the departure process $\mathcal{D}$ is a Poisson process of intensity λ. In the storage model, the sequence $(r_n)_n$ of the amounts of product P departing at successive slots is a sequence of i.i.d. exponential RVs of parameter μ.*

4.2. Output theorem in the geometric case

Let $\mathcal{A}$ be a Bernoulli point process of parameter $p \in (0, 1)$; that is, all the points are integer valued, there is a point at a given integer with probability p, and the occurrences of points at different integers are independent. As before, set $\mathcal{A} = (A_n)_n$ with $A_0 \leq 0 < A_1$ and $a_n = A_{n+1} - A_n$. Then the sequence $(a_n)_{n \geq 1}$ is a sequence of i.i.d. geometric RVs with parameter p (meaning that, for all $k \in \mathbb{N}^*$, $\mathrm{P}\{a_1 = k\} = (1-p)^{k-1}p$). Let $(s_n)_n$ be a sequence of i.i.d. geometric RVs, independent of $\mathcal{A}$, with parameter $q \in (0, 1)$. We assume that $p < q$ (the stability condition). Let the marked point process $(\mathcal{A}, s)$ be the input of the model of Section 3. As in Section 4.1, the model is stable and the output $(\mathcal{D}, r)$ is a marked point process. We prove a Burke-type theorem using an original approach based on a symmetrization of the workload process.

Define the (random) càdlàg maps $f_n, g_n : \Omega \times \mathbb{R} \to \mathbb{R}$ by

$$f_n(t) = \begin{cases} 0 & \text{if } t < A_n, \\ D_n - t & \text{if } A_n \leq t \leq D_n, \\ 0 & \text{if } t > D_n, \end{cases} \qquad g_n(t) = \begin{cases} 0 & \text{if } t < A_n, \\ t - A_n & \text{if } A_n \leq t \leq D_n, \\ 0 & \text{if } t > D_n. \end{cases}$$

Set

$$W(t) = \bigvee_{n \in \mathbb{Z}} f_n(t), \qquad \bar{W}(t) = \bigvee_{n \in \mathbb{Z}} g_n(t). \tag{4.1}$$
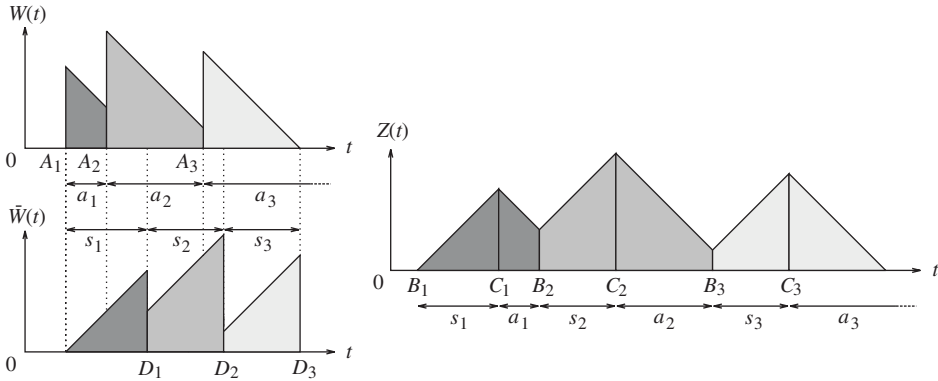
FIGURE 3: The workload process, $W(t)$, the dual workload process, $\bar{W}(t)$, and the zigzag process, $Z(t)$.

Since $D_n = A_n + w_n + s_n$, the workload just before the arrival of the nth customer is given by $W(A_n^-) = w_n$. For the queueing model, $W(t)$ is the total amount of service remaining to be done by the server at instant t, and $W = (W(t))_t$ is called the *workload* process. The process $\bar{W} = (\bar{W}(t))_t$ is the dual of the process W; see Figure 1 or Figure 3 for an illustration.

Looking at W and $\bar{W}$ together is a way to symmetrize the workload process. Consider the ordered pair of processes $X = (W, \bar{W})$, and define

$$\mathcal{R}(X) = (\mathcal{R}(\bar{W}), \mathcal{R}(W))$$

(observe that the first element of the pair is now $\mathcal{R}(\bar{W})$, as opposed to W for the pair X). Obviously, if we know X then we can recover the input of the model: $(\mathcal{A}, s) = \Psi(X)$. It is also clear (see Figure 1 or Figure 3) that $\mathcal{R}(\mathcal{D}, r) = \Psi \circ \mathcal{R}(X)$.

The following result is the crux of the argument in proving a Burke-type theorem.

Lemma 4.1. *The ordered pair of the workload process and its dual is reversible, that is,* $(W, \bar{W}) \stackrel{\mathrm{D}}{=} (\mathcal{R}(\bar{W}), \mathcal{R}(W))$.

Proof. Clearly, W and $\bar{W}$ are (time-)stationary and ergodic processes. An *idle period* is a maximal nonempty time interval during which $W(t) = 0$ (or, equivalently, $\bar{W}(t) = 0$). The *busy periods* are the time intervals between the idle periods. Hence, the time domain, $\mathbb{R}$, is partitioned into an alternating sequence of busy and idle periods.

Let $(i_n)_{n \in \mathbb{Z}}$ and $(b_n)_{n \in \mathbb{Z}}$ respectively be the lengths of the successive idle periods and busy periods. Clearly, the sequences $(i_n)_n$ and $(b_n)_n$ are independent and i.i.d. Moreover, it follows from the memoryless property of the geometric distribution that i_n is geometrically distributed with parameter p. Therefore, to prove the reversibility of $(W, \bar{W})$, it is enough to prove the reversibility within a single busy period.

To do so, it is easiest to introduce an auxiliary process. Consider a busy period that is assumed to consist, for simplicity, of the k customers numbered from 1 to k. For $n \in \{1, \ldots, k\}$, let

$$C_n = B_n + s_n, \qquad B_{n+1} = C_n + a_n.$$

This defines $(B_n)_n$ and $(C_n)_n$ up to a translation. Define

$$Z(B_1) = 0, \qquad Z(t) = \begin{cases} Z(B_n) + (t - B_n) & \text{for } t \in [B_n, C_n), \\ [Z(C_n) - (t - C_n)]^+ & \text{for } t \in [C_n, B_{n+1}). \end{cases}$$

On an interval of type $[B_n, C_n)$ we have $dZ/dt = 1$, and on an interval of type $[C_n, B_{n+1})$ we have $dZ/dt = (-1)\,\mathbf{1}_{\{Z>0\}}$.

Now, in a busy period, the process Z is in one-to-one correspondence with $(W, \bar{W})$, and the time-reversed process $\mathcal{R}(Z)$ is in one-to-one correspondence with $(\mathcal{R}(\bar{W}), \mathcal{R}(W))$. This is illustrated in Figure 3, where we display a trajectory of $(W, \bar{W})$ and the corresponding trajectory of Z. The process Z is obtained by gluing together the slices of $\bar{W}$ and W alternately. Hence, proving the reversibility of $(W, \bar{W})$ in a busy period is equivalent to proving the reversibility of Z. We call Z the *zigzag process*. This process was considered (with a different motivation) in [12] and, in particular, was there proved to be reversible (although not Markovian). We recall the argument for completeness. Identify Z with the finite sequence of the lengths of its intervals of increase and decrease. Let $(l_1, \ldots, l_{2k}) \in (\mathbb{N}^*)^{2k}$, with $\sum_i l_{2i} = \sum_i l_{2i+1} = L$, be a possible value for Z. (The total length of the periods of increase has to be equal to the total length of the periods of decrease.) Recall that the variables $(s_n)_n$ and $(a_n)_n$ are geometrically distributed with respective parameters q and p. We find that

$$
\begin{aligned}
\mathrm{P}\{Z = (l_1, \ldots, l_{2k})\} &= (1-p)^{l_1-1}p(1-q)^{l_2-1}q \cdots (1-p)^{l_{2k-1}-1}p(1-q)^{l_{2k}-1} \\
&= (1-p)^{L-k}p^k(1-q)^{L-k}q^{k-1}.
\end{aligned}
$$

Hence, given L and k, all the different possible trajectories for Z are equally probable. This implies that the same is true for the time-reversed process $\mathcal{R}(Z)$, which completes the proof.

We are now ready to prove the following result.

Theorem 4.2. *In the geometric case, the marked point process $(\mathcal{D}, r)$ has the same law as the marked point process $(\mathcal{A}, s)$.*

Proof. The proof is similar to that of Theorem 4.1, with $(W, \bar{W})$ now playing the role of Q. Recall that $(\mathcal{A}, s) = \Psi(W, \bar{W})$ and $\mathcal{R}(\mathcal{D}, r) = \Psi \circ \mathcal{R}(W, \bar{W})$. According to Lemma 4.1, we have

$$
(\mathcal{A}, s) = \Psi(W, \bar{W}) \overset{\mathrm{D}}{=} \Psi \circ \mathcal{R}(W, \bar{W}) = \mathcal{R}(\mathcal{D}, r).
$$

We conclude that $(\mathcal{D}, r) \overset{\mathrm{D}}{=} \mathcal{R}(\mathcal{A}, s) \overset{\mathrm{D}}{=} (\mathcal{A}, s)$, where the final equivalence comes from the fact that $(\mathcal{A}, s)$ is a Bernoulli process with i.i.d. marks and hence is reversible. □

4.3. Comments on the different proofs of Burke's theorem

Reflecting on the above, there are three different ways to prove Burke's theorem, be it in the exponential or geometric case. The first way is by using analytic methods, the second is by using the reversibility of the queue length process Q, and the third is by using the reversibility of the zigzag process Z.

Burke's original proof is for the exponential model and uses analytic methods [7]. For the geometric model, an analytic proof was given by Bedekar and Azizoğlu [4]. For the exponential model, the idea of using the reversibility of Q to obtain the result is due to Reich [30]. This proof, which has become a cornerstone of queueing theory, has been extended to various contexts and has given birth to the concept of *product-form networks* [5], [15]. Reich's proof does not translate directly to the geometric model. Of course, $(Q(n))_{n \in \mathbb{Z}}$ is a reversible birth-and-death Markov chain. However, a difficulty arises: it is not possible to reconstruct $\mathcal{A}$ and s from Q. Indeed, on the event $\{Q(n-1) = Q(n) > 0\}$, two cases may occur: there is either no departure and no arrival at instant n, or one departure and one arrival. It is not possible to distinguish between them knowing only Q. One feasible solution is to add an auxiliary sequence that contains the information lacking, but the details become quite intricate. This programme has

been carried out in [14] (see also [18, Theorem 4.1]) for a variant, the geometric model with unused services. The above idea of using the zigzag process to prove Burke's theorem is original. Observe that the zigzag process is also clearly reversible in the exponential model. Hence, the proof used for Theorem 4.2 could also be used to prove Theorem 4.1.

In the original references [7], [30] and in all the classical textbooks presenting the result [5], [15], [31], the statement proved is $\mathcal{A} \overset{\mathrm{D}}{=} \mathcal{D}$ ('Poisson input, Poisson output'). In [4], or [14] for the variant with unused services, the result proved is $s \overset{\mathrm{D}}{=} r$. The complete version, $(\mathcal{A}, s) \overset{\mathrm{D}}{=} (\mathcal{D}, r)$, first appeared in [25, Theorem 3] for the exponential model with unused services. This was extended to the geometric model with unused services in [18]. Brownian analogues were proved in [13] and [24].

4.4. Noncolliding random walks

Following the lines of thought in [25], [18], and [12], it is possible to use Theorems 4.1 and 4.2 to obtain representation results for noncolliding random walks. We start by stating a preliminary lemma.

Lemma 4.2. *The Lindley equation (3.6) admits a backward analogue:*

$$w_n = (w_{n+1} + r_n - d_{n-1})^+. \tag{4.2}$$

Proof. Recall (3.7): $r_n = s_n + w_n - w_{n+1}$. Using (3.4), we obtain

$$d_n = a_n + s_{n+1} - r_n = a_n - w_n + w_{n+1} - s_n + s_{n+1}. \tag{4.3}$$

Therefore,

$$\begin{aligned}
w_{n+1} + r_n - d_{n-1} &= w_{n+1} + s_n + w_n - w_{n+1} - a_{n-1} - w_n + w_{n-1} - s_n + s_{n-1} \\
&= w_{n-1} + s_{n-1} - a_{n-1}.
\end{aligned}$$

By taking the maximum with 0 on both sides of the equality and using (3.6), we obtain

$$(w_{n+1} + r_n - d_{n-1})^+ = w_n.$$

Proposition 4.1. *Let the sequences of RVs $(a_n)_{n \in \mathbb{N}^*}$ and $(s_n)_{n \in \mathbb{N}^*}$ be as in Section 4.1 or 4.2. The conditional law of $(\sum_{i=1}^n a_i, \sum_{i=1}^n s_i)$, given that $\{\sum_{i=1}^k a_i \geq \sum_{i=1}^{k+1} s_i \text{ for all } k \geq 0\}$ occurs, is the same as the unconditional law of*

$$\left(\max_{1 \leq j \leq n} \left[\sum_{i=1}^j a_i + \sum_{i=j+1}^{n+1} s_i \right], \min_{1 \leq j \leq n} \left[\sum_{i=2}^j s_i + \sum_{i=j+1}^n a_i \right] \right).$$

Proof. Denote by $v_n = w_n + s_n = w_{n+1} + r_n$ the *sojourn time* of customer n in the queue. By developing (4.2), we obtain

$$v_n = w_{n+1} + r_n = \sup_{k \geq n} \left[\sum_{i=n}^k r_i - \sum_{i=n}^{k-1} d_i \right]. \tag{4.4}$$

By summing the equalities (4.3), on the event $\{v_1 = 0\}$ we obtain

$$\sum_{i=1}^{n} d_i = \sum_{i=1}^{n} a_i + w_{n+1} + s_{n+1}$$

$$= \sum_{i=1}^{n} a_i + \max_{1 \leq j \leq n}\left[\sum_{i=j+1}^{n+1} s_i - \sum_{i=j+1}^{n} a_i\right]$$

$$= \max_{1 \leq j \leq n}\left[\sum_{i=1}^{j} a_i + \sum_{i=j+1}^{n+1} s_i\right],$$

and, by summing (3.7), on the event $\{v_1 = 0\}$ we obtain

$$\sum_{i=1}^{n} r_i = \sum_{i=2}^{n} s_i - w_{n+1}$$

$$= \sum_{i=2}^{n} s_i - \max_{1 \leq j \leq n}\left[\sum_{i=j+1}^{n} s_i - \sum_{i=j+1}^{n} a_i\right]$$

$$= \min_{1 \leq j \leq n}\left[\sum_{i=2}^{j} s_i + \sum_{i=j+1}^{n} a_i\right].$$

To summarize, we have

$$\sum_{i=1}^{n} d_i = \max_{1 \leq j \leq n}\left[\sum_{i=1}^{j} a_i + \sum_{i=j+1}^{n+1} s_i\right], \qquad \sum_{i=1}^{n} r_i = \min_{1 \leq j \leq n}\left[\sum_{i=2}^{j} s_i + \sum_{i=j+1}^{n} a_i\right]. \qquad (4.5)$$

By applying Theorem 4.1 or 4.2, we find that the distribution of $(\sum_{i=1}^{n} a_i, \sum_{i=1}^{n} s_i)$, given that $\{\sum_{i=1}^{k} a_i \geq \sum_{i=1}^{k+1} s_i$ for all $k \geq 0\}$ occurs, is the same as the law of $(\sum_{i=1}^{n} d_i, \sum_{i=1}^{n} r_i)$, given that $\{\sum_{i=1}^{k} d_i \geq \sum_{i=1}^{k+1} r_i$ for all $k \geq 0\}$ occurs. By (4.4), the event

$$\left\{\sum_{i=1}^{k} d_i \geq \sum_{i=1}^{k+1} r_i \text{ for all } k \geq 0\right\}$$

is equal to $\{v_1 = 0\}$. Using (4.5), we see that the distribution of $(\sum_{i=1}^{n} d_i, \sum_{i=1}^{n} r_i)$, given that $\{v_1 = 0\}$ occurs, is the same as the law of

$$\left(\max_{1 \leq j \leq n}\left[\sum_{i=1}^{j} a_i + \sum_{i=j+1}^{n+1} s_i\right], \min_{1 \leq j \leq n}\left[\sum_{i=2}^{j} s_i + \sum_{i=j+1}^{n} a_i\right]\right),$$

given that $\{v_1 = 0\}$ occurs. This last conditional law is the same as the unconditional law of

$$\left(\max_{1 \leq j \leq n}\left[\sum_{i=1}^{j} a_i + \sum_{i=j+1}^{n+1} s_i\right], \min_{1 \leq j \leq n}\left[\sum_{i=2}^{j} s_i + \sum_{i=j+1}^{n} a_i\right]\right).$$

Indeed, the variables $(a_i)_{i \geq 1}$ and $(s_i)_{i \geq 2}$ are independent of the event $\{v_1 = 0\}$.

By combining these arguments, we obtain the desired result.

It is possible to extend Proposition 4.1 to the limiting case $E[a_1] = E[s_1]$, and also to higher dimensions, by adapting the methods of [25], [18], and [12] to the present setting.

Customer/Time slot	Queue 2/(Virtual) Store 1	(Virtual) Queue 1/Store 2
1	$u(1, 2) = s_1$	$u(1, 1) = a_0$
2	$u(2, 2) = s_2$	$u(2, 1) = a_1$
$\vdots$	$\vdots$	$\vdots$
n	$u(n, 2) = s_n$	$u(n, 1) = a_{n-1}$

5. Transient behaviour and the RSK representation

In Sections 5.1 and 5.3, below, the model described and the results obtained are of a purely combinatorial nature. In particular, the results are true pathwise, without any probabilistic assumption. In Section 5.4, these results are then applied to geometric queues or stores: here the pathwise arguments are coupled with probabilistic ones.

5.1. The saturated tandem

Here we consider another aspect of the dynamics of queues and stores, namely the transient evolution for the model when initially empty. More precisely, we consider the model of Section 3 under the assumption that $w_0 = A_0 = s_0 = 0$ (which implies that $D_0 = r_0 = 0$) and focus on the customers, or time slots, from number 1 onwards.

It is convenient to describe such a model from a different perspective. We first do so for the queue. View the arrivals as being the departures from a *virtual queue* having at instant 0 an infinite number of customers (labelled in $\mathbb{N}^*$) in its buffer. The service time of customer n in the virtual queue is a_{n-1}. We describe this as a *saturated tandem of two queues*. Now consider the store. View the supplies as being the departures from a virtual store having an infinite stock at the end of time slot 0. In the virtual store, the request (or departure) at time slot n is s_n. This is a *saturated tandem of two stores*.

We now want to fit these two descriptions together. Denote the virtual queue/store as queue/store 1 and the other one as queue/store 2. For convenience, set $u(n, 1) = a_{n-1}$ and $u(n, 2) = s_n$ for all $n \geq 1$. The saturated tandem is completely specified by the family $(u(n, i), \ n \in \mathbb{N}^*, \ i = 1, 2)$ of input variables. These variables are the services or requests when the model is seen as a tandem of queues or, respectively, stores. Table 1 gives the input variables in the saturated tandems of two queues/stores.

A couple of observations are in order. Observe that $u(n, i)$ is the service of customer n in queue i and the request at time slot n at store $3 - i$. In other words, the elements (queues or stores) associated with a given sequence $(u(\cdot, i), \ i = 1, 2)$ are crossed in reverse order in the queueing/storage tandem. This is illustrated in Figure 4 (set $K = 2$). Observe also that there is a shift in the time slots for the storage model: the departure from store 1 at time slot n is the supply of store 2 at time slot $n + 1$ (contrast this with the situation for the queues). This is consistent with a model in which we view a time slot as being decomposable into three consecutive stages: first the supply arrives, then the request is made, and finally the departure occurs (see Figure 2).

The above setting is naturally extended to define the *saturated tandem of K queues/stores*. Such a model is entirely defined by a family of $\mathbb{N}$-valued variables

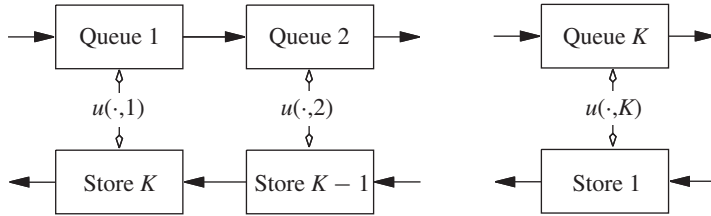$$(u(i, j), \ i \in \mathbb{N}^*, \ j \in \{1, \ldots, K\}).$$

FIGURE 4: Queues and stores are gone through in opposite orders.

(In Section 3, the variables were valued in $\mathbb{N}^*$. This restriction is not necessary here.) For the queueing model,

(i) at instant 0, queue 1 has an infinite number of customers (labelled in $\mathbb{N}^*$) in its buffer, and the other queues are empty;

(ii) $u(i, j)$ is the service of customer i at queue j; and

(iii) the instant of departure of customer n from queue i is the instant of arrival of customer n in queue $i + 1$.

For the storage model,

(i) at the end of time slot 0, store 1 has an infinite stock and the other stores have zero stock;

(ii) $u(i, K + 1 - j)$ is the request at time slot i at store j; and

(iii) the departure at time slot n from store i is the supply at time slot $n + 1$ at store $i + 1$.

The models are depicted in Figure 4.

5.2. The interacting particle representation

It is helpful to describe the above tandem models as interacting particle models (see, for instance, [19]). Commonly, interacting particle systems are defined as continuous-time Markov processes. In contrast, the description below is pathwise and the time is discrete.

Classically, the tandem queues may be viewed as the following *zero-range process*. The set of sites is $\{1, \ldots, K\}$ and a site is occupied by a number of particles in $\mathbb{N} \cup \{\infty\}$. At each site, the first particle (if any) has a clock whose value is decremented by 1 at each time slot. When it reaches 0, the particle jumps to the site on its right. The left-most site has an infinite number of particles. This is illustrated in Figure 5. The variables $(u(i, j))_i$ correspond to the initial values of the clocks of the successive first particles at site j.

We now describe the tandem stores as an original zero-range process: the *bus stop* model. At the beginning of each time slot, a bus arrives at each site. The size of the bus varies from site to site and from time slot to time slot. During each time slot, the particles take places in the bus until it is full and are transported to the next site on the left. The right-most site has an infinite number of particles. This mechanism is illustrated in Figure 6. Here the variables $(u(i, j))_i$ correspond to the successive bus sizes at site j.

Alternatively, the two models above can be viewed as *exclusion processes*. The set of sites is embedded in $\mathbb{N}$ and each site now contains either 0 or 1 particle. The tandem queues are associated with the classical totally asymmetric exclusion process. For the tandem stores, the exclusion process is as follows. Particles take jumps of integer length to the right but are

constrained not to pass over other particles. In one time step, particles jump in order from left to right. This model is similar, but not identical, to the exclusion processes considered in [28] and [33].

The mapping between the zero-range process and the exclusion process, both for queues and stores, is illustrated in Figure 7.
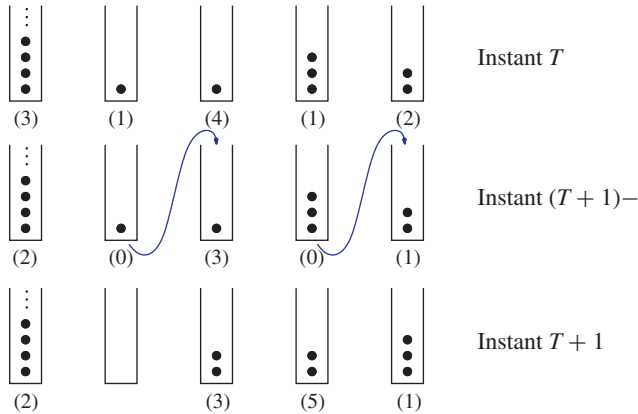
FIGURE 5: The zero-range process for tandem queues. The clock reading of the first particle at each site is shown in parentheses.
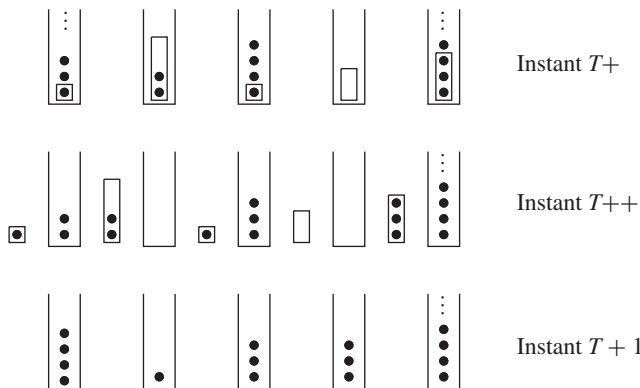
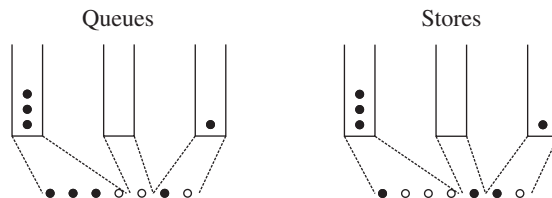FIGURE 6: The zero-range process for tandem stores. The boxes represent the buses.

FIGURE 7: Exclusion processes for queues and stores. The black and white circles represent occupied and empty sites, respectively.
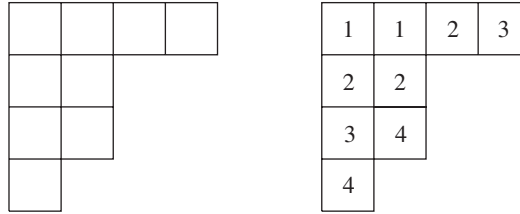
FIGURE 8: A Ferrers diagram (*left*) and a semistandard Young tableau (*right*) of shape $(4, 2, 2, 1) \vdash 9$.

5.3. The Robinson–Schensted–Knuth representation

A *partition* of $n \in \mathbb{N}^*$ is a sequence of integers $\lambda = (\lambda_1, \dots, \lambda_k)$ such that $\lambda_1 \geq \cdots \geq \lambda_k \geq 0$ and $\lambda_1 + \cdots + \lambda_k = n$, denoted by $\lambda \vdash n$. By convention, we identify partitions having the same nonzero components. The *(Ferrers) diagram* of $(\lambda_1, \dots, \lambda_k) \vdash n$ is a collection of n boxes arranged in left-justified rows, the ith row (counting from the top) consisting of λ_i boxes. A *(semistandard Young) tableau* on the alphabet $\{1, \dots, \ell\}$ is a diagram in which each box is filled in by a label from $\{1, \dots, \ell\}$ in such a way that the entries are weakly increasing from left to right along the rows and strictly increasing down the columns. The *shape* of a diagram or tableau is the underlying partition. A *standard tableau* of size n is a tableau of shape $\lambda \vdash n$ whose entries are from $\{1, \dots, n\}$ and are distinct. In Figure 8 we display a diagram on the left and a tableau of the same shape on the right.

The Robinson–Schensted–Knuth row-insertion algorithm (RSK algorithm) takes a tableau T and an $i \in \mathbb{N}^*$ and constructs a new tableau denoted by $T \leftarrow i$. The tableau $T \leftarrow i$ has one more box than T and is constructed as follows. If i is at least as large as the labels of the first (upper) row of T, a box labelled i is added to the end of the first row of T and the procedure halts. Otherwise, the left-most entry in the first row that is strictly larger than i is found, the corresponding box is relabelled by i, and the same procedure is recursively applied to the second row and to the 'bumped' label. By convention, an empty row has label 0. With this convention, the above procedure halts.

Consider a word $v = v_1 \cdots v_n$ over the alphabet $\{1, \dots, k\}$. The tableau associated with v is by definition

$$P = (\cdots ((T_0 \leftarrow v_1) \leftarrow v_2) \cdots \leftarrow v_n),$$

where T_0 is the empty tableau. Observe that P has at most k nonempty rows. Classically, the length of the first (and longest) row of P is equal to the longest weakly increasing subsequence in v.

Remark 5.1. While building the tableau P, it is possible to build another tableau of the same shape in which the entries, labelled from 1 to $|v|$, record the order in which the boxes are added. This *recording tableau* is a standard tableau of size $|v|$. By doing this, one defines a bijection between words of $\{1, \dots, k\}^n$ and ordered pairs of tableaux of the same shape, the first being semistandard over the alphabet $\{1, \dots, k\}$ and the second being standard and of size n. This bijection is often referred to as the Robinson–Schensted–Knuth correspondence. Here, we do not need this result and we do not consider the recording tableau.

Consider a family

$$U = (u(i, j), \ (i, j) \in \{1, \dots, N\} \times \{1, \dots, K\})$$

of variables valued in $\mathbb{N}$. (Here we do not make any assumption on the variables, which may or may not be random.) We associate with U the word $w(U)$ over the alphabet $\{1, \ldots, K\}$, defined by

$$w(U) = w_1 \cdots w_N \quad \text{with } w_i = \underbrace{1 \cdots 1}_{u(i,1)} \underbrace{2 \cdots 2}_{u(i,2)} \cdots \underbrace{K \cdots K}_{u(i,K)}. \tag{5.1}$$

Set

$$M = |w(U)| = \sum_{i=1}^{N} \sum_{j=1}^{K} u(i, j).$$

For $i = 1, \ldots, K$, define

$$x_i(n) = |\{j \leq n \colon w(U)_j = i\}| \quad \text{for all } n \leq M.$$

Given two maps $x, y \colon \{1, \ldots, N\} \to \mathbb{N}$, define the maps $x \triangledown y, x \triangle y \colon \{1, \ldots, N\} \to \mathbb{N}$ as follows: for all $n \leq N$,

$$x \triangledown y(n) = \max_{0 \leq m \leq n} [x(m) + y(n) - y(m)],$$

$$x \triangle y(n) = \min_{0 \leq m \leq n} [x(m) + y(n) - y(m)].$$

Denote by $P(U)$ the tableau obtained from $w(U)$ by applying the RSK algorithm and let $(\lambda_1, \ldots, \lambda_K)$ be its shape. Then

$$\lambda_1 = x_1 \triangledown x_2 \triangledown \cdots \triangledown x_K(M), \qquad \lambda_K = x_K \triangle \cdots \triangle x_2 \triangle x_1(M), \tag{5.2}$$

where the (nonassociative) operations '$\triangledown$' and '$\triangle$' are performed from left to right. The expression for λ_1 follows from the fact that λ_1 is the longest weakly increasing subsequence in $w(U)$. The expression for λ_K is proved in [22, Theorem 3.1]. In fact, the result of [22] is more general: there exists a min–max-type operator Γ_K such that $(\lambda_1, \ldots, \lambda_K) = \Gamma_K(x_1, \ldots, x_K)$.

A *lattice path* is a sequence $\pi = ((i_1, j_1), \ldots, (i_l, j_l))$ with $(i_k, j_k) \in \mathbb{Z}^2$. The *steps* of π are the differences

$$(i_{k+1} - i_k, j_{k+1} - j_k), \qquad k = 1, \ldots, l - 1.$$

Let Π be the set of lattice paths from $(1, 1)$ to (N, K) with steps of the type $(1, 0)$ or $(0, 1)$. All the paths in Π have the same number of nodes, $N + K - 1$. Let $\widetilde{\Pi}$ be the set of lattice paths from one of the points in

$$\{(1 + i, K - i), i = 0, \ldots, K - 1\}$$

to one of the points in

$$\{(N - j, 1 + j), j = 0, \ldots, N - 1\}$$

with steps of the type $(1 + i, -i)$, $i \geq 0$. Observe that all the paths in $\widetilde{\Pi}$ also have the same number of nodes, $(N - K + 1)^+$. In particular, $\widetilde{\Pi}$ is empty when $N < K$. An example of a path from each of the two sets is provided in Figure 9.

When interpreting the following result, recall that queues and stores are arranged in opposite orders in the tandems (see Figure 4). It is also fruitful to compare the statements of Proposition 4.1 and Theorem 5.1 for $K = 2$.

FIGURE 9: A path from Π (*solid line*) and a path from $\widetilde{\Pi}$ (*dashed line*).

Theorem 5.1. *Consider a saturated tandem of K queues/stores with variables $u(i, j)$, $i \in \mathbb{N}^*$, $j \in \{1, \dots, K\}$, as defined in Section 5.1. Fix an $N \in \mathbb{N}^*$. Let D be the instant of departure of customer N from queue K and let R be the cumulative departures from store K over the time slots 1 to N. Set*

$$U = (u(i, j), 1 \le i \le N, 1 \le j \le K)$$

and define $w(U)$ as in (5.1). Let $(\lambda_1, \dots, \lambda_K)$ be the shape of the tableau associated with $w(U)$. Then

$$\lambda_1 = \max_{\pi \in \Pi} \left[\sum_{(i,j) \in \pi} u(i, j) \right] = D, \tag{5.3}$$

$$\lambda_K = \min_{\pi \in \widetilde{\Pi}} \left[\sum_{(i,j) \in \pi} u(i, j) \right] = R. \tag{5.4}$$

Before proving Theorem 5.1, we state and prove a preliminary result.

Lemma 5.1. *We have*

$$x_K \triangle \cdots \triangle x_2 \triangle x_1(M) = \min_{\pi \in \widetilde{\Pi}} \left[\sum_{(i,j) \in \pi} u(i, j) \right].$$

Proof. By definition,

$x_K \triangle \cdots \triangle x_1(M)$
$$= \min_{1 \le m_K < \cdots < m_1 = M} [x_K(m_K) + x_{K-1}(m_{K-1}) - x_{K-1}(m_K) + \cdots + x_1(m_1) - x_1(m_2)].$$

Let $(m_K^*, \dots, m_2^*)$ be a sequence of integers realizing the minimum in the above equation. Consider the set of integers

$$I = \left\{ \sum_{i=1}^{k} |w_i| + \sum_{j=1}^{K-1} u(k + 1, j), \ k \in \{0, \dots, K-1\} \right\}.$$

$$\overbrace{}^{w_{i_K}}$$

$$w(U): \quad |111222\cdots\cdots|111122\cdots\cdots(K-1)(K-1)(K-1)KKK\cdots K|111\cdots\cdots KKK|$$

$$\underset{n_K}{\uparrow} \qquad\qquad\qquad\qquad \underset{m_K^*}{\uparrow}$$

FIGURE 10: Illustration of the proof of Lemma 5.1.

Without loss of generality, we may assume that $m_K^* \in I$. Indeed, assume that $m_K^* \notin I$ and let n_K be the smallest integer such that $n_K \in I$ with $n_K > m_K^*$. (See Figure 10.) Then we clearly have $x_K(n_K) = x_K(m_K^*)$ and

$$
\begin{aligned}
x_K &\triangle \cdots \triangle x_1(M) - x_K(n_K) \\
&\leq \min_{n_K < m_{K-1} < \cdots < m_2} [x_{K-1}(m_{K-1}) - x_{K-1}(n_K) + \cdots + x_1(M) - x_1(m_2)] \\
&\leq \min_{m_K^* < m_{K-1} < \cdots < m_2} [x_{K-1}(m_{K-1}) - x_{K-1}(m_K^*) + \cdots + x_1(M) - x_1(m_2)] \\
&= x_K \triangle \cdots \triangle x_1(M) - x_K(m_K^*).
\end{aligned}
$$

Thus, we can assume that $m_K^* \in I$. Let i_K be such that

$$m_K^* = \sum_{i=1}^{i_K-1} |w_i| + \sum_{j=1}^{K-1} u(i_K, j).$$

Then $x_K(m_K^*) = \sum_{j=1}^{i_K-1} u(j, K)$ and, by the same type of argument, we can assume that

$$m_{K-1}^* \in \left\{ \sum_{i=1}^{k} |w_i| + \sum_{j=1}^{K-2} u(k+1, j), \ k \in \{i_K, \ldots, K-1\} \right\},$$

without loss of generality.

Since $m_K^* \in I$, we have $x_K(m_K^*) = \sum_{j=1}^{i_K} u(K, j)$ for some i_K. Then, again by the same type of argument, we can assume that

$$m_{K-1}^* \in \left\{ \sum_{i=1}^{k} |w_i| + \sum_{j=1}^{K-2} u(k, j), \ k \in \{i_K + 1, \ldots, K-1\} \right\},$$

without loss of generality. This implies that

$$x_{K-1}(m_{K-1}) - x_{K-1}(m_K) = \sum_{j=i_K+1}^{i_{K-1}-1} u(K-1, j) \quad \text{for some } i_{K-1}.$$

By repeating the argument, we obtain the desired result.

Proof of Theorem 5.1. The result for λ_1 is essentially due to Schensted. The left-hand equality in (5.3) follows from (5.2) and the right-hand equality appears, for instance, in [21], [37], and [11]. The left-hand equality in (5.4) follows from Lemma 5.1, whence

$$\lambda_K = x_K \triangle \cdots \triangle x_2 \triangle x_1(M) = \min_{\pi \in \widetilde{\Pi}} \left[\sum_{(i,j) \in \pi} u(i, j) \right].$$

It remains to prove the right-hand equality in (5.4). Let us denote by $r(i-1, j)$ the amount departing at time slot i from store j and by $w(i, j)$ the quantity remaining at the end of time slot i in store j. According to (3.7), we have

$$r(n, k) = r(n-1, k-1) + w(n, k) - w(n+1, k). \tag{5.5}$$

By applying (3.6) to our tandem model, we obtain

$$w(n+1, k) = [w(n, k) + r(n-1, k-1) - u(n, K+1-k)]^+$$

$$= \max_{1 \le m \le n} \left[\sum_{i=m}^{n-1} r(i, k-1) - \sum_{i=m+1}^{n} u(i, K+1-k) \right].$$

Using (5.5), for $n \ge K$ we have

$$r(n, K) = r(n-1, K-1) + w(n, K) - w(n+1, K)$$
$$= r(n-2, K-2) + w(n-1, K-1) - w(n, K-1) + w(n, K) - w(n+1, K)$$
$$= r(n-K+1, 1) + \sum_{i=0}^{K-2} [w(n-i, K-i) - w(n-i+1, K-i)]$$
$$= u(n-K+1, K) + \sum_{i=0}^{K-1} [w(n-i, K-i) - w(n-i+1, K-i)].$$

Hence,

$$R = \sum_{n=1}^{N} r(n, K) = \sum_{n=K}^{N} r(n, K)$$

$$= \sum_{n=K}^{N} u(n-K+1, K) + \sum_{n=K}^{N} \sum_{i=0}^{K-1} [w(n-i, K-i) - w(n-i+1, K-i)]$$

$$= \sum_{n=1}^{N-K+1} u(n, K) + \sum_{i=0}^{K-1} \sum_{n=K}^{N} [w(n-i, K-i) - w(n-i+1, K-i)]$$

$$= \sum_{n=1}^{N-K+1} u(n, K) + \sum_{i=0}^{K-1} [w(K-i, K-i) - w(N-i+1, K-i)].$$

Now recall that $w(k, k) = 0$ for all $k \ge 1$. This gives

$$R = \sum_{n=1}^{N-K+1} u(n, K) + \sum_{i=1}^{K} w(i, i) - \sum_{i=1}^{K} w(N-K+i+1, i)$$

$$= \sum_{n=1}^{N-K+1} u(n, K) - \sum_{i=1}^{K} w(N-K+i+1, i). \tag{5.6}$$

Moreover, if we let $A = w(N + 1, K) + w(N, K - 1)$, we have

$$
A = \max_{1 \le i_1 \le N} \left[\sum_{n=i_1}^{N-1} r(n, K-1) - \sum_{n=i_1+1}^{N} u(n, 1) \right] + w(N, K-1)
$$

$$
= \max_{1 \le i_1 \le N} \left[\sum_{n=i_1}^{N-1} [r(n-1, K-2) + w(n, K-1) - w(n+1, K-1)] - \sum_{n=i_1+1}^{N} u(n, 1) \right]
$$
$$
+ w(N, K-1)
$$

$$
= \max_{1 \le i_1 \le N} \left[\sum_{n=i_1-1}^{N-2} r(n, K-2) + w(i_1, K-1) - w(N, K-1) - \sum_{n=i_1+1}^{N} u(n, 1) \right]
$$
$$
+ w(N, K-1)
$$

$$
= \max_{1 \le i_1 \le N} \left[\sum_{n=i_1-1}^{N-2} r(n, K-2) + w(i_1, K-1) - \sum_{n=i_1+1}^{N} u(n, 1) \right]
$$

$$
= \max_{1 \le i_1 \le N} \left[\sum_{n=i_1-1}^{N-2} r(n, K-2) + \max_{1 \le i_2 \le i_1-1} \left[\sum_{n=i_2}^{i_1-2} r(n, K-2) - \sum_{n=i_2+1}^{i_1-1} u(n, 2) \right] \right.
$$
$$
\left. - \sum_{n=i_1+1}^{N} u(n, 1) \right]
$$

$$
= \max_{1 \le i_2 < i_1 \le N} \left[\sum_{n=i_2}^{N-2} r(n, K-2) - \sum_{n=i_2+1}^{i_1-1} u(n, 2) - \sum_{n=i_1+1}^{N} u(n, 1) \right].
$$

Applying this repeatedly yields

$$
\sum_{i=1}^{K} w(N - K + i + 1, i)
$$
$$
= \max_{1 \le i_{K-1} < \cdots < i_1 \le N} \left[\sum_{n=i_{K-1}}^{N-K+1} u(n, K) - \sum_{n=i_{K-1}+1}^{i_{K-2}-1} u(n, K-1) + \cdots + \sum_{n=i_1+1}^{N} u(n, 1) \right].
$$
$$
\tag{5.7}
$$

By combining (5.6) and (5.7), we obtain

$$
R = \min_{1 \le i_{K-1} < \cdots < i_1 \le N} \left[\sum_{n=1}^{i_{K-1}-1} u(n, K) + \sum_{n=i_{K-1}+1}^{i_{K-2}-1} u(n, K-1) + \cdots + \sum_{n=i_1+1}^{N} u(n, 1) \right]
$$
$$
= \min_{\pi \in \widetilde{\Pi}} \left[\sum_{(i,j) \in \pi} u(i, j) \right].
$$

Let us comment on Theorem 5.1. First, it is interesting to observe that the six quantities appearing in (5.3) and (5.4) all come from the same model. Now consider (5.3) and (5.4) separately. The first is classical and has led to several interesting developments for queues in series. A by-product of the newly proved identity (5.4) is therefore that it paves the way to obtaining analogous results for stores in tandem. We set out this programme in more detail in the next section.

5.4. Stochastic models, interchangeability, and hydrodynamic limits

If T is a tableau over the alphabet $\{1, \ldots, k\}$, we write $x^T = \prod_{i=1}^{k} x_i^{\alpha_i}$, where α_i is the number of occurrences of the label i in the tableau. The *Schur function* s_λ associated with the partition $\lambda = (\lambda_1, \ldots, \lambda_k)$ is defined by

$$s_\lambda(x_1, \ldots, x_k) = \sum_{\{T \, : \, \mathrm{sh}(T) = \lambda\}} x^T,$$

where $\mathrm{sh}(T)$ is the shape of the tableau T. We refer the reader to the books [32] and [34] for more about Schur functions and their connection to the RSK algorithm.

Suppose that the random variables $U = (u(i, j), (i, j) \in \{1, \ldots, N\} \times \{1, \ldots, K\})$ are independent, that $u(i, j)$ is geometrically distributed with parameter q_j, and that $q = (q_1, \ldots, q_K) \in (0, 1)^K$ is fixed. Then the law of the random partition $\lambda = (\lambda_1, \ldots, \lambda_K)$ associated with $w(U)$ is given by (see [23])

$$\mathrm{P}\{\lambda = l\} = a(q)^N s_l(q) s_l(1, \ldots, 1), \qquad l \in P_K,$$

where P_K is the set of integer partitions with at most K nonzero parts, $a(q) = \prod_{j=1}^{K} (1 - q_j)$, and s_l is the Schur function associated with the integer partition l. In particular, the law of λ is symmetric in the parameters $q_1, \ldots, q_K$.

This extends to the level of processes: if we write $\lambda = \lambda^{(N)}$ to emphasize its dependence on N, it is easy to check (see [23]) that the random sequence $\lambda^{(N)}$ is a Markov chain in P_K with transition probabilities given by

$$\mathrm{P}\{\lambda^{(N+1)} = l \mid \lambda^{(N)} = m\} = a(q) \frac{s_l(q)}{s_m(q)}$$

for all m and l such that $l_1 \geq m_1 \geq l_2 \geq m_2 \geq \cdots$. In particular, we can see that the law of the sequence $\lambda^{(N)}$ is symmetric in the parameters $q_1, \ldots, q_K$.

These remarks give the laws of the departure processes of both of the stochastic tandem queueing and storage models described above. Consider a saturated tandem of K queues/stores as defined in Section 5.1. Assume that the random variables $u(i, j)$, $i \in \mathbb{N}^*$, $j \in \{1, \ldots, K\}$, are all independent and that the $u(i, j)$ are geometrically distributed with parameter $q_j \in (0, 1)$. Let $D^{(N)}$ be the instant of departure of customer N from queue K and let $R^{(N)}$ be the cumulative number of departures from store K over the time slots 1 to N. By Theorem 5.1, $D^{(N)}$ and $R^{(N)}$ are respectively the lengths of the longest and shortest rows, $\lambda_1^{(N)}$ and $\lambda_K^{(N)}$, of the random tableau obtained by the RSK algorithm.

5.4.1. Interchangeability. A result of Weber [38] (also see [1]) states that a series of $\cdot$/M/1 queues arranged in tandem are interchangeable, i.e. the output process has the same distribution under any reordering of the queues. This is exactly the fact that the law of D is symmetric in the parameters and can now be seen as a direct consequence of the symmetry of the Schur functions. We can extend this result as follows.

Theorem 5.2. *The law of $(D^{(N)}, R^{(N)})_N$ is unchanged if the parameters of the geometric laws are modified from $(q_1, \ldots, q_K)$ to $(q_{\sigma(1)}, \ldots, q_{\sigma(K)})$, for any permutation σ of $\{1, \ldots, K\}$.*

In particular, stores arranged in tandem are interchangeable.

5.4.2. *Hydrodynamic limits.* Proving *shape theorems* (also known as *hydrodynamic limits*) and fluctuation results is an important issue in interacting particle systems; see, for instance, [16]. The law of $D^{(N)}$ is well known and the connection with the RSK algorithm has been exploited before to obtain detailed asymptotic and fluctuation results for the corresponding interacting particle system. The tandem storage (and bus stop) model, however, has not been discussed in the literature. A by-product of our analysis is that this model is amenable to some very precise analysis and development.

Indeed, consider Theorem 5.2 and the important special case of when the parameters are equal, with $q_j = q$, say. Then the random variables D and R are jointly distributed in the same way as the largest and smallest 'eigenvalues' of the Meixner ensemble (or, in the exponential case, of the Laguerre ensemble). For example, if the $u(i, j)$ are exponentially distributed with parameter 1, then the cumulative number of departures, $R^{(N)}$, over the time slots 1 to N in the corresponding tandem of K stores is distributed in the same way as the smallest eigenvalue of the $K \times K$ random matrix AA^*, where A is a $K \times N$ matrix with i.i.d. standard normal complex entries. This distribution is known; see, for instance, [29]. When $N = K$, it is exponential with mean K. This can be used as a starting point for the analysis of the fluctuations and hydrodynamic limit corresponding to the bus stop model of Section 5.2, a programme which is beyond the scope of this paper.

5.4.3. *Fixed point results.* Finally, we remark that the identity

$$\min_{\pi \in \widetilde{\Pi}} \left[\sum_{(i, j) \in \pi} u(i, j) \right] = R$$

can also be used as a cornerstone for proving fixed point and related weak convergence results for tandem stores. In the context of tandem queues, these questions have been pursued for some time, and the identity

$$\max_{\pi \in \Pi} \left[\sum_{(i, j) \in \pi} u(i, j) \right] = D$$

was a key ingredient in solving them [8], [2], [20], [26].

References

[1] ANANTHARAM, V. (1987). Probabilistic proof of the interchangeability of ·/M/1 queues in series. *Queueing Systems Theory Appl.* **2,** 387–392.
[2] BACCELLI, F., BOROVKOV, A. AND MAIRESSE, J. (2000). Asymptotic results on infinite tandem queueing networks. *Prob. Theory Relat. Fields* **118,** 365–405.
[3] BARYSHNIKOV, YU. (2001). GUEs and queues. *Prob. Theory Relat. Fields* **119,** 256–274.
[4] BEDEKAR, A. S. AND AZIZOĞLU, M. (1998). The information-theoretic capacity of discrete-time queues. *IEEE Trans. Inf. Theory* **44,** 446–461.
[5] BRÉMAUD, P. (1981). *Point Processes and Queues. Martingale Dynamics*. Springer, Berlin.
[6] BRUNEEL, H. AND KIM, B. G. (1993). *Discrete-Time Models for Communication Systems Including ATM*. Kluwer, Boston, MA.
[7] BURKE, P. (1956). The output of a queueing system. *Operat. Res.* **4,** 699–704.
[8] CHANG, C. S. (1994). On the input–output map of a G/G/1 queue. *J. Appl. Prob.* **31,** 1128–1133.
[9] DALEY, D. J. AND VERE-JONES, D. (1988). *An Introduction to the Theory of Point Processes*. Springer, New York.
[10] DRAIEF, M., MAIRESSE, J. AND O'CONNELL, N. (2003). Joint Burke's theorem and RSK representation for a queue and a store. In *Discrete Random Walks* (Paris, 2003), eds C. Banderier and C. Krattenthaler, Association of Discrete Mathematics and Theoretical Computer Science, Nancy, pp. 69–82.
[11] GLYNN, P. W. AND WHITT, W. (1991). Departures from many queues in series. *Ann. Appl. Prob.* **1,** 546–572.
[12] HAMBLY, B., MARTIN, J. AND O'CONNELL, N. (2001). Pitman's 2M-X theorem for skip-free random walks with Markovian increments. *Electron. Commun. Prob.* **6,** 73–77.

[13] HARRISON, J. AND WILLIAMS, R. (1990). On the quasireversibility of a multiclass Brownian service station. *Ann. Prob.* **18,** 1249–1268.

[14] HSU, J. AND BURKE, P. (1976). Behavior of tandem buffers with geometric input and Markovian output. *IEEE Trans. Commun.* **24,** 358–361.

[15] KELLY, F. (1979). *Reversibility and Stochastic Networks*. John Wiley, New York.

[16] KIPNIS, C. AND LANDIM, C. (1999). *Scaling Limits of Interacting Particle Systems*. Springer, Berlin.

[17] KNUTH, D. (1970). Permutations, matrices, and generalized Young tableaux. *Pacific J. Math.* **34,** 709–727.

[18] KÖNIG, W., O'CONNELL, N. AND ROCH, S. (2002). Non-colliding random walks, tandem queues, and discrete orthogonal polynomial ensembles. *Electron. J. Prob.* **7,** 1–24.

[19] LIGGETT, T. M. (1999). *Stochastic Interacting Systems: Contact, Voter and Exclusion Processes*. Springer, Berlin.

[20] MAIRESSE, J. AND PRABHAKAR, B. (2003). The existence of fixed points for the $\cdot/GI/1$ queue. *Ann. Prob.* **31,** 2216–2236.

[21] MUTH, E. (1979). The reversibility property of production lines. *Manag. Sci.* **25,** 152–158.

[22] O'CONNELL, N. (2003). A path-transformation for random walks and the Robinson–Schensted correspondence. *Trans. Amer. Math. Soc.* **355,** 3669–3697.

[23] O'CONNELL, N. (2003). Conditioned random walks and the RSK correspondence. *J. Phys. A* **36,** 3049–3066.

[24] O'CONNELL, N. AND YOR, M. (2001). Brownian analogues of Burke's theorem. *Stoch. Process. Appl.* **96,** 285–304.

[25] O'CONNELL, N. AND YOR, M. (2002). A representation for non-colliding random walks. *Electron. Commun. Prob.* **7,** 1–12.

[26] PRABHAKAR, B. (2003). The attractiveness of the fixed points of a $\cdot/GI/1$ queue. *Ann. Prob.* **31,** 2237–2269.

[27] PRABHAKAR, B. AND GALLAGER, R. (2003). Entropy and the timing capacity of discrete queues. *IEEE Trans. Inf. Theory* **49,** 357–370.

[28] RAJEWSKY, N., SANTEN, L., SCHADSCHNEIDER, A. AND SCHRECKENBERG, M. (1998). The asymmetric exclusion process: comparison of update procedures. *J. Statist. Phys.* **92,** 151–194.

[29] RATNARAJAH, T., VAILLANCOURT, R. AND ALVO, M. (2005). Eigenvalues and condition numbers of complex random matrices. *SIAM J. Matrix Anal. Appl.* **26,** 441–456.

[30] REICH, E. (1957). Waiting times when queues are in tandem. *Ann. Math. Statist.* **28,** 768–773.

[31] ROBERT, P. (2003). *Stochastic Networks and Queues* (Appl. Math. **52**). Springer, Berlin.

[32] SAGAN, B. (2001). *The Symmetric Group. Representations, Combinatorial Algorithms, and Symmetric Functions* (Graduate Texts Math. **203**), 2nd edn. Springer, Berlin.

[33] SEPPÄLÄINEN, T. (2000). A variational coupling for a totally asymmetric exclusion process with long jumps but no passing. In *Hydrodynamic Limits and Related Topics* (Fields Inst. Commun. **27**; Toronto, ON, 1998), American Mathematical Society, Providence, RI, pp. 117–130.

[34] STANLEY, R. (1999). *Enumerative Combinatorics* (Camb. Studies Adv. Math. **62**), Vol. 2. Cambridge University Press.

[35] SZCZOTKA, W. AND KELLY, F. (1990). Asymptotic stationarity of queues in series and the heavy traffic approximation. *Ann. Prob.* **18,** 1232–1248.

[36] TAKAGI, H. (1993). *Queueing Analysis: A Foundation of Performance Evaluation*, Vol. 3, *Discrete-Time Systems*. North-Holland, Amsterdam.

[37] TEMBE, S. AND WOLFF, R. (1974). The optimal order of service in tandem queues. *Operat. Res.* **24,** 824–832.

[38] WEBER, R. (1979). The interchangeability of $\cdot/M/1$ queues in series. *J. Appl. Prob.* **16,** 690–695.