A Survey on Performance Analysis of Warehouse Carousel Systems

Nelly Litvak*, Maria Vlasiou**

February 27, 2010

* Faculty of Electrical Engineering, Mathematics & Computer Science,
Department of Applied Mathematics, University of Twente,
7500 AE Enschede, The Netherlands.

** Eurandom and Department of Mathematics & Computer Science,
Eindhoven University of Technology,
P.O. Box 513, 5600 MB Eindhoven, The Netherlands.

n.litvak@ewi.utwente.nl, m.vlasiou@tue.nl

Abstract

This paper gives an overview of recent research on the performance evaluation and design of carousel systems. We discuss picking strategies for problems involving one carousel, consider the throughput of the system for problems involving two carousels, give an overview of related problems in this area, and present an extensive literature review. Emphasis has been given on future research directions in this area.

Keywords: order picking, carousels systems, travel time, throughput

AMS Subject Classification: 90B05, 90B15

Acknowledgements: The authors would like to thank Ivo Adan for the suggestion to write this survey paper and for his active involvement in the research described in Sections 2 and 3. In the Netherlands, the three universities of technology have formed the 3TU Federation. This article is the result of joint research in the 3TU Centre of Competence NIRICT (Netherlands Institute for Research on ICT).

1 Introduction

A carousel is an automated storage and retrieval system, widely used in modern warehouses. It consists of a number of shelves or drawers, which are linked together and are rotating in a closed loop. It is operated by a picker (human or robotic) that has a fixed position in front of the carousel. A typical vertical carousel is given in Figure 1.

Carousels are widely used for storage and retrieval of small and medium-sized items, such as health and beauty products, repair parts of boilers for space heating, parts of vacuum cleaners and sewing machines, books, shoes and many other goods. In e-commerce companies use carousel to store small items and manage small individual orders. An order is defined as a set of items that must be picked together (for instance, for a single customer).

Carousels are highly versatile, and come in a huge variety of configurations, sizes, and types. They can be horizontal or vertical and rotate in either one or both directions. Although both unidirectional (one-way rotating) or bidirectional (two-way rotating) carousels are encountered in practice, the bidirectional types are the most common (as well as being the most efficient) [53]. One of the main advantages of carousels is that, rather than having the picker travel to an item (as is the case in a warehouse where items are stored on shelves), the carousel rotates the items to the picker. While the carousel is travelling, the picker has the time to perform other tasks, such as pack or label the retrieved items, or serve another carousel. This practice enhances the operational efficiency of the warehouse.

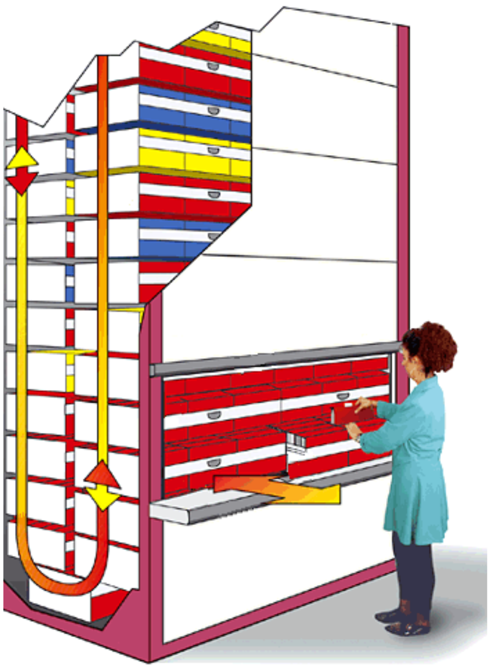

Figure 1: A typical vertical carousel.

Carousel models have received much attention in the literature and continue to pose interesting problems. There is a rich literature on carousels that dates back to 1980 [121]. In Section 6 we shall review some of the main research topics that have been of interest to the research community so far. To name a few, one may wish to study various ways of storing the items on a carousel (storage arrangements) so as to minimise the total time needed until an order is completed (response time) or the strategy that should be followed in rotating the carousel so as the total time the carousel travels between items of one order is minimised (travel time for a single order). One may also consider design issues, for instance, the problem of pre-positioning the carousel in anticipation of storage or retrieval requests (choosing a *dwell point*) in order to improve the average response time of the system. The list of references presented here is by no means exhaustive; it rather serves the purpose of indicating the continuing interest in carousels.

In this review paper we focus on the modelling and the performance of carousel systems. Usually a carousel is modelled as a circle, either as a discrete model [6, 60, 102, 127], where the circle consists of a fixed number of locations, or as a continuous one [43, 76, 105, 116], where the circle has unit length and the locations of the required items are represented as arbitrary points on the circle. Throughout this paper we shall view the carousel as a continuous loop of unit length. Beyond this initial assumption, we shall examine modelling issues such as how to model travel times or picking times of items in a system of several carousels so as to be able to derive approximations of various performance characteristics. Under "performance" one may understand a variety of notions. For example, in single-carousel single-order problems (cf. Section 2), the performance measure under consideration is the travel time of the carousel until all items in an order are picked. On the other hand, in Section 3, performance may be measured by the time the picker is idle between picking items from various carousels, i.e. by the picker's utilisation.

In this paper we consider two research topics in detail. In Section 2, we discuss the problem of choosing a reasonable picking strategy for one order and a single carousel, where the order is represented as a list of items, and by order pick strategy we mean an algorithm that prescribes in which sequence the items are to be retrieved. We present a general probabilistic approach developed by Litvak *et al.* [76, 79, 80, 81] to analytically derive the probability distribution of the travel time in case when items locations are independent and uniformly distributed. This line of research seems to be the only example in the literature where exact statistical characteristics of the travel time have been obtained by means of a systematic mathematical approach. The presented technique is based on properties of uniform spacings and their relations to exponential distributions. We demonstrate the effectiveness of this method by considering several relevant order-picking strategies, such as the greedy nearest-item strategies and so-called m-step strategies that provide a good approximation for the optimal (shortest) route.

In Section 3 we consider the second topic that relates to multiple-carousel settings and the modelling challenges that appear in such problems. Having optimised the travel time of a single carousel for a single order, one wonders if optimising locally every time each order on each carousel leads to the best solution (fastest, cheapest, or with the largest picker utilisation) for a complicated system. As is mentioned later on, multiple-carousel problems become too complicated too quickly, and often exact analysis is not possible. Therefore, we discuss which concessions have to be made in order to be able to obtain estimates of the performance measures we are interested in, and we give in detail the impact that these concessions have on our estimations. There exist a few exact results for two-carousel models and related models in healthcare logistics; see Boxma and Vlasiou [21] and Vlasiou *et al.* [111]–[119]. However, to the best of our knowledge, no exact results exist for systems involving more than two carousels.

Preferably, these two research topics that we consider in this paper should be studied in parallel. However, establishing any exact results, say on determining the optimal retrieval and travelling strategy for a multiple-carousel model, without any restrictions to the sequence the items in an order are picked or the sequence the carousels are served, seems to be intractable. Nonetheless, quite a few research opportunities related to the optimal design and control of carousel systems are still available. We elaborate on further research topics in Section 5. We conclude with Section 6, which outlines the problems examined so far on carousels and related storage and retrieval systems.

2 Picking a single order on a single carousel

Performance analysis of single units is a necessary step in structural design of order pick systems [128]. In a setting of a single order on a single carousel, the major performance characteristic is the response time, that is, the total time it takes to retrieve an order. The response time consists of pick times needed to collect the items from their locations by an operator, and the travel (rotation) time of the carousel. While pick times can hardly be improved, the travel time depends on the location of each item and the order picking sequence, and thus, it is subject to analysis and optimisation. Therefore, in this section, we discuss properties of the travel time needed to collect an order of n items. In this section, our focus is on the case when the item locations are randomly distributed on a carousel circumference. This model allows one to compute statistical characteristics of the travel time such as the average travel time or the travel time distribution. Later on, in Section 6.2 we discuss some results from the literature on evaluating the travel times under different assumptions on the items locations, in particular, the case when the pick positions are fixed.

We note that in case of a single carousel, it is natural to assume that the pick times and the travel time are independent. The situation, however, is quite different in the systems of two or more carousels, where pick times on one carousel affect the travel times on other carousels. This issue will be discussed in detail in Section 3.

The model addressed in this section is as follows. We model a carousel as a circle of length 1. The order is represented by the list of n items whose positions are independent and uniformly distributed on $[0, 1)$. For ease of presentation, we act as if the picker travels to the pick positions instead of the other way around. Also, we assume that the acceleration/deceleration time of the carousel is negligible or that it is assigned to the pick time, and that the carousel rotates at unit speed. Therefore the travel distance can be identified with the travel time (see also Section 6.4).

Obviously, the travel time depends heavily on the pick strategy. Here by *order pick strategy* we mean an algorithm that prescribes the sequence in which the items are collected. For example, assume that the picker always proceeds in the clockwise (CW) direction and denote by T_n^{CW} the time needed to collect n items under this simple strategy. Then, clearly, the distribution function $\mathbb{P}(T_n^{CW} \leq t)$ of T_n^{CW} simply equals t^n, $0 < t \leq 1$. However, we would like to study strategies that provide smaller travel times. In this sense, a better algorithm that one can think of is the 'greedy' strategy, also called the *nearest-item heuristic*: always travel to the nearest item to be picked (as in Figure 2). The nearest-item strategy indeed performs very well and is often used in practice, but the question is: "what is the distribution of the travel time under the nearest-item heuristic?". This problem is not at all trivial. For example, straightforward methods, such as conditioning on possible item locations, do not lead to feasible calculations. The same ap-

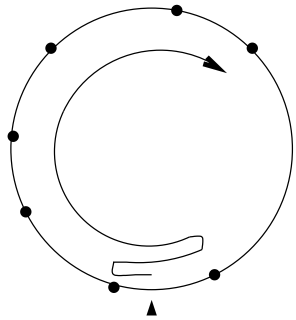

Figure 2: A route under the nearest-item heuristic.

plies to the optimal strategy. Bartholdi and Platzman [6] showed that the shortest route admits at most one turn. Intuitively, this follows merely by observing Figure 2, where the displayed route can be shortened by collecting the first item in the counterclockwise direction and then collecting the rest of the items rotating clockwise. Thus, the shortest route is merely the minimum among the $2n$ candidate routes than have at most one turn. However, in spite of this simple structure of the shortest route, its distribution function is hard to derive.

Below we discuss in detail a general methodology developed by Litvak *et al.* [76, 79, 80, 81] to obtain the distribution of the travel time under various order pick strategies. The proposed technique is based on properties of uniform spacings and their connection with exponential random variables. We show how this approach allows us to derive exact and often counterintuitive results on several relevant order pick strategies. Some other methods from the literature are described in Section 6.2.

We start with introducing the notation and presenting some background results. Let the random variable $U_0 = 0$ be the picker's starting point and the random variable U_i, where $i = 1, 2, \ldots, n$, be the position of the ith item. We suppose that the U_i's, $i = 1, 2, \ldots, n$, are independent and uniformly distributed on $[0, 1)$. Let $U_{1:n}, U_{2:n}, \ldots U_{n:n}$ denote the order statistics of $U_1, U_2, \ldots U_n$ and set $U_{0:n} = 0$, $U_{n+1:n} = 1$. Then the uniform spacings are defined as

$$D_{i,n} = U_{i:n} - U_{i-1:n}, \quad 1 \leq i \leq n+1. \tag{1}$$

If we consider n items randomly located on a circle, then the spacings $D_{2,n}, D_{3,n}, \ldots, D_{n,n}$ are the distances between two neighbouring items, and the spacings $D_{1,n}$ and $D_{n+1,n}$ are the distances between the starting point and the two items adjacent to it. Whatever strategy the picker takes, he always has to cover one or more uniform spacings on his way from one location to another. Hence, in general, the travel time can be expressed as a function of the uniform spacings.

Uniform spacings have been analysed extensively in two classical review papers by Pyke [96, 97]. The author gives four useful constructions that establish a connection between uniform spacings and exponential random variables. We will use such a connection in the following form. Let X_1, $X_2, \ldots$ be independent exponential random variables with mean 1. Moreover, define the random variables

$$S_0 = 0, \qquad S_i = X_1 + X_2 + \cdots + X_i, \quad i \geq 1.$$

Then, according to Pyke [96], uniform spacings can be represented as follows:

$$(D_{1,n}, D_{2,n}, \ldots, D_{n+1,n}) \stackrel{d}{=} (X_1/S_{n+1}, X_2/S_{n+1}, \ldots, X_{n+1}/S_{n+1})\,. \tag{2}$$

Here and throughout this paper $a \stackrel{d}{=} b$ means that a and b have the same probability distribution. Linear combinations of uniform spacing have nice properties. In particular, the moments of linear combinations with non-negative coefficients can be easily computed, and their distribution function has been derived by Ali [2], Ali and Obaidullah [3].

Now, let X and Y be independent exponential random variables with parameters λ and μ, respectively. We write $X = X_1/\lambda$, $Y = Y_1/\mu$, where X_1 and Y_1 are independent exponential random variables with parameter 1. Then, given the event $[X < Y]$, we obtain the following useful statements: (i) the distribution of $X = \min\{X, Y\}$ is exponential with parameter $\lambda + \mu$ (property of the minimum of two exponentials), which is distributed as $X_1/(\lambda + \mu)$; (ii) since $[Y > X]$, then, according to the memoryless property, Y can be written as a sum of two terms: $\min\{X, Y\}$ and another independent exponential with parameter μ, so Y is distributed as $X_1/(\lambda + \mu) + Y_1/\mu$. (iii) it is easy to check that the distribution of $S = \lambda X + \mu Y = X_1 + Y_1$ is independent of the event $[X < Y]$ because according to (i) and (ii), given $[X < Y]$, S is again distributed as $X_1 + Y_1$ (see also Chapter 2 of [76]).

Based on the above-mentioned properties of exponential random variables, and their connections to uniform spacings and travel times, one may adopt the following methodology for analysing the travel times under various strategies [76, 79, 80, 81]:

1. Express the travel time under a given strategy as a function of uniform spacings.

2. By conditioning on linear inequalities between the spacings and employing the above mentioned properties of exponential random variables, rewrite the travel time as a linear combination of uniform spacings or as a probabilistic mixture of such linear combinations.

3. Use the results from [2, 3] to obtain the moments and the distribution of the travel time.

Below we show how this approach works in case of the nearest-item heuristic [79, 81] and so-called m-step strategies [80].

2.1 The nearest-item heuristic

Under the nearest-item heuristic, the picker always moves towards the nearest item to be retrieved. The positions of the items partition the circle in $n + 1$ uniform spacings $D_{1,n}, D_{2,n}, \ldots, D_{n+1,n}$ defined by (1). Under the nearest-item heuristic, the picker first considers the two spacings adjacent to his starting position and then travels to the nearest item. Next he also looks at the other spacing adjacent to that item and compares the distance to the item located at the endpoint of that spacing and the distance to the first item in the other direction, which is the sum of the spacings previously considered. Then he travels again to the nearest item, and so on. Furthermore, by employing (2), we may act as if the picker faces non-normalised exponential spacings, and afterwards divide the travel time (which is equal to the travel distance) by the sum of all spacings. Then it is clear that each new spacing faced by the picker is independent of the ones already observed. Now let X_i, $i = 1, \ldots, n+1$, denote the i-th non-normalised exponential spacing faced by the picker. That is, the spacings are numbered as observed by the picker operating under the nearest-item heuristic (see Figure 3). Then T_n^{NI} can be expressed as

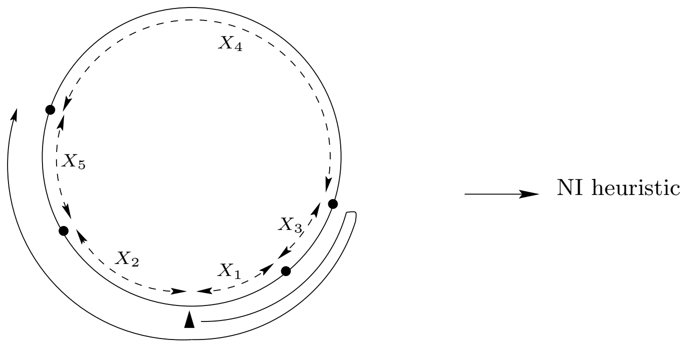

Figure 3: The nearest-item route of the picker facing 5 exponential spacings.

$$T_n^{NI} = \sum_{i=2}^{n+1} \frac{\min\{X_i, S_{i-1}\}}{S_{n+1}}. \tag{3}$$

We first provide an informal explanation of how the proposed methodology can be applied to (3). To start with, note that first term in the right-hand side of (3) is $\min\{X_1, X_2\}/S_{n+1}$, which is distributed simply as $(1/2)X_1/S_{n+1}$. Moreover, under the event $[X_1 < X_2]$ the rest of the sum remains unaltered. Further, consider the term

$$(1/2)X_1 + \min\{X_3, S_2\} = (1/2)X_1 + \min\{X_3, X_1 + X_2\}. \tag{4}$$

Let X_1', X_2', X_3' be auxiliary independent exponential random variables with mean 1. Given $[X_3 < X_1]$, the random variable X_3 is distributed as $(1/2)X_1'$, X_1 is distributed as $(1/2)X_1' + X_2'$ and X_2 is distributed as X_3'. Then the term in (4) is distributed as $(3/4)X_1' + (1/2)X_2'$. Furthermore, given the event $[X_3 > X_1, X_3 < X_1 + X_2]$, we obtain that X_1 is distributed as $(1/2)X_1'$, X_3 is distributed as $(1/2)X_1' + (1/2)X_2'$ and X_2 is distributed as $(1/2)X_2' + X_3'$. Substituting the above in (4), we obtain again $(3/4)X_1' + (1/2)X_2'$! Remarkably, under the event $[X_3 > X_1 + X_2]$, (4) again transforms into $(3/4)X_1' + (1/2)X_2'$. Furthermore, the sum $S_3 = X_1 + X_2 + X_3$ becomes simply $S_3 = X_1' + X_2' + X_3'$. We may now rename (X_1', X_2', X_3') back to (X_1, X_2, X_3) since the two 3-dimensional vectors are identically distributed. Then the term (4) becomes $(3/4)X_1 + (1/2)X_2$, and the rest of the terms in the right-hand side of (3) remain unaltered in all three cases. Proceeding further, we obtain the next statement which is proved rigorously in [79].

Theorem 1 (Litvak and Adan [79]). *For all $n = 1, 2, \ldots,$*

$$T_n^{NI} \stackrel{d}{=} \sum_{i=1}^{n} \left(1 - \frac{1}{2^i}\right) D_{i,n} \tag{5}$$

and

$$\mathbb{P}(T_n^{NI} \le t) = \sum_{i=0}^{n} \left(2^i t - 2^i + 1\right)_+^n \prod_{\substack{j=0 \\ j \ne i}}^{n} \frac{2^j}{2^j - 2^i}, \quad 0 < t \le 1, \tag{6}$$

where $x_+ = x$ if $x > 0$ and $x_+ = 0$ otherwise.

Here (6) follows directly from (5) and the result by Ali [2], which we applied in the form given by Theorem 2 in [3].

The above theorem is surprising because it provides an elegant solution for a problem that looks intractable at first. An interesting by-product is the distribution of the number of turns under the nearest-item heuristics and the counterintuitive result that *the travel time and the number of turns are independent* [76]! The latter can be seen directly from (3). Indeed, a turn after step i is equivalent to the event $[X_{i+1} > S_i]$. However, as we saw earlier, the form of the distribution of the travel time is given by (5) and it is independent of this sort of events.

2.2 The m-step strategy

Under the m-step strategies, the picker chooses the shortest route among the $2(m+1)$ routes that change direction at most once, and only do so after collecting no more than m items. Note that the optimal strategy is in fact an $(n-1)$-step strategy since it is never optimal to turn more than once, and the maximal possible number of items collected before a turn is $n - 1$. The m-step strategies give a good approximation for the shortest travel time. In fact, they often provide the optimal route even for moderate values of m, as in Figure 4. Rouwenhorst *et al.* [100] were the first to propose these strategies as an upper bound for the optimal route. In case of independent uniformly distributed pick positions, they obtained the distribution of the travel time under the m-step strategy for $m \le 2$ using analytical methods. Later on, Litvak and Adan [80] applied the described methodology based on the properties of uniform spacings to completely analyse the travel time under the m-step strategies, provided $2m < n$. The travel under the m-step strategy can be expressed as follows

$$T_n^{(m)} = 1 - \max\left\{ \max_{1 \le j \le m+1} \left\{ D_{j,n} - \sum_{l=1}^{j-1} D_{l,n} \right\}, \max_{1 \le j \le m+1} \left\{ D_{n+2-j,n} - \sum_{l=1}^{j-1} D_{n+2-l,n} \right\} \right\}.$$

Indeed, the term $D_{j,n} - \sum_{l=1}^{j-1} D_{l,n}$ is the gain in travel time (compared to one full rotation) obtained by skipping the spacing $D_{j,n}$ and going back instead, ending in a clockwise direction. On the other hand, $D_{n+2-j,n} - \sum_{l=1}^{j-1} D_{n+2-l,n}$ is the gain obtained by skipping the spacing $D_{n+2-j,n}$ and going back ending counterclockwise. Under the m-step strategy the picker skips the spacing that provides the largest possible gain (see Figure 4). Using

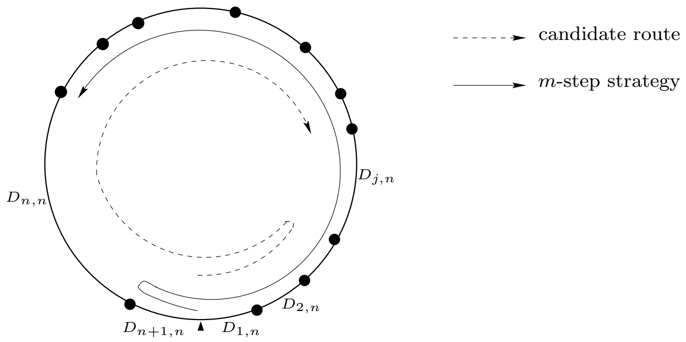

Figure 4: A route under the m-step strategy.

property (2), and after appropriate manipulations of exponential random variables, one can prove the following result.

Theorem 2 (Litvak and Adan [80]). *For any $m = 0, 1, \ldots,$ with $2m < n$,*

$$T_n^{(m)} \stackrel{d}{=} 1 - \frac{1}{S_{n+1}} \max\left\{ \sum_{j=1}^{m+1} \frac{1}{2^j - 1} X_j, \sum_{j=1}^{m+1} \frac{1}{2^j - 1} X_{n+2-j} \right\}. \tag{7}$$

The maximum in the right-hand side of (7) implies that $T_n^{(m)}$ is distributed as a complicated probabilistic mixture of linear combinations of uniform spacings [80]. The number of terms in this mixture is the well-known Catalan number

$$\frac{1}{m+2}\binom{2m+2}{m+1},$$

which grows extremely fast with m. Computing the expectations, we conclude that on average, the m-step strategy performs better than the nearest-item heuristic already for $m = 2$ provided $n \geq 5$.

Again, as a by-product, we can obtain the distribution of the number of steps before the turn. Moreover, the latter random variable turns out to be independent of the travel time. This surprising statement follows from a similar reasoning as the independence of the travel time and the number of turns under the nearest-item heuristic. Furthermore, when n goes to infinity, the number of steps before the turn converges to a shifted geometric distribution with parameter $1/2$. That is, in the limit, with probability $1/2$ there will be no turn, with probability $1/4$ there will be one step before a turn, etc. Also, in the limit, the m-step strategy with $2m < n$ coincides with the optimal strategy since the probability of achieving the minimal travel time by making more than $n/2$ steps before a turn will converge to zero. Thus, for large enough n, the probability that a 2-step strategy provides an optimal route is about 7/8. This explains the remarkably good performance of the m-step strategies.

As a side remark, we would like to note that [77] provides slightly more general results than those presented in (5) and (7).

2.3 Optimal route

Since the optimal strategy simply coincides with the $(n-1)$-step strategy (at most one turn after collecting at most $n-1$ items) it can be analysed by methods from Section 2.2. However, the condition $2m < n$ is violated for $m = n-1$, and hence, (7) does not hold. In fact, the proposed methodology applied to the optimal travel time T_n^{Opt} very soon results in analytically infeasible calculations. Litvak and van Zwet [82] analysed the optimal route. They employed the results on the m-step strategy to derive a recursive expression for the distribution of the minimal travel time.

We would like to also note that the process of comparing the lengths of the spacings and deriving corresponding linear combinations of normalised exponentials can be easily translated into a computer program. Then, for moderate values of n the exact distribution of the optimal travel time can be obtained numerically. The result will be a complicated mixture of linear combinations of uniform spacings. For large values of n such exact calculations will require too much computer capacity. However, in this case, the knowledge of the exact distribution is not very important since one can apply approximations based on asymptotic results discussed in the next section.

2.4 Asymptotic results

When the order is large, we can model this situation by letting $n \to \infty$. Then the expressions in (5) and (7) for the travel time allow us to obtain asymptotic results that are of independent mathematical interest. Obviously, if $n \to \infty$ then the travel time under any strategy goes to one with probability 1. However, with linear scaling, we obtain non-trivial distributions that we present below for the nearest-item heuristic and for the optimal travel time.

Theorem 3. *Let $X_1, X_2, \ldots, X_1', X_2', \ldots,$ be independent exponentials with mean 1. Then*

$$(n+1)\left(1-T_n^{NI}\right) \xrightarrow{d} \sum_{j=1}^{\infty} \frac{1}{2^{j-1}} X_j \qquad \text{(Litvak and Adan [80])}, \tag{8}$$

$$(n+1)\left(1-T_n^{Opt}\right) \xrightarrow{d} \max\left\{\sum_{j=1}^{\infty} \frac{1}{2^j-1} X_j, \sum_{j=1}^{\infty} \frac{1}{2^j-1} X_j'\right\} \qquad \text{(Litvak and van Zwet [82])} \tag{9}$$

as $n \to \infty$.

Litvak [78] generalises (9) to the case when items positions are independent and have some positive density f.

The expression in the right-hand side of (8) is a well-known functional of the Poisson process, which has been extensively studied in the literature. We will briefly discuss this topic in Section 4.2.

3 Multiple carousels: modelling challenges

The problems examined so far relate to one-carousel models. In industry though, one rarely meets a facility where only one carousel is used. Multiple-carousel systems tend to have a higher level of throughput; however, they increase the investment cost due to the extra driving and control mechanisms [55, 57]. A natural question is how much the throughput of a standard carousel can be improved by the corresponding multiple-carousel system that has the same number of shelves as the standard carousel. Thus, the question we would like to examine in this section is the following: given a setup, i.e. a specific storage scheme of the items stored on the carousel *and* a specific travelling strategy, such as those described in the previous section, how much can we increase the utilisation of the picker (by assigning to him more carousels to handle) without increasing the response time of an order above some chosen level? In other words, how do we reach a *quality and efficiency* regime in a real situation?

To illustrate things better, consider the following simple example. A facility assigns n carousels to a single picker. Each carousel is assigned to an order of a single customer, and each order consists of exactly one item. Moreover, each carousel rotates independently until the desired item reaches the picker, who is standing at a fixed point, the origin. Once this position is reached, the carousel stops until the item is picked. Only then will the next order be given to the carousel, which will start rotating the new order to the origin. The picker serves the carousels in a fixed order, visiting each carousel only once in every cycle. Clearly, as n goes to infinity, the utilisation of the picker in steady state tends to one, since almost surely he will never have to wait. The carousels will have brought each of their respective items to the origin by the time the picker is ready to serve them. On the other hand, the time until the picker returns to the first carousel tends to infinity; i.e. each individual customer suffers long waiting times.

Multiple carousel problems differ intrinsically from single-carousel problems in a number of ways. Such systems tend to be more complicated. The system cannot be viewed as a number of independently operating carousels (cf. [84] and Section 6.4), since there may be some interaction between two separate carousels by means of the picker that is assigned to them. Namely, if the number of pickers is less than the number of carousels, then the picking strategy that is chosen for an isolated carousel may affect significantly the waiting time of another carousel. Thus, one cannot guarantee that minimising the travel time of a single carousel maximises the total throughput of the system; the outcome may be quite the contrary because of the system's interdependency. Another point is that in multiple-carousel problems, the i.i.d. assumption of the time needed to pick each of two consecutive orders with random item storage is in principle invalid. Characteristics such as the time needed to reach the optimal point or the travel time for each carousel depend on one another through the picker's movements. For all these reasons, multiple-carousel systems merit a special reference.

Ideally, the problems of minimising the travel time of all carousels and maximising the picker's utilisation without surpassing certain levels of each order's response time should be studied together. However, the interdependence that appears in multiple carousel problems usually leads to complicated mathematical structures that can hardly be analysed exactly. One will have to resort to simplifications.

One technique that can help overcome some of these difficulties is the setting proposed in Vlasiou *et al.* [116]. The system we consider below consists of two carousels operated by a single picker. Given a setting, i.e. a storage scheme and a travel strategy, one first needs to obtain an estimate of the travel time needed in order to collect all items under this setting. For example, if the items are stored in random positions on the carousel, then the distribution of the travel time under the nearest-item heuristic is given by (6). In most settings though, this distribution cannot be computed analytically, in which cases the empirical distribution or simulation may provide a partial answer. Subsequently, one may need to approximate this distribution by a phase-type distribution; see e.g. [90]. Then, the following modelling assumption is made. We aggregate all items in one. That is, we consider an order that consists of exactly one item. It is assumed that the travel time of the carousel until that single item is reached is uniformly distributed (i.e. it is assumed that the item is located randomly on the carousel), while the distribution of the pick time for that item is taken to be equal to the phase-type distribution computed previously. Under these assumptions, one can compute the utilisation of the picker by applying the results developed in Vlasiou *et al.* [116]. This procedure can be repeated until the desired quality and efficiency regime is reached.

To describe things concretely, we consider a system consisting of two identical carousels and one picker. At each carousel there is an infinite supply of pick orders that need to be processed. The picker alternates between the two carousels, picking one order at a time. There are two ways one can view this. Either, as mentioned above, one aggregates all items in an order in one super-item (i.e. we consider an order that consists of exactly one item) or under the term "picking time" we understand the total time needed for the actual picking and travelling from the moment the picker is about to pick the first item in an order until the time the last item is picked. For ease of presentation, we will opt for the first solution, considering orders consisting of exactly one item.

As in Section 2, we model a carousel as a circle of length 1 and we assume that it rotates in one direction at a constant speed. The picking process may be visualised as follows. When the picker is about to pick an item at one of the carousels, he may have to wait until the item is rotated in front of him. In the meantime, the other carousel rotates towards the position of the next item. After completion of the first pick the carousel is instantaneously replenished and the picker turns to the other carousel, where he may have to wait again, and so on. Let the random variables A_n, B_n and W_n, $n \geq 1$, denote the pick time, rotation time and waiting time for the n-th item. Clearly, the waiting times

W_n satisfy the recursion

$$W_{n+1} = \max\{0, B_{n+1} - A_n - W_n\}, \qquad n = 0, 1, \ldots \tag{10}$$

where $A_0 = W_0 \stackrel{\text{def}}{=} 0$. We assume that both $\{A_n\}$ and $\{B_n\}$, $n \geq 1$, are sequences of independent identically distributed random variables, also independent of each other. The pick times A_n follow a phase-type distribution and the rotation times B_n are uniformly distributed on $[0, 1)$ (which means that the items are randomly located on the carousels). Then $\{W_n\}$ is a Markov chain, with state space $[0, 1)$. Moreover, it can be shown that $\{W_n\}$ is an aperiodic, recurrent Harris chain, which possesses a unique equilibrium distribution. In equilibrium, equation (10) becomes

$$W \stackrel{d}{=} \max\{0, B - A - W\}. \tag{11}$$

Once the distribution of W is computed from (11), we can compute $\mathbb{E}[W]$ and thus also the throughput of the system τ from

$$\tau = \frac{1}{\mathbb{E}[W] + \mathbb{E}[A]}. \tag{12}$$

Equation (11) with a plus sign instead of minus sign in front of W at the right-hand side, is precisely Lindley's equation for the stationary waiting time in a PH/U/1 queue. The equation for the standard PH/U/1 queue has no simple solution, but in Vlasiou *et al.* [116] we show that the waiting time of the picker in our problem can be solved for explicitly.

For example, assume that the service times follow an Erlang distribution with scale parameter λ and n stages; that is,

$$F_A(x) = 1 - \mathrm{e}^{-\lambda x} \sum_{i=0}^{n-1} \frac{(\lambda x)^i}{i!}, \qquad x \geq 0$$

and define $\pi_0 = \mathbb{P}[W = 0]$. Then, for the Laplace transform $\omega(s)$ of W, i.e.

$$\omega(s) = \int_0^1 e^{-sx} f_W(x) dx,$$

where $f_W(x)$ is the density of W, the following theorem holds (recall that the domain of integration is bounded by the length of the carousel).

Theorem 4 (Vlasiou *et al.* [116]). *For all s, the transform $\omega(s)$ satisfies*

$$\omega(s)R(s) = -\mathrm{e}^{-s} s(\lambda + s)^n T(-s) - \lambda^n T(s), \tag{13}$$

where

$$R(s) = s^2(\lambda^2 - s^2)^n + \lambda^{2n},$$

$$T(s) = \pi_0 \left(\lambda^n + \mathrm{e}^{-(\lambda+s)} \sum_{i=0}^{n-1} \sum_{j=0}^{i} \frac{s\lambda^i(\lambda+s)^{n-i-1+j}}{j!} \right) - \mathrm{e}^{-s}(\lambda+s)^n + $$
$$+ \mathrm{e}^{-(\lambda+s)} \sum_{i=0}^{n-1} \sum_{j=0}^{i} \sum_{\ell=0}^{j} \binom{j}{\ell} \frac{s\lambda^i(\lambda+s)^{n-i-1+j}}{j!} \omega^{(\ell)}(-\lambda).$$

In (13) we still need to determine the $n+1$ unknowns π_0 and $\omega^{(\ell)}(-\lambda)$ for $\ell = 0, \dots, n-1$. Note that for any zero of the polynomial R, the left-hand side of (13) vanishes (since ω is analytic everywhere). This implies that the right-hand side should also vanish. Hence, the zeros of R provide the equations necessary to determine the unknowns. In [116] it is explained how to determine these unknown parameters (which incidentally form the unique solution to a linear system of equations) and how to invert the transform. Qualitatively, the result is as follows.

Theorem 5 (Vlasiou *et al.* [116]). *The density of W on $[0, 1]$ is given by*

$$f_W(x) = \sum_{i=1}^{2n+2} c_i \mathrm{e}^{r_i x}, \qquad 0 \le x \le 1, \tag{14}$$

and

$$\pi_0 = \mathbb{P}[W = 0] = 1 - \sum_{i=1}^{2n+2} \frac{c_i}{r_i}(\mathrm{e}^{r_i} - 1), \tag{15}$$

where r_i is a zero of the polynomial R appearing in Theorem 13, and where the coefficients c_i are known explicitly.

As a by-product, we have that

Corollary 1. *The throughput τ satisfies*

$$\tau^{-1} = \mathbb{E}[A] + \mathbb{E}[W] = \frac{n}{\lambda} + \sum_{i=1}^{2n+2} \frac{c_i}{r_i^2}[1 + (r_i - 1)\mathrm{e}^{r_i}].$$

Remark 1. The same qualitative result holds in case the pick times follow a mixed-Erlang distribution. In this case, the waiting time density is again a mixture of exponentials, where all parameters can be computed explicitly; cf. [116].

In a series of papers, Vlasiou *et al.* [21, 111, 112, 113, 114, 116, 118, 119] have relaxed several of the assumptions made above for the two-carousel setting. For example, the travel time needed to pick all items in an order can have any general distribution (e.g. depending on the pick strategy that is followed). In such cases, one can compute the distribution of the waiting time of the picker by approximating the distribution of the travel time by an appropriate phase-type distribution. Phase-type distributions may be used to approximate any given distribution on $[0, 1)$ for the travel times arbitrarily close; see for example Asmussen [4]. As an illustrative example, we give below the steady-state distribution of the waiting time of the picker in case the pick times follow some general distribution with Laplace-Stieltjes transform (LST) α, and the travel times follow an Erlang distribution with parameter μ and n stages. Here, ω denotes the (unknown) LST of the waiting time of the picker. In this case we have the following:

Theorem 6 (Vlasiou and Adan [112]). *The waiting-time distribution has a mass π_0 at the origin, which is given by*

$$\pi_0 = \mathbb{P}[B < W + A] = 1 - \sum_{i=0}^{n-1} \frac{(-\mu)^i}{i!} \phi^{(i)}(\mu)$$

and has a density f_W on $[0, \infty)$ that is given by

$$f_W(x) = \mu^n \mathrm{e}^{-\mu x} \sum_{i=0}^{n-1} \frac{(-1)^i}{i!} \phi^{(i)}(\mu) \frac{x^{n-1-i}}{(n-1-i)!}. \tag{16}$$

In the above expression, we have that

$$\phi^{(i)}(\mu) = \sum_{k=0}^{i} \binom{i}{k} \omega^{(k)}(\mu)\, \alpha^{(i-k)}(\mu)$$

and that the parameters $\omega^{(i)}(\mu)$ for $i = 0, \ldots, n-1$ are the unique solution to the system of equations

$$\begin{aligned} \omega(\mu) &= 1 - \sum_{i=0}^{n-1} (-\mu)^i \Big(1 - \frac{1}{2^{n-i}}\Big) \sum_{k=0}^{i} \frac{\omega^{(k)}(\mu)\, \alpha^{(i-k)}(\mu)}{k!\,(i-k)!} \\ &\text{and for } \ell = 1, \ldots, n-1 \\ \omega^{(\ell)}(\mu) &= \sum_{i=0}^{n-1} \mu^{i-\ell} \frac{(-1)^{i+\ell}}{2^{n-i+\ell}} \frac{(n-i+\ell-1)!}{(n-i-1)!} \sum_{k=0}^{i} \frac{\omega^{(k)}(\mu)\, \alpha^{(i-k)}(\mu)}{k!\,(i-k)!}. \end{aligned} \tag{17}$$

As a final curiosity, we present Figure 5. For single-server queuing models it is well-known that the mean waiting time depends (approximately linearly) on the squared coefficients of variation of the interarrival (and service) times; see also Section 4.3 for connections of this model to the classical single-server queue. The results in Figure 5, however, show that for this two-carousel model, the throughput τ, and thus the mean waiting time, is not very sensitive to the squared coefficient of variation of the pick time; it indeed decreases as c_A^2 increases, but very slowly. This phenomenon may be explained by the fact that the waiting time of the server is bounded by one, that is, the time needed for a full rotation of the carousel.

We refrain from giving all results derived for the waiting time distribution in this setting, as they can be found in the papers mentioned so far. One point needs to be stressed though. This technique makes usage of several simplifications (e.g. aggregating orders in one item) and approximations (e.g. modelling various distributions as a phase-type distribution). Some of them are almost unavoidable. For example, a carousel storing items in separate drawers should be evidently modelled with a discrete travel-time distribution; for the application of these results though, one should approximate this distribution by a (continuous) phase-type distribution. However, the effect that some of these assumptions have to the final result is marginal, or at least fully controlled. As was shown in Vlasiou and Adan [113], the error made in computing the distribution of the time the picker has to wait (is not utilised) is bounded.

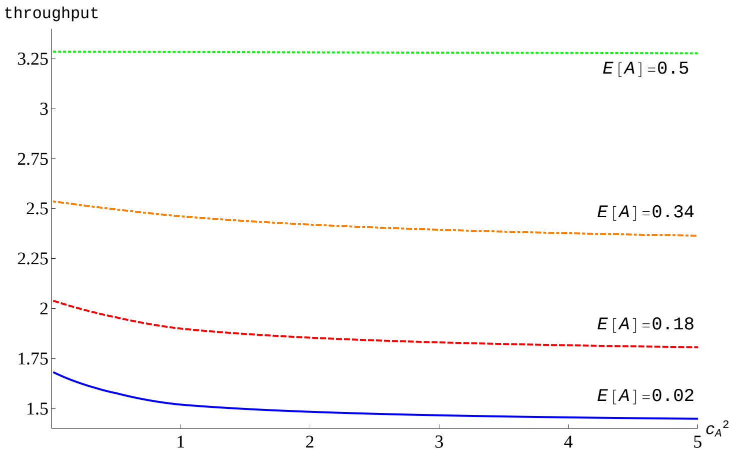

Figure 5: The throughput is almost insensitive to c_A^2.

Error bounds have been studied widely. The main question is to define an upper bound of the distance between the distribution in question and its approximation, that depends on the distance between the governing distributions.

For our model, recall that A, B, and W denote respectively the pick time needed for an item, the travel time of the carousel until this item is reached, and the waiting time of the picker until the carousel stops for the pick. Moreover, F_B represents the distribution of B (and similarly also for W) and $\widehat{F}_B$ is its approximation (such as the phase-type approximation mentioned above). Using this approximation, $\widehat{F}_B$, one can derive analytically an exact solution that is obtained for this case for the distribution of W. Denote this solution by $\widehat{F}_W$. Then the following error bound holds.

Theorem 7 (Vlasiou and Adan [113]). *Let* $\|F_B - \widehat{F}_B\| = \varepsilon$. *Then* $\|F_W - \widehat{F}_W\| \leq \varepsilon/(1 - \mathbb{P}[B > A])$.

In the theorem above, the norm under consideration is the uniform norm. The main ingredient of the proof relies on the fact that the density for the stationary waiting time of (10) can be described in terms of an integral equation that is a contraction mapping. As a result, approximation errors can be bounded.

An almost identical result can be derived in case one approximates the pick time, rather than the travel time. Thus, as this theorem indicates, resorting to approximations yields results of validity that can be controlled, provided that one has an estimation of the error that is being made by the original approximation.

Other results derived for the two-carousel setting include the study of the conditions under which there exists a steady-state distribution [111], the study of the tail behaviour of this distribution under general assumptions for the pick and travel times [111], the derivation of the steady-state distribution for various cases for the distributions of the pick and travel times [111, 112, 116], as well as the time-dependent distribution of the waiting times of the picker for a specific setting for the distributions of the pick and travel times [119]. Moreover, certain types of dependencies between the pick and travel times have also been studied, and the steady-state distribution has been derived for these cases as well [118].

It is worth a mention that such multiple-carousel systems, their mathematical peculiarities, their results and the way those are derived are not limited only to carousel, warehousing, or manufacturing problems. The same equation describing the dynamics of a two-carousel setting describes also the dynamics of a queuing model with two nodes that is applied to situations varying from a university canteen to a surgeon's operating room. For a description of such systems and detailed analysis see Vlasiou *et al.* [21, 111, 112, 119].

What we have discussed so far on multiple-carousel problems is summarised as follows. Multiple-carousel problems are intrinsically different from their single-carousel counterparts. What is of interest in such problems is to strike a balance between the utilisation of the picker and the response time of an order. To date, not much is known about such systems; see Section 6.5 for an exhaustive literature review. A few of these results are simulation studies. However, it is almost inevitable to make use of some simulation or approximations in these problems. The results developed in Vlasiou *et al.* [113, 116] help predict the performance of two-carousel systems and ultimately, combined with the results on e.g nearest-item heuristic or m-step strategies discussed in Section 2, they help design a facility having a specific quality and efficiency target. However, such results are still far from accurate. More research is needed on the subject; specific directions are provided in the next section.

4 Related research areas

The mathematics and models involved in the research regarding carousel systems have surprisingly many connections to broader areas in queuing theory and applied probability. Other than the relation to polling systems which will be explained in detail in Section 5.6, the subjects we have presented so far are connected to the classical single-server queue, to rendezvous networks and layered queues and even to graph theory. In the following, we highlight few of these connections.

4.1 Uniform spacings

The uniform spacings defined in (1) constitute a classical mathematical construction which is very well studied. Uniform spacings have been analysed extensively in two classical review papers by Pyke [96, 97]. In particular, [96] discusses the connections between uniform spacings and exponential random variables that are a main concept in the methodology presented in Section 2. The Markovian property (which is also called the memoryless property) of the exponential distribution is systematically exploited in Operations Research and in particular in queuing theory [4].

Uniform spacings play an important role in mathematical statistics. Mainly, they are used for goodness-of-fit tests which examine how well a sample of data agrees with a given distribution F_0 as its population. The idea of using uniform spacings is based on the integral transformation $x \to F_0(x)$ which reduces the problem to the testing of uniformity of the transformed sample. There is a vast literature on the distributions, limiting behaviour, approximations and bounds for various goodness-of-fit test statistics and empirical processes based on uniform spacings. These investigations are of great mathematical and practical interest. Considerable progress in the area has been achieved in the eighties, but there are still many open problems motivating new studies.

In his detailed review, Pyke [96] distinguishes two main types of goodness-of-fit statistics based on a function of uniform spacings, namely a sum of the form

$$G_n = \sum_{i=1}^{n} g_n(D_i),$$

or a function of the ordered spacings and their ranks. The analysis of the first kind of tests goes back to Le Cam [68] and gives rise to an extensive literature, see e.g. [42, 97, 122] and references therein. Recent progress on multivariate spacings has been reported in [71]. The second type of tests requires the knowledge of the properties of ordered spacings. This subject has been extensively studied; we refer the interested reader to the work by Deheuvels [27] and Devroye [30, 31]. An original discrete version of the problem is analysed by Henze [54] who derives the distribution of the maximal and minimal spacings in lottery tickets.

Apart from the tests mentioned above, there are also tests based on m-spacings which are the gaps between the order statistics $U_{i:n}$ and $U_{i+m:n}$. For the analysis of such test statistics and their asymptotic properties as the number of observations goes to infinity, see, e.g., Del Pino [29], Hall [50], and references therein. The tests based on ordered m-spacings have been also analysed, see, e.g., [7, 28]. More references on this subject and results on the approximations for m-spacings can be found in [44]. For further analysis and applications of various empirical processes based on spacings see Pyke [97], Beirlant et al. [7, 8], Einmahl and Van Zuijlen [34, 35] and references therein.

4.2 Exponential functionals of Poisson processes

Let $X_1, X_2, \dots$ be i.i.d. exponential random variables with mean 1. For any $q \in (0,1)$, define

$$J^{(q)} = (q^{-1} - 1) \sum_{j=1}^{\infty} (q^{-j} - 1)^{-1} X_j,$$

$$I^{(q)} = \sum_{j=1}^{\infty} q^{j-1} X_j.$$

Note that the right-hand side of (8) is exactly $I^{(q)}$ with $q = 1/2$. Likewise, the right-hand side of (9) is the minimum of two independent random variables distributed as $J^{(1/2)}$. We see that the sums of independent exponentials with exponentially decreasing coefficients play an important role in the limiting results for the travel time in carousel systems as the number of items goes to infinity. Specifically, these random variables appear if we consider the difference between the travel time and one complete carousel rotation, and then scale this quantity linearly with the number of items.

Now let $N(t)$ be a standard Poisson process. Then we can write $I^{(q)}$ as an exponential functional associated with $N(t)$:

$$I^{(q)} = \int_0^{\infty} q^{N(t)}\, dt.$$

The functional $I^{(q)}$ has been intensively studied in the literature. Its density was obtained independently in [11, 32], and in [81] for $q = 1/2$. Carmona *et al.* [23] derived a density of $\int_0^\infty h(N(t))\,dt$ for a large class of functions $h : \mathbb{N} \longrightarrow \mathbb{R}_+$, in particular, for $h(n) = q^n$. Bertoin and Yor [11] found the fractional moments of $I^{(q)}$. If $i^{(q)}(t)$ is a density of $I^{(q)}$, then $i^{(q)}(t)$ and all its derivatives equal 0 at point $t = 0$. This implies that all moments of $1/I^{(q)}$ are finite. However, for $q = 1/e$, it was proved in [10] that $1/I^{(1/e)}$ is not determined by its moments. Guillemin *et al.* [48] found the distribution and the fractional moments of the exponential functional

$$I(\xi) = \int_0^\infty e^{-\xi(t)}\,dt, \tag{18}$$

where $(\xi(t), t \geq 0)$ is a compound Poisson process.

The distribution function of $I^{(q)}$ and $J^{(q)}$ has an interesting asymptotic behaviour in the neighbourhood of zero. Bertoin and Yor [10] obtained the following logarithmic asymptotics:

$$\log i(t) \sim -\frac{1}{2}(\log(1/t))^2 \quad \text{as} \quad t \to +0,$$

where $i(t)$ is a density of

$$I = \int_0^\infty e^{-N(t)}\,dt = \sum_{j=1}^\infty e^{-j} X_j.$$

The exact asymptotic behaviour has been derived by Litvak and van Zwet [82]. Compared to the logarithmic asymptotics, their formula contains several additional terms and reveals an unexpected oscillating behaviour involving theta-functions. The explanation of why the oscillations appear seems to lie in the sort of a 'binary tree structure' of the functional I, whose coefficients are negative powers of e. Later on, Robert [98] and Mohamed and Robert [86] found that such oscillating asymptotic behaviour is a typical feature of algorithms with a tree structure. This phenomenon is compelling and deserves further studies.

Exponential functionals of Poisson process and, more generally, of Lévy processes, appear in a number important applications. For instance, they are relevant to the analysis of randomised algorithms [38] and in mathematical finance [12]. In [32] and [48] the exponential functionals of Poisson processes, and, respectively, of compound Poisson processes, play a key role in the analysis of the limiting behaviour of a Transmission Control Protocol connection for the Internet. We refer to the survey [12] for further applications, results and references. The study of exponential functionals of Lévy processes are a current subject of research, see e.g. [75], [94].

4.3 Lindley's recursion

One of the most intriguing mathematical observations that arise when studying the two-carousel model presented in Section 3 is that Recursion (10) differs from the original Lindley's recursion [74], which is $W_{n+1} = \max\{0, B_n - A_n + W_n\}$, only in the change of a plus sign into a minus sign. At the right-hand side of these two recursions, the sign in front of W_n is reversed. Lindley's recursion describes the waiting time W_{n+1} of a customer in a single-server queue in terms of the waiting time of the previous customer, his or her service time B_n, and the interarrival time A_n between them. It is one of the fundamental and most well-studied equations in queuing theory. For a detailed study of Lindley's equation we refer to Asmussen [4], Cohen [24], and the references therein.

In the applied probability literature there has been a considerable amount of interest in generalisations of Lindley's recursion, namely the class of Markov chains, which are described by the recursion $W_{n+1} = g(W_n, X_n)$. The model in Section 3 is a special case of this general recursion and it is obtained by taking $g(w, x) = \max\{0, x - w\}$. Many structural properties of the recursion $W_{n+1} = g(W_n, X_n)$ have been derived. For example Asmussen and Sigman [5] develop a duality theory, relating the steady-state distribution to a ruin probability associated with a risk process. For more references in this domain, see for example Borovkov [18] and Kalashnikov [61]. An important assumption which is often made in these studies is that the function $g(w, x)$ is non-decreasing in its main argument w. For example, in [5] this assumption is crucial for their duality theory to hold. Clearly, in the special case of $g(w, x) = \max\{0, x - w\}$ which is discussed in Section 3, this assumption does not hold. This fact produces some surprising results when analysing the equation.

The implications of this 'minor' difference in sign are rather far reaching. For example, in Section 3 we have presented two results in Theorems 4 and 5, where we have seen that the waiting time of the picker can be solved for explicitly. For Lindley's recursion, i.e. with a plus sign instead of minus sign for W in stationarity, this case correspond to the stationary waiting time in a classical single-server PH/U/1 queue. However, this equation has no simple solution for Lindley's recursion, while we have derived a closed-form expression for the carousel recursion. Furthermore, numerical results (see also Figure 5) show that for this carousel model the mean waiting time is not very sensitive to the coefficient of variation of the pick time, which is in complete contrast to Lindley's recursion. For these reasons, we believe that it is interesting to investigate in detail the impact on the analysis of such a 'slight' modification to the original equation. In this section, we highlight some of the differences of these two models.

4.3.1 Stability

For the single-server queue, i.e. Lindley's recursion, it is well-known [4, Ch. III.6] that the random variables representing waiting times of customers converge in distribution (and in total variation) when the mean of the associated random walk is less than zero, or equivalently when the traffic intensity ρ is less than 1; i.e., when $\mathbb{E}[B] < \mathbb{E}[A]$, where we recall that B is the generic service-time random variable, and A is the generic interarrival-time random variable.

For the two-carousel model, though, which is given by Recursion (10), the situation is slightly different. In case $\mathbb{P}[B < A] > 0$, the stochastic process $\{W_n\}$ is an aperiodic, (possibly delayed) regenerative process with the time points where $W_n = 0$ being the regeneration points. Moreover the process has a finite mean cycle length. To see this, let

$X_n = B_n - A_{n-1}$, define the stopping time $\tau = \inf\{n \geqslant 1 : X_{n+1} \leqslant 0\}$, and observe that a generic cycle length is stochastically bounded by τ and that

$$\mathbb{P}[\tau > n] \leqslant \mathbb{P}[X_k > 0 \text{ for all } k = 2, \dots, n+1] = \mathbb{P}[X_2 > 0]^n.$$

Moreover, we have that $\mathbb{P}[X_2 > 0] < 1$ because of the condition $\mathbb{P}[B < A] > 0 \Leftrightarrow \mathbb{P}[X < 0] > 0$. Therefore, from the standard theory on regenerative processes it follows that the limiting distribution exists and the process converges to it in total variation. Through coupling, stability can be shown also for the case where $\mathbb{P}[X < 0] = 0$; see [111] for details. We see thus that while for Lindley's recursion the stability condition is given by $\mathbb{E}[X] < 0$, for Recursion (10) *stability always holds*; moreover, excluding the deterministic case, we have convergence in total variation.

4.3.2 Tail behaviour

For Lindley's recursion, there has been a substantial amount of investigations on the behaviour of $\mathbb{P}[W > x]$ as $x \to \infty$, the state of the art can be found in [67]. Results of this type for Recursion (10) have been derived in [111]. If the right tail of e^X is regularly varying of index $-\gamma$ (see [14] for background), then

$$P(W > x) \sim \mathbb{E}[e^{-\gamma W}]\mathbb{P}[X > x].$$

If the right tail of e^X is of rapid variation (see again [14]), then

$$P(W > x) \sim \mathbb{P}[W = 0]\mathbb{P}[X > x].$$

In both equations, we use the notational convention $f(x) \sim g(x)$ to denote $f(x)/g(x) \to 1$ as $x \to \infty$. Note that the class of distributions covering these results include all phase-type distributions, as well as the Weibull, Gamma, Lognormal and Pareto distributions. Moreover, these results indicate that large values of W are caused by a single large value of X. This is contrasting with the qualitative picture for Lindley's recursion, where a large value of W is most likely caused only by a single big jump only in the case where X is heavy-tailed. If X is light-tailed (for example phase type), then a large value of W is the cause of a more intricate event involving a change of measure; see [4] for background.

A natural question is whether it is possible to unify the results for Lindley's recursion and Recursion (10). This is possible by considering a recursion that has a minus before W_n (cf. Recursion (10) too) only with probability $1-p$, $p \in [0, 1]$, and has a plus before W_n (i.e. equal to Lindley's recursion) with probability p. For this recursion, the tail behaviour has been studied in [117] under assumptions similar to the ones made in [67]. To summarise the qualitative picture emerging from that paper, the tail behaviour for the unified recursion with $p \in [0, 1]$ converges continuously to the results for Recursion (10) (i.e. if $p = 0$) for the heavy-tailed case, while it has a discontinuity for $p = 1$; for the so-called Cramér case the result is reversed: the unified recursion is continuous for $p = 1$ and discontinuous for $p = 0$, while for the intermediate case (where X is light tailed but does not satisfy the Cramér condition) the results for the unified recursion are continuous at both end-points.

4.3.3 Time-dependent behaviour

It is well known that for Lindley's recursion, the time-dependent waiting-time distribution is determined by the solution of a Wiener-Hopf problem, see for example [4] and [24]. Recursion (10) though, regularly gives rise to *generalised* Wiener-Hopf equation. For example, in [111] we have derived a generalised Wiener-Hopf equation for the density of the stationary waiting time, while [119] contains an integral equation for the generating function of the distribution of W_n that is equivalent to a generalised Wiener-Hopf equation, which cannot be solved in general. In Noble [88] it is shown that such equations can sometimes be solved, but a general solution, as is possible for the classical Wiener-Hopf problem (arising in Lindley's recursion), seems to be absent.

This makes it appear that (10) may have a more complicated time-dependent behaviour than Lindley's recursion. However, a point we make in [119] is that this is not necessarily the case. Thus, Equation (10) is a rare example of a stochastic model which allows for an *explicit* time-dependent analysis. The reason is that, if B_1 has a phase-type distribution, we can completely describe (10) in terms of the evolution of a finite-state Markov chain.

We shall refrain from giving all results on the time-dependent behaviour of (10) or their differences from the classical Lindley recursion for the single-server queue, as these results have been well documented elsewhere [111]. Here, we simply list the major findings.

Other than deriving the time-dependent waiting time distribution for (10) under the assumption that the random variables B_i are phase-type distributed, one can derive explicit expressions for the correlation between two waiting times. It results that the covariance function $c(k)$ between two waiting times with lag k converges to zero geometrically fast in k. This is consistent with the fact that the distribution of W_n converges geometrically fast to that of W, cf. Vlasiou [109]. One of the properties of $c(k)$ is that it is non-negative if k is even and non-positive if k is odd. If in addition, the random variable $X = B - A$ has a strictly positive density on an arbitrary interval, then the inequalities are strict. In contrast, the literature on the covariance function of the waiting times for the single-server queue seems to be sporadic. For the G/G/1 queue, Daley [26] and Blomqvist [16, 17] give some general properties. In particular, in [26] it is shown that the serial correlation coefficients of a stationary sequence of waiting times are non-negative and decrease monotonically to zero.

As we have mentioned before, $\{W_n\}$, as given by (10), is a regenerative process; regeneration occurs at times when $W_n = 0$. Other transient results relate to the length of a generic regeneration cycle C. For Recursion (10), we do not need to resort to the usage of generating functions, as is necessary when analysing the corresponding quantity in Lindley's recursion. Note that the interpretation of C for the carousel model is completely different from the corresponding quantity for Lindley's recursion. There, C represents the number of customers that arrived during a busy period. In the carousel setting, C represents the number of pauses a picker has until he needs to pick two consecutive orders without any pause. In this sense C can be seen as a "non-busy period".

4.4 The machine repair problem

When deriving Equation (10), one of the main assumptions we have made, which led to this particular form for the equation is that the picker is not allowed to pick two consecutive orders at the same carousel and must alternate between the two carousels (thus picking all odd-numbered orders from one carousel and all even-numbered orders from the other). This condition is crucial. If we remove this condition, then under certain distributional assumptions, the problem turns out to be the classical machine repair problem, and certain analogies between these two models arise.

In the machine repair problem, there is a number of machines working in parallel (two in our situation, corresponding to the two carousels) and one repairman (corresponding to the picker), who serves the machines when they fail. The machines are working independently and as soon as a machine fails, it joins a queue formed in front of the repairman where it is served in order of arrival. A machine that is repaired is assumed to be as good as new. The machine repair problem, also known as the computer terminal model (see for example Bertsekas and Gallager [13]) or as the time sharing system (see, e.g., Asmussen [4, p. 79] or Kleinrock [64, Section 4.11]) is a well studied problem in the literature. It is one of the key models to describe problems with a finite input population. A fairly extensive analysis of the machine repair problem can be found in Takács [104, Chapter 5]. In [112] we compare the two models and discuss their performance.

The issue that is usually investigated in the machine repair problem is the waiting time of a machine until it becomes again operational. In the situation described in Section 3 though, we are concerned with the waiting time of the repairman. It is quite surprising that although the machine repair problem under general assumptions is thoroughly treated in the literature, this question remains unanswered. In the machine repair problem the operating time of the machine is usually more valuable than the utilisation of the repairman, which might explain why the classical literature has been mainly focused on performance measures related to the machines.

In [112] the waiting time of the repairman is derived under the assumption that 'rotation' times follow a phase-type distribution while 'pick' times are generally distributed. Moreover, it is shown that the random variables for the waiting time for the picker/repairman in the two models *are not stochastically ordered*. However, on average, the alternating strategy connected to the two-carousel model leads to longer waiting times for the picker, which readily implies that the *throughput of the machine repair model is bigger*. Furthermore it is shown that the probability that the picker does *not* have to wait is *larger* in the two-carousel alternating system than in the machine repair (i.e. non-alternating) model one. This result is perhaps counterintuitive, since the inequality for the mean waiting times of the picker in the two situations is reversed. Regarding the relationship between the i-th waiting time of the picker in the two-carousel alternating model (denote this by W_i^{A}), and that of the repairman in the machine repair problem (let this be given by W_i^{NA}), an immediate corollary of the results stated above is as follows.

Corollary 2. *For all i, $\sum_j^i W_j^{\mathrm{A}} \geqslant_{st} \sum_j^i W_j^{\mathrm{NA}}$.*

So, although the stationary random variables W^{A} and W^{NA} are not stochastically ordered, the partial sums of the sequences W_i^{A} and W_i^{NA} are. Moreover, a conjecture stated in [112] suggests that a direct application of the Karlin-Novikoff cut-criterion (cf. Szekli [103]) leads to an *increasing convex ordering*, namely:

Conjecture. *For all increasing convex functions ϕ, for which the mean exists, we have that*

$$\mathbb{E}[\phi(W^{\mathrm{NA}})] \leqslant \mathbb{E}[\phi(W^{\mathrm{A}})].$$

4.5 Rendezvous networks and layered queues

The essence of layered queueing (a special case of which is rendezvous networks) is a form of simultaneous resource possession [89].

In its most simple form in computer science applications, in a rendezvous network, a task may serve requests in two rounds (phases) of service. In computer applications, tasks or applications may act both as customers that needs service from other tasks and as servers to other tasks too. As a naive example, think of an application that adds up numbers. It acts both as a server, accepting requests from other applications that need numbers added, and as a customer, requiring service from the central processing unit. One can imagine that tasks are ordered in several levels or layers. Tasks have directed arcs to other tasks at lower layers to represent service requests. A request from an task (client) to a lower-layered task (server) may return a reply to the requester (a synchronous request, or rendezvous). While in the first phase (i.e. in the rendezvous) the client is blocked and the server merely continues the thread of control of its client. However, in the second phase the client has an independent thread of control of its own. For example, Task A makes a request to Task B which then makes a request to Task C. While Task C is servicing the request from Task B, Tasks A and B are both blocked [39]. Among the advantages of the rendezvous is efficiency, since it provides communication without the effort of buffer management and the message copying associated with asynchronous communication. However, some potential for concurrency is lost, and there may be performance-impairing bottlenecks when a key task spends long periods send-blocked [87]. Special approximations are needed to solve queueing models which contain a two-phase server, because the second phase effectively creates a new customer in the queueing network, violating the conditions of product form queueing [39].

Distributed or parallel software with synchronous communication via rendezvous is found in client-server systems and in proposed Open Distributed Systems, in implementation environments such as Ada, V, Remote Procedure Call systems, in Transputer systems, and in specification techniques such as CSP, CCS and LOTOS. The delays induced by rendezvous can cause serious performance problems, which are not easy to estimate using conventional models which focus on hardware contention, or on a restricted view of the parallelism which ignores implementation constraints. Stochastic Rendezvous Networks are queueing networks of a new type which have been proposed as a modelling framework for these systems. They incorporate the two key phenomena of *included service* and the *second phase of service* mentioned above. The main work on rendezvous networks focuses on Mean Value Analysis and gives approximate performance estimates. This method has been applied to moderately large industrial software systems [126].

A Layered Queuing Network (LQN) model is a canonical form for extended queueing networks that represent layered service. In a layered queue a server, while executing a service, may request a lower layer service and wait for it to complete. Thus, in LQNs there exist entities that have a dual role; they act as servers to other entities of a lower layer and as customers to higher layered entities. The service time of the upper server includes the queueing delay and service time of the lower server, and this may extend through multiple layers. LQN was developed for modelling software servers, with for example blocking remote procedure calls to lower layer software servers, however it applies to any extended queueing network in which resource usages are nested, lower layer usages within higher layer usages [89].

The two-carousel model we have presented in Section 3 is a layered queue, and in particular a rendezvous network. To see this, organise the system as follows. The items that are stored on the carousel and have to be picked comprise the lowest layer. Carousels are in the middle layer, while the picker is put in the highest layer. One may view the rotation time of a carousel as a first phase of service for the item that will be picked. The carousel (middle layer) acts in this case as a server. However, the second phase of service (the actual picking) does not necessarily happen immediately (rendezvous). The item might have to wait for the picker to return from the previous carousel – cf. Recursion (10). At this stage, the carousels act as customers waiting to be served by the higher layer, the picker. We see thus that each carousel acts both as a server (rotating items to the picking location) and as a customer (waiting until the picker completes his task before the carousel can resume its role as a server, bringing the next item to the picking location).

Layered systems are quite unknown outside the computer-science community. E.g., in [95] it is mentioned that "this paper presents a model, never studied before in the queueing literature, of a system of two connected queues where customers of one queue act as the servers of the other queue" – a comment that may very well be valid outside the computer-science literature. The analysis of Recursion (10), as it developed in [21, 111, 112, 113, 116, 115, 119] as well as [95] are the only papers we are aware of that deal with LQNs using analytic and probabilistic tools, and admittedly all the aforementioned work on the two-carousel model had not made the connection between this model and layered queues.

4.6 Maximum weight independent sets in sparse random graphs

However unusual it might be in queuing theory to encounter a non-increasing Lindley-type recursion, Recursion (10) appears in problems involving the computation of the distribution of the maximum weight of an independent set in a sparse random graph.

Consider an n-node sparse random r-regular graph (i.e. a graph selected uniformly at random from the set of all graphs on n nodes in which every node has degree r). An independent set is a set of nodes of the graph where no two nodes in the set are connected by an edge. Suppose that the nodes of the graph are equipped with some nonnegative weights w_i which are generated independently according to some common distribution F_w. One may be interested for example in the limits of maximum weight independent sets and matchings in sparse random graphs for some types of i.i.d. weight distributions. Then Recursion (10) corresponds exactly to the one related to the weight distribution in an 1-regular graph; see [40]. Moreover, if one considers r-regular graphs, then the corresponding recursion giving the weight distribution in this case is similar to the one corresponding to the waiting time of a picker serving r carousels; see (19). The crucial difference in this case is in (19) the random variables W_{n+1} and W_n appearing at the right-hand side of the recursion are not independent, while the corresponding variables in the recursion related to r-regular graphs are independent; see [40, Eq. (3)]. It would be interesting to investigate the connections between the research areas of warehouse logistics and graph theory.

5 Further research

5.1 Considering different item storage schemes

As mentioned in Section 2, as of yet the case of independent uniformly distributed items locations is the only known scenario where the travel time can be evaluated analytically by applying a systematic mathematical approach. It is important to develop methods to obtain statistical characteristics of the travel time under more realistic assumptions on the items locations. As we discuss below in Section 6.2, there are not many results in this direction in the literature. The non-uniform distributions of pick positions and especially the correlations between the items in an order lead to challenging mathematical problems. We believe that no feasible analytical solutions can be obtained in most of the realistic models. Thus, the problem calls for well justified heuristics and efficient numerical methods.

5.2 Further topics in two-carousel problems

The model we have considered in Section 3 applies to a two-carousel system that is operated by a single picker. Two-carousel systems have received some attention in the literature (cf. Section 6.5) but many questions remain open. A line of research is directed towards studying the performance of two-carousel systems under various storage-assignment policies (randomised or not), for various pick/travel time strategies and heuristics (sequential picking, nearest-item heuristic, m-step strategies, etc.), for single- or dual-command cycles, and for open- and closed-loop strategies. Here a single command cycle assumes a single operation, such as only storage or only retrieval. In a dual-command cycle, a storage and retrieval are combined to efficiently use the time of the operator. Furthermore, an open-loop strategy implies that the carousel remains stationary at the point where the last item was retrieved (awaiting the next order to be fed), while under the closed-loop strategy the carousel returns to a predefined point after the retrieval of an order is completed.

As explained in Section 3, two-carousel systems differ in nature and in analysis from the corresponding one-carousel problems even when studied under the same assumptions on the various storage, pick, cycle, and starting-point strategies that are followed. Since two-carousel systems perform in broad terms better than single-carousel systems [57], studying the expected increase of the throughput of the system can help answer questions of financial nature, such as whether the benefits from the increased throughput justify the increased cost of building and operating a two-carousel system.

5.3 Extensions to multiple carousels

The model discussed in Section 3 can be extended to the case of multiple carousels as follows. For instance, consider the situation where a single picker operates three carousels. Apart from the number of carousels, all other characteristics of the model remain the same as in Section 3. That is, we consider again an infinite queue of orders that need to be picked, we have again a rotation stage and a picking stage for each item. Moreover, as before, the picker serves all carousels cyclically. For three carousels, this leads to the recursion

$$W_{n+2} = \max\{0, B_{n+2} - W_{n+1} - A_{n+1} - W_n - A_n\}, \tag{19}$$

where now the variables appearing at the right-hand side are not independent of one another, as was the case for all variables appearing at the right-hand side of Recursion (10). We may assume for convenience that the sequences $\{A_n\}$ and $\{B_n\}$ are independent among them and between them. Furthermore, we note that the waiting times W_n and W_{n+1} are not independent. The state of the system can be modelled e.g. as a two-dimensional Markov chain, where apart from the waiting time of the picker for the n-th item that will be picked we also need to incorporate the remaining rotation time of the next carousel to be served. Evidently, if the rotation times are assumed to be exponentially distributed, the system (for three or more carousels) can be analysed explicitly by similar techniques as the ones applied in Chapter 4 of [110], although it is doubtful how realistic such an assumption is.

Naturally, if one considers a system with multiple carousels or stations, one can think about optimisation questions. Namely, as the number of carousels increases, the waiting time of the picker is expected to decrease. After serving a long series of carousels cyclically, when you return to the beginning of the cycle, with high probability the item to be picked will have reached the origin. This implies that an item will have to wait for the picker at the origin more frequently than in the two-carousel system, which means that the throughput of a single carousel decreases. Intuitively, as the number of carousels increases to infinity, the utilisation of the picker increases to one, while the throughput of each individual carousel decreases to zero. Given a setting, one might wonder how many carousels a single picker can operate so that we maximise both the throughput of the carousels and the utilisation of the picker simultaneously.

5.4 Incorporating picking strategies to multiple carousel problems

The ultimate goal of the analysis of carousel systems is to provide a mathematical model that adequately describes the reality and, at the same time, can be efficiently evaluated either analytically or numerically. At the moment, the literature on a single carousel has advanced enough to characterise the travel time with great precision, at least for independent uniform items locations. However, as mentioned above, single carousel systems are rarely used in modern warehouses. Clearly, multiple carousel models are more relevant from a practical point of view. The drawback is that such models tend to become extremely complex. Until now the studies of multiple carousel systems were either solely based on simulations or employed analytical models that involved simplifying assumption on the order picking time. For instance, in Section 3 we assumed that each order is collected within a random time that has the same distribution for each order. This is definitely a simplifying assumption, because, for instance, the orders may differ in size, and as we saw in Section 2, the distribution of the travel time depends on the number of items to be collected. Further literature on multiple carousels discussed in detail in Section 6.5 also involves significant simplifications of the real-life situation.

In this respect, a major challenge for future studies is to develop a unified approach for rigorous studies of real-life automated storage and retrieval systems. Such an approach is expected to involve the methods proposed so far for single and multiple carousels. In Sections 2 and 3 we presented well-developed methodologies for analytical studies of order picking in one and two carousel units. Thus, an important topic for further research is to combine these two problems in one integrated study of multiple carousel systems. One may hope to obtain interesting analytical results in this direction because of the analytical nature of both methodologies. However, the problems of combining these two settings are challenging. In Section 2 we have seen that the travel time distribution can be of a complicated form, while the results in Section 3 often rely on assumption such as exponential or phase-type pick times (recall that the travel time needed to pick all items corresponds to the pick time for orders aggregated in one item). Also, as mentioned above, the travel time depends on the size of the order, while the technique of aggregating orders in one item has made use of the assumed independence between pick times and rotation times (while one might expect that in orders with multiple items, long travel times might be correlated to orders with multiple items and thus to shorter rotation times to the first item in the order). Eventually, one will have to resort to the development of reasonable algorithms rather than the derivation of exact distributions. In this respect, we emphasise again that algorithmic studies of realistic carousel models constitute an important part of further research.

5.5 Considering the order arrival process

It is also interesting to study if single or multiple carousel systems can be analysed in case there is an arrival process according to which the orders arrive. If orders arrive according to a Poisson process in front of the carousel, what can be said for the waiting time of the picker? This question can also be combined with a non-alternating system, where the picker serves the first carousel that has completed the rotation to the next item on that carousel that needs to be picked, or with Bernoulli-type requests, where the picker has to serve with a certain probability the "first" carousel and with the complementary probability the "other" carousel (potentially waiting for an item if none is present at the designated carousel). For each case, one should also consider the stability of the system in case the arrival rate of the orders is less than the throughput of the system with an infinite queue of orders.

5.6 Polling systems

A polling system is comprised of a number of customer queues that are served in an order by a single server. In the literature on polling systems, the polling system with two queues where at each queue the server serves exactly one customer before switching to the other queue is often referred to as the *1-limited alternating-service* model. The model described in Section 3 is closely related to such polling systems. The two main differences are the existence of an extra stage, the rotation time of the carousel, and the absence of an arrival process for the orders. In polling systems one deals only with one stage, which in the terminology of Section 3 is represented by the picking stage. Extending the model of Section 3 by introducing an arrival process of the orders as suggested above, is equivalent to studying an 1-limited alternating-service model with switch-over times between the stations (which can be seen as being equivalent to the rotation time towards the single item).

The polling model with two queues, Poisson arrivals, and no switch-over times has first been studied by Eisenberg [36], where the main question studied is the queue-length distribution, as is often the case in the literature on polling systems. Eisenberg [36] gives the generating function for the stationary joint distribution of the two queue sizes. Cohen and Boxma [25] study the single server queue with two Poissonian arrival streams and no switch-over times. The server handles alternatingly a customer of each queue if the queues are not empty and it is assumed that customers of the same arrival stream have the same service time distribution. It is shown that the determination of the joint queue-length distribution at the departure epoch can be formulated as a Riemann-Hilbert boundary problem that can be completely solved for general service time distributions. Introducing switch-over times increases the complexity of the problem. In Boxma [19] the analysis is extended to include switch-over times of the server between queues, under the restriction that both queues have identical characteristics. This work is further extended in Boxma and Groenendijk [20], where the authors no longer request that both queues have identical characteristics. It is assumed that service times and switch-over times are generally distributed.

The literature on polling systems with alternating service is not limited to the references above but is rather extensive; see [46, 59, 91] for some references. It seems though, that the question regarding the waiting time of the picker for the 1-limited polling system with two carousels has not been considered outside the scope of [110]. Thus, introducing an arrival process for the orders in the model of Section 3 complements the existing literature on polling systems and forms a challenging problem. The interesting feature then is that the switch-over time between two queues depends on the current picking time. Again, the results from Section 2 can be incorporated into the model for adequate description of order picking times.

An extension considered in polling systems is the *k-limited* service policy, where the server switches queues after having served at most k customers in one queue. For an extensive list of references on k-limited polling systems see Van Vuuren and Winands [107]. The main focus of the existing literature is again on the queue-length distribution of all stations. As the authors note in [107], "to this very day, not only hardly any exact results for polling systems with the k-limited service policy have been obtained, but also their derivations give little hope for extensions to more realistic systems". It is worth considering the k-limited service discipline under the exact setting we have established in Section 3, where now the focus is on the distribution of the waiting time of the server.

6 Literature overview

In the following, we classify the literature on carousels according to the main theme handled. This taxonomy allows for a better overview of the variety of the subjects examined. A crucial distinction is made between systems that involve a single carousel and systems with multiple carousels. The first four categories presented relate to single-carousel systems, while systems with multiple carousels are examined later on.

6.1 Storage

The performance of a carousel system depends greatly upon the way it is loaded and the demand frequency of the items placed on it. An effective storage scheme may reduce significantly the travel time of the carousel. Several strategies have been followed in practice to store items on a carousel. The simplest strategy is to place the items randomly on the carousel. *Randomised policies* have been examined extensively [55, 76], and various performance characteristics have been derived under the assumption that the items are uniformly distributed on the carousel.

One way to improve the throughput of a carousel system is to adopt a storage policy other than the randomised assignment policy. Ha and Hwang [49] have studied what they call the "two-class-based storage", which is a storage scheme that divides the items in two classes based on their demand frequency. The items with a higher turnover are randomly assigned to one continuous region of the carousel, while the less frequently asked items occupy the complementary region. The authors show by simulation that the two-class-based storage can reduce significantly the expected cycle time, both in the case where a cycle is a single pick or storage of an item (single-command cycle), and in the case where a cycle consists of the paired operations of storing and retrieving (dual-command cycle). The same authors in [56] examine the effects of the two-class-based storage policy on the throughput of the system, and present a case where there is a 16.29% improvement of this policy over the randomised policy.

Another storage scheme is suggested by Stern [102]. Assignments are made using a *maximal adjacency principle*, that is, two items are placed closely if their probability of appearing in the same order is high. The author evaluates this storage assignment analytically by using a Markov chain model he develops.

The *organ pipe arrangement* for a carousel system is introduced in Lim *et al.* [73] and is proven to be optimal in Bengü [9] and in Vickson and Fujimoto [108] under a wide variety of settings. The organ pipe arrangement has been widely used in storage units, such as magnetic tapes [15] and warehouses [83]. This arrangement is based on the classical mathematical work of Hardy, Littlewood and Polya [51]. Their concept is used in [15] to minimise the expected distance travelled by an access head as it travels from one record to another. Various optimality properties of this arrangement have been proven; see for example Keane *et al.* [62] and references therein.

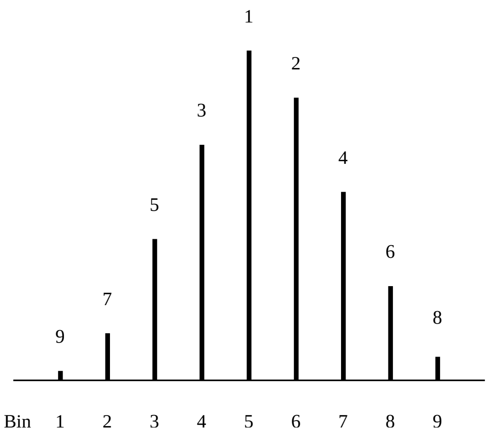

Figure 6: Illustration of the organ pipe arrangement, where the upper numbers indicate the frequency ranking of an item.

In carousel systems, the organ pipe arrangement places the item with the highest demand in an arbitrary bin, the items with the second and third highest demands in the bin next to the first one but from opposite sides, and sequentially all other items next to the previous ones, where the odd-numbered items according to their frequency are grouped together and placed next to one another in a decreasing order from the one side of the most frequent item (and similarly the even-numbered items are grouped together and placed to the other side). Figure 6 illustrates the organ pipe arrangement. The numbers at the top indicate the ranking of an item in a decreasing order of frequency.

Park and Rhee [93] study the system throughput and the job sojourn times under the organ pipe arrangement, where independent one-item orders arrive according to a Poisson process. They explicitly quantified the gain of the organ pipe arrangement compared to random assignment and showed that this gain grows with the 'skewness' in the demand distribution.

Abdel-Malek and Tang [1] study the travel times in carousels with N bins and the organ pipe arrangement under the assumption that each order consists of one item and a sequence of orders forms a Markov chain: if the current order requires bin p then the next order requires bin q with probability P_{pq}. The objective is to find the optimal assignment, which minimises the average travel time. Their extensive numerical experiments show that although the organ-pipe arrangement is not optimal in this setting, it performs very close to optimality in a wide range of system parameters. The optimal solution in [1] is determined by solving a *quadratic assignment problem*. The quadratic assignment problem is a well-known optimisation problem on choosing an optimal permutation of

n coordinates of a vector $\mathbf{x} = (x_1, \ldots, x_n)$ in order to minimise $\mathbf{x}C\mathbf{x}^T$, where C is a cost matrix. Such problems have a long history started with the work of Koopmans and Beckmann [66]. Litvak [78] shows by experimental studies and by providing asymptotic results for large orders that in general the optimal storage depends on the order size. Moreover, the organ-pipe storage is disadvantageous when an order is large.

Another question related to storage is about the number of items of each type that should be stored on the carousel in order to maximise the number of orders that can be retrieved without having to reload. This question is examined in Jacobs *et al.* [60], where the authors propose a heuristic that yields a reasonable solution, the error of which can be bounded. This method has been improved by Yeh [127], where a more accurate solution is obtained, and further on by Kim [63], where it is observed that the heuristic described in [127] does not always lead to the optimal solution. The author constructs an algorithm that yields the optimal solution. This algorithm is further improved in Li and Wan [72]. This line of research has been continued in the recent paper by Hassini [52]. In the formulation used in Jacobs *et al.* [60], the author determines the optimal allocation. Along with exact optimal solutions for deterministic and stochastic demand, [52] also provides heuristics that perform close to optimal.

6.2 Picking a single order

One of the most important performance characteristics of a carousel system is the total time to pick an order. The total time to retrieve all items of an order may be expressed as a sum of the total time that the carousel is travelling plus the total time that the carousel is stopped for picking. The latter is effectively the total pick time, and it is not affected by the sequence in which we choose to retrieve the objects. However, the total travelling time greatly depends upon the retrieval sequence. The analysis of the travel time under various strategies is, in general, a non-trivial problem. This problem, however, has been resolved for independent and uniformly distributed item locations [76], as we discussed in detail in Section 2.

Various picking strategies have been proposed. Bartholdi and Platzman [6] assume a discrete model and study the performance of an algorithm and three heuristics that determine an efficient, but not necessarily optimal, sequence of retrieving all items. A heuristic is a simpler, non-optimal procedure that is based on a specific strategy. The heuristic methods proposed are the *nearest-item heuristic*, where the next item to be picked is always the one that is closer to the picker at any given moment, the *shorter-direction heuristic*, where the carousel chooses the shortest direction between the route that simply rotates clockwise and the route that rotates counterclockwise, and the *monomaniacal heuristic*, that always chooses to rotate to the right and pick items sequentially. The *optimal retrieval* algorithm that is presented enumerates all possible paths; therefore, it is guaranteed to find the quickest sequence in which to retrieve a single order.

In [6] the authors prove among other things that the travel time under the nearest-item heuristic is never greater than one rotation of the carousel. Litvak *et al.* [81] provide the upper bound of $1 - 1/2^n$ full rotations, where n is the number of items in the order, and show that the new upper bound is tight. Litvak and Adan [79] obtained the distribution and the asymptotic properties of the travel time under the nearest-item heuristic for uniformly distributed independent items locations. These results, based on properties of uniform spacings, have been discussed in detail in Sections 2.1, 2.4. In [81], the first two moments of the travel time and the distribution of the number of turns are computed recursively by conditioning on the event that there is an empty space of size x on one side of the picker's current position. We presume that such methods may lead to the travel time distribution in some special cases with non-uniform items locations.

Another interesting picking strategy that has been already discussed in Section 2.2 is the so-called *m-step strategy*, where the carousel chooses the shortest route among the ones that change direction at most once, and only do so after collecting at most m items. In case of independent uniformly distributed items locations the average travel time under the m-step strategy is smaller than the one under the nearest-item heuristic already for $m = 2$; see [80]. The results by Litvak and Adan [80] on the m-step strategies have been presented in Section 2.2. In an earlier paper, Rouwenhorst *et al.* [100] apply analytical methods to study the case when $m \leq 2$. This means that the carousel changes direction after collecting at most two items. They interpret m-step strategies as stochastic upper bounds for the minimal travel time and present convincing numerical results on the excellent performance of such strategies.

Wen and Chang [124] model the carousel as a discrete bidirectional loop and assume that the time to move between the bins of a shelf is *not* negligible. They propose three heuristic solution procedures and compare their performance. An earlier version of this work can be found in Wen [123].

Ghosh and Wells [43] model the carousel as a continuum of *clusters* and *gaps*, where a cluster is a segment on the circle that corresponds to a series of locations that have to be visited for the retrieval of an order, while a gap is the segment of the circle between two clusters. The authors develop two algorithms to find optimal retrieval strategies. In particular, to find an optimal path, they avoid a complete enumeration by noticing that a turn can never be made after covering more than 1/3 length of the carousel.

Stern [102] studies properties of the optimal, i.e. minimal, picking sequence both for the *open-loop* strategy, where the carousel remains stationary at the point where the last item was retrieved (awaiting the next order to be fed), and for the *closed-loop* strategy, where the carousel returns to a predefined point after the retrieval of an order is completed. He formally shows that under the open-loop strategy the carousel will change its direction at most once when following the optimal picking sequence, while under the closed-loop strategy the carousel will turn at most twice. A recursive expression for the distribution of the minimal travel time needed to collect one order of n randomly distributed items in the open-loop scenario is given explicitly by Litvak and Van Zwet [82].

The case when positions of the items in an order are dependent has not received much studies. One way to model the dependencies is described by Abdel-Malek and Tang [1] who assume that the positions of successive items form a Markov chain. In this setting, they study the performance of the organ-pipe storage rule. Stern [102] introduces correlations between items in an order by considering several order types, where each type corresponds to a fixed list of items. The work of Wan and Wolff [120] focuses on minimising the travel time for “clumpy” orders, that is, orders concentrated on a relatively small segment of the carousel, and introduces the nearest-endpoint heuristic for which they obtain conditions for it to be optimal. In this setting, one can no longer assume that the items locations are uniformly distributed. Moreover, there is clearly a strong dependence between items positions.

The model with non-uniform items locations reflects a relevant situation when some of the items are required more frequently than others. Most of the papers that assume distinct frequencies assume the orders of one item (see e.g. [9]). An interesting work on non-uniformly distributed items is given by Litvak [78], where the focus is on the length of the shortest rotation time needed to collect a single order when the order size is large and the items locations have a non-uniform continuous distribution with a positive density f on $[0, 1]$.

6.3 Picking multiple orders

A popular strategy for reducing the mean travel time per order in carousel storage and retrieval systems is batching together a number of orders and then picking them sequentially. A *batch* is a set of orders that is picked in a single tour. Two consecutively picked items do not necessarily belong to the same order. An excellent literature survey by Van den Berg [106] on planning and control of warehousing systems addresses this issue and the problems that arise if large batches are formed. Apart from the questions mentioned before, Stern [102] also considers the performance of a carousel for a fixed set of order types (for example, big orders with many items, and small ones).

Bartholdi and Platzman [6] are mainly concerned with sequencing batches of requests in a bidirectional carousel. They specify the number of orders to be retrieved (ignoring any new arrivals) and propose three heuristic methods to solve this static problem. Orders may be picked in any sequence (and not necessarily at the order they arrive), but picks within the same order are performed consecutively. They define the *minimum spanning interval*, which is the shortest interval containing all the items of an order and, by assuming that the picker always begins and finishes retrieving an order at one of the endpoints of this interval, they construct the shortest matching chain by ordering the orders accordingly. This procedure may fail to give an uninterrupted sequence in which to pick the orders; therefore, they propose the following heuristics. The first one, called the *hierarchical* heuristic, picks any order that happens to have a common endpoint with another order, and then travels clockwise until an unpicked endpoint is encountered, and repeats the procedure. The *nearest-order* heuristic is practically an extension of the nearest-item heuristic described earlier in the paper, as is the case with the *second monomaniacal* heuristic they propose. Under these heuristics, they obtain upper bounds for the travel time.

Ghosh and Wells [43] assume that the orders have to be picked under a FIFO sequencing restriction, which means that the first order to arrive at the warehouse is the first order that will be picked, and so on. Since the orders are retrieved in a FIFO fashion, the problem is reduced to finding how to retrieve each individual order so that the best overall retrieval is achieved. They develop an algorithm for the optimal retrieval path of n orders via dynamic programming, and show how to update dynamically the solution when new orders arrive.

Rouwenhorst *et al.* [100] model the carousel as an M/G/1 queuing system, where the orders are the "customers" that require service, and the service they get depends on the pick strategy that is followed. This approach permits the derivation of various queuing characteristics such as the mean response time and the waiting time when orders arrive randomly. The authors mention that the tight upper bounds for the mean response time can be further exploited to obtain also good approximations for excess probabilities of the response time.

Van den Berg [105] assumes either a fixed or an arbitrary sequence of orders. When the sequence of the orders is given, he presents an efficient dynamic programming algorithm that finds an optimum path that visits all orders in the specified sequence. Furthermore, when there is no given order sequence, he simplifies the problem to a *rural postman problem* on a circle and solves this problem to optimality. The rural postman problem is the problem of finding the shortest route in an undirected graph which includes all edges at least one time. Van den Berg [105] concludes that the obtained solution requires at most 1.5 revolutions more than a lower bound of an optimal solution to the original problem. Simulation results suggest that the average rotation time may be reduced up to 25% when allowing a free order sequence. Lee and Kuo [70] formulate the problem of optimal sequencing of items and orders as a *multi-travelling salesman problem*. In the multi-travelling salesman problem, there are several salesmen in a home city, and each of the other cities has to be visited only by one salesman. Using this formulation, Lee and Kuo [70] provide efficient heuristics for optimal picking of several orders consisting of multiple items.

6.4 Design issues

All research papers mentioned so far that deal with travel time models of carousel systems assume average uniform velocity of the carousel. In other words, the main assumption is that the carousel travels with constant speed and the acceleration from the stationary position (when a pick is performed) to its full speed, as well as the deceleration from the maximum speed to zero speed, are negligible factors when computing the travel time of the carousel. Guenov and Raeside [47] give some empirical evidence that the error induced when neglecting acceleration and deceleration of an order picking vehicle is indeed negligible. Thus the problem of minimising retrieval times can be considered to be equivalent to the problem of minimising the average distance travelled by the carousel per retrieval.

Hwang *et al.* [58], however, develop strategies for picking that take into consideration the variation in speed of the carousel. For unit-load automated storage and retrieval systems there are several travel-time models that consider the speed profiles of the storage and retrieval robot. In [58] some relevant references are given. Unlike the unit-load automated storage and retrieval systems, almost all the existing travel-time models for carousel systems assume that the effects of the variation in speed are negligible. In [58] the authors try to bridge this gap in the literature. They assume that the items are randomly distributed on the carousel and derive the expected travel time both in the case of a single command cycle and in the case of a dual command cycle. They verify the accuracy of the proposed models by comparing the results to results directly obtained from discrete racks.

Egbelu and Wu [33] study the problem of pre-positioning the carousel in anticipation of storage or retrieval requests in order to improve the average response time of the system. Choosing the right starting point of a carousel in anticipation of an order is also referred as the *dwell point selection problem.* This strategy becomes relevant when the items are stored under the organ pipe arrangement. In this situation the dwell point should be chosen to be the location of the most popular item; see, e.g., [9].

Spee [101] is concerned with developing design criteria for carousels. He states the basic conditions for designing an automatic order picking system with carousels and comments on the optimal storage design. Namely, he is interested in finding the right number of picking robots and the right number and dimensions of a carousel so that the investment is minimised, provided that the size of the orders that need to be retrieved is given.

McGinnis [84] studies some of the design and control issues relevant to *rotary racks*. A rotary rack is an automated storage and retrieval system that strongly resembles carousels. In fact, conceptually, a rotary rack is simply a carousel, where the only difference is that each level or shelf of this carousel can rotate independently of the others. The author concludes that, while rotary racks appear to be a simple generalisation of conventional carousels, the control strategies that have been shown effective for carousels do not appear to be as effective for these systems. Rotary racks can be viewed as a multiple-carousel system (where each level is considered as a sub-carousel) with a single picker.

6.5 Problems involving multiple carousels

While almost all work mentioned in this section concerns one-carousel models, real applications have triggered the study of models involving multiple carousels. The study of such models is not as developed yet as the study of models involving a single carousel. The list of references that follows seems to be complete.

Perhaps the first academic study that investigates the performance of a system involving several carousels is that of Emerson and Schmatz [37]. The authors simulated the operation of the warehouse of Rockwell's Collins Telecommunications Products. The system consists of twenty-two carousels, where each pair of carousels had a single-operator station (so there are in total eleven operator stations). The questions they are concerned with are how big the batch size of orders should be so as to complete the week's work (which is used as a performance measure) and keep all operators busy, what happens when a carousel or a station is down, and how is an overload or an imbalance (for example, unequal operator performance, unequal carousel loading, or large orders) handled. In order to investigate potential solutions to these three imbalance conditions, the authors investigate two operating rules.

The first operating rule studies six different storage schemes with seven carousel pairs (and thus seven operators). It uses simulation models to study simple storage schemes such as random storage, sequential alternating storage, and storage in the carousel with the largest number of openings. The aim in [37] is to study the degree of carousel usage. The authors find that there is no significant difference between the carousel loads among the storage schemes. However, they do not treat the problem of optimally assigning items to carousel bins, and do not present any analytical models to help investigate the problem. The second operating rule they investigate is a floating operator. This is an operator who is trained to work at any station, and who is moving to different stations according to specific needs (for example, depending on the size of the queue at a particular station). They conclude that this solution seems advantageous for the purposes of the warehouse they investigate.

Koenigsberg [65] presents analytic solutions for evaluating the performance of a single carousel, and discusses the ways in which his approach can be extended to a system involving two unidirectional carousels both served by a single robotic operator. The carousels are related only through the state of the robot, which means that each carousel is independent of the other except for the time it waits for an operation to commence (such as pick, storage, or repair) because the robot is busy at the other carousel. The author concludes that under some conditions, it is often more advantageous to have two carousels of identical length instead of one carousel of double the length. Furthermore, going to three carousels of equal length (i.e. one third of the length of the single carousel) will offer little further improvement.

Hwang and Ha [55] study the throughput performance both of a single and of a double carousel system. Based on a randomised storage assignment policy, cycle time models are developed for single and dual commands. Furthermore, they examine the value of the information on the succeeding orders in terms of system efficiency, which may lead to better scheduling of the orders to be processed.

In a later work, Hwang *et al.* [57] attempt to measure analytically the effects of double shuttles of the storage and retrieval machine (i.e. the robotic picker) on the throughput both of the standard and of the double carousel system. Storage and retrieval machines with double shuttles are machines that have space for two items. Thus, for example, an item can be retrieved from the carousel and stored on one shuttle, while the other shuttle has an item that needs to be stored to the carousel. After this item is stored, a second item can be retrieved from the carousel and placed on the empty shuttle. All these operations occur during a single cycle of the carousel operation. For the double carousel system, a retrieval sequence rule is proposed which utilises the characteristics of the two independently rotating carousels. From the test results, double shuttles are shown to have a substantial improvement over single shuttles. This improvement tends to be more prominent in the double carousel system. Due to cost concerns, the authors note that an economic evaluation will be needed to justify the extra cost of double carousel systems and double shuttles before implementing them in real world situations.

Wen *et al.* [125] consider a system comprised of two carousels and a single retrieval machine. Their main assumption is that every order must be picked in a single tour, i.e., an order cannot be divided into two or more sub-tours. Batching orders together is also not allowed. They analyse the retrieval time and propose four heuristic algorithms for the scheduling sequence of retrieving items from the system to satisfy an order. Their method is an extension of the algorithm presented in [6] and [102].

Meller and Klote study the throughput of a group of several carousels, a so-called carousel pod [85]. They use approximations to evaluate the order pick time in one carousel and then evaluate the throughput of a pod by plugging in the average response times of each unit and modelling the pod of c carousels as a queuing system where $1/c$ picker operates one carousel. Further, they derive an approximation for the system's throughput using a diffusion approximation by Gelenbe [41] which was earlier applied by Bozer and White [22] in the analysis of end-of-aisle order-picking systems.

Recently, Hassini and Vickson [53] studied storage locations for items, aiming to minimise the long-run expected travel time in a two-carousel setting with a single picker. They assume that the products are available at all times (so as to be able to ignore possible delays due to lack of stock), and that orders are not batched; that is, the carousel system processes only single-item orders. This is applicable in situations where individual product orders are processed in a first-come-first-served policy, or when the next item to be retrieved is known only after the present one has been picked. The authors compare the performance of three heuristic storage schemes and a genetic algorithm [45] that for small-sized problems completely enumerates the solution space. They conclude that none of the heuristic approaches leads to a solution that outperforms the algorithmic solution they provide.

The same model is also studied by Park *et al.* [92]. As is the case in [53], in [92] the basic assumptions are that there is an infinite number of items to be picked and that an order consists of a single item. The authors, however, are not interested in storage issues. They further assume that the single operator, the picker, is alternately serving the two carousels. This may cause the picker to have to wait for an amount of time until the item at the carousel he is currently serving is rotated in front of him. They derive the distribution of the waiting time of the picker under specific assumptions for the pick times. This allows them to derive expressions for the system throughput and the picker utilisation.

The model presented in [92] has been extended further in Vlasiou *et al.* [110, 113, 118, 114, 116] by removing all assumptions related to the pick times or rotation times. In related work, Vlasiou *et al.* [111, 112, 119] have shown that the two-carousel model studied in [53, 92] is equivalent to an alternating service queue, if one allows for rotation times with an infinite support. Some of these results have been presented in Section 3.

Finally, we would like to mention that there is a broad literature on automated storage and retrieval systems (see e.g. the survey by Le-Duc [69]). An extensive list of references has been also made available on-line by Roodbergen [99].

References

[1] Abdel-Malek, L. and Tang, C. (1994). A heuristic for cyclic stochastic se-

quencing of tasks on a drum-like storage system. *Computers & Operations Research* **21,** 385–396.

[2] Ali, M. M. (1973). Content of the frustum of a simplex. *Pacific Journal of Mathematics* **48,** 313–322.

[3] Ali, M. M. and Obaidullah, M. (1982). Distribution of linear combination of exponential variates. *Communications in Statistics A. Theory and Methods* **11,** 1453–1463.

[4] Asmussen, S. (2003). *Applied Probability and Queues.* Springer-Verlag, New York.

[5] Asmussen, S. and Sigman, K. (1996). Monotone stochastic recursions and their duals. *Probability in the Engineering and Informational Sciences* **10,** 1–20.

[6] Bartholdi, III, J. J. and Platzman, L. K. (1986). Retrieval strategies for a carousel conveyor. *IIE Transactions* **18,** 166–173.

[7] Beirlant, J. and van Zuijlen, M. C. A. (1985). The empirical distribution function and strong laws for functions of order statistics of uniform spacings. *Journal of Multivariate Analysis* **16,** 300–317.

[8] Beirlant, J., Deheuvels, P., Einmahl, J. H. J. and Mason, D. M. (1991). Bahadur-Kiefer theorems for uniform spacings processes. *Akademiya Nauk SSSR. Teoriya Veroyatnosteĭ i ee Primeneniya* **36,** 724–743.

[9] Bengü, G. (1995). An optimal storage assignment for automated rotating carousels. *IIE Transactions* **27,** 105–107.

[10] Bertoin, J. and Yor, M. (2002). On the entire moments of self-similar Markov processes and exponential functionals of Lévy processes. *Annales de la Faculté des Sciences de Toulouse. Mathématiques. Série 6* **11,** 33–45.

[11] Bertoin, J., Biane, P. and Yor, M. (2004). Poissonian exponential functionals, q-series, q-integrals, and the moment problem for log-normal distributions. In *Seminar on Stochastic Analysis, Random Fields and Applications IV.* vol. 58 of *Progress in Probability.* Birkhäuser, Basel pp. 45–56.

[12] Bertoin, J. and Yor, M. (2005). Exponential functionals of Lévy processes. *Probability Surveys* **2,** 191–212 (electronic).

[13] Bertsekas, D. and Gallager, R. (1992). *Data Networks.* Prentice-Hall, Englewood Cliffs, New Jersey.

[14] Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1987). *Regular Variation* vol. 27 of *Encyclopedia of Mathematics and its Applications.* Cambridge University Press, Cambridge.

[15] Bitner, J. R. and Wong, C. K. (1979). Optimal and near-optimal scheduling algorithms for batched processing in linear storage. *SIAM Journal on Computing* **8,** 479–498.

[16] Blomqvist, N. (1968). Estimation of waiting time parameters in the GI/G/1 queueing system, part I: General results. *Skandinavisk Aktuarietidskrift* **51,** 178–197.

[17] Blomqvist, N. (1969). Estimation of waiting time parameters in the GI/G/1 queueing system, part II: Heavy traffic approximations. *Skandinavisk Aktuarietidskrift* **52,** 125–136.

[18] Borovkov, A. A. (1998). *Ergodicity and Stability of Stochastic Processes.* Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., Chichester.

[19] Boxma, O. J. (1985). Two symmetric queues with alternating service and switching times. In *Performance '84 (Paris, 1984).* North-Holland, Amsterdam pp. 409–431.

[20] Boxma, O. J. and Groenendijk, W. P. (1988). Two queues with alternating service and switching times. In *Queueing Theory and its Applications – Liber Amicorum for J.W. Cohen.* ed. O. Boxma and R. Syski. vol. 7 of *CWI monographs.* North-Holland, Amsterdam pp. 261–282.

[21] Boxma, O. J. and Vlasiou, M. (2007). On queues with service and interarrival times depending on waiting times. *Queueing Systems. Theory and Applications* **56,** 121–132.

[22] Bozer, Y. and White, J. (1996). A generalized design and performance analysis model for end-of-aisle order-picking systems. *IIE Transactions* **28,** 271–280.

[23] Carmona, P., Petit, F. and Yor, M. (1997). On the distribution and asymptotic results for exponential functionals of Lévy processes. In *Exponential functionals and principal values related to Brownian motion*, Bibl. Rev. Mat. Iberoamericana, Rev. Mat. Iberoamericana, Madrid pp. 73–130.

[24] Cohen, J. W. (1982). *The Single Server Queue.* North-Holland Publishing Co., Amsterdam.

[25] Cohen, J. W. and Boxma, O. J. (1981). The M/G/1 queue with alternating service formulated as a Riemann-Hilbert problem. In *Performance '81 (Amsterdam, 1981).* North-Holland, Amsterdam pp. 181–199.

[26] Daley, D. J. (1968). The serial correlation coefficients of waiting times in a stationary single server queue. *Journal of the Australian Mathematical Society* **8,** 683–699.

[27] Deheuvels, P. (1982). Strong limiting bounds for maximal uniform spacings. *The Annals of Probability* **10,** 1058–1065.

[28] Deheuvels, P. and Devroye, L. (1984). Strong laws for the maximal k-spacing when $k \leq c \log n$. *Z. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **66,** 315–334.

[29] Del Pino, G. E. (1979). On the asymptotic distribution of k-spacings with applications to goodness-of-fit tests. *The Annals of Statistics* **7,** 1058–1065.

[30] Devroye, L. (1981). Laws of the iterated logarithm for order statistics of uniform spacings. *The Annals of Probability* **9,** 860–867.

[31] Devroye, L. (1982). A log log law for maximal uniform spacings. *The Annals of Probability* **10,** 863–868.

[32] Dumas, V., Guillemin, F., and Robert, P. (2002). A Markovian analysis of additive-increase multiplicative-decrease algorithms. *Advances in Applied Probability* **34,** 85–111.

[33] Egbelu, P. J. and Wu, C.-T. (1998). Relative positioning of a load extractor for a storage carousel. *IIE Transactions* **30,** 301–317.

[34] Einmahl, J. H. J. and van Zuijlen, M. C. A. (1988). Strong bounds for weighted empirical distribution functions based on uniform spacings. *The Annals of Probability* **16,** 108–125.

[35] Einmahl, J. H. J. and van Zuijlen, M. C. A. (1992). Glivenko-Cantelli-type theorems for weighted empirical distribution functions based on uniform spacings. *Statistics & Probability Letters* **13,** 411–419.

[36] Eisenberg, M. (1979). Two queues with alternating service. *SIAM Journal on Applied Mathematics* **36,** 287–303.

[37] Emerson, C. R. and Schmatz, D. S. (1981). Results of modeling an automated warehouse system. *Industrial Engineering* **13,** 28–32, cont. on p. 90.

[38] Flajolet, P. (2004). Counting by coin tossings. In *Proceedings of the 9th Asian Computing Science Conference.* ed. M. J. Maher Springer, Berlin/Heidelberg. volume 4, pp. 1–12.

[39] Franks, G., Al-Omari, T., Woodside, M., Das, O. and Derisavi, S. (2009). Enhanced modeling and solution of layered queueing networks. *IEEE Transactions on Software Engineering* **35,** 148–161.

[40] Gamarnik, D., Nowicki, T. and Swirszcz, G. (2006). Maximum weight independent sets and matchings in sparse random graphs. Exact results using the local weak convergence method. *Random Structures & Algorithms* **28,** 76–106.

[41] Gelenbe, E. (1975). On approximate computer system models. *Journal of the ACM (JACM)* **22,** 261–269.

[42] Ghosh, K. and Jammalamadaka, S. R. (2001). A general estimation method using spacings. *Journal of Statistical Planning and Inference* **93,** 71–82.

[43] Ghosh, J. B. and Wells, C. E. (1992). Optimal retrieval strategies for carousel conveyors. *Mathematical and Computer Modelling* **16,** 59–70.

[44] Glaz, J., Naus, J., Roos, M. and Wallenstein, S. (1994). Poisson approximations for the distribution and moments of ordered m-spacings. *Journal of Applied Probability* **31A,** 271–281.

[45] Goldberg, D. E. (1989). *Genetic Algorithms in Search, optimisation, and Machine Learning.* Addison-Wesley Professional, Reading, MA.

[46] Groenendijk, W. P. (1990). Conservation laws in polling systems. *PhD thesis.* Utrecht University.

[47] Guenov, M. and Raeside, R. (1989). Real time optimisation of man on board order picking. In *Proceedings of the 10th International Conference on Automation in Warehousing.* ed. J. White. IFS(Publications), Dallas, Texas. pp. 87–93.

[48] Guillemin, F., Robert, P. and Zwart, B. (2004). AIMD algorithms and exponential functionals. *The Annals of Applied Probability* **14**, 90–117.

[49] Ha, J.-W. and Hwang, H. (1994). Class-based storage assignment policy in carousel system. *Computers & Industrial Engineering* **26,** 489–499.

[50] Hall, P. (1986). On powerful distributional tests based on sample spacings. *Journal of Multivariate Analysis* **19,** 201–224.

[51] Hardy, G. H., Littlewood, J. E. and Pólya, G. (1952). *Inequalities* second ed. Cambridge University Press.

[52] Hassini, E. (2008). Storage space allocation to maximize inter-replenishment times. *Computers & Operations Research* **35,** 2162–2174.

[53] Hassini, E. and Vickson, R. G. (2003). A two-carousel storage location problem. *Computers & Operations Research* **30,** 527–539.

[54] Henze, N. (1995). The distribution of spaces on lottery tickets. *The Fibonacci Quarterly. The Official Journal of the Fibonacci Association* **33,** 426–431.

[55] Hwang, H. and Ha, J.-W. (1991). Cycle time models for single/double carousel system. *International Journal of Production Economics* **25,** 129–140.

[56] Hwang, H. and Ha, J.-W. (1994). An optimal boundary for two class-based storage assignment policy in carousel system. *Computers & Industrial Engineering* **27,** 87–90.

[57] Hwang, H., Kim, C.-S. and Ko, K.-H. (1999). Performance analysis of carousel systems with double shuttle. *Computers & Industrial Engineering* **36,** 473–485.

[58] Hwang, H., Song, Y.-K. and Kim, K.-H. (2004). The impacts of acceleration/deceleration on travel time models for carousel systems. *Computers & Industrial Engineering* **46,** 253–265.

[59] Ibe, O. C. (1990). Analysis of polling systems with mixed service disciplines. *Communications in Statistics. Stochastic Models* **6,** 667–689.

[60] Jacobs, D. P., Peck, J. C. and Davis, J. S. (2000). A simple heuristic for maximizing service of carousel storage. *Computers & Operations Research* **27,** 1351–1356.

[61] Kalashnikov, V. (2002). Stability bounds for queueing models in terms of weighted metrics. In *Analytic Methods in Applied Probability.* ed. Y. Suhov. vol. 207 of *American Mathematical Society Translations Ser. 2.* American Mathematical Society, Providence, RI pp. 77–90.

[62] Keane, M., Konheim, A. G. and Meilijson, I. (1984). The organ pipe permutation. *SIAM Journal on Computing* **13,** 531–540.

[63] Kim, B. (2005). Maximizing service of carousel storage. *Computers & Operations Research* **32,** 767–772.

[64] Kleinrock, L. (1976). *Queueing Systems, Vol. 2: Computer Applications.* Wiley, New York.

[65] Koenigsberg, E. (1986). Analysis of the efficiency of carousel and tote-stacker performance. In *Proceedings of the 7th International Conference on Automation in Warehousing.* ed. J. White. Springer, San Francisco, California. pp. 173–183.

[66] Koopmans, T. C. and Beckmann, M. (1957). Assignment problems and the location of economic activities. *Econometrica* **25,** 53–76.

[67] Korshunov, D. (1997). On distribution tail of the maximum of a random walk. *Stochastic Processes and their Applications* **72,** 97–103.

[68] Le Cam, L. (1958). Un théorème sur la division d'un intervalle par des points pris au hasard. *Publications de l'Institut de Statistique de l'Université de Paris* **7,** 7–16.

[69] Le-Duc, T. (2005). Design and control of efficient order picking processes. *PhD thesis.* Erasmus University Rotterdam.

[70] Lee, S. D. and Kuo, Y. C. (2008). Exact and inexact solution procedures for the order picking in an automated carousal conveyor. *International Journal of Production Research* **46,** 4619–4636.

[71] Li, J. and Liu, R. Y. (2008). Multivariate spacings based on data depth. I. Construction of nonparametric multivariate tolerance regions. *The Annals of Statistics* **36,** 1299–1323.

[72] Li, C.-L. and Wan, G. (2005). Improved algorithm for maximizing service of carousel storage. *Computers & Operations Research* **32,** 2147–2150.

[73] Lim, W. K., Bartholdi, J. J. and Platzman, L. K. (1985). Storage schemes for carousel conveyors under real time control. Material Handling Research Center Technical Report MHRC-TR-85-10. Georgia Institute of Technology.

[74] Lindley, D. V. (1952). The theory of queues with a single server. *Proceedings Cambridge Philosophical Society* **48,** 277–289.

[75] Lindner, A. and Sato, K. (2009). Continuity properties and infinite divisibility of stationary distributions of some generalized Ornstein-Uhlenbeck processes. *The Annals of Probability* **37,** 250–274.

[76] Litvak, N. (2001). Collecting n items randomly located on a circle. *PhD thesis.* Eindhoven University of Technology Eindhoven, The Netherlands. Available at http://alexandria.tue.nl/extra2/200210141.pdf.

[77] Litvak, N. (2001). Some peculiarities of exponential random variables. *Journal of Applied Probability* **38,** 787–792.

[78] Litvak, N. (2006). Optimal picking of large orders in carousel systems. *Operations Research Letters* **34,** 219–227.

[79] Litvak, N. and Adan, I. J.-B. F. (2001). The travel time in carousel systems under the nearest item heuristic. *Journal of Applied Probability* **38,** 45–54.

[80] Litvak, N. and Adan, I. J.-B. F. (2002). On a class of order pick strategies in paternosters. *Operations Research Letters* **30,** 377–386.

[81] Litvak, N., Adan, I. J.-B. F., Wessels, J. and Zijm, W. H. M. (2001). Order picking in carousel systems under the nearest item heuristic. *Probability in the Engineering and Informational Sciences* **15,** 135–164.

[82] Litvak, N. and Van Zwet, W. R. (2004). On the minimal travel time needed to collect n items on a circle. *The Annals of Applied Probability* **14,** 881–902.

[83] Malmborg, C. J. and Bhaskaran, K. (1990). A revised proof of optimality for the cube-per-order index rule for stored item location. *Applied Mathematical Modelling* **14,** 87–95.

[84] McGinnis, L. F., Han, M. H. and White, J. A. (1986). Analysis of rotary rack operations. In *Proceedings of the 7th International Conference on Automation in Warehousing*. ed. J. White. Springer, San Francisco, California. pp. 165–171.

[85] Meller, R. D. and Klote, J. F. (2004). A throughput model for carousel/VLM pods. *IIE Transactions* **36,** 725–741.

[86] Mohamed, H. and Robert, P. (2005). A probabilistic analysis of some tree algorithms. *The Annals of Applied Probability* **15,** 2445–2471.

[87] Neilson, J. E., Woodside, C. M., Petriu, D. C. and Majumdar, S. (1995). Software bottlenecking in client-server systems and rendezvous networks. *IEEE Transactions on Software Engineering* **21,** 776–782.

[88] Noble, B. (1958). *Methods Based on the Wiener-Hopf Technique for the Solution of Partial Differential Equations* vol. 7 of *International Series of Monographs on Pure and Applied Mathematics*. Pergamon Press, New York.

[89] Omari, T., Franks, G., Woodside, M. and Pan, A. (2007). Efficient performance models for layered server systems with replicated servers and parallel behaviour. *The Journal of Systems and Software* **80,** 510–527.

[90] Osogami, T. (2005). Analysis of multi-server systems via dimensionality reduction of markov chains. *PhD thesis*. School of Computer Science, Carnegie Mellon University Pittsburgh, PA 15213. Available at http://www.cs. cmu.edu/∼osogami/.

[91] Ozawa, T. (1990). Analysis of a single server model with two queues having different service disciplines. *Electronics and Communications in Japan. Part III: Fundamental Electronic Science* **73,** 18–27.

[92] Park, B. C., Park, J. Y. and Foley, R. D. (2003). Carousel system performance. *Journal of Applied Probability* **40,** 602–612.

[93] Park, B. C. and Rhee, Y. (2005). Performance of carousel systems with 'organ-pipe' storage. *International Journal of Production Research* **43,** 4685–4695.

[94] Patie, P. (2009). Exponential functional of a new family of Lévy processes and self-similar continuous state branching processes with immigration. *Bulletin des Sciences Mathématiques* **133,** 355–382.

[95] Perel, E. and Yechiali, U. (2008). Queues where customers of one queue act as servers of the other queue. *Queueing Systems. Theory and Applications* **60,** 271–288.

[96] Pyke, R. (1965). Spacings. (With discussion). *Journal of the Royal Statistical Society. Series B. Methodological* **27,** 395–449.

[97] Pyke, R. (1972). Spacings revisited. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Berkeley, Calif., 1972. Univ. California Press, Vol. I: Theory of statistics*, pp. 417–427.

[98] Robert, P. (2005). On the asymptotic behaviour of some algorithms. *Random Structures Algorithms* **27,** 235–250.

[99] Roodbergen, K. Warehousing literature. Available at http://www.roodbergen.com/literature/wh_literature_version8.pdf Last visited in October 2007.

[100] Rouwenhorst, B., Van den Berg, J. P., Van Houtum, G. J. and Zijm, W. H. M. (1996). Performance analysis of a carousel system. In *Progress in Material Handling Research: 1996.* ed. R. J. Graven, L. F. McGinnis, D. J. Medeiros, R. E. Ward, and M. R. Wilhelm. The Material Handling Institute, Charlotte, NC pp. 495–511.

[101] Spee, D. (1996). Automatic order picking system with horizontal racks. In *Progress in Material Handling Research: 1996.* ed. R. J. Graven, L. F. McGinnis, D. J. Medeiros, R. E. Ward, and M. R. Wilhelm. The Material Handling Institute, Charlotte, NC pp. 545–550.

[102] Stern, H. I. (1986). Parts location and optimal picking rules for a carousel conveyor automatic storage and retrieval system. In *Proceedings of the 7th International Conference on Automation in Warehousing.* ed. J. White. Springer, San Francisco, California. pp. 185–193.

[103] Szekli, R. (1995). *Stochastic Ordering and Dependence in Applied Probability.* Springer, New York.

[104] Takács, L. (1962). *Introduction to the Theory of Queues.* University Texts in the Mathematical Sciences. Oxford University Press, New York.

[105] Van den Berg, J. P. (1996). Multiple order pick sequencing in a carousel system: A solvable case of the rural postman problem. *Journal of the Operational Research Society* **47,** 1504–1515.

[106] Van den Berg, J. P. (1999). A literature survey on planning and control of warehousing systems. *IIE Transactions* **31,** 751–762.

[107] Van Vuuren, M. and Winands, E. M. M. (2006). Iterative approximation of k-limited polling systems. Technical Report 2006-06. Eindhoven University of Technology. Available at http://www.win.tue.nl/math/bs/spor/.

[108] Vickson, R. G. and Fujimoto, A. (1996). Optimal storage locations in a carousel storage and retrieval system. *Location Science* **4,** 237–245.

[109] Vlasiou, M. (2005). A non-increasing Lindley-type equation. Technical Report 2005-015. Eurandom Eindhoven, The Netherlands. Available at http://www.eurandom.nl.

[110] Vlasiou, M. (2006). Lindley-type recursions. *PhD thesis*. Eindhoven University of Technology Eindhoven, The Netherlands.

[111] Vlasiou, M. (2007). A non-increasing Lindley-type equation. *Queueing Systems. Theory and Applications* **56,** 41–52.

[112] Vlasiou, M. and Adan, I. J.-B. F. (2005). An alternating service problem. *Probability in the Engineering and Informational Sciences* **19,** 409–426.

[113] Vlasiou, M. and Adan, I. J.-B. F. (2007). Exact solution to a Lindley-type equation on a bounded support. *Operations Research Letters* **35,** 105–113.

[114] Vlasiou, M., Adan, I. J.-B. F., Boxma, O. J. and Wessels, J. (2003). Throughput analysis of two carousels. Technical Report 2003-037. Eurandom Eindhoven, The Netherlands. Available at http://www.eurandom.nl.

[115] Vlasiou, M., Adan, I. J.-B. F. and J, B. O. (2009). A two-station queue with dependent preparation and service times. *European Journal of Operational Research* **195,** 104–116.

[116] Vlasiou, M., Adan, I. J.-B. F. and Wessels, J. (2004). A Lindley-type equation arising from a carousel problem. *Journal of Applied Probability* **41,** 1171–1181.

[117] Vlasiou, M. and Palmowski, Z. (2009). Large deviations for a random sign lindley recursion. Technical Report 2009-003. Eindhoven University of Technology The Netherlands.

[118] Vlasiou, M. and Yechiali, U. (2008). M/G/∞ polling systems with random visit times. *Probability in the Engineering and Informational Sciences* **22,** 81–106.

[119] Vlasiou, M. and Zwart, B. (2007). Time-dependent behaviour of an alternating service queue. *Stochastic Models* **23,** 235–263.

[120] Wan, Y.-W. and Wolff, R. W. (2004). Picking clumpy orders on a carousel. *Probability in the Engineering and Informational Sciences* **18,** 1–11.

[121] Weiss, D. J. (1980). Computer controlled carousels. In *Proceedings of the 3rd International Conference on Automation in Warehousing*. IFS(Publications), Stratford-upon-Avon, UK. pp. 413–418.

[122] Wells, M. T., Jammalamadaka, S. R. and Tiwari, R. C. (1993). Large sample theory of spacings statistics for tests of fit for the composite hypothesis. *Journal of the Royal Statistical Society. Series B. Methodological* **55,** 189–203.

[123] Wen, U.-P. (1986). The order picking problem in the carousel systems. In *Proceedings of the 7th International Conference on Automation in Warehousing.* ed. J. White. Springer, San Francisco, California. pp. 195–199.

[124] Wen, U.-P. and Chang, D.-T. (1988). Picking rules for a carousel conveyor in an automated warehouse. *OMEGA: The International Journal of Management Science* **16,** 145–151.

[125] Wen, U.-P., Lin, J. T. and Chang, D.-T. (1989). Order picking for a two-carousel-single-server system in an automated warehouse. In *Proceedings of the 10th International Conference on Automation in Warehousing.* ed. J. White. IFS(Publications), Dallas, Texas. pp. 87–93.

[126] Woodside, C. M., Neilson, J. E., Petriu, D. C. and Majumdar, S. (1995). The stochastic rendezvous network model for performance of synchronous client-server-like distributed software. *IEEE Transactions on Computers* **44,** 20–34.

[127] Yeh, D.-H. (2002). A note on "A simple heuristic for maximizing service of carousel storage". *Computers & Operations Research* **29,** 1605–1608.

[128] Yoon, C. and Sharp, G. (1996). A structured procedure for analysis and design of order pick systems. *IIE Transactions* **28,** 379–389.