Chapter 2
Storage Systems and Policies

Marc Goetschalckx

Abstract A storage system is an engineered system with the function to store materials, in other words to hold materials until they are needed. These materials come in an enormous variety from consumer products such as TVs in local distribution centers, emergency drug doses for battling a biological attack on a city, vertical storage silos for grain, to the strategic reserve of main battle tanks parked in the dessert. Storage systems are an essential component of virtually every supply chain. While most storage systems are stationary, some are moveable such as the truck of a repair man that holds a ready inventory of service parts, a bunker ship that resupplies a navy fleet, an intercity long-haul truck, or an ocean-going intermodal container ship. Irrespective of the type of material, the geographical location of the storage system, or the size of the storage system, storage systems have three main processes: put-away, holding, and retrieval. The first process put materials into storage; the second process holds the materials in a stationary position inside the storage system, and the third process removes materials from storage and is often also called order picking.

2.1 Introduction

As any engineered system storage systems have structure and behavior. The structure concerns the physical form of the storage system, while the behavior is concerned with the management of the storage system. The performance of a storage system in its physical environment is based in part on the efficiency with which it executes its basic function. The performance is dependent on its design,

M. Goetschalckx (✉)
Georgia Institute of Technology, 765 Ferst Drive, Georgia 30332-0205, USA
e-mail: marc.goetschalckx@isye.gatech.edu

R. Manzini (ed.), *Warehousing in the Global Supply Chain*,
DOI: 10.1007/978-1-4471-2274-6_2,

which determines its structure, and on the management policies, which determine its behavior. Given the complexity and importance of storage systems, a large body of literature exists that describes engineering methodology to either determine or optimize the structure or behavior of storage systems.

The performance of a storage system depends on four internal characteristics and their interrelations: (1) storage capacity or equivalently storage density; (2) ease of access to storage locations; (3) complexity of the internal structure; and (4) level of information technology. The performance of a storage system also depends on external characteristics such as number of products, type of products, total inventory to be stored, and type and balance of the input and output flows. External characteristics are always assumed to be given parameters in the following methodology. An example of how external product characteristics influence the storage system is a grocery warehouse with products in three temperature classes: ambient, refrigerated, and frozen. The products in one temperature section of the warehouse cannot be mixed with products in the other sections. An example of where the design methodology partitions the warehouse is a forward-reserve distribution center where the most commonly retrieved products are stored together in the forward area and the relatively rarely retrieved products are picked from the reserve area. Splitting the products in those two classes and the warehouse into two sections allows both sections of the warehouse to perform better than a joint warehouse, but a cost of increased internal complexity. In general, determining the number of sub warehouses under one roof of the overall warehouse is part of the warehouse design process. Such partitioning inside a single warehousing system is a common occurrence in practice. In the remainder of this research it is also assumed that all the products to be stored are compatible with respect to their physical characteristics such as temperature, weight, chemical type, and hazardousness.

The focus of this paper is two fold. The first emphasis is on the design of the structure of a storage system, specifically determining the size or holding capacity of the system. This approach ignores many other important characteristics of the storage system such as safety, structural soundness, energy efficiency, and environmental impact. The second focus is on storage policies, which are the management policies that determine where an incoming item is placed in the warehouse. Even for this restricted problem, a single paper can only touch on one or more examples of storage systems. This paper focuses on unit load storage systems since the simplicity of their physical form allows for a clearer illustration of the underlying principles. The methodology is based on the construction of mathematical models of the storage system.

2.2 Literature Review

Providing a comprehensive literature review of warehousing systems, storage systems, and storage policies is not possible given the large variety of systems and the large body of research. Earlier research reviews were compiled by Cormier and Gunn (1992) and Rouwenhorst et al. (2000). Some of the most recent reviews are given in Gu et al. (2005, 2010). Most of the results are concerned with modelling or optimizing the behavior of warehousing systems. A comprehensive, modeling-based engineering methodology for warehouse design does not yet appear to exist.

It should be noted that the methodology presented here is independent of how the expected travel time to a storage location is computed, be it by physical observation, simulation, mathematical model, or closed form equation. There exists a large variety of research results to determine these travel times for various hardware and storage policy combinations.

2.3 Unit Load Storage Systems and Policies

2.3.1 Introduction

In unit load warehouses it is assumed that all the items in the warehouse are aggregated into units of the same dimensions that can be moved, stored, and controlled as a single entity. Typical examples of unit loads are pallets, intermodal containers, and wire baskets. Furthermore, it is assumed that no incompatibilities exist between the materials to be stored. It is also assumed that all the storage locations are the same size and each location can hold any unit load. Hence unit load storage system eliminates many of the material, material-to-location, and location specifics and complications.

One of the most common examples of a unit load storage system is the single-deep pallet rack. In this storage system the warehouse contains a number of parallel travel aisles; on each side of an aisle is a steel rack capable of holding a single pallet in each of its storage locations. The storage locations, also called pallet openings, are conceptually arranged in horizontal rows and vertical stacks or columns. The storage locations are completely inert, i.e. their location cannot be changed and the locations cannot move the unit loads stored in them. The put-away and retrieval operations are performed by a material handling truck such as a forklift truck or a turret truck. A single-deep unit load pallet rack storage system and its material handling truck are shown in the next three figures.

Unit load systems can be further divided based on the access allowed to individual stored loads. Random access unit load systems allow access to all unit loads in all storage locations with the same amount of effort except for the travel to the storage location. Examples are the single-deep pallet rack, person-aboard tote stacker, and side-loaded intercity trucks (Europen model). If a unit load system is not random access, then the effort of accessing a load is different (again excluding the travel to reach the storage location). Examples are double-deep pallet rack, deep lane storage, block stacking, intermodal container storage yard, an intermodal container ship, or a rear-loaded intercity truck (United States model). Reaching the unit load at the back end of a lane or at the bottom of stack requires much more effort than reaching the most accessible unit load since other unit loads may block the access.

A second important characteristic of storage system is their command cycle. A warehouse is said to operate under single command when on each trip of the material handling truck or device a single operation is performed, be it either storing a load or retrieving a load. During a dual command cylce first a unit load is put away in the system and then a unit load is retrieved from the system. Finally, individual items such as individual cartons can also be retrieved from the unit load using a order picking truck. The order picking truck is said to operate under multi-command. In the latter storage systems, typically the unit load is stored by a different material handling truck operating in single command mode.

Unit load systems exhibit the standard advantages and disadvantages of standardization and unitization. Some of the major advantages are the standardized material handling and storage operations and equipment, the reduced effort of control, and the high macro storage volume utilization. Macro storage space utilization is high because a storage location is never left unused or empty because a storage load does not fit. If the unit load storage system does not have hardware that holds the loads, then a loss of storage volume occurs when only a few unit loads remain in a lane or a stack but these few unit loads still consume the full space for that lane or stack. This loss is also called the honeycombing loss, see White and Francis (1971). Some of the major disadvantages of unit load storage systems are the effort required to assemble and disassemble the unit load, the cost of the containers to hold the unit load and the empty container movement, storage, and management, and low micro storage volume utilization. The low micro storage volume utilization is illustrated by the top pallet on the right-hand side of the aisle in Fig. 2.1; only a few boxes remain stored on the pallet and most of the volume of this storage location is empty (Figs. 2.2 and 2.3).

2.3.2 Single Command Operation Cycles

Because of the elegant simplicity of the problem and because such unit load systems have been implemented frequently in practice, the problems of designing and managing unit load storage systems have been studied extensively in the literature, especially when the storage system hardware implementation is an automated storage and retrieval system (ASRS) for unit loads such as pallets. Most of the early results were for storage systems operating in a single command cycle mode, where on each trip of the material handling truck a single operation is performed, be it either storing a load or retrieving a load.

2.3.3 Individual Unit Load Storage Model

A (unit load) storage model is a mathematical optimization model that determines in which storage location the (unit) loads are to be held in order to minimize the material handling effort and cost of put-way and retrieval, while observing the

Fig. 2.1 Single-deep pallet rack (photos courtesy of LogisticsCAD, used with permission)

storage capacity of the storage system. A storage policy is a management rule that makes the same determination. A storage policy may not generate the optimal storage plan, but it is easier to implement since it does not require the solution of an optimization model. Some of the earliest results were developed by Heskett (1963, 1964) and some the earliest mathematical models were developed in

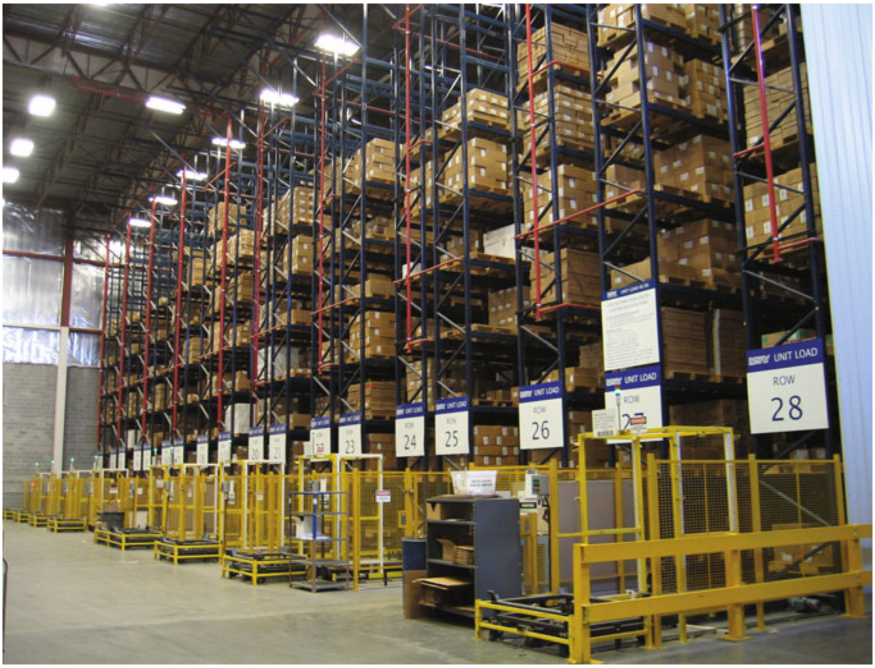

Fig. 2.2 Parallel aisles of a pallet rack (photos courtesy of LogisticsCAD, used with permission)

Malette and Francis (1972) and White and Francis (1971). Hausman et al. (1976) studied the performance of storage system under dual command cycles, where on each trip of the material handling truck first a storage and then a retrieval operation are performed. Malette and Francis (1972) and Hausman et al. (1976) proved that for single command product turnover-based dedicated storage minimizes the total travel time required to execute all the tasks. In a series of articles Malmborg together with several co-authors, starting with Malmborg and Deutsch (1988), extended these results to dual command cycles.

2.3.3.1 Individual Load Storage Model Notation

H	The number of equal-sized periods in the planning horizon. The length of a period may range from minutes to months depending on the application.
arr_i, dep_i	Arrival and departure period of individual unit load i in the storage system; the unit load occupies a storage location starting with the arrival period and up to and including the departure period.
b_i	Residence vector of individual unit load i in the storage system, this vector has elements equal to one from the arrival to the departure

Fig. 2.3 Turret truck accessing a unit load (photo courtesy of LogisticsCAD, used with permission)

period of the unit load i and elements zero anywhere else. Note that the residence vector exhibits the consecutive ones property, i.e. all non-zero elements are equal to one and they appear in a single contiguous section of the vector. The dimension of the vector is H. This vector is sometimes also called occupancy vector.

c_{ij} Expected one-way travel time or cost for a unit load i stored in location j.

x_{ij} Decision variable equal to one if unit load i is stored in storage location j.

2.3.3.2 Factoring Condition

If the cost of the material handling operation for a location is independent of the stock keeping unit (SKU) stored in that location, then the cost elements in the objective function can be simplified and can be computed in advance of the solution of the model. If this travel independence or factoring condition is satisfied, then it is assumed that all the items in the warehouse have the same probability mass function for selection of a dock or input/output point. The travel independence condition is equivalent to

$$p_{pk} = p_k \; \forall p \text{ or } e_j = e_{pj} \quad \forall p. \tag{2.1}$$

The expected one-way distance for each location can then be computed as

$$e_j = \sum_k p_k \cdot c_{jk} \tag{2.2}$$

The travel independence condition was first specified by Malette and Francis (1972) under the name factoring condition, since, if this condition is satisfied, the expected travel time to a particular location holding a particular product can be factored or computed as the product of the expected travel time of the location and the frequency of access of the product. In practice it is most often satisfied if the storage system has only one input/output point, e.g. a single aisle in an automated storage and retrieval system, or if all SKUs follow the same path through the warehouse, e.g. the receiving dock is on one side of the facility and the shipping dock is at the opposite side.

2.3.3.3 Individual Load Storage Model

$$\min \quad \sum_{i=1}^{M} \sum_{j=1}^{N} 4c_{ij}x_{ij} \tag{2.3}$$

$$\text{s.t.} \qquad \sum_{j=1}^{N} x_{ij} = 1 \qquad \forall i \tag{2.4}$$

$$\sum_{i=1}^{M} b_i x_{ij} \leq 1 \qquad \forall j \tag{2.5}$$

$$x_{ij} \in \{0, 1\} \tag{2.6}$$

The objective is to minimize the access cost associated with the storage of an item in particular storage location. Constraint (4) ensures that every unit load is stored exactly once. Constraint (5) ensures that the residences of unit loads do not overlap for a particular location since the location has a capacity of a single unit, i.e. multiple items can be stored subsequently in the same storage location but not at the same time.

The individual load storage model is a vector assignment problem (VAP). The VAP is an extension of the standard assignment problem (AP), where the constraints that prohibit sharing of the storage location are not for a single time period but for a vector of time periods. Since the occupancy vector for every item exhibits the consecutive ones property, the vector of constraints (5) can be converted into network flow constraints by the following transformation.

The transformation has two steps. Recall that there are H time periods in the planning horizon.

1. Add a row of zeroes corresponding to the flow balance constraint of node or period $H + 1$. Add a zero element for row $H + 1$ to the right-hand size vector.
2. Execute iteratively the following linear row operation (including the right-hand size column)
   for $r = H$ *down to* 1

$$row[r+1] = row[r+1] - row[r].$$

Each transformed column of the occupancy vector now contains a single positive and negative one, indicating that this column corresponds to a directed arc in a network with (H+1) nodes. The positive one occurs at the arrival period and the negative one occurs at the period following the departure. The positive one corresponds to the start node of the arc in a network with (H+1) nodes and the negative one corresponds to the terminal node of the arc. The right-hand side vector has been transformed into a vector with a positive one in the first period (row) and a negative one in the last row, which corresponds to period $H + 1$. The VAP has been transformed into problem with the block diagonal structure shown in the following figure and where each block column corresponds to a storage location and matrix B corresponds to the directed arcs in the network.

$$A = \begin{bmatrix} I & I & I & .. & I \\ B & 0 & 0 & .. & 0 \\ 0 & B & 0 & .. & 0 \\ .. & .. & .. & .. & .. \\ 0 & 0 & 0 & .. & B \end{bmatrix},$$

Unlike the AP the VAP does not posses the integrality property, so constraint (6) must be explicitly enforced and the individual load storage model requires a mixed integer programming solver. But its linear relaxation is very close to the binary formulation and many contemporary solvers recognize the special block diagonal structure of the problem and the network structure of each block row. Large instances can be solved relatively quickly. For even larger problem instances Lagrangean relaxation or other decomposition techniques can be used to reduce the problem size and the required memory for the problems that have to solved as a whole by the solver, but this increases the solution time. The model is thus useful for storage systems where a modest number of unit loads are to be stored during the planning horizon. However, for standard warehouses, the instance size makes solving the model with mathematical programing solvers impossible. Consider the example where there are 20,000 storage locations and 100,000 unit loads need to be stored during the planning horizon, the number of variables for this instance would be 2,000,000,000. The storage problem needs to be aggregated to yield a smaller and solvable problem instance.

2.3.3.4 Individual Load Storage Policy

A heuristic storage policy for individual unit loads can be determined if the expected travel time for a storage location is independent of the unit load that is stored in that location. A warehouse that satisfies this condition is said to satisfy the factoring condition. The storage policy sorts all the unit loads to held during the planning horizon by increasing departure time and sorts all storage locations by increasing expected travel time. The storage policy then assigns the unit loads by increasing index to the storage location with the lowest index that feasibly hold the unit load, i.e. the storage location that is empty during the residence time of the unit load. Since the storage policy only requires two sorting operations it can be applied to large problem instances. Since it is a heuristic policy no information on its performance for a particular case is known, either with respect to the required warehouse size or with respect to the travel cost.

2.3.4 Aggregate Unit Load Storage Model

Aggregation of the individual load storage model is based on the assumption that the requirements on the storage system are stationary. The most common assumption is that each SKU has a constant demand rate over the planning horizon and is replenished periodically in batches of identical quantities. Stochastic behavior that differs from the deterministic mean values is assumed to be absorbed by the safety stock for that SKU.

The most common aggregation of individual unit loads is to consider the unit loads as inventory of a particular product or SKU. This is called product based aggregation. Two unit loads of the same SKU are always treated identically. The corresponding storage model is then based on product characteristics and the resulting solution yields product based storage policies. This product based aggregation is shown next. However, the individual unit loads can also be aggregated based on how long they stay or reside in the system, i.e. how long they occupy their storage location. This duration-of-stay based aggregation is developed later on. One particular duration-of-stay based storage policy that divides the unit loads into two classes, i.e. the short-term-residents and the long-term-residents, is better known as cross docking. In cross docking the short-term-resident unit loads are never placed or held in the main storage area but moved directly from receiving to shipping or moved from receiving into a short term order picking area and from there to shipping.

2.3.4.1 Storage Model Notation

I_{pt} On-hand inventory of product p during time period t expressed in unit loads.

N_{pol} Required number of storage locations in the warehouse when using a particular storage policy such as dedicated, shared, or maximum, which is indicated by the subscript.

q_p Replenishment quantity of product p in unit loads, also called the cycle inventory of product p.
s_p Safety inventory quantity of product p in unit loads.
r_p Demand rate for product p.
r_{pk} Demand rate for product p entering or leaving through warehouse dock k.
p_{pk} Probability for a unit of product p to enter or leave through warehouse dock k. If the product enters through one dock on one side of the warehouse and departs through another dock at the opposite site, then each of those docks would have a probability equal to 0.5.
f_p Frequency of access of a location in zone p or assigned to product p.
e_j Expected one-way travel cost to location j.
c_{jk} Travel cost to location j from warehouse dock k
e_{pj} Expected one-way travel time or cost for a unit load of product p stored in location j.
T_p Total travel time or cost for product p or zone p during the planning horizon.

By definition, the following relationships exist since the storage of one unit load requires two material handling operations.

$$r_p = \frac{1}{2}\sum_k r_{pk} \text{ and } p_{pk} = \frac{r_{\mathrm{pk}}}{2r_p} \tag{2.7}$$

$$e_{pj} = \sum_k p_{pk} \cdot c_{jk}. \tag{2.8}$$

2.3.4.2 Warehouse Sizing Problem

The storage policies partly determine the behavior of the storage system but also impact the required warehouse size, which is an element of the structure of the storage system. In other words the storage policy which is a management policy must be taken in consideration when designing the structure of the warehouse. Consider the class of product dedicated storage policies where each product is assigned a dedicated section of the warehouse. Such policies are often implemented because of their simplicity. Under such policies a particular SKU will always be located in the same section of the warehouse which facilitates put-away and order picking operations and inventory management. However, each section dedicated to a product must be large enough to hold the maximum inventory of that product during the planning horizon. The required warehouse size for this policy is the largest of all the warehouse sizes required by any policy.

$$N_{\mathrm{DED}} = \sum_p \left(s_q + q_p\right) = \sum_p \max_t \left\{I_{pt}\right\} = N_{\mathrm{MAX}}. \tag{2.9}$$

Now consider a storage policy which assigns an arriving unit load randomly to one of the open storage locations of the warehouse. This policy is called random storage and denoted by RAN. A similar policy assigns an arriving unit load to the open location with the lowest expected travel cost for that product. This policy is called closest open location and denoted by closest open location (COL). Since no internal restriction is placed on the location of a unit load, the warehouse must be large enough to hold the maximum aggregate inventory over the time horizon. This required warehouse size is the smallest warehouse size among all the storage policies.

$$N_{\mathrm{RAN}} = \max_t \left\{ \sum_p I_{pt} \right\} = N_{\mathrm{MIN}}. \tag{2.10}$$

A warehouse is said to be balanced if the aggregate number of arriving unit loads is equal to the aggregate number of departing unit loads for every time period. In a balanced warehouse the aggregate number of unit loads stored remains constant. The smallest possible warehouse size for a balanced warehouse is the average aggregate inventory. If the warehouse were to operate without safety inventory for any product and if the classic triangular inventory pattern is assumed then the average aggregate inventory is equal to half the sum of the maximum product inventories.

The policy warehouse size ratio of the required warehouse size to the maximum required size under dedicated storage is denoted by α. Alpha has a range of [0.5, 1]. The upper bound is achieved by any product dedicated storage policy since they all require the maximum warehouse size. The lower bound is achieved by any storage policy that fully shares the storage locations among the unit loads without any restrictions and when the warehouse is balanced and when no safety inventory is stored. Both random storage and closest open location are examples of such fully sharing policies. The real value of α can most easily be determined by simulation. It depends on the function of the warehouse and daily or seasonal material flow patterns. A conservative choice for the value of α is 0.85 or higher.

$$\alpha = \frac{N}{N_{\mathrm{DED}}} \in [0.5, 1]. \tag{2.11}$$

The warehouse balance β can then be computed as

$$\beta = 2(1 - \alpha) \tag{2.12}$$

$$\alpha = 1 - \frac{\beta}{2} \tag{2.13}$$

The warehouse balance indicates how well balanced the input and output flows of the warehouse are. A value of $\alpha = 1$ or $\beta = 0$ indicates that the flows are not balanced at all. A value of $\alpha = 0.5$ or $\beta = 1$ indicates that the flows are perfectly balanced.

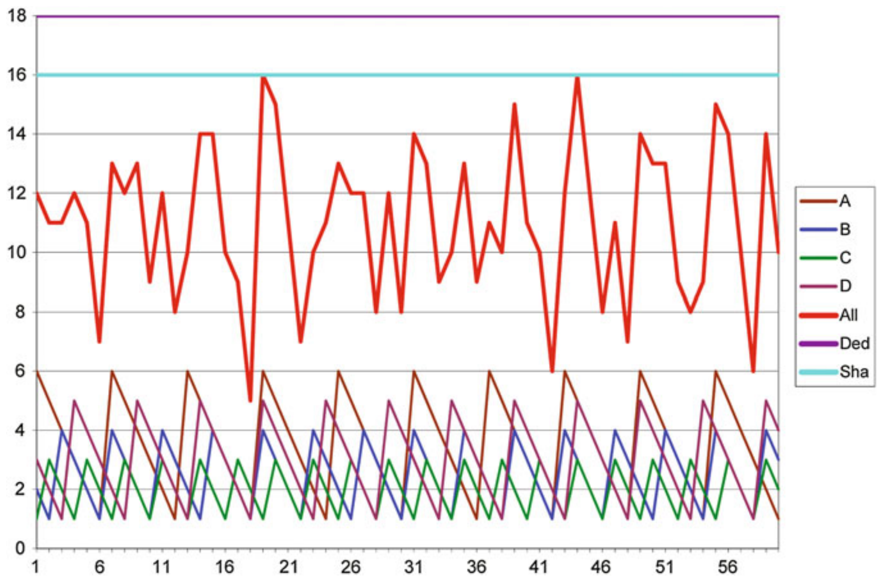

Fig. 2.4 Inventory patterns for the warehouse example with four products

Consider the following warehouse example. The warehouse stores four products A, B, C, and D with replenishment quantities equal to 5, 4, 3, and 2, respectively. Each product has a safety inventory of one unit load and a demand rate of one unit load per time period. The inventory patterns for the individual products are shown in Fig. 2.4. The aggregate inventory is indicated by All. The required warehouse size by a perfectly sharing storage policy is equal to the maximum of the aggregate inventory, which is equal to 16 and indicated by Sha. The required warehouse size by a product based dedicated storage policy is the sum of the maximums of the product inventories, i.e. the sum of 6, 5, 4, and 3, respectively, which is equal to 18 and indicated by Ded. The warehouse size ratio for this example is then

$$\alpha = \frac{16}{18} = 0.889$$

$$\beta = 2(1 - 0.889) = 0.22$$

Observe that for this particular example the average inventory is 11 and the minimum inventory during the time horizon shown in the figure is five units (Fig. 2.4).

2.3.5 Product-Based Storage Model

Product based storage policies determine the storage location of an arriving unit load based on the characteristics of the product or SKU this unit load holds.

Product characteristics that are commonly used are the demand rate, the maximum inventory over the planning horizon, or the turnover ratio. Product based storage policies also partition the storage locations into a number of sections each of which is dedicated to hold the unit loads of a group of SKUs. The numbers of sections that are commonly used are two, three, or the number of products when each section hold unit loads belonging to a single SKU. A very common product based storage policy is to divide the products in fast, medium, and slow movers depending on their demand rate.

2.3.5.1 Product Dedicated-Based Storage Model

If the warehouse is partitioned in sections that each are dedicated to hold unit loads of a single SKU then the following storage model determines the optimal product dedicated storage policy. Note that e indicates the one-way travel cost, so it has to be doubled for the trip to either store or retrieve the unit load.

Each unit load that is held in a particular storage location requires two material handling operations, one to store the load and one to retrieve the load. The frequency of access f of a storage location is thus twice the number of times this location is used to hold a unit load. Most warehouses operate on a first-in first-out (FIFO) basis with respect to the unit loads of a particular SKU to avoid spoilage and obsolescence of the product held in the unit load. This implies that there is no difference between the storage locations that hold safety stock and cycle stock for that SKU since in the long run they all will be accessed the same number of times. The frequency of access of a storage location can then be approximated by the following expression.

$$f_i = \frac{2r_i}{q_i + s_i} \tag{2.14}$$

$$\text{Min} \quad \sum_{i=1}^{M} \sum_{j=1}^{N} \left(2f_i e_{ij}\right) x_{ij} \tag{2.15}$$

$$\text{s.t.} \quad \sum_{j=1}^{N} x_{ij} = (q_i + s_i) \quad \forall i \tag{2.16}$$

$$\sum_{i=1}^{M} x_{ij} \leq 1 \qquad \forall j \tag{2.17}$$

$$x_{ij} \geq 0 \tag{2.18}$$

In the product dedicated storage model, the cost of a material handling operation depends both on the storage location and on which SKU is stored in that location. In practice this often occurs when the different SKUs have different interface patterns with the storage systems docks. For example, some SKUs may arrive by truck while other SKUs arrive by rail and all SKUs depart through the truck shipping dock.

2.3.5.2 Product Turnover Storage Policy

If the storage system satisfies the travel independence or factoring condition, then the travel cost to or from a storage location can be computed without knowledge of which unit load will be stored in this location. This computation can occur outside the model.

$$\text{Min} \quad \sum_{i=1}^{M} \sum_{j=1}^{N} \left(2 f_i e_j\right) x_{ij}. \tag{2.19}$$

This objective can be minimized by sorting the products by decreasing frequency of access and the storage locations by increasing expected access cost. The storage policy then iteratively assigns a number of storage locations equal to the maximum inventory of the next product to the next storage locations. If in the frequency of access vector each product is represented by a number of elements equal to its maximum inventory, then the objective value is equal to the inner product of the frequency of access and expected roundtrip travel cost vectors. This storage policy is known as product dedicated storage and also denoted as the cube-per-order index (COI) by Heskett (1963, 1964). It has been studied extensively since then. It is the optimal storage policy among the class of product dedicated storage policies with respect to the travel time or cost provided the factoring condition is satisfied. But because it belongs to the class of product dedicated storage policies it also requires the maximum warehouse size to store the unit loads. The conditions for the optimality of the product turnover-based storage or COI have not always been listed explicitly, especially in the material handling trade literature. This has led to some confusion on the optimality of this policy.

2.3.5.3 Two Class Product Turnover Storage Policy

If the storage policy allows sharing of the storage locations among the different SKUs then the required warehouse size can be reduced. This reduction may outweigh the increase in travel times because the unit loads are no longer perfectly ordered. This trade-off also has been repeatedly studied, mostly by simulation based analysis; see for instance the original paper by Hausman et al. (1976) and Goetschalckx and Ratliff (1990). The warehouse is divided into a number of sections dedicated to a group of SKUs. Each section is called a (storage) zone. The group of SKUs assigned to a single zone is called a (product) class. Most often SKUs are assigned to a product class based on their frequency of access, but assignment based on the demand rate is also commonly used. The most common numbers of classes used in practice range from two to four. Inside a storage zone, the unit loads of the different SKUs that belong to this class are assumed to be stored randomly.

A fundamental problem is to determine the size of each zone in function of the products that are to be stored in this zone. If the replenishment patterns of the SKUs are correlated then sharing of the storage locations may generate only a limited reduction of the required warehouse size. Typically, it is assumed that the replenishment patterns of the SKUs are independent processes. In that case, sharing the storage locations can take advantage of the fact that during a particular time period some SKUs will have a high inventory will other SKUs will have a low inventory. If the replenishment patterns are independent, statistical analysis can be used to determine the probability that the storage zone size will be sufficiently large to hold the unit loads of the product class. When more unit loads of product class are to be stored than fit in the corresponding storage zone, it is assumed that the excess unit loads will be stored in the next storage zone.

Assume initially that no safety inventory of the SKU is stored. A constant demand rate and an instantaneous replenishment batch are also assumed. The cycle inventory for that SKU is then uniformly distributed over time between q down to one unit. The inventory pattern is shown in the next figure.

Note that the expected value and variance of a uniformly distributed random variable between the boundary values a and b is equal to

$$\bar{x} = \frac{b+a}{2} \tag{2.20}$$

$$\sigma^2 = \frac{(b-a)^2}{12}. \tag{2.21}$$

Assuming there is no safety inventory present, the mean and standard deviation of the inventory of product p is then (Fig. 2.5)

$$\bar{I}_p = \frac{q_p + 1}{2} \tag{2.22}$$

$$\sigma_p = \frac{(q_p - 1)}{\sqrt{12}} \tag{2.23}$$

Based on the central limit theorem, if the class K contains a reasonable large number of products, the total inventory in class K is normally distributed with mean and standard deviation equal to

$$\bar{I}_K = \sum_{p \in K} \bar{I}_p \tag{2.24}$$

$$\sigma_K = \sqrt{\sum_{p \in K} \sigma_p^2}. \tag{2.25}$$

The required zone size dedicated to a particular class can then be determined given the acceptable probability that the zone will be full when a unit load of the class arrives. Let α be the maximum acceptable probability that a zone will be full, then

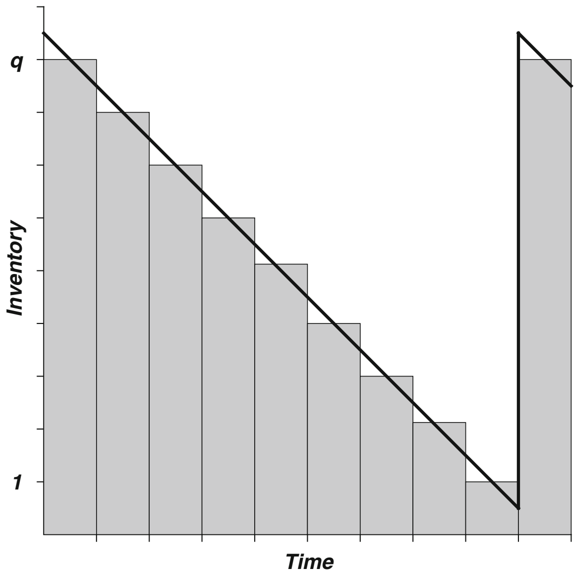

Fig. 2.5 Inventory patterns for a constant demand rate

$$P\left[\frac{x-\bar{x}}{\sigma} \geq z\right] \leq \alpha \tag{2.26}$$

$$Z_K = \bar{I}_K + z \cdot \sigma_K. \tag{2.27}$$

If the size of the storage zone is computed with the formulas above, then α percent of the accesses is executed with the expected travel of zone k and $(1 - \alpha)$ percent is executed with the expected travel of zone $k + 1$. It is assumed that no unit load will have to be located in a zone more than one removed from its home zone.

The performance of the storage system depends on the interactions of the number of classes, the size of the corresponding zones, the number of SKUs in the warehouse, the size of the replenishment batch sizes of the SKUs, and the balance or correlation of the input and output flows of different SKUs. Simulation appears the only analysis method than can accommodate all of these factors to predict the warehouse performance. Travel time models for specific hardware configurations such as an ASRS rack with I/O station in one corner have been developed.

2.3.6 Residence-Based Storage Model

The second major type of aggregation for individual unit loads is based on the duration of their residence in the storage system, which is also called their duration of stay. Again it is assumed that the warehousing processes are stationary.

The aggregation based on residence time is based on the observation that the first and last unit loads of a replenishment batch for a particular SKU have a very different behavior in the warehouse. Assume again initially that there is no safety stock for this SKU. The first unit is put away and almost immediately removed from the storage system and thus spends most of its residence time in the warehouse in motion. The last unit is put away and remains in storage until the whole replenishment batch has been removed through the demand process and then finally is removed. This unit spends most of its time in the warehouse immobilized in storage. To exploit this difference, the storage policy for the first unit should emphasize movement costs while the storage policy for the last unit should emphasize storage efficiency. Two external factors impact the performance of duration-of-stay-based policies. First, the larger the replenishment batch size, the more different the first and last unit load of a batch are and the more impact duration-of-stay policies will have. Second, the more safety inventory for an SKU is held, the more similar the first and last unit load of batch will behave because of the FIFO requirement and the least impact duration of stay storage policies will have.

Two cases can be further distinguished. In the first case each storage zone is exclusively occupied by unit loads that the same duration-of-stay. In the second case storage zones are assigned to a group of duration of stays.

2.3.6.1 Duration-of-Stay Storage Policy for Perfectly Balanced Warehouses

$n_{\text{DOS}}(t)$	Number of unit loads arriving during time period t that have a duration-of-stay equal to DOS periods.
z_{DOS}	Size of the zone, expressed as a number of storage locations, reserved for the storage of unit loads with duration of stay equal to DOS periods

Previously it was shown that a balanced warehouse will require the smallest possible warehouse size to hold all the unit loads during the planning horizon.

$$n_{\text{in}}(t) = n_{\text{out}}(t) \qquad \forall\, t \tag{2.28}$$

For a perfectly balanced warehouse, the dedicated duration-of-stay storage policy will both required the smallest possible warehouse size and simultaneously minimize the transportation time or cost. However, the satisfaction of the requirement that the warehouse is perfectly balanced is exceedingly rare in real warehouse operations, so the dedicated duration-of-stay storage policy and its corresponding performance is more to be considered as a theoretical limit rather than an actually attainable performance. Finally, it should be observed that a perfectly balanced warehouse implies that the warehouse is also balanced.

$$n_{\text{DOS}}(t) = n_{\text{DOS}}(t + \text{DOS}) \qquad \forall\, t, \forall\, \text{DOS} \tag{2.29}$$

$$z_{\mathrm{DOS}} = \sum_{i=1}^{\mathrm{DOS}} n_{\mathrm{DOS}}(i) \tag{2.30}$$

$$T_{\mathrm{DOS}} = 4\frac{1}{\mathrm{DOS}} z_{\mathrm{DOS}} t_{\mathrm{DOS}} = \frac{4}{\mathrm{DOS}} z_{\mathrm{DOS}} \left(\frac{\sum_{j \in \mathbf{Z}_{\mathrm{DOS}}} e_j}{z_{\mathrm{DOS}}} \right) = \frac{4}{\mathrm{DOS}} \left(\sum_{j \in \mathbf{Z}_{\mathrm{DOS}}} e_j \right) \tag{2.31}$$

The total required warehouse size and total transportation cost per period is compute by the following expression.

$$N_{\mathrm{DOS}} = \sum_{\mathrm{DOS}} z_{\mathrm{DOS}} \tag{2.32}$$

$$T = \sum_{\mathrm{DOS}} T_{\mathrm{DOS}} \tag{2.33}$$

The optimal storage policy among duration-of-stay based storage policies is to sort the storage locations by increasing expected travel time. The policy then assigns the storage zones by increasing duration of stay to storage locations by increasing travel time. Since this is again equivalent to the inner product of two vectors, this policy minimizes the travel cost. Since the warehouse is balanced and the policy maintains a constant level of inventory, the policy also minimizes the required warehouse size.

2.3.6.2 Duration-of-Stay Storage Policy

Virtually no real world warehouse will satisfy the perfectly balanced condition. A more practical storage policy is to assign all unit loads within a range of duration of stay to a zone in the warehouse. This is similar to class turnover-based storage, but the criterion to divide the unit loads is their residence time rather than the frequency of access of their product.

Again the performance of this storage policy depends on the interactions of the number of classes, the size of the corresponding zones, the number of SKUs in the warehouse, the skewness of the Pareto curve of the duration of stays of the unit loads, the size of the replenishment batch sizes of the SKUs, and the balance or correlation of the input and output flows of different SKUs. Simulation appears the only analysis method than can accommodate all of these factors to predict the warehouse performance. There is some evidence presented in Goetschalckx and Ratliff (1990) that policies with few zones and based on the product frequency of access perform slightly better than policies with few zones and based on duration of stay of unit loads.

2.3.7 Random Storage and Closest Open Location Storage Policies

The simplest storage policy is the random storage policy since it uses no information about the unit load. It ignores both the product characteristics of the SKU to which the unit load belongs or the residence time characteristics of the unit load. Since no internal structure or partitioning of the storage locations is imposed, the random storage policy requires the smallest possible warehouse size of all storage policies. The COL storage policy is equivalent to the pure random storage policy if all locations in the rack are used.

$$T = \sum_i 4f_i(s_i + q_i)\frac{1}{N}\left(\sum_{j=1}^{N} e_j\right) = \sum_i 4r_i\bar{t}. \tag{2.34}$$

Random storage and COL require the smallest storage size. A smaller storage size in turn reduces the travel cost to a particular location. However, under these policies the location of unit loads belonging to different SKUs is constantly changing in the warehouse. The reductions in travel associated with COL may be completely negated if the put-away or retrieval operations have to search for either an open storage location or a unit load belonging to a particular SKU. Maintaining an accurate inventory map that is accessible in real time to the material handling operators is an absolute requirement for the efficient operation of the storage system under these policies. This almost always requires a fully automated system such as an ASRS or a rack system with automated stacker cranes.

2.4 Conclusion

Storage systems are ubiquitous elements of supply chains. Storage models and storage policies are essential components during the design of a warehouse in order to ensure it has the required capacity and during the operation of the warehouse to maximize its operational performance. The characteristics of storage models or storage policies are the clearest exposed in storage systems for unit loads, since many of the practical complexities of other storage systems are eliminated. This clarification is even stronger for random access unit load storage systems.

For these systems an individual unit load storage model has been developed. But the problem instances of this model can only be solved for small storage systems. For practical sizes of storage systems, aggregation has to be applied so that the storage problem can be solved. This aggregation requires that the storage system has stationary process. However, the storage model is also useful to compute the relative performance of other policies.

The two most prominent aggregation methods are either based on the product or residence time characteristics of the unit load. The optimal policy for full product dedicated storage policies is shown to be based on the product frequency of access also known as COI. However, the optimality of this policy requires the factoring condition and maximum warehouse size. If storage locations can be shared based on product characteristics, the required warehouse size is reduced. For random storage, closest open location, or product class-based storage (with few classes), this space reduction may be sufficient to overcome mixing the products assigned to a class.

For aggregation based on duration of stay, in the extreme case of a perfectly balanced warehouse, the duration of stay dedicated storage policy is optimal with respect to both warehouse size and cost of the material handling operations. However, practical warehouse operations are far from perfectly balanced. When unit loads with similar durations of stay share a zone dedicated to them, the performance of the warehouse system is similar to the class based product turnover storage policies.

There exist many directions for future research, but two appear to be particularly promising. The first direction extends the unit load storage policies to cases where the locations in the warehouse cannot be randomly accessed. Container yards and container ships for intermodal containers are very important examples of such system. The second direction is to extend analysis to conditions that incorporate more stochastic conditions than the current research which is mostly based on mean value analysis. Current methodology for such stochastic systems is almost exclusively based on simulation.

References

Cormier G, Gunn E (1992) A review of warehouse models. Eur J Oper Res 58:1–13

Goetschalckx M, Ratliff HD (1990) Shared versus dedicated storage policies. Manage Sci 36(9):1120–1132

Gu J, Goetschalckx M, McGinnis L (2005) A comprehensive review of warehouse operation. Eur J Oper Res 177(1):1–21

Gu J, Goetschalckx M, McGinnis LF (2010) Research on warehouse design and performance evaluation: a comprehensive review. Eur J Oper Res 203(3):539–549

Hausman WH, Schwarz LB, Graves SC (1976) Optimal storage assignment in automatic warehousing systems. Manag Sci 22(6):629–638

Heskett JL (1963) Cube-per-order index: a key to warehouse stock location. Trans Distrib Manag 3:27–31 April

Heskett JL (1964) Putting the cube-per-order index to work in warehouse layout. Trans Distrib Manag 4:23–30

Malette AJ, Francis RL (1972) Generalized assignment approach to optimal facility layout. AIIE Trans 4(2):144–147

Malmborg CJ, Deutsch SJ (1988) A stock location model for dual address order picking systems. IIE Trans 20(1):44–52

Rouwenhorst B, Reuter B, Stockrahm V, Van Houtum GJ, Mantel RJ, Zijm WHM (2000) Warehouse design and control: framework and literature review. Eur J Oper Res 122:515–533

White JA, Francis RL (1971) Normative models for some warehouse sizing problems. AIIE Trans 9(3):185–190

http://www.springer.com/978-1-4471-2273-9

Warehousing in the Global Supply Chain
Advanced Models, Tools and Applications for Storage Systems
Manzini, R. (Ed.)
2012, XXII, 486 p., Hardcover
ISBN: 978-1-4471-2273-9