Invited Review

A survey of literature on automated storage and retrieval systems

Kees Jan Roodbergen a,*, Iris F.A. Vis b,1

a RSM Erasmus University, P.O. box 1738, 3000 DR Rotterdam, The Netherlands
b VU University Amsterdam, Faculty of Economics and Business Administration, De Boelelaan 1105, Room 3A-31, 1081 HV Amsterdam, The Netherlands

Received 27 June 2006; accepted 22 January 2008
Available online 5 February 2008

Abstract

Automated Storage and Retrieval Systems (AS/RSs) are warehousing systems that are used for the storage and retrieval of products in both distribution and production environments. This paper provides an overview of literature from the past 30 years. A comprehensive explanation of the current state of the art in AS/RS design is provided for a range of issues such as system configuration, travel time estimation, storage assignment, dwell-point location, and request sequencing. The majority of the reviewed models and solution methods are applicable to static scheduling and design problems only. Requirements for AS/RSs are, however, increasingly of a more dynamic nature for which new models will need to be developed to overcome large computation times and finite planning horizons, and to improve system performance. Several other avenues for future research in the design and control of AS/RSs are also specified.

© 2008 Elsevier B.V. All rights reserved.

Keywords: Logistics; Automated Storage and Retrieval Systems; Warehouse; System design; Control policies

1. Introduction

Automated storage and retrieval systems have been widely used in distribution and production environments since their introduction in the 1950s. An automated storage and retrieval system (AS/RS) usually consists of racks served by cranes running through aisles between the racks. An AS/RS is capable of handling pallets without the interference of an operator, thus the system is fully automated. Both in production and distribution environments AS/RSs are used for putting products (e.g., raw materials or (semi-) finished products) in storage and for retrieving those products from storage to fulfill an order. Between 1994 and 2004, there has been a significant increase in the number of AS/RSs used in distribution environments in the United States (Automated Storage Retrieval Systems Production Section of the Material Handling Industry of America, 2005). The usage of AS/RSs has several advantages over non-automated systems. Examples are savings in labour costs and floor space, increased reliability and reduced error rates. Apparent disadvantages are high investments costs (approximately $634,000 for a single aisle AS/RS, Zollinger, 1999), less flexibility and higher investments in control systems (about $103,000, Zollinger, 1999).

In designing an AS/RS, many physical design and control issues have to be addressed in the right way to fully take advantage of all its pros. This paper intends to present a critical overview of all important issues concerning AS/RS design and control in both production and distribution environments while studying recent and past literature. Previously, several overview papers have been published that discuss part of the AS/RS literature. Almost all of these overview papers, however, have a focus different from AS/RSs, for example, general warehouse design. Because of this, only a limited number of aspects of AS/RSs and a limited number of references with respect to AS/RSs are presented in those papers. Matson and White (1982) review a number of material handling research areas one of which is concerned with AS/RSs. Kusiak (1985)
describe design and operational decision problems for flexible manufacturing systems with a focus on automated guided vehicles and AS/RSs. The author discusses design, storage and batching (i.e., consolidation of orders) policies for AS/RSs. Johnson and Brandeau (1996) discuss stochastic models for the design and control of automated guided vehicles and AS/RSs. Manda and Palekar (1997) discuss some papers on travel time estimation for AS/RSs and storage assignment rules.

General overviews of warehouse design and control include Cormier and Gunn (1992), Van den Berg (1999), Rouwenhorst et al. (2000), De Koster et al. (2007) and Gu et al. (2007). Due to their broad scope, these five papers only discuss a fraction of the AS/RS issues and literature. To our knowledge, Sarker and Babu (1995) is the only paper discussing exclusively AS/RSs, however, this paper only reviews some design aspects of AS/RSs while focussing on travel time models. We conclude that our paper seems to be the first overview paper in over 10 years devoted exclusively to AS/RSs, and the first ever to give a broad overview of all design and control issues in AS/RSs.

The main structure of the paper can be described as follows. First, we present a general description of AS/RSs and a classification of related design and control issues. The main body of the paper consists of an overview of literature discussing solution methods for these design and control problems. At the end, we indicate relevant open research questions. In more detail, Section 2 defines various types of AS/RSs and describes some important technical characteristics. Section 3 presents a classification of both physical design and control issues. This broad introduction will be followed in Section 4 by a more detailed description of methods that support each of the discussed physical design issues. Papers addressing individual control policies for storage assignment (Section 5), batching (Section 6), parking of idle AS/RSs (Section 7) and sequencing (Section 8) will be discussed in subsequent sections. Travel time estimates and other performance measures will be treated in Section 9. Section 10 presents conclusions and issues for further research.

2. AS/RS types

An AS/RS system is defined as a storage system that uses fixed-path storage and retrieval machines running on one or more rails between fixed arrays of storage racks (Automated Storage Retrieval Systems Production Section of the Material Handling Industry of America (2005)). AS/RSs are used to store and retrieve loads in various settings. The main components of an AS/RS are racks, cranes, aisles, I/O-points, and pick positions. Racks are typically metal structures with locations that can accommodate loads (e.g., pallets) that need to be stored. Cranes are the fully automated storage and retrieval machines that can autonomously move, pick up and drop off loads. Aisles are formed by the empty spaces between the racks, where the cranes can move. An input/output point (I/O-point) is a location where retrieved loads are dropped off, and where incoming loads are picked up for storage. Pick positions (if any) are places where people are working to remove individual items from a retrieved load before the load is sent back into the system.

A large number of system options exist for AS/RSs. The most basic version of an AS/RS has in each aisle one crane, which cannot leave its designated aisle (aisle-captive) and which can transport only one unit-load at a time (single shuttle). Product handling in this case is by unit-load (e.g., full pallet quantities) only; no people are involved to handle individual products. The racks in the basic version are stationary and single-deep, which means that every load is directly accessible by the crane. This AS/RS type is referred to as a single unit-load aisle-captive AS/RS. Numerous variations exist of this basic AS/RS. An overview of the main options is presented in Fig. 1. We will briefly discuss some of the options below.

One possible variation of the basic AS/RS is when cranes are capable of changing aisles. In this case, it is possible to have fewer cranes than aisles in the system. This may be useful if the amount of requests does not justify the purchase of a crane for each aisle. To overcome the restriction of the crane’s unit-load capacity, multi-shuttle cranes exist. Such a crane can transport two or more loads at a time. Cranes which can transport two loads are also referred to as dual-shuttle cranes; cranes capable of transporting more than two loads are still rarely seen. The increased transport capacity enables a crane, for example, to first retrieve one load and then store another load in the same location without having to go to the I/O-point in between.

Often AS/RS is installed for handling unit-loads only (typically, pallets). Unit-loads arrive at the I/O-point of the AS/RS from other parts of the warehouse by means of, for example, automated guided vehicles, conveyors, or forklift trucks. The unit-loads are stored in the AS/RS and after a period of time they are retrieved again, for example, to be shipped to a customer. In many cases, however, only part of the unit-load may be needed to fulfill a customer’s order. This can be resolved by having a separate picking area in the warehouse; in which case the AS/RS serves to replenish the picking area. Alternatively, the picking operation can be integrated with the AS/RS. One option is to design the crane such that a person can ride along (person-on-board). Instead of retrieving a full pallet automatically from the location, the person can pick one item from the location. A more common option to integrate item picking is when the AS/RS drops off the retrieved unit loads at a workstation. A picker at this workstation takes the required amount of products from the unit-load after which the AS/RS moves the remainder of the load back into the storage rack. This system is often referred to as an end-of-aisle system. If the unit-loads are bins, then the system is generally called a miniload AS/RS.

Storage in the racks may occur single or double deep. In a double-deep rack, each rack location has space for two
unit-loads; one load is stored in front of the other load. A load can only be put into or retrieved from the second position if there is no load in the first position. Double-deep storage might be beneficial if the variety of loads is relatively low and the turnover rate of these loads is high (Tompkins et al., 2003). Modifications to the crane may be required to be able to store and retrieve loads from both positions. Carousel systems (horizontal or vertical, single or double) are suitable for storing small and medium-sized products at different levels. A crane is used to store and retrieve items from the rotating carousel. The lower and upper part of a double carousel can rotate independently of each other.

Finally, worthy of mentioning is a special type of AS/RSs called autonomous vehicle storage and retrieval systems. This system separates horizontal and vertical travel. Vehicles travel horizontally over rails through aisles, while lifts are used to transfer loads vertically.

3. Overview of design decisions

It is crucial to design an AS/RS in such a way that it can efficiently handle current and future demand requirements while avoiding bottlenecks and overcapacity. Due to the inflexibility of the physical layout and the equipment, it is essential to design it right at once. In Fig. 2 a schematic view of design issues and their interdependence is presented.

It is important to realise that the AS/RS is usually just one of many systems to be found in a warehouse. The true performance of the AS/RS is typically influenced by the other systems as are the other systems’ performances influenced by the AS/RS. This is most visible at, but not restricted to, the interplay of systems at the AS/RS’ I/O-points. Loads are picked up and dropped off at an I/O-point by the AS/RS. It is the task of, for example, a conveyor system or a set of vehicles to make the connection from the I/O-point to the rest of the warehouse. Delays in one system can cause delays in the other system. Thus, when deciding on the number of I/O-points, their location and their buffer capacity it may be necessary to also look at the other systems’ characteristics.

Furthermore, the requirements for the AS/RS may depend on the general environment of the system. In manufacturing environments the AS/RS primarily needs to provide all required materials in time to make sure that production can continue. Production is leading and should never be halted to wait for the AS/RS. In distribution envi-
The AS/RS performs or supports the order retrieval process, to make sure that customers' orders leave the facility in time.

Part of the actual design of an AS/RS consists obviously of determining its physical appearance. The physical design consists of two aspects which together determine the physical manifestation of the system. First, we have the choice of the AS/RS type (system choice). Second, the chosen system must be configured, for example, by deciding on the number of aisles and the rack dimensions (system configuration). These interrelated choices can be made based on, among others, historical and forecasted data, product characteristics, the available budget, required throughput, required storage space and available land space. Various options for AS/RS types are displayed in Fig. 1, however, little research is available to select the best type of system from the available options. A more elaborate overview of various AS/RS types along with selection criteria can be found in Allen (1992).

A list of configuration decisions for any given type of AS/RS is displayed in Table 1. For a typical design problem, total capacity is given beforehand. This essentially means that the mathematical product of the number of aisles, rack height, and rack length is constant. Increasing the number of aisles thus implies reducing rack length and/or height to maintain the desired storage capacity. Because of this relation, having more aisles indirectly results in shorter response times, due to the decreased rack length and height. Furthermore, design changes often have an impact in multiple ways at the same time. In a standard system with one crane per aisle, having more aisles also means having more cranes, which in turn results in a higher throughput and higher investment costs.

<table>
<thead>
<tr>
<th>Class of problems</th>
<th>Decisions to be made</th>
</tr>
</thead>
<tbody>
<tr>
<td>System configuration</td>
<td>• Number of aisles</td>
</tr>
<tr>
<td></td>
<td>• Height of the storage racks</td>
</tr>
<tr>
<td></td>
<td>• Length of the aisles</td>
</tr>
<tr>
<td></td>
<td>• Equally sized or modular storage locations</td>
</tr>
<tr>
<td></td>
<td>• Number and location of the I/O-points</td>
</tr>
<tr>
<td></td>
<td>• Buffer capacity at the I/O-points</td>
</tr>
<tr>
<td></td>
<td>• Number of cranes per aisle</td>
</tr>
<tr>
<td></td>
<td>• Number of order pickers per aisle (if any)</td>
</tr>
<tr>
<td>Storage assignment</td>
<td>• Storage assignment method</td>
</tr>
<tr>
<td></td>
<td>• Number of storage classes</td>
</tr>
<tr>
<td></td>
<td>• Positioning of the storage classes</td>
</tr>
<tr>
<td>Batching</td>
<td>• Type of batching (static or dynamic)</td>
</tr>
<tr>
<td></td>
<td>• Batch size (capacity or time based)</td>
</tr>
<tr>
<td></td>
<td>• Selection rule for assignment of orders to batches</td>
</tr>
<tr>
<td>Sequencing</td>
<td>• Sequencing restrictions (e.g., due dates)</td>
</tr>
<tr>
<td></td>
<td>• Type of operation (single or dual command)</td>
</tr>
<tr>
<td></td>
<td>• Scheduling approach (block or dynamic)</td>
</tr>
<tr>
<td>Dwell-point</td>
<td>• Sequencing method</td>
</tr>
<tr>
<td></td>
<td>• Type of positioning (static or dynamic)</td>
</tr>
<tr>
<td></td>
<td>• Location where idle cranes will be placed.</td>
</tr>
</tbody>
</table>

Even when the number of aisles is given, there is still a trade-off between rack height and length. Since cranes can travel vertically and horizontally simultaneously, the actual travel time equals the maximum of the horizontal and vertical travel time (Chebyshev distance metric). Horizontal travel speeds are typically up to 3 m² compared to vertical travel speeds of up to 0.75 m² (Tompkins et al., 2003). A good balance between rack height and length can help to reduce travel times. A common, yet not necessarily optimal, configuration is one where the racks are square-in-time, which means that the time needed to reach the highest row equals the time needed to reach the farthest column. Any rack that is not square-in-time is called rectangular.

Often racks have equally sized storage locations. However, to meet highly varying customers' demand, it is also interesting to allow the storage of different shaped loads within a single rack. Also, an AS/RS may have more than one I/O-point per aisle. Instead of only placing an I/O-point at the front of an aisle, another one might be located at the middle and/or rear of the aisle. In that way, for example, flows of incoming and outgoing loads can easily be separated. Research in the field of physical design will be treated in Section 4.

Just as important as the physical design are the software controls needed to get the AS/RS operational (see e.g., Fohn et al., 1994; Terry et al., 1988). A good design procedure should simultaneously address both physical design and control issues of the system. Regardless of the actual optimisation procedure, a system of performance measurement is needed to evaluate the overall performance of the resulting system at every stage. This underlines the importance of performance measurement in the field. Many publications have appeared on performance measurement, which will be discussed in Section 9.

Control policies are methods which determine the actions performed by the AS/RS. Typically, the operation of an AS/RS is governed by a coherent set of such control policies, which each take care of a specific subset of the activities. A storage assignment policy serves to determine which products are assigned to which locations. The position where an idle crane (i.e., a crane that has no jobs to perform) waits is determined by a dwell-point policy. The dwell-point is best chosen to minimise the expected time to travel to the next (still unknown) request.

A unit-load AS/RS can operate in two ways, namely in a single command cycle or in a dual command cycle. In a single command cycle the crane performs either a single storage or a single retrieval request. The storage cycle time then is equal to the sum of the time to pick-up a load at the input station, the time to travel to the storage location, the time to place the load in the rack and the time to return to the input station. The retrieval cycle time can be defined similarly. If an AS/RS performs both a storage and a retrieval request in a single cycle, we call this a dual command cycle. In this case, the cycle time is defined as the sum of the time to pick-up the load, the time to travel to
the storage location and store the load, the empty travel time (interleaving time) from the storage location to the retrieval location and the time to pick the unit-load and transport it to the output station. Clearly, the total time to perform all storage and retrieval requests reduces if dual commands are performed. A tour of an AS/RS consists of a sequence of single or dual command cycles, starting at the origin of the first request and ending at the destination of the last request. Sequencing rules can be used to create tours such that the total time to handle all request is minimised or the due times are least violated.

As a final control policy, batching considers how one can combine different customer orders into a single tour of the crane (mainly applicable to person-on-board AS/RS). Table 1 provides an overview of all control decision problems that may need to be selected. From literature it is known that certain combinations of control policies work better than other combinations. In Sections 5–8 we present an extensive discussion of all AS/RS control policies that have been treated in literature.

4. Physical design

Only a few papers address the design of AS/RSs in combination with the design of other material handling systems in the facility. Most of these papers consider manufacturing environments. Chincholkar and Krishnaiah Chetty, 1996 use approaches as Petri Nets and the Taguchi method to simultaneously address the scheduling of jobs to an AS/RS system and the scheduling of jobs to machines in a flexible manufacturing system. The AS/RS is both responsible for storing and retrieving loads and for transferring them between machines. Inman (2003) studies the usage of AS/RSs in the automotive industry. The function of the AS/RS is to restore the sequence in which jobs are handled at the various processes in the facility. A model is proposed to determine the capacity of the AS/RS based on the number of jobs that need to be rescheduled. As a result, the design of the AS/RS is completely subordinate to the assembling processes in the facility.

Hwang et al. (2002) consider the design of miniload AS/RSs in combination with Automated Guided Vehicles. Both a non-linear model and heuristics have been proposed to determine the optimal number of loads to be transferred by each AGV to machines in combination with an optimal design of the AS/RS. Park and Webster (1989a) address the design of warehouses by proposing an approach that simultaneously selects the used storage equipment, that might be an AS/RS, and the overall size and shape of the storage area.

In order to deal with one or more design issues for AS/RSs, methods ranging from simulation, analytical models, artificial intelligent approaches (e.g., Knapp and Wang, 1992; Chincholkar et al., 1994; Hsieh et al., 1998) to experimental approaches (e.g., Lim et al., 1996) have been proposed in the literature. We will use the classification as indicated in Table 1 in our discussion on solution procedures that assist in decision making for one or multiple physical design issues in combination with one or more control issues. Table 2 presents an overview of papers and indicates which issues the authors address. Note that we do not mark a control policy if it is only used as a fixed input factor to the model. Only decision variables are marked. Furthermore, papers that only focus on control issues, but not on physical design, have been excluded since they will be discussed in one of the subsequent sections.

For all simulation models, we can conclude that they only address some of the physical design aspects and that configurability of control policies is very limited. Furthermore, only few configurations and types of AS/RSs have been tested in combination with fixed values for various input factors. In this way, it can never be guaranteed that a (near-) optimal design has been found. Randhawa and Shroff (1995) have been performing the most extensive simulation study. These authors examine the effect of different sequencing rules on six layout configurations (with a varying I/O-point, item distribution over racks, rack configuration and rack dimensions). Based on a limited number of experiments the authors conclude, among other things, that locating the I/O-point at the middle of the aisle, instead of at the end of the aisle, results in a higher throughput.

In our opinion, in general the strength of simulation could be better exploited in AS/RS research to compare numerous designs, while taking into account more design aspects, especially in combination with control policies. Also, sensitivity analyses on input factors should be performed such that a design can be obtained which can perform well in all applicable scenarios. As a result more general information could be obtained on good design practices.

Rosenblatt et al. (1993) already tried to overcome this drawback of existing simulation models by alternatingly using optimisation and simulation in order to reach the best design given a certain service level. From several experiments the authors conclude that an optimal design requires fewer cranes than aisles. Hwang and Ko (1988) use their travel time model for a crane operating in multiple aisles to conclude that multiple-aisles AS/RSs might be interesting if the number of storage and retrieval requests are low. Contrary to these results, most other analytical models included in Table 2 assume exactly one crane per aisle. Mathematical programming models, queuing theory and heuristics have been mostly used in developing analytical design models. From the overview in Table 2 we conclude that all models address some, but certainly not all physical design aspects. None of the physical analytical design models includes any of the mentioned control decisions problems.

Summarising, most research is performed on determining the layout of a single storage rack. In this context, no attention has been paid to the storage capacity (single or multiple deep) of the storage locations themselves. Furthermore, hardly any attention has been paid to the location
<table>
<thead>
<tr>
<th>Solution approach</th>
<th>Authors</th>
<th>Type of AS/RS</th>
<th>System configuration</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Storage racks</td>
<td>Storage locations</td>
</tr>
<tr>
<td>Simulation</td>
<td>Ashayeri et al. (1983)</td>
<td>Unit-load</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Houshyar and Chung (1991)</td>
<td>Unit-load</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Taboun and Bhote (1993)</td>
<td>Unit-load</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Randhawa and Shroff (1995)</td>
<td>Unit-load</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lee et al. (1996)</td>
<td>Unit-load</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Potré et al. (2004)</td>
<td>Unit-load and multi-shuttle</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>Simulation and analytical</td>
<td>Rosenblatt et al. (1993)</td>
<td>Unit-load</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Analytical</td>
<td>Karasawa et al. (1980)</td>
<td>Unit-load</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Zollinger (1982)</td>
<td>Unit-load</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ashayeri et al. (1985)</td>
<td>Unit-load</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Azadivar (1987)</td>
<td>Unit-load</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bozer and White (1990)</td>
<td>End-of-aisle miniload</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Van Oudheusden and Zhu (1992)</td>
<td>Person-on-board</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bozer and White (1996)</td>
<td>End-of-aisle miniload</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chang and Wen (1997)</td>
<td>Unit-load</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hwang and Ko (1988)</td>
<td>Unit-load</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Park et al. (1999)</td>
<td>End-of-aisle miniload</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Malmborg (2001a)</td>
<td>Unit-load</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Malmborg (2001b)</td>
<td>Multi-shuttle</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Koh et al. (2005)</td>
<td>End-of-aisle miniload</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lee et al. (2005)</td>
<td>Unit-load</td>
<td>×</td>
<td></td>
</tr>
</tbody>
</table>
and the number of I/O-points. In some types of distribution centres (e.g., cross-docking areas), it might, for example, be interesting to locate I/O-points at both ends of a storage rack. Also, only little research has been performed on the configuration of multi-shuttle AS/RSs, which already have been proven to be efficient in terms of throughput. From the literature studied, it can be concluded that most authors combine only one or a few control rules with their physical design research. Batching and dwell-point locations have not been considered in physical design. Thus, despite the fact that decisions in both categories (physical design and control rules) are highly interrelated, they are usually addressed separately. In the next sections we discuss papers that propose policies for the various control issues indicated in Table 2.

5. Storage assignment

Several methods exist for assigning products to storage locations in the racks. Five often used storage assignment policies for AS/RSs are described here in more detail (see e.g., Hausman et al., 1976 or Graves et al., 1977). These rules are:

- dedicated storage assignment
- random storage assignment
- closest open location storage assignment
- full-turnover-based storage assignment
- class-based storage assignment

For the dedicated storage method each product type is assigned to a fixed location. Replenishments of that product always occur at this same location. The main disadvantage of dedicated storage are its high space requirements and consequent low space utilisation. This is due to the fact that locations are reserved even for products that are out of stock. Furthermore, for each product type sufficient space must be reserved to accommodate the maximum inventory level that may occur. Most advantages of dedicated storage, such as locating heavy products at the bottom or matching the layout of stores, are related to non-automated order-picking areas and are not as interesting for AS/RSs. For random storage all empty locations have an equal probability of having an incoming load assigned to it. If the closest open location storage is applied, the first empty location that is encountered will be used to store the products. This typically leads to an AS/RS where racks are full around the I/O-points and gradually more empty towards the back (if there is excess capacity).

The full-turnover storage policy determines storage locations for loads based on their demand frequency. Frequently requested products get the easiest accessible locations, usually near the I/O-points. Slow-moving products are located farther away from the I/O-point. An important assumption for this rule is that the turnover frequencies need to be known beforehand. Heskett (1963, 1964) presents the cube-per-order index (COI) rule, which is a form of full-turnover storage. The COI of a load is defined as the ratio of the load’s required storage space to the number of request for this product per period. The COI rule assigns loads with the lowest COI to the locations closest to the I/O-point. Malmborg and Bhaskaran (1990) give a proof of optimality for this rule while taking into account the non-uniqueness of the COI layout if dual command scheduling is used. Malmborg and Krishnakumar (1989) show that the COI-rule is optimal for person-aboard AS/RSs with respect to order-picking costs if there are fixed inventory levels and a fixed balanced assignment of order pickers to items. However, according to Lee (1992) the COI-rule cannot be applied for person-on-board systems due to the fact that an order usually consists of more than two independent items at different locations. Therefore, the author develops a new heuristic that outperforms the COI-rule.

For practical purposes it is easier if a full-turnover policy is combined with dedicated storage. The problem is that demand frequencies change constantly and also the product assortment is usually far from constant. Any change in frequency and any addition of a new product to the system may require a large amount of repositioning of loads to bring it back in line with the full-turnover rule. To prevent excessive repositioning, a new storage allocation is in practice typically calculated once per period. To reduce space requirements and periodic repositioning while maintaining most of the efficiency gains, class-based storage can be used, which will be discussed next.

5.1. Class-based storage

This storage assignment method divides the available warehouse space into a number of areas. Each item is subsequently assigned to one of the areas, based on the item’s demand frequency. Random storage is applied within an area. Actually, the full-turnover storage policy can be seen as a class-based policy with only one item per class. Often class-based storage with three classes is referred to as ABC storage. This reflects the common practice to call the fastest-moving items the A-items, the next fastest-moving items the B-items, and so on. The main advantage of class-based storage is an increased efficiency due to storing the fast-moving items near the I/O-point, while at the same time the low storage space requirements and flexibility of the random storage method apply.

A designer faces three major decisions when implementing class-based storage in an AS/RS.

1. Zone divisioning (i.e., determining number of classes).
2. Zone sizing (i.e., number of products to be assigned to each zone).
3. Zone positioning (i.e., where to locate each of the zones).

From Table 3 it can be seen that several types of zone sizing procedures have been developed to derive boundaries
for each of the classes. For dual command scheduling, research has mainly focused on rectangular racks. Clearly, the resulting procedures can also be used in square-in-time racks which are a special case of rectangular racks. Both Rosenblatt and Eynan (1989) and Eynan and Rosenblatt (1994) conclude that a relatively small number of classes, usually less than 10, is to be preferred to obtain most of the potential savings in travel times as compared to full-turnover storage. In practice, often the number of classes is restricted to three. Van den Berg (1996) proposes a dynamic programming algorithm that assigns both locations and products to classes.

Several strategies exist for zone positioning, varying from optimal solutions for single command scheduling to rules of thumb for dual command scheduling. Hausman et al. (1976) prove that a \(L\)-shaped configuration with square-in-time boundaries for classes A, B and C is optimal when single command scheduling is applied in square-in-time racks. Graves et al. (1977) demonstrate through simulation that this \(L\)-shaped configuration is close to optimal for dual command scheduling in square-in-time racks. Guenov and Raeside (1992) also study dual command scheduling and compare three different zone shapes. Their conclusion is that the performance of each of the proposed shapes depends on the location of the I/O-point and that none is superior to the others. Eynan and Rosenblatt (1994) present a layout for \(n\) classes in rectangular racks while single command scheduling is used. This layout combines \(n - 2L\) zones, a transient region for class \(n - 1\) and a rectangular zone for class \(n\). Fig. 3 presents examples of zone shapes for both types of racks for three classes.

Some variations of the class-based storage policy have been studied. Park and Webster (1989b) propose a class-based storage method in a three-dimensional system which minimises total travel times. Hsieh and Tsai (2001) suggest a class-based method for production facilities based on the bill of materials. Thonemann and Brandeau (1998) alter the algorithms of Hausman et al. (1976) such that they can be

### Table 3

<table>
<thead>
<tr>
<th>Class sizes</th>
<th>Single command</th>
<th>Dual command</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 classes</td>
<td>Numerical procedure (Hausman et al., 1976)</td>
<td>Recursive procedure (Rosenblatt and Eynan, 1989)</td>
</tr>
<tr>
<td>3 classes</td>
<td>Numerical procedure (Hausman et al., 1976)</td>
<td>Recursive procedure (Van den Berg, 1996)</td>
</tr>
<tr>
<td>(n) classes</td>
<td>No paper is specifically written about this problem can be solved using (n = 2) solution approaches for (n) classes</td>
<td>Recursive procedure (Van den Berg, 1996)</td>
</tr>
</tbody>
</table>

No paper is specifically written about this problem can be solved using solution approaches for \(n\) classes. Kouvelis and Papanicolaou (1995) suggest to use their formulas in combination with the recursive procedure of Eynan and Rosenblatt (1993) for this problem.

![Fig. 3. Typical zone positioning for three classes in respectively square-in-time (upper part) and rectangular racks (lower part).](image-url)
used in an environment with stochastic demand. Several authors (e.g., Hwang and Ha, 1994; Vickson and Fujimoto, 1996; Vickson and Lu, 1998) examine class-based policies for carousels. The racks in the carousel are partitioned into two regions with an appropriate boundary shape (above or next to each other). Bengü (1995) studies an organ pipe arrangement for carousels in which the items with the highest access probability are located in the middle and the rest is positioned at the left or right side according to a decreasing access probability.

5.2. Performance of storage assignment rules

Travel time estimates (see Section 9) for both single and dual command scheduling in different types of AS/RS configurations are an appropriate analytical tool for comparing control rules (e.g., Hausman et al., 1976; Graves et al., 1977). With simulation more extensive experiments under stochastic conditions can be performed (e.g., Schwarz et al., 1978; Kaylan and Medeiros, 1988; Kulturel et al., 1999; Van den Berg and Gademann, 2000). Malmberg (1996) and Malmberg and Altassan (1998) develop a storage policy trade-off for respectively unit-load and small systems to roughly compare different policies in a short time prior to a simulation study.

Results from both analytical and simulation studies show that full-turnover-based and class-based storage assignment outperform random storage. Kulturel et al. (1999) compare a three class-based policy with a duration of stay policy, which was originally introduced by Goetschalckx and Ratliff (1990). While applying the duration of stay policy, products with the shortest duration in the warehouses are assigned to storage locations closest to the I/O-point. The three-class-based policy only outperforms the duration of stay policy if the number of product types is small. Methods as Petri Nets (e.g., Lin and Wang, 1995) are able to update the system to cope with rapidly changing dynamic environments and to compare different policies.

Updating, reshuffling of items in idle periods of the AS/RS and reconsidering storage assignment decisions can be vital in current dynamic environments to maintain the desired performance level (e.g., Muralidharan et al., 1995; Jaikumar and Solomon, 1990). The COI policy is usually well applicable in static environments with independent demand of products. Moon and Kim (2001) show by means of a simulation study that re-location of items is required if the production quantity of each item changes over time. Sadiq et al. (1996) propose a dynamic storage assignment policy to reassign items to storage locations in systems with a rapidly changing product mix and short product life cycles. By using predicted future product mix, correlated demand of products and demand forecasts, the dynamic policy intends to minimise total order processing times, which consist of order-picking times and relocating times. It is shown that this dynamic policy outperforms the static COI rule.

Summarising, in literature various storage assignment policies have been developed and were compared both through simulation and analytical methods. Most authors address single aisle AS/RSs with one I/O-point. Storage assignment policies for other types of configurations (e.g., multiple I/O-points) or other types of AS/RSs (e.g., multiple shuttle AS/RSs) hardly have been formulated.

6. Batching

Suppose we have a number of orders that need to be retrieved from storage in a person-on-board item-picking AS/RS. We could retrieve the orders one at a time or we could try to combine several orders in a single tour of the crane. The latter approach is called batching. Batching problems for person-on-board AS/RSs are quite similar to batching problems for order pickers in warehouses. For a detailed review of this type of research, refer to De Koster et al. (2007). In this section, we will only focus on papers that deal with batching problems for AS/RSs with cranes that operate in a single aisle.

An advantage of batching is that the length of a tour for a batch of orders is shorter than the sum of the individual orders’ tour lengths. However, more effort is needed to keep track of which item belongs to which order or to sort items later on. A limit on the size of a batch is usually determined by the capacity of the crane or an upper limit on the required response time. As a result, an important decision problem in batching is the determination of the size of each batch in combination with the assignment of orders to these batches such that travel times are minimised. One of the first papers referring to batching of orders for person-on-board AS/RSs is Barrett (1977). Elsayed (1981) concludes that this problem, that can be formulated as a mixed integer programming model (Armstrong et al., 1979) is NP-hard.

To obtain solutions for large problems in acceptable computation times, numerous heuristics have been developed. As presented in Fig. 4, most heuristic batching methods basically follow the same three steps: a method of initiating batches by selecting a seed, a method of allocating orders to batches, and a stopping rule to determine when a batch has been completed. An important assumption in all heuristics is the fact that a single order cannot be split over various batches, but needs to be picked as a whole. Table 4 indicates several seed selection, order addition rules and stopping rules.

Contrary to a single seeding rule a cumulative seeding rule (e.g., Elsayed and Stern, 1983) uses all orders that are already in the batch as the seed. Hwang et al. (1988) and Hwang and Lee (1988) develop heuristics based on cluster analysis. As common in cluster analysis, both attribute vectors related to storage locations of an order and similarity measures, based on, for example, the boundaries of the area in which the crane needs to travel to reach all locations of an order, are used in formulating heuristics. A seed order will be selected and the most similar order will
be added. These two orders can be merged and can be used as a new seed in the next step. This process will be repeated until a cluster of orders has been obtained.

Pan and Liu (1995) perform a comparative analysis of batching heuristics based on average travel times. In the experiments the shape and the capacity of the cranes as well as the storage assignment policies have been varied. The authors conclude that only the storage capacity of a crane highly impacts the choice for a certain rule. Based on the results of 42 experiments with each 30 orders, the authors advice to use the heuristic of Hwang and Lee (1988).

The methods included in Table 4 do not address time constraints on retrievals. Both Elsayed et al. (1993) and Elsayed and Lee (1996) discuss the batching problem with due dates in combination with the scheduling problem of orders. A penalty function has been introduced which measures both the earliness and tardiness of orders. Furthermore, all mentioned papers assume that the arrival patterns of orders are known before the start of the operations. However, in many warehouses there is a batch-arrival component (for example, orders left from the previous day that did not make the departure time) and an on-line component. It is known that batching becomes more difficult when orders arrive on-line (see e.g., Bhaskaran and Malmberg, 1989). This is due to the fact that there is a trade-off for on-line arrivals between reducing waiting times (by calculating batches frequently based on few available orders) and reducing travel times (by calculating batches less frequently to obtain more possibilities for efficient combinations).

7. Dwell-point location

Several methods have been proposed to deal with the decision where to position an idle crane, i.e., how to determine the crane’s dwell-point. Bozer and White (1984) introduced four simple static dwell-point strategies. Table 5 summarises these rules and indicates the resulting one-dimensional parking location. Park (1991) showed that the input station rule returns an optimal dwell-point if the probability, that the first request after an idle period is a storage request, is at least 1/2.

Egbelu (1991) proposes linear programming models that are capable of responding to fluctuations in types of requests. A two-dimensional dwell-point location (including the height in the rack) can be determined such that the response time to the location of need is minimised. Important drawbacks of this dynamic approach are the facts that LP solution techniques need to be implemented in the AS/RS control system and that computation times may be too high to be of practical value. Hwang and Lim (1993) propose an efficient algorithm for the minimum expected travel time model of Egbelu (1991) by transforming it into a single facility location model.

Egbelu and Wu (1993) compare the four rules of Bozer and White (1984) and the two dynamic rules of Egbelu (1991) by means of a simulation study. A ranking of the alternatives is difficult to make due to the fact that the rules are compared with a small number of replications for just one layout with five different arrival rates of requests under random and dedicated storage policies.

Several static approaches have been proposed for other types of unit-load AS/RSs. Chang and Egbelu (1997) consider a single crane serving multiple aisles. A mathematical

Table 4
Overview of seed selection, order addition and stopping rules for batching

<table>
<thead>
<tr>
<th>Type of rule</th>
<th>Rule</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seed selection</td>
<td>Order with largest number of locations to be visited</td>
<td>Elsayed (1981)</td>
</tr>
<tr>
<td></td>
<td>Order with smallest number of locations to be visited</td>
<td>Elsayed (1981)</td>
</tr>
<tr>
<td></td>
<td>Order with largest volume</td>
<td>Elsayed (1981)</td>
</tr>
<tr>
<td></td>
<td>Order with smallest volume</td>
<td>Elsayed (1981)</td>
</tr>
<tr>
<td></td>
<td>Order with highest percentage of capacity of crane</td>
<td>Elsayed and Unal (1989)</td>
</tr>
<tr>
<td></td>
<td>Cumulative rule</td>
<td>Elsayed and Stern (1983)</td>
</tr>
<tr>
<td></td>
<td>Clustering rule</td>
<td>Hwang et al. (1988) and Hwang and Lee (1988)</td>
</tr>
<tr>
<td>Order addition</td>
<td>Largest number of common locations</td>
<td>Egbelu (1991)</td>
</tr>
<tr>
<td></td>
<td>Geometric similarities</td>
<td>Hwang and Lee (1988)</td>
</tr>
<tr>
<td>Stopping rules</td>
<td>Capacity constraint</td>
<td>Several</td>
</tr>
<tr>
<td></td>
<td>Time constraint</td>
<td>Several</td>
</tr>
<tr>
<td></td>
<td>All orders are completed</td>
<td>Several</td>
</tr>
</tbody>
</table>
programming model has been developed to determine a dwell-point from a three-dimensional point of view. Except for the location and height, the exact aisle also needs to be specified. Peters et al. (1996) developed closed-form analytical models to determine dwell-points in continuous square-in-time and rectangular racks under random storage based on the travel time estimations of Bozer and White (1984). Park (1999) developed an optimal dwell-point policy for square-in-time racks under dedicated storage. Park (2001) shows that for rectangular racks under random storage the dwell-point location can be determined in terms of the probability of occurrence of a certain type of request. Van den Berg (2002) presents analytical expressions for determining dwell-points under randomised and class-based storage policies. Contrary to Peters et al. (1996), these methods can be used for various AS/RS configurations with varying locations of the I/O-point instead of situating it at the lower corners of the rack.

Summarising, simple rules of thumb, closed-form analytical expressions and mathematical programming approaches have been formulated to position idle AS/RS cranes at a dwell-point in various AS/RS configurations. Most approaches address static positioning of single unit-load capacity AS/RSs and do not study other types of AS/RSs. So far, no extensive and statistically sound simulation study has been performed on dwell-point policies in combination with other physical design and control issues (see also Table 2) which makes it difficult to provide designers with an advice when to use which policy.

8. Sequencing of storage and retrieval requests

Storage requests in distribution or production environments are usually not time-critical. The exact point in time at which loads are stored is not of much importance for the performance of the system. Therefore, storage requests are usually stored according to the first-come-first-served principle. In sequencing retrievals usually due times of retrievals should be met, which makes it necessary to look beyond simple first-come-first-served. Furthermore, by sequencing the retrievals in a smart way, improvements in the overall throughput of the AS/RS can be obtained.

The list of retrievals is continuously changing over time. Performed retrievals are deleted from the list and new retrieval jobs are added. Han et al. (1987) suggest two ways to deal with this dynamic problem. Firstly, select a block of the most urgent storage and retrieval requests, sequence them and when they are completed select the next block, and so on. This is called block sequencing. Secondly, we can resequence the whole list of requests every time a new request is added and use due times or priorities. We refer to this kind of sequencing as dynamic sequencing. The performance of both approaches differs per situation. For example, Eben-Chaime (1992) concludes that in a specified non-deterministic environment, the block sequencing strategy might be inappropriate. However, a block sequencing approach is more transparent and simpler with respect to implementation.

Various algorithms and heuristics can be used to schedule storage and retrieval requests within a block. The main objectives in those approaches are to minimise total travel times or total travel distances. Most literature focuses on single and dual command scheduling of unit load AS/RSs with one input/output station and one crane per aisle. Therefore, we first discuss sequencing methods for this basic sequencing problem. In Section 8.2, we discuss extensions of the basic sequencing problem. The sequencing problem of other types of AS/RSs is treated in Section 8.3.

8.1. The basic sequencing problem

For the sequencing problem of unit-load AS/RSs, the two common types of cycles to be addressed are single command and dual command cycles (see also Section 3). In a single command cycle only one unit-load is moved (either a storage or a retrieval) before the crane returns to the I/O-point. A dual command cycle consists of two moves, one storage and one retrieval. The possibility of performing dual command cycles depends on the availability of storage and retrieval requests. If both types of requests are available, dual command cycles give advantages with respect to travel times (Graves et al., 1977).

An alternative might be to perform dual commands whenever possible and single commands otherwise. Eben-Chaime and Pliskin (1996, 1997) show that systems that operate under this more hybrid mode might achieve more stable waiting lines and can use less cranes. Some warehouses, however, have patterns in arriving and leaving loads. For example, trucks with incoming loads arrive in the morning and trucks transporting outgoing loads arrive in the evening. In this case, cranes might perform single command cycles. If arriving and leaving trucks overlap in time, dual command cycles can be performed.

Table 5

<table>
<thead>
<tr>
<th>Rule</th>
<th>Dwelling-point</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input station</td>
<td>Always at input station</td>
</tr>
<tr>
<td>Midpoint</td>
<td>Always at midpoint location of racks</td>
</tr>
<tr>
<td>Input/output</td>
<td>If a single command storage request has been performed then positioning at input station</td>
</tr>
<tr>
<td></td>
<td>If a single command retrieval request or a dual command has been performed then positioning at output station</td>
</tr>
<tr>
<td>Last location</td>
<td>If a single command storage request has been performed then positioning at last storage location</td>
</tr>
<tr>
<td></td>
<td>If a single command retrieval request or a dual command has been performed then positioning at output station</td>
</tr>
</tbody>
</table>
Bozer et al. (1990) explain that the dual command scheduling of AS/RSs can be formulated as a Chebyshev travelling salesman problem, which is known to be NP-complete. Han et al. (1987) indicate that in general the problem of optimally sequencing a given list of requests is NP-hard. One of the reasons for this complexity is the fact that the set of locations available for storage depends on where previous loads are stored and which retrievals have already been performed. However, some special cases of the sequencing problem can be solved in polynomial time. A summary of these results is shown in Table 6. Computation times are high for solving the problem in which non-dedicated storage is used (Lee and Schaefer, 1996).

For the dynamic sequencing problem various methods, as summarised in Table 7 have been formulated. Several studies, including simulation studies, have been performed to compare the performance of these heuristics in combination with several storage assignment policies (see, e.g., Linn and Wysk, 1987). Han et al. (1987) show that the nearest-neighbour heuristic can provide a lower average cycle time than the first come first served rule. Furthermore, it is shown that the nearest-neighbour heuristic has a better performance than the shortest leg heuristic over the long run due to the fact that the storage locations close to the input station are filled up first and, thereafter, only locations far from the input station remain open.

Table 6
Solvable special cases of the basic sequencing problem for unit-load AS/RS

<table>
<thead>
<tr>
<th>Characteristics of case</th>
<th>Steps of solution method</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Non-dedicated storage</td>
<td>1. Solve assignment problem to combine storage (with location) with retrieval requests</td>
<td>Lee and Schaefer (1996)</td>
</tr>
<tr>
<td>• The number of storages equals the number of retrievals</td>
<td>2. Apply Murthy’s ranking algorithm (Murthy, 1968) to search assignment solutions</td>
<td></td>
</tr>
<tr>
<td>• A single I/O point</td>
<td>3. The optimal solution equals the first assignment solution where no load needs to be stored in a non-empty location (this may occur when the location will be empty at a later time due to a retrieval)</td>
<td></td>
</tr>
<tr>
<td>• Dedicated storage</td>
<td>Can be translated into an assignment problem</td>
<td>Lee and Schaefer (1997)</td>
</tr>
<tr>
<td>• The number of storages is smaller than or equal to the number of retrievals</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• A single I/O-point</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Dedicated storage</td>
<td>Transportation problem that combines departure positions of empty trips (end of loaded trip) with arrival positions of empty trips (start of loaded trip) such that total empty travel distances are minimised</td>
<td>Van den Berg and Gademann (1999)</td>
</tr>
<tr>
<td>• Any number of storage and retrieval requests</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• A separate input and a separate output point that might be positioned at different locations of the rack</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Eynan and Rosenblatt (1993) conclude that quite significant savings in interleaving times can be obtained by combining nearest-neighbour scheduling with class based storage assignment. Schwarz et al. (1978) investigate the performance and predictions of previously developed deterministic models in a stochastic environment. It is found that the results of the models hold in this environment. However, the predictions in improvements are generally larger than the actual improvements.

Other approaches to find solutions to the sequencing problem include neural networks (e.g., Wang and Yih, 1997), expert systems (e.g., Linn and Wysk, 1990a,b), artificial intelligence (e.g., Seidmann, 1988), genetic algorithms (e.g., Krishnaiah Chetty and Sarveswar Reddy, 2003) and the Taguchi method (e.g., Lim et al., 1996). These methods can be applied in situations with high uncertainty and little information. Furthermore, these methods are capable of learning and adapting to changes in the environment, such as fluctuations in demand. The output can consist of combinations of storage assignment, retrieval location selection, queue selection and job sequencing.

8.2. Extensions of the basic sequencing problem

An extension of the previously described basic scheduling problem is the problem in which storage and retrieval requests with release and due times need to be scheduled.
Models and heuristics have been proposed that intend to minimise the sum of earliness and tardiness penalties in the case that all requests or a group of requests have one common due time (e.g., Lee and Kim, 1995; Elsayed and Lee, 1996; Elsayed et al., 1993; Linn and Xie, 1993).

Other configurations of an AS/RS such as multiple I/O-points per aisle might require new solution approaches. Randhawa et al. (1991) use simulation to evaluate various scheduling rules for systems with one input/output station at each aisle and systems with two input/output stations at each aisle. It is shown that reductions in expected crane round trip times and in throughput times can be obtained for systems with two input/output stations.

Kanet and Ramirez (1986) add location selection to the scheduling problem in case products are stored at multiple locations. This decision is incorporated in the integer programming model by including costs per retrieval operation and costs to separate items at a location into different parts related to different retrieval orders. Jaikumar and Solomon (1990) study the prepositioning of pallets in periods the AS/RS would otherwise be idle. By positioning loads that are expected to be needed in the next time period, closer to the I/O-point it is possible to reduce travel times during the actual handling of retrieval requests.

8.3. Sequencing for other types of AS/RSs

Some research has been done to the scheduling of storage and retrieval requests for other types of AS/RSs. Several authors address the problem of sequencing storage and retrieval requests in a twin-shuttle AS/RS. Due to the double capacity of the crane, more routing options emerge since cycles can be performed in which at most four locations are visited before returning to the I/O-point. Simple nearest-neighbour heuristics (Sarker et al., 1991, 1994) and a minimum perimeter heuristic (Keslerla and Peters, 1994) have been developed.

In a miniload AS/RS current retrieval requests become future storage requests, since loads are returned into the system after items have been picked. Therefore, the problem can be reformulated such that only a queue of retrieval requests exists, which result in a less complicated problem. The picker at the end of the aisle also needs to be incorporated in the model. Retrieval requests are rearranged such that successive requests are located close to each other. Storages and retrievals which are close to each other can be paired by a nearest-neighbour heuristic (Mahajan et al., 1998).


Summarising, various methods are described in the literature to schedule storage and retrieval requests such that the total (empty) distance travelled is minimised. The situation considered in the majority of the literature concerns an unit-load AS/RS working in one aisle with one input/output station. For some specific instances optimal sequencing methods exist. However, the general dynamic sequencing problem is hard to solve and, therefore, heuristics have been developed to find feasible schedules. Hardly any attention has been paid to methods for the scheduling of storage and retrieval requests when each aisle has two or more input/output stations or when a single crane operates in multiple aisles. Although multi-shuttle cranes have proven to be successful, only a few heuristics have been developed for quadruple (or more) command scheduling.

9. Performance measurement

In evaluating the design and control rules of an AS/RS several performance measures can be used. Based on the literature overview presented in this paper, we can at least consider the following performance measures:

- travel time per request,
- number of requests handled per time period (e.g., Azadivar, 1986; Foley et al., 2002),
- total time required to handle a certain number of requests,
- waiting times of cranes of the AS/RS,
- waiting times of products to be stored/retrieved,
- number of requests waiting to be stored/retrieved (e.g., Hur et al., 2004).

Lee (1997), Malmborg and Altassan (1997) and Bozer and Cho (2005) propose throughput performance models. The models of Lee (1997) and Malmborg and Altassan (1997) are similar but published independently in the same year (Eldemir et al., 2003). Eldemir et al. (2003) concludes that the more time-efficient model of Bozer and Cho (2005) slightly overestimates the throughput and that the other model slightly underestimates the throughput. Eldemir et al. (2004) propose more time-efficient throughput models which can be used to estimate space requirements for both random and class-based dedicated storage.

Clearly, throughput estimates are the inverse of the expected travel times of an AS/RS. As a result, estimating travel times is very important in designing AS/RSs. Numerous research has been done in this area. Sarker and Babu (1995) presented a short review of travel time models for AS/RSs. Here we extend this overview and present a categorisation of all literature in Table 8 by discussing the main characteristics of each paper. We discuss some of these papers in more detail to provide a rough line of research in this area. First we discuss travel time models for single unit-load aisle-captive AS/RSs. Thereafter, we discuss relevant literature for other types of AS/RSs.

9.1. Travel time models for single-shuttle unit-load AS/RSs

Hausman et al. (1976) were one of the first to present travel time models for single-shuttle unit-load AS/RSs.
Table 8
Overview of research in travel time models for different types of AS/RS for different types of layouts, racks, location of I/O points, storage assignment methods, scheduling approaches and operational characteristics

<table>
<thead>
<tr>
<th>AS/RS characteristics</th>
<th>Authors</th>
<th>Layout</th>
<th>Rack</th>
<th>I/O point(s)</th>
<th>Storage</th>
<th>Schooling</th>
<th>Operational characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Single aisle</td>
<td>Multiple aisles</td>
<td>Square-in-time</td>
<td>Non-square-in-time</td>
<td>Continuous</td>
<td>Discrete</td>
</tr>
<tr>
<td>Unit-load</td>
<td>Hausman et al. (1976)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>Graves et al. (1977)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>Boone and White (1984)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>Han et al. (1987)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>Hwang and Ko (1988)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>Rosenblatt and Eynan (1989)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>Kim and Soledmann (1990)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>Hwang and Lee (1990)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>Eynan and Rosenblatt (1993)*</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>Eynan and Rosenblatt (1994)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>Chang et al. (1995)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>Kenvelo and Papamichalou (1993)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>Pan and Wang (1996)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>Mansari (1997)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>Thieme and Brandeau (1998)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>Lee et al. (1999)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>Wen et al. (2001)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>Ashayeri et al. (2002)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>Eldemir et al. (2004)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Multiple shuttle</td>
<td>Sarker et al. (1991)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>Koostra and Poter (1994)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>Meller and Mungarawarina (1997)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>Malmborg (2000)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>
The authors proposed estimates for single command scheduling in square-in-time continuous racks. Random, full-turnover, two- and three-class-based storage assignment policies were considered. Graves et al. (1977) extended those results by also considering interleaving times resulting from a first-come-first-served (FCFS) dual command scheduling policy.

Bozer and White (1984) relaxed some of the assumptions by proposing travel time models for rectangular racks with alternative single I/O-points. The authors considered FCFS dual command scheduling and random storage as control policies. The authors introduced $b$ as the shape factor of the rack to deal with rectangular racks. Here $b = \min(t_h/T, t_v/T)$, where $t_h$ indicates the horizontal travel time to the farthest column, $t_v$ the vertical travel time to the farthest row and $T = \max(t_h, t_v)$. Based on empirical tests, the authors conclude that the model’s performance is satisfactory.

Other authors since then mainly continue the research of Hausman et al. (1976), Graves et al. (1977) and Bozer and White (1984) by studying different control policies, configurations of AS/RSs and/or operational characteristics. Instead of a FCFS-policy a nearest-neighbour (NN) policy can be used to schedule requests. Recursive procedures (Rosenblatt and Eynan, 1989; Eynan and Rosenblatt, 1994) and closed-form expressions (Kim and Seidmann, 1990; Kouvelis and Papanicolaou, 1995) have been proposed for $n$-class based storage and full turnover storage. From Table 8 it can be concluded that only a few papers address dedicated storage as storage assignment policy (Mansuri, 1997; Eldemir et al., 2004). Instead of addressing discrete values in applying their control policies several authors (Thonemann and Brandeau, 1998; Pan and Wang, 1996; Ashayeri et al., 2002) study stochastic environments with varying demand.

Different configurations of single-shuttle unit-load AS/RSs that have been studied include multi-aisle AS/RSs (see Hwang and Ko, 1988) and racks with unequal sized cells (see Lee et al., 1999). The results of the model of Hwang and Ko (1988) can be used to determine the minimum number of cranes and number of aisles served by each crane.

Almost all papers mentioned so far assumed that the operational characteristics of an AS/RS could be ignored. Hwang and Lee (1990) incorporate both the maximum velocity of a crane and the time required to reach the peak velocity or to come to a halt. Chang et al. (1995) extend the work of Bozer and White (1984) by including acceleration and deceleration rates instead of assuming constant speed. Wen et al. (2001) extend the work of Chang et al. (1995) by considering class-based and full-turnover-based storage assignment policies.

9.2. Travel time models for other types of AS/RSs

New travel time estimates are required for multi-shuttle AS/RSs to deal with quadruple and even sextuple (e.g.,
Meller and Mungwattana, 1997) command scheduling. Both the work of Bozer and White (1984) and Han et al. (1987) have been extended by various authors (Sarker et al., 1991; Kesler and Peters, 1994; Malmborg, 2000) for multi-shuttle AS/RSs while examining different control policies.

In miniload AS/RSs picker and crane are dependent on each other; delays of one influence the performance of the other. Foley and Frazelle (1991) consider a miniload square-in-time AS/RS operating under a FCFS dual command scheduling policy and random storage. Also closed-form travel time expressions for miniload AS/RSs with one- and two-class-turnover based storage have been formulated (Park et al., 2003, 2006).

Person-on-board AS/RSs usually handle multiple locations in one tour. Authors (e.g., Elsayed and Unal, 1989; Guenov and Raeside, 1992; Chiang et al., 1994) typically derive expressions for upper and lower bounds for travel times based on the number of locations to be visited given a storage assignment and sequencing policy.

Hwang and Ha (1991) present travel time models for single and double carousel systems. The authors investigate the effects on performance of information availability of upcoming orders. Clearly, an increase in throughput can be obtained with this additional information, however, the increase in throughput is lower than one might expect.

Finally, some authors present travel time estimates for less frequently used types of AS/RSs, such as a rotating tower cranes (Koh et al., 2002), heavy load platform based AS/RSs (Hu et al., 2005) and flow-rack AS/RSs (Sari et al., 2005).

Summarising, it can be concluded that travel time models for both square-in-time and rectangular racks have been developed for unit-load AS/RSs operating under most common storage assignment policies and two sequencing heuristics derived from practice. So far no studies have been performed that include dwell-point rules in travel time estimates. Compared to unit-load AS/RSs, many issues, such as N-class based storage or operational characteristics, have not been addressed in travel time models for other types AS/RSs.

10. Conclusions and further research issues

In both manufacturing and distribution environments, AS/RSs are used to store products and to retrieve products from storage in response to production orders or customers’ orders. In designing an AS/RS, various physical design problems and control problems need to be solved (see Section 3). Literature related to the various interrelated physical design and control problems, such as storage assignment, batching, dwell-point location and request scheduling was treated in Sections 4–8. To evaluate the performance of AS/RSs, we can use, for example, the travel time estimates we discussed in Section 9.

From the literature survey, we conclude that most of the literature addresses design and control problems in static environments. However, in today’s world of rapidly changing customers’ demand, small internet orders, tight delivery schedules, high competition and high service level requirements, it will be increasingly difficult to maintain a good performance when using existing static solution techniques. The research in the field of AS/RSs should now move towards developing models, algorithms and heuristics that include the dynamic and stochastic aspects of current business. In this context, one can think of self-adaptive storage assignment methods, on-line-batching policies and dynamic dwell-point rules. Also algorithms for physical design may need to focus more on robustness of the design than on perfect optimality to ensure that the system will be capable of remaining efficient in yet unknown future situations.

Furthermore, almost all existing papers just address one or two decision problems simultaneously, instead of jointly optimising a combination of physical design problems and control problems (including batching, dwell-point rules and I/O-point decisions). Obviously, it is not a simple exercise to include a multitude of design and control aspects in one model. However, we would like to encourage the development of simulation models which compare numerous designs while taking combinations of design aspects and control policies into account.

Little attention has been paid so far to the relationship between AS/RSs and other material handling systems in production or distribution facilities. Especially in situations where the AS/RS is just one of many systems, total warehouse performance cannot be assessed by simply adding up performances of all individual systems. An integrated approach would be desirable. Therefore, we advice – as a first step – to develop approaches which simultaneously optimise the design of an AS/RS and another material handling system. For example, by explicitly considering the interface between an AS/RS and a conveyor system, or by analysing the impact of replenishments by the AS/RS to a separate order-picking area.

Except for those general issues, further detailed research can also be advised for each of the following issues.

- Models to assist in AS/RS type selection.
- Analytical and simulation models for the design of non-traditional AS/RSs (e.g., multi-shuttle AS/RSs).
- Storage assignment policies for multi-shuttle AS/RSs.
- Storage assignment policies for AS/RSs working in multiple aisles and/or multiple I/O-points.
- Policies which simultaneously address storage assignment and batching of orders.
- Superior heuristics for batching that outperform all existing rules in various settings.
- Dwell-point rules for non-traditional AS/RSs.
- Algorithms or heuristics to schedule AS/RSs in a single aisle with multiple I/O-points.
- Travel-time models for AS/RSs operating in a single aisle with multiple I/O-points.
- Travel-time models which incorporate operational characteristics of non-traditional AS/RSs.
References


