International Journal of Production Research,
Vol. 43, No. 8, 15 April 2005, 1515–1536

Taylor & Francis
Taylor & Francis Group

A comprehensive review of lot streaming

JEN HUEI CHANG†* and HUAN NENG CHIU‡

†Department of Industrial Management, Tung Nan Institute of Technology, 152 PeiShen Road, Section 3, Shenkeng, Taipei, Taiwan 222, Republic of China
‡Department of Industrial Management, National Taiwan University of Science and Technology, 43 Keelung Road, Section 4, Taipei, Taiwan 106, Republic of China

(*Received August 2004*)

Lot streaming combined lot splitting with operations overlapping is one of the effective techniques used to implement the time-based strategy in today's era of global competition. Therefore, this technique has been studied extensively over the past few decades. In this paper, we first propose a uniquely categorized structure to characterize the existing lot streaming problems in terms of three main dimensions, seven subdimensions, and 17 levels. Then, a notation set is defined to systematically express each existing lot streaming problem by seven ordered elements corresponding to the seven subdimensions. Based on the classification structure and the defined notation set, a comprehensive survey is presented to make up for the lack of a literature review on this subject. The objective is to help the reader gain a clear understanding of the evolution of previous research on lot streaming. This paper concludes with some constructive suggestions for future research directions.

Keywords: Lot streaming; Lot splitting; Operations overlapping

1. Introduction

In the past few years, intensive global competition and dramatic changes have led many managers to seek better strategies that can improve company performance. One of these is the time-based strategy. In fact, many world-class manufacturing companies (e.g., Dell and Toyota) have adopted this strategy to quickly produce and deliver goods/services. It can be referred to as time-based competition (TBC) (Blackburn 1991, Bockerstette and Shell 1993). One important tactic when implementing TBC in manufacturing is to compress the manufacturing lead time (MLT). To achieve the goal of compressing MLT, some well-known techniques (e.g., just-in-time (JIT) and optimized production technology (OPT)) have been widely adopted. While the Japanese JIT system has received considerable attention from both practitioners and researchers, OPT (initiated by Goldratt (1980) in Israel) has received comparatively little attention in the literature (Fawcett and Pearson 1991). More recently, it has been popularly referred to as the theory of constraints (TOC) (Goldratt 1990).

*Corresponding author. E-mail: jenhuei.chang@msa.hinet.net

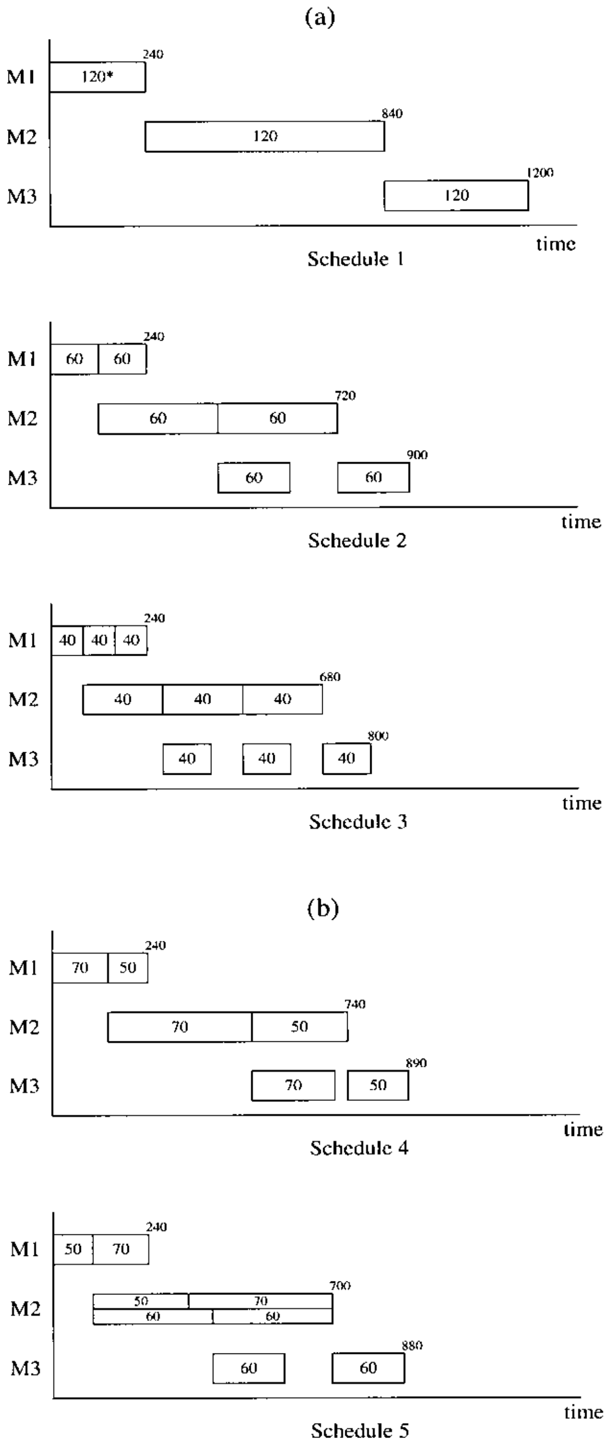
International Journal of Production Research
ISSN 0020-7543 print/ISSN 1366-588X online © 2005 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/00207540412331325396

The term *lot streaming* was first introduced by Reiter (1966). The concept was then embodied by Lundrigan (1986). He summarized nine rules of OPT, among which the seventh rule states that a transfer batch (i.e., a sublot) may not, and often should not, be equal to a processing batch (i.e., a lot). Clearly, lot streaming is a technique in which a processing lot is split into several sublots and overlapping operations in different manufacturing workstations (i.e., stages) are performed. In this way, production can be accelerated. Until the late 1980s, lot streaming was researched independently of OPT. Since then, this technique has been extensively studied in academic as well as industrial fields and has been shown to be an effective technique for compressing MLT (e.g., Ronen and Starr 1990, Smunt *et al.* 1996, Silver *et al.* 1998, and Kher *et al.* 2000).

A simple example is used here to illustrate the benefits of lot streaming. Suppose that there are three stages in a flow shop. Each stage has only one machine. Machines 1, 2, and 3 take 2, 5, and 3 min per unit, respectively. Now, a lot of size 120 units will be processed in the shop. For simplicity, the setup time and the sublot transportation time are ignored. If we adopt the following three splitting policies: no splitting, two equal sublots splitting, and three equal sublots splitting, denoted as Schedules 1, 2, and 3, respectively, then the Gantt charts obtained by using the three splitting policies are those presented in figure 1(a). The makespan of Schedule 1 is 1200 min, while that of Schedule 2 is 900 min. The reduction in MLT is 25%. Furthermore, the makespan of Schedule 3 is 800 min, and the reduction in MLT is 33.33% compared with Schedule 1 and 11.11% compared with Schedule 2. Apparently, the makespan with lot streaming (i.e., Schedules 2 and 3) is significantly shorter than that without lot streaming (i.e., Schedule 1).

Although lot streaming has received much more attention during the last ten years, few papers have reviewed this topic. In fact, this topic has only been preliminarily surveyed by Potts and Van Wassenhove (1992) and Trietsch and Baker (1993). Potts and Van Wassenhove (1992) discussed the problem of integrating scheduling with batching and lot-sizing in terms of minimizing the time-related instead of the cost-related objective functions. However, only two papers pertaining to lot streaming were listed in that review paper. We also find that the number of studies reviewed by Trietsch and Baker (1993) was very small (i.e., only five papers, two working papers, and two technical reports). They provided a narrower range of lot streaming problems and solution methods so that a common time-related performance measurement (i.e., the makespan) could be minimized. To make up for the lack of review papers on lot streaming, this paper comprehensively reviews the existing lot streaming models. In this paper, we divide the previous papers into four categories according to the number of products (i.e., single product or multi-product) and the performance measurements (i.e., time-related or cost-related).

The remainder of this paper is organized as follows. In section 2, we use a structure to explain and categorize the existing lot streaming problems and define a notation set to systematically describe each existing problem characterized by seven ordered elements corresponding to seven subdimensions. The single-product and multi-product time-related problems are discussed in sections 3 and 4. Section 5 reviews the single-product cost-related problems. The problems that cannot be suitably placed in the above-mentioned three categories are discussed in section 6. In section 7, we conclude this paper with some constructive suggestions for future research.



Note: an asterisk (*) represents a sublot size

Figure 1. Gantt charts of Schedules 1, 2, 3, 4, and 5.

2. Structure and classification of lot streaming problems

2.1 *The problem structure*

To appropriately describe each existing lot streaming problem, three main dimensions and seven subdimensions are used here to present a structure for classifying the existing lot streaming problems, as shown in table 1. In fact, each subdimension in table 1 consists of one or more levels. To help casual readers easily understand this structure, clear definitions must be provided for all levels. The first three production types (i.e., flow shop (F_m), job shop (J_m), and open shop (O_m) with m stages) assumed in this paper are identical to those assumed in traditional scheduling. An arborescent shop (T_m) is an m-stage production system, in which each stage has at least one immediate successor except for the last stage (i.e., the finished goods stage), and has only one immediate predecessor except for the first stage (i.e., the raw materials or purchased part stage). The multi-product (L_n) indicates that n products are manufactured in a production system. Here, we use the term 'product' to refer to a 'job', 'lot', or 'item', a term which is also used in the literature.

Table 1. Detailed classification of lot streaming problems.

Main dimension	Subdimension	Level[a]
System configuration (α)	Production type (α_1)	Flow shop (F_m) Job shop (J_m) Open shop (O_m) Arborescent shop (T_m)
	Number of products (α_2)	Single product (L_1) Multi-product (L_n)
Sublot-related feature (β)	Sublot type (β_1)	Equal sublots (E) Consistent sublots (C) Variable sublots (V)
	Divisibility of sublot size (β_2)	Continuous version (R) Discrete version (A)
	Operation continuity (β_3)	No idling case (I_{no}) Idling case (I)
	Activities involved (β_4)	Setup ($S_{(-,-)}$, $S_{(A,-)}$, $S_{(-,D)}$, or $S_{(A,D)}$ for the time model) (S for the cost model) Production (P for the time model) ($P_{(-,IP,-)}$, $P_{(-,IP,FG)}$, or $P_{(IM,IP,FG)}$ for the cost model) Transportation ($M^{\rightarrow}_{\leftarrow(w)}$)
Performance measurement (γ)	Performance criterion (γ_1)	For the time model: Makespan (C_{max}) Mean flow time ($\overline{F}$) Total flow time ($\sum F$) Mean tardiness ($\overline{T}$) Number of tardy jobs (n_T) Total deviation from due date ($\sum \lvert C-d \rvert$) For the cost model: Total cost (TC) Total cost ($TC(C_{max}, k)$)

[a]Notation in parentheses represents a specific level of a subdimension.

The three types of sublots are equal sublots (E), consistent sublots (C), and variable sublots (V). Variable sublots are sublots with sizes between stages i and $i+1$ that are not equal to those between stages $i+1$ and $i+2$, given the same sublot count. That is, in an m-stage production system with k sublots, $q_{ij} \neq q_{i(j+1)}$, $i=1,\ldots,m$, $j=1,\ldots,k$–1 and $q_{ij} \neq q_{(i+1)j}$, $i=1,\ldots,m$–1, $j=1,\ldots,k$, where q_{ij} is the size of sublot j at stage i. The sizes of consistent sublots between any two adjacent stages are identical, given the same sublot count. Symbolically, $q_{ij}=q_j$, $i=1,\ldots,m$, $j=1,\ldots,k$. The sizes of equal sublots between any two adjacent stages are the same for different sublot counts (i.e., $q_j=q$, $j=1,\ldots,k$). Obviously, equal sublots is a special case of consistent sublots, which is also a special case of variable sublots. Hence, the type of sublot has a significant impact on MLT. To further demonstrate this impact, an example, which is identical to that given in section 1, is illustrated. Besides the three above-mentioned splitting policies, we also use two other splitting policies: two consistent sublots splitting ($q_1=70$ and $q_2=50$) and two variable sublots splitting ($q_{11}=50$, $q_{12}=70$, $q_{21}=60$, and $q_{22}=60$), denoted by Schedule 4 and Schedule 5 in figure 1(b), respectively. The makespan (880 min) of Schedule 5 is shorter than those of Schedule 2 (900 min) shown in figure 1(a) and Schedule 4 (890 min) shown in figure 1(b).

Two types of sublot size divisibility are the continuous version (R) and discrete version (A). A sublot size can be a real number or an integer. The no idling case (I_{no}) is the case where there is no production interruption time (i.e., the idle time) between any two adjacent sublots at the same stage, whereas the idling case (I) allows idle time. As is widely known, the makespan based on the idling case is shorter than that based on the no idling case under the same sublot type.

Three kinds of activities are needed to complete production. One is a production activity that involves a series of operations. We use the notation P to represent the production activity for the time model. For the cost model, this activity can be represented by one of three cases (i.e., $P_{(-,\ \mathrm{IP},-)}$, $P_{(-,\ \mathrm{IP},\ \mathrm{FG})}$, and $P_{(\mathrm{IM},\ \mathrm{IP},\ \mathrm{FG})}$, where '–' in the subscript parentheses indicates that the corresponding inventory type is excluded and IM, IP, and FG denote raw materials, work-in-process (WIP), and finished goods, respectively). The other two cases are complementary activities, including setups and transportation. Setups can be divided into two types: attached setup and detached setup (the latter setup type has also been called anticipatory setup or before-job-arrival setup in the literature). Attached setup means that a machine at a given stage cannot be set up until at least one unit is received from its immediately upstream stage. Four setup cases for the time model are $S_{(-,-)}$, $S_{(\mathrm{A},-)}$, $S_{(-,\mathrm{D})}$, and $S_{(\mathrm{A},\mathrm{D})}$, where '–' in the subscript parentheses indicates that the corresponding setup type is not considered and A and D represent attached setup and detached setup, respectively. On the other hand, since most of the existing cost models only consider the number of setups and ignore the setup type, the notation S is a succinct representation. Transportation activity includes the transportation of a sublot and the return of an empty transporter. The transportation activity is expressed by $M^{\rightarrow}_{\leftarrow(w)}$, where the superscript $\rightarrow$ represents the sublot transportation and the subscripts $\leftarrow$ and w represent the empty transporter return and the number of capacitated transporters, respectively. Note that the extent to which the transportation activity affects the makespan depends on the number of capacitated transporters.

Because most of the existing models only consider a single performance measurement, the measurement can be chosen from among the makespan ($C_{\max}$), the mean flow time ($\overline{F}$), the total flow time ($\sum F$), the mean tardiness ($\overline{T}$), the number of tardy jobs (n_T), and the total deviation from the due date ($\sum |C - d|$) for the time model. The definitions of the six performance measurements mentioned above are identical to those that have appeared in the traditional scheduling literature. As for the cost model, if the total cost consists of the setup cost, the inventory holding cost, and the sublot transportation cost, then the notation TC is adopted; otherwise, the notation $TC(C_{\max}, k)$ is used when the total cost includes the makespan cost and the total sublot transportation cost.

2.2 *Notation set definition*

The notation given by Potts and Van Wassenhove (1992), that is, the three-field problem descriptor, was used in their paper to review the scheduling problem, into which batching and lot-sizing were incorporated. In this paper, each existing lot streaming problem can be expressed by a notation set that includes seven ordered elements with respect to seven subdimensions, as shown in table 1. This notation set is defined as $\{\alpha_1, \alpha_2, \beta_1, \beta_2, \beta_3, \beta_4, \gamma_1\}$. It is noted here that if a '–' appears in the position of any element in a notation set, then it indicates that that specific level is not considered in the lot streaming problem. Additionally, if one element (e.g., β_1) involves more than one level (e.g., equal sublots (E), consistent sublots (C), and variable sublots (V)), then $\beta_1 = E/C/V$.

According to the above-defined notation set, two examples are given here to describe two lot streaming problems studied in the literature. One problem explored by Szendrovits (1975) can be expressed by $\{F_m, L_1, E, R, I_{no}, S/P_{(-,IP,FG)}, TC\}$, which involves an m-stage flow shop, single-product, equal sublots, continuous version, no-idling case, and includes setup and production for the cost model in terms of minimizing the total cost. Another problem studied by Vickson (1995) is $\{F_2, L_n, C, A, I, S_{(A,D)}/P/M^{\rightarrow}{}_{(-)}, C_{\max}\}$, which implies a two-stage flow shop, multi-product, consistent sublots, discrete version, idling case, and involves setup, production, and transportation for the time model in which the makespan is minimized. In the following three sections, the defined notation set will be used to describe in detail the existing lot streaming problems.

3. Single-product time-related problems

Table 2 summarizes the existing single-product time-related problems. The main goal in solving these problems is to find the best allocation of sublots (i.e., to determine the size of each sublot) that minimizes one time-related performance measurement. For the flow shop, Potts and Baker (1989) provided some important properties related to the optimal sublot allocation. They showed that a lot streaming problem and its inverse problem (i.e., $t_i = t_{m-i+1}$, $i = 1, \ldots, m$, where t_i is the processing time per unit at stage i) were equivalent, and that there existed an optimal schedule in which $q_{1j} = q_{2j}$ and $q_{m-1,j} = q_{mj}$, $j = 1, \ldots, k$. Hence, in the two- and three-machine cases, the shortest makespan was obtained by using the consistent sublots splitting policy. Moreover, they presented the optimal sublot allocation with a geometric series of $q_{j+1} = rq_j$ and $q_j = r^{j-1}/\sum_{l=1}^{k} r^{l-1}$, $j = 1, \ldots, k$, where $r = t_2/t_1$ (note that the processing lot size was set to be one in their paper), and also gave several worst-case performance bounds in terms of the ratio of M^{ES} to M^{CS}, where M^{ES} and M^{CS}

Table 2. A summary of single-product time-related problems.

Problem	Author(s)
$\{F_2, L_1, C, R, I_{no}, P, C_{max}\}$	Potts and Baker 1989, Trietsch and Baker 1993
$\{F_2, L_1, C, A, I_{no}, P, C_{max}\}$	Trietsch and Baker 1993
$\{F_2, L_1, C, R, I_{no}, P/M^{\rightleftarrows}_{(w)}, C_{max}\}$	Trietsch and Baker 1993
$\{F_2, L_1, C, A, I_{no}, P/M^{\rightleftarrows}_{(w)}, C_{max}\}$	Trietsch and Baker 1993
$\{F_2, L_1, E, A, I, P, C_{max}\}$	Sen *et al.* 1998
$\{F_2, L_1, C, A, I, P, C_{max}\}$	Sen *et al.* 1998, Chen and Steiner 1999
$\{F_2, L_1, C, R, -, S_{(-,D)}/P, C_{max}\}$	Sriskandarajah and Wagneur 1999
$\{F_2, L_1, C, A, -, S_{(-,D)}/P, C_{max}\}$	Sriskandarajah and Wagneur 1999
$\{F_2, L_1, C, R, I, S_{(A,-)}/P, \overline{F}\}$	Bukchin *et al.* 2002
$\{F_3, L_1, C, R, I, P, C_{max}\}$	Trietsch and Baker 1993, Glass *et al.* 1994
$\{F_3, L_1, V, R, I, P, C_{max}\}$	Trietsch and Baker 1993
$\{F_3, L_1, V, A, I, P, C_{max}\}$	Trietsch and Baker 1993
$\{F_3, L_1, E/C/V, R, I/I_{no}, P, C_{max}\}$	Baker and Jia 1993
$\{F_3, L_1, C, A, I, S_{(-,D)}/P, C_{max}\}$	Chen and Steiner 1996
$\{F_3, L_1, C, A, I, S_{(A,-)}/P, C_{max}\}$	Chen and Steiner 1998
$\{F_m, L_1, C, A, I, P, C_{max}\}$	Glass *et al.* 1994, Chen and Steiner 1997, 2003, Glass and Potts 1998
$\{F_m, L_1, E, R, I_{no}, S_{(A,D)}/P/M^{\rightleftarrows}_{(-)}, C_{max}\}$	Truscott 1985
$\{F_m, L_1, C, R, I_{no}, S_{(A,D)}/P/M^{\rightleftarrows}_{(1)}, C_{max}\}$	Truscott 1986
$\{F_m, L_1, C, R, I, P, C_{max}\}$	Baker and Pyke 1990, Chiu and Chang 1996a, b, c, Williams *et al.* 1997
$\{F_m, L_1, C, R, I, S_{(A,-)}/P/M^{\rightarrow}_{(-)}, C_{max}\}$	Kropp and Smunt 1990
$\{F_m, L_1, C, R, I, S_{(A,-)}/P/M^{\rightarrow}_{(-)}, \overline{F}\}$	Kropp and Smunt 1990
$\{F_m, L_1, E, R, I_{no}, P/M^{\rightarrow}_{(-)}, C_{max}\}$	Steiner and Truscott 1993
$\{F_m, L_1, E, R, I_{no}, P/M^{\rightarrow}_{(-)}, \sum F\}$	Steiner and Truscott 1993
$\{F_m, L_1, E, R, I, P/M^{\rightarrow}_{(-)}, C_{max}\}$	Kalir and Sarin 2001a
$\{F_m, L_1, E, R, I, S_{(A,-)}/P, C_{max}\}$	Kalir and Sarin 2001a
$\{F_m, L_1, V, A, I_{no}, S_{(A,D)}/P/M^{\rightleftarrows}_{(w)}, C_{max}\}$	Chiu and Lee 2002, Chiu *et al.* 2004
$\{F_m, L_1, C, R, -, S_{(-,D)}/P, C_{max}\}$	Kumar *et al.* 2000
$\{F_m, L_1, C, A, -, S_{(-,D)}/P, C_{max}\}$	Kumar *et al.* 2000
$\{O_m, L_1, C, R, I, P, C_{max}\}$	Glass *et al.* 1994, Sen and Benli 1999

Note that '–' indicates that this specific level was not considered in the problem.

are the makespans based on equal sublots and consistent sublots, respectively. Trietsch and Baker (1993) proposed a problem classification scheme and a dominance relationship for a multistage flow shop. In the two- and three-machine cases, they provided some important structural insights and presented several solution procedures. Two linear programming models (i.e., one for minimizing the makespan and another for minimizing the total idle time) were developed. Sen *et al.* (1998) proposed a sublot completion time model to determine the optimal sublot allocation. The sublot completion time in their paper was defined as the time at which a sublot was completed at the second machine. Chen and Steiner (1999) used a network representation to analyse the structure of the optimal sublot allocation.

They proposed an efficient solution method based on the structural properties. Bukchin *et al.* (2002) presented an analytical procedure based on a single-machine bottleneck property to solve a two-machine problem, in which sublot-attached setups (i.e., an attached setup occurs before each sublot is started) were involved. Baker and Jia (1993) conducted several computational experiments to analyse the effects of the sublot type, the number of sublots, and the operation continuity on the makespan based on whether or not the second machine was dominant. Glass *et al.* (1994) used a network representation to analyse the structure of optimal solutions for an m-machine case. In the three-machine case, they found that the optimal sublot allocation depended on the relationship between $(t_2)^2$ and t_1t_3, which was similar to that found by Trietsch and Baker (1993). They developed an analytical solution procedure based on that relationship. Chen and Steiner (1996) extended the results of Glass *et al.* (1994) to further consider the detached setup case. Chen and Steiner (1998) investigated the attached setup case.

Truscott (1985) presented a simple model to describe the impact of two complementary activities (i.e., setups and transportation) on the makespan. Truscott (1986) extended the problem of Truscott (1985) to consider constraints on the transportation resources by using a single capacitated transporter. He proposed a binary mixed linear programming (BMIP) model and an algorithm based on the branch and bound method to solve the segmental adjacent-stage production system problem. This problem was decomposed into $m-1$ subproblems. Each subproblem was solved independently. Baker and Pyke (1990) proposed a computationally efficient algorithm based on a network representation for finding the optimal allocation of two sublots. They also developed five heuristic methods for solving the case of more than two sublots. Kropp and Smunt (1990) developed a linear and a quadratic programming model. Based on the results they obtained by conducting a full factorial (three-factor) experiment, two heuristic methods were proposed. Steiner and Truscott (1993) presented two expressions and found the optimality conditions for minimizing the makespan and the total flow time, respectively. Chiu and Chang (1996a, b, c) were the first researchers to formally show the diminishing marginal effect of lot streaming. Namely, under the same sublot type, although increasing the number of sublots will reduce the makespan, the marginal reduction in the makespan will decrease with the increase of the number of sublots. They gave five t_i combinations, $i = 1, 2, \ldots, m$, based on the constant, increasing and/or decreasing unit processing times of m machines in a flow shop. Then, through a computational experiment, they applied a linear programming (LP) model developed by Kropp and Smunt (1990) to discuss the optimal sublot allocation for each combination. Chen and Steiner (1997) presented two quickly obtainable heuristic methods with very good quality, as demonstrated by analysing the error bounds of the worst case. Williams *et al.* (1997) reviewed the $m \times 2$ (i.e., m-machine, two-sublot) problem discussed by Baker and Pyke (1990) and developed an optimal algorithm for the $m \times 3$ problem on the basis of the properties of the $m \times 2$ problem. Moreover, a generalization concerning the $m \times k$ problem was made based on the properties of the $m \times 2$ and $m \times 3$ problems. Glass and Potts (1998) extended the work of Glass *et al.* (1994) to an m-stage case and developed a two-phase method to find the optimal allocation of sublots. In the first phase, a powerful relaxation algorithm that used the machine dominance property was derived to reduce the number of machines so that only dominant machines needed be considered. In the second phase, an optimal sublot allocation in terms of the critical path structure was characterized by applying the network representation. Kalir and Sarin (2001a) used an optimization method to determine the optimal number of sublots when the sublot transportation time was included. In addition, they developed an algorithm, in which the sublot-attached setup time was considered, to determine the optimal number of sublots. Chen and Steiner (2003) studied the no-wait problem, in which each sublot must be continuously processed from the beginning at the first machine to its completion at the last machine, without interruption at machines and without any waiting time between consecutive machines. They proposed an optimal algorithm to solve the $m \times 2$ problem and presented two approximation methods for the $m \times k$ problem. Chiu and Lee (2002) and Chiu *et al.* (2004) extended the single-transporter problem studied by Truscott (1986) to the multi-transporter problem. They also extended the two-stage method developed by Trietsch and Baker (1993) to obtain a solution procedure used in an m-stage flow shop.

For the open shop, the purpose is to simultaneously deal with routing and sublot-sizing. Sen and Benli (1999) stated that one of the important characteristics of the open shop was the one-to-one correspondence between an m-machine, n-product problem and an n-machine, m-product problem, where machines were considered as products and products as machines. For example, a four-machine, three-product problem is equivalent to a three-machine, four-product problem. Glass *et al.* (1994) investigated the multi-route model, in which all sublots allowed different production processes (i.e., routes). Sen and Benli (1999) divided the open shop problem into two subproblems: two sublots and more than two sublots. They defined the single-route model as that in which all sublots were forced to follow a common route. In the case of the single-route model with two sublots, once the route was determined, the two sublot sizes could be obtained by using Baker and Pyke's (1990) method. In the case with more than two sublots, the optimal sublots were of equal size. For the multi-route model, the results were identical to those found by Glass *et al.* (1994).

Here, we summarize the results of previous studies on this area. Researches have mainly focused on discussing the sublot allocation associated with deterministic processing, setup, and sublot transportation times. Basically, three main approaches (i.e., the network analysis, heuristic, and mathematical programming approaches) have been used. However, in the future, meta-heuristic procedures (i.e., artificial intelligence algorithms) will be applied in this field. The stochastic situation will be considered, and even fuzzy set theory will be applied to mathematical programming. In addition, the limited buffer size between any two adjacent stages and the variable sublot transportation time (i.e., the time that depends on the sublot size) will be considered.

4. Multi-product time-related problems

In this area, the primary purpose is to simultaneously obtain the best sublot allocation and the best production sequence. These problems, also collectively called the lot streaming sequencing problem (LSSP) by Kalir and Sarin (2001b), can be further divided into intermingled and non-intermingled cases. The former case is similar to the preemption case used in traditional scheduling. Additionally, in essence, the intermingled LSSP can also be regarded as the JIT mixed-type problem (e.g., Monden 1983, Miltenburg 1989, Philipoom *et al.* 1996, McMullen 2002, Ventura and Radhakrishnan 2002). So far, most of the related papers have been based on the

Table 3. A summary of the multi-product time-related problems solved using analytical and/or heuristic methods.

Problem	Author(s)
$\{F_2, L_n, C, R, I, P, C_{\max}\}$	Potts and Baker 1989
$\{F_2, L_n, E, R, I, P, C_{\max}\}$	Vickson and Alfredsson 1992
$\{F_2, L_n, E, R, I, S_{(-,D)}/P, C_{\max}\}$	Cetinkaya and Kayaligil 1992
$\{F_2, L_n, C, R, I, S_{(-,D)}/P/M^{\rightarrow}{}_{(-)}, C_{\max}\}$	Vickson 1995
$\{F_2, L_n, C, A, I, S_{(-,D)}/P/M^{\rightarrow}{}_{(-)}, C_{\max}\}$	Vickson 1995
$\{F_2, L_n, C, R, I, S_{(A,-)}/P/M^{\rightarrow}{}_{(-)}, C_{\max}\}$	Cetinkaya 1994, Vickson 1995
$\{F_2, L_n, C, A, I, S_{(A,-)}/P/M^{\rightarrow}{}_{(-)}, C_{\max}\}$	Cetinkaya 1994, Vickson 1995
$\{F_2, L_n, E, R, I, S_{(A,D)}/P, C_{\max}\}$	Baker 1995
$\{F_2, L_n, C, R, -, S_{(-,D)}/P, C_{\max}\}$	Sriskandarajah and Wagneur 1999
$\{F_2, L_n, C, A, -, S_{(-,D)}/P, C_{\max}\}$	Sriskandarajah and Wagneur 1999
$\{F_3, L_n, E, R, I, P, C_{\max}\}$	Vickson and Alfredsson 1992
$\{F_m, L_n, C, R, -, S_{(-,D)}/P, C_{\max}\}$	Kumar *et al.* 2000
$\{F_m, L_n, C, A, -, S_{(-,D)}/P, C_{\max}\}$	Kumar *et al.* 2000
$\{F_m, L_n, E, R, I, P, C_{\max}\}$	Kalir and Sarin 2001b
$\{F_m, L_n, C, R, I, P, \sum\lvert C-d\rvert\}$	Yoon and Ventura 2002a, b
$\{J_m, L_n, C, A, I, P, C_{\max}\}$	Stephane and Lasserre 1993, 1997
$\{J_m, L_n, E, R, I, S_{(-,D)}/P, \sum\lvert C-d\rvert\}$	Jin *et al.* 1999
$\{J_m, L_n, E/C, R, I, S_{(-,D)}/P, C_{\max}\}$	Jeong *et al.* 1999
$\{O_2, L_n, C, R, I, P, C_{\max}\}$	Sen and Benli 1999

non-intermingled case except for those by Stephane and Lasserre (1997) and Sen and Benli (1999). Table 3 summarizes the multi-product time-related problems, to which analytical and/or heuristic methods have been applied.

For the flow shop, Potts and Baker (1989) used a simple two-machine numerical example to show that, in the intermingled case, an optimal solution generally cannot be found when the sequencing approach and the splitting approach are used independently. They suggested that the two approaches should be used simultaneously. The problem studied by Vickson and Alfredsson (1992) could be regarded as a simple sequence problem since unit-sized sublots (i.e., each sublot size is only one) were given. Consequently, the optimal sequences of the two-machine and special three-machine problems could be obtained by using Johnson's rule (Johnson 1954). Cetinkaya and Kayaligil (1992) also adopted the unit-sized sublot splitting policy and performed an attached setup at the first machine and a detached setup at the second machine. Cetinkaya (1994) and Vickson (1995) showed that optimal sequencing of the products and splitting of each product into optimal sublots under a specified number of sublots could be performed separately. The optimal sequence was obtained using Johnson's rule. As for the continuous version, the optimal sublot allocation had a geometric series that was identical to that presented by Potts and Baker (1989). They also developed an algorithm for the discrete version. Baker (1995) extended Vickson and Alfredsson's (1992) model to obtain one model with attached and detached setups, and another model with time lags. Sriskandarajah and Wagneur (1999) assumed that the processing of sublots was performed on a no-wait basis. For the case where the number of sublots for each product was fixed, three optimal and three heuristic procedures were developed. A tabu search (TS) algorithm was designed to determine the best number of sublots for each product. Kumar *et al.* (2000) extended the work of Sriskandarajah and Wagneur (1999) to an m-machine case. A genetic algorithm (GA) was used to solve problems in which fixed and variable numbers of sublots for each product were included. Kalir and Sarin (2001b) presented an efficient heuristic procedure, called the bottleneck minimal idleness (BMI) heuristic. The BMI heuristic constructed a schedule in an attempt to minimize the idle time at the bottleneck machine. Yoon and Ventura (2002a) developed an LP model to find the optimal sublot allocation for a given sequence. Sixteen pairwise interchange methods were employed to generate the best sequences. The 16 methods were derived from combining four rules used to generate the initial sequences with four neighbourhood search mechanisms. In principle, the problem was divided into two independent subproblems and was solved separately. For the same problem, Yoon and Ventura (2002b) proposed a hybrid genetic algorithm.

For the job shop, Stephane and Lasserre (1993b, 1997) presented a mixed integer programming (MIP) model. A two-level iterative heuristic was proposed to solve the MIP model hierarchically. The first level was used to find the best sublot allocation for each product, given a fixed number of sublots and a fixed sequence of sublots. At the second level, given the sublot sizes determined at the first level, a sequence of sublots for each product was rearranged by solving a classical job shop problem, where each sublot was treated as an independent product (i.e., the intermingled case). Since solving a classical job shop problem is also NP-hard, a modified shifting bottleneck procedure proposed by Stephane and Lasserre (1993a), which, in fact, was an improved version of the shifting bottleneck procedure originally developed by Adams *et al.* (1988), was used. Jin *et al.* (1999) developed an integer programming (IP) model, a solution methodology based on combining Lagrangian relaxation with backward dynamic programming (BDP), and a heuristic method. An experiment involving four test cases was conducted to analyse the benefit obtained by adopting lot streaming, and to evaluate the performance of their methods. Jeong *et al.* (1999) proposed a three-step algorithm. The first step was used to obtain an initial solution by employing a modified shifting bottleneck procedure, in which the detached setup time, release time, transportation time, and due date were considered. The second step aimed to determine which product should be split. The final step was used to achieve schedule improvement. Moreover, they proposed methods for handling some dynamic situations (e.g., machine failure and rush orders).

For the open shop, Sen and Benli (1999) considered the non-intermingled single-route model and the intermingled multi-route model. In the former model, one property, namely, the existence of at most one dominant product along with Potts and Baker's (1989) geometric sublot sizes, was derived to obtain the optimal makespan. In the latter model, another property, namely, that at most one dominant product with two sublots of equal size could be obtained, was also derived.

From the above discussion, it can be seen that three approaches applied to the literature of this area are the analytical approach (i.e., the exact or optimization method), the heuristic approach, and the experimental approach (i.e., the simulation method). The papers listed in table 3 focused on the first two approaches. In the last part of this section, we will summarize the papers that focused on the experimental approach and are listed in table 4. Jacobs and Bragg (1988) introduced the repetitive lots (RL) method with equal sublots. In their method, once a machine was set up to process a sublot of a product, the machine remained dedicated to that product.

Table 4. A summary of the multi-product time-related problems solved using the experimental approach.

Author	Problem	Environment	Lot splitting method
Jacobs and Bragg 1988	$\{J_m, L_n, E, R, -, S_{(-,-)}/P, \overline{F}\}$ $\{J_m, L_n, E/C, R, -, S_{(-,-)}/P, C_{\max}\}$	Ten-machine, ten-job	Repetitive lots
Sassani 1990	$\{J_m, L_n, E/C, R, -, S_{(-,-)}/P, \sum \lvert C-d \rvert\}$	Three group technology cells	Transfer ratio
Wagner and Ragatz 1994	$\{O_m, L_n, E, R, -, S_{(-,-)}/P, \overline{T}\}$ $\{O_m, L_n, E, R, -, S_{(-,-)}/P, n_T\}$ $\{J_m, L_n, E, R, -, S_{(-,-)}/P, \overline{F}\}$	Five-machine open shop	Modified repetitive lots
Kannan and Lyman 1994	$\{J_m, L_n, E, R, -, S_{(-,-)}/P, \sum \lvert C-d \rvert\}$	Five-machine cell	Equal sublots
Smunt *et al.* 1996	$\{F_m, L_n, E/C, R, -, S_{(-,-)}/P, \overline{F}\}$ $\{J_m, L_n, E/C, R, -, S_{(-,-)}/P, \overline{F}\}$	Five-machine flow shop Ten-machine job shop	Modified repetitive lots
Ruben and Mahmoodi 1998	$\{T_m, L_n, E, R, -, S_{(-,-)}/P, \overline{F}\}$ $\{T_m, L_n, E, R, -, S_{(-,-)}/P, \overline{T}\}$	Six-machine A system Six-machine V system	Unit-sized sublots

When all the sublots of the product were completed, the machine began to process another product. Thus, those sublots of the product that used the same setup type were given the highest priority (i.e., this could be viewed as the non-intermingled case). When a new setup type was needed, a choice was made among the products waiting for processing based on a given criterion (i.e., the first-in, first-served (FIFS) rule). Sassani (1990) proposed a deterministic computer simulation model to study the effects of using lot streaming on the makespan, the percentage of late jobs, and the machine utilization. It was found that the performance of the group technology (GT) cells could be improved by using lot streaming, although the extent to which improvements could be achieved depended on the processing characteristics of each job in each cell. Wagner and Ragatz (1994) focused on the impact of adopting lot streaming on the due date performance. They used the concept behind the RL method presented by Jacobs and Bragg (1988) to develop a modified RL method, into which six dispatching rules were incorporated. Kannan and Lyman (1994) examined the effect of introducing lot streaming when three group (i.e., family-based) scheduling rules were considered. Although increasing the number of sublots could improve the performance, increasing the setup frequency retarded the effect of shortening the flow time. A full factorial (four-factor) experiment was carried out. They reported that employing the family-based scheduling rules was an effective way to reduce the negative impact on the flow time. On the other hand, lot streaming had little positive effect on the due date performance even when family-based scheduling rules were used. Smunt *et al.* (1996) conducted two sets of full factorial (seven-factor) experiments to analyse the impact on the stochastic flow shop and job shop, where the mean and variance of the flow time were the two performance measurements employed. Ruben and Mahmoodi (1998) investigated the impact of using unit-sized sublot splitting on two unbalanced production systems (i.e., the two simplest arborescent shops). An unbalanced system was one in which there was a long-term bottleneck in the production system. Similar to Wagner and Ragatz's (1994) RL method, they also used the concept behind Jacobs and Bragg's (1988) RL method to develop five scheduling rules and conducted a full factorial (five-factor) experiment.

It can easily be found from the above literature review on this area that, first, the multi-product lot streaming problem was divided into two independent subproblems (i.e., the sublot allocation subproblem and the sequence subproblem). In the former subproblem, the best sublot allocation was generally determined by using the analytic approach. In the latter subproblem, some traditional scheduling methods, such as Johnson's rule, meta-heuristic procedures, and the shifting bottleneck procedure, were used. As stated by Potts and Baker (1989), this hierarchical methodology could not be used to obtain a global solution since two independent subproblems were involved. In addition, most of the studies concentrated on flow shops, job shops, and open shops. In the future, arborescent shops and parallel-machine shops can be explored. Some methods (i.e., the branch and bound method and dynamic programming) used to solve the JIT mixed-type problem can be suitably applied to the intermingled LSSP since it is analogous to the JIT mixed-type problem. Second, the simulation method was used to solve the whole problem directly. Here, we suggest that lot streaming can be applied to more complex and real-life production systems (e.g., an integrated petrochemical and plastic production system) through the use of this method. Note that sequence-dependent setups will be worth consideration in the first and second parts.

5. Single-product cost-related problems

To deeply explore the evolution of lot streaming in this area, the research on lot-sizing can be traced back to the classical economic order quantity (EOQ) model (Harris 1913). This is the earliest and best known inventory model, which is referred to as the Wilson formula (Chikan 1990). Since some of the strong assumptions that were made to derive the EOQ model appeared to be very restrictive, some extended models, such as the economic production quantity (EPQ) model for the continuous-time stationary demand case and the Wagner–Whitin algorithm and other heuristics for the discrete time-varying demand case (e.g., Wagner and Whitin 1958, Nydick and Weiss 1989, Coleman 1992, Chiu 1997, Chiu and Chen 1997, Chiu *et al.* 2003, Chiu and Chen (forthcoming)), were developed. Later, some variations of the EPQ model were also developed to handle problems involving MLT in a multistage production system (e.g., Eilon 1962, Taha and Skeith 1970). In Eilon's (1962) model, only the finished-product inventory was considered, and the multistage production system was roughly treated as a single-stage system. Taha and Skeith (1970) recognized the relationship between MLT and WIP, and developed a lot size model in which variable processing lot sizes at different stages were permitted. On the other hand, several interesting papers related to the multistage production system but not involving MLT have been published by one of the present authors (Chiu and Lin 1988, Chiu 1993, Chiu and Huang 2003). However, these papers did not take lot streaming into account. In the literature, Taha and Skeith's (1970) model has been called the variable lot size (VLS) model. Alternatively, the uniform lot size (ULS) model, in which lot streaming is incorporated and the processing lot sizes at all stages are the same, is called the ULSLS model in this paper. Szendrovits (1975) was the first to present a ULSLS model which was essentially based on the classical EPQ model. A summary of the single-product cost-related problems is given in table 5.

As shown in table 5, Goyal (1976) and Szendrovits (1976) stated that the sublot transportation cost should be added to the total cost in Szendrovits's ULSLS model,

Table 5. A summary of single-product cost-related problems.

Problem	Author(s)
$\{F_2, L_1, C, R, I_{no}, S/P_{(-,IP,-)}/M^{\rightarrow}_{(-)}, TC\}$	Goyal 1977
$\{F_2, L_1, E, R, I_{no}, S/P_{(IM,IP,FG)}/M^{\rightarrow}_{(-)}, TC\}$	Graves and Kostreva 1986
$\{F_m, L_1, E, R, I_{no}, S/P_{(-,IP,FG)}, TC\}$	Szendrovits 1975
$\{F_m, L_1, E, R, I_{no}, S/P_{(-,IP,FG)}/M^{\rightarrow}_{(-)}, TC\}$	Goyal 1976, 1978, Szendrovits 1976, Drezner *et al.* 1984, Szendrovits and Golden 1984, Ranga *et al.* 2000
$\{F_m, L_1, E, R, I_{no}, S/P_{(-,IP,FG)}, TC\}$	Szendrovits and Drezner 1980
$\{F_m, L_1, V, R, I_{no}, S/P_{(-,IP,FG)}/M^{\rightarrow}_{(-)}, TC\}$	Goyal and Szendrovits 1986
$\{F_m, L_1, E, R, I, S/P_{(-,IP,-)}/M^{\rightarrow}_{(-)}, TC\}$	Szendrovits 1987
$\{F_m, L_1, E, R, I_{no}, P_{(-,IP,-)}/M^{\rightarrow}_{(-)}, TC\}$	Steiner and Truscott 1993
$\{F_m, L_1, C, R, I, P_{(-,IP,-)}/M^{\rightarrow}_{(-)}, TC\}$	Chiu and Xiang 1994, 1996
$\{F_m, L_1, E, R, I_{no}, S/P_{(IM,IP,FG)}/M^{\rightarrow}_{(-)}, TC\}$	Chiu and Chang (forthcoming)
$\{F_m, L_1, C, R, I_{no}, S/P_{(-,IP,-)}/M^{\rightarrow}_{(-)}, TC\}$	Bogaschewsky *et al.* 2001
$\{F_m, L_1, V, A, I_{no}, S_{(A,D)}/P/M^{\rightleftarrows}_{(w)}, TC(C_{max}, k)\}$	Chiu and Lee 2002, Chiu *et al.* 2004
$\{T_m, L_1, E, R, -, S/P_{(IM,IP,FG)}, TC\}$	Moily 1986

and that a solution procedure could be employed to simultaneously obtain the solution values of the processing lot size and the number of sublots. Goyal (1977) presented a model in which sublot sizes with a geometric series of $\{q, rq, r^2q, \ldots, r^{k-1}q\}$, where $r = t_2/t_1$, were assumed (note that the processing lot size was set to be more than one in his paper). He showed that the total amount of WIP when sublots of the geometric series were assumed to be consistent was less than that when equal sublots were assumed, as in Szendrovits's ULSLS model. Goyal (1978) extended Szendrovits's ULSLS model to obtain a more generalized model to accurately measure the amount of WIP. He obtained a processing lot size for each predetermined number of sublots and then chose the one with the lowest total cost as the optimum size. Drezner *et al.* (1984) proposed a variable lot size with lot streaming (VLSLS) model. Due to the considerable complexity of optimally solving this model, a heuristic procedure was proposed and was shown to be close-to-optimal based on a lower bound. Szendrovits and Golden (1984) compared the VLS model with the ULSLS model based on the results of a large number of simulated practical problems. The results revealed that using the ULSLS model yielded a lower total cost. The cost savings obtained by using the ULSLS model could increase as the number of stages increased, but it depended on the magnitude of the sublot transportation cost. Graves and Kostreva (1986) applied the concept behind Szendrovits's ULSLS model to overlapping operations in a material requirements planning (MRP) framework. They studied only one manufacturing segment consisting of two stages and constructed a more complicated model in which the setup cost, the sublot transportation cost, and the holding costs of raw materials, WIP, and finished goods were included. Moily (1986) developed a cost model for an arborescent production system in which a lot splitting policy (called integer split lot requirements (ISLR) lot-sizing in his paper) was assumed. In fact, it was an equal sublots splitting policy. An optimal procedure and a heuristic procedure were proposed. Note that Moily's model was not the ULSLS or VLSLS model. Szendrovits (1987) presented a complicated ULSLS model, in which machine idle time was allowed. He developed an iterative multivariate heuristic procedure based on some intuitive observations about the model to determine the processing lot size, the number of sublots, and a vector associated with the machine idle time. Steiner and Truscott (1993) applied the concept behind Szendrovits's ULSLS model to develop two cost models in order to calculate the amount of WIP on the basis of the makespan and the total flow time, respectively. The optimality conditions were given to minimize the total cost (called the total processing cost in their paper). Chiu and Xiang (1994, 1996) derived a total cost function and proposed an iterative two-phase procedure. In the first phase, the LP model developed by Kropp and Smunt (1990) was used to find the optimal allocation of consistent sublots, given a known number of sublots. The second phase was used to determine the total cost, given the optimal sublot sizes obtained in the first phase. This process was repeated until the number of sublots with the lowest total cost was found. It should be noted that the optimal number of sublots and the optimal sublot sizes obtained by using their procedure were local optimums, not global optimums. Ranga *et al.* (2000) proposed a model in which the setup time, the wait time, and the sublot transportation time that constituted a part of the makespan were also considered, while Szendrovits (1975) used an arbitrary factor to multiply the technological lead time based only on the processing time. They used a method similar to that developed by Szendrovits (1975) to measure the amount of WIP. Because their model was very complex and had no closed-form solution, the optimal solutions could be obtained only in two special cases. Chiu and Chang (forthcoming) generalized the work of Goyal (1978) to develop two more realistic models. The main difference between the two studies was that shipment of sublots rather than of the whole lot was assumed in the former paper.

The above papers assumed that the number of sublots was the same across stages (i.e., equal sublots). For the case where the number of sublots might differ between any two adjacent stages (i.e., consistent or variable sublots), Szendrovits and Drezner (1980) presented an optimization method for solving a ULSLS model. Goyal and Szendrovits (1986) integrated the works of Goyal (1977) and Szendrovits and Drezner (1980) to develop a ULSLS model. In their paper, a heuristic procedure was proposed for obtaining equal and/or unequal sublot sizes. Bogaschewsky *et al.* (2001) extended Goyal's (1977) two-stage model to obtain an m-stage model. They assumed that the sublot sizes formed a geometric series. Chiu and Lee (2002) and Chiu *et al.* (2004) developed a BMIP model to minimize the total cost, including the makespan cost and the transportation cost. Two efficient heuristic methods were proposed since solving the BMIP model was extremely difficult.

From the above review, it is clear that Szendrovits (1975) was the first to extend the classical EPQ model to a multistage lot streaming model. Then, subsequent researchers directly or indirectly extended Szendrovits's concept to more general models in two ways. In the case of equal sublots, some researchers developed more accurate methods for measuring the inventory value during production. Others considered different numbers of sublots between two adjacent stages. Three issues were not considered in these studies. First, they did not try to maximize the total revenue or the total profit. Second, the learning and forgetting effects based on sublots or units did not appear in most of these papers. Third, the variable sublot transportation cost was not assumed. Thus, we suggest that these three issues can be considered in future research.

6. Other problems

Problems that do not properly fall into the three research areas discussed in sections 3, 4, and 5 will be examined in this section. Hancock (1991) studied the impact of applying a simple lot splitting rule to a job shop. He found that job lateness could be significantly reduced at the expense of slightly increasing the processing cost under three different routing strategies. Sekhar Naidu *et al.* (1995) dealt with the problem of determining the optimal lot sizes for a two-stage production system, which involved a single-sampling plan for rectifying inspections. They developed an iterative method for obtaining the processing lot size (called the 'whole lot' in their paper) at the first stage and several equal sublots (called 'split lots' in their paper) at the second stage. Eynan and Li (1997) considered a single-product lot splitting model with learning effects in a single-stage production system, where a long-term contract between a producer and a customer was assumed. In their model, the net present value (NPV) collected by determining each sublot size or, equivalently, the delivery time of each sublot shipment was maximized. They developed an optimal and two near-optimal solution procedures. Kalir and Sarin (2000) presented, for the first time, analytic results related to the potential benefits of using lot streaming in single-product and multi-product flow shops. The upper bounds on the achievable benefits with respect to three performance measurements (i.e., makespan, mean flow time, and average WIP level) were provided in terms of the ratio of the performance measure with lot streaming to that without lot streaming. In the multi-product case, a common bottleneck, the non-intermingled case, and unit-sized sublot splitting were assumed for the sake of analytical simplicity. Kher *et al.* (2000) discussed the effects of their push and pull lot splitting policies on lot tracing, which referred to activities that were performed so as to clearly identify the component material and the individual process attributes used in the production of specific lots or batches, and on the material handling costs in stochastic flow shop environments similar to those studied by Wagner and Ragatz (1994) and Smunt *et al.* (1996). Serin and Kayaligil (2003) presented two models that were extensions of Eynan and Li's (1997) model. Bukchin and Masin (2004) investigated a bi-objective lot splitting problem, in which the makespan and mean flow time were minimized simultaneously. Riezebos (2004) conducted a simulation study to analyse the extent to which each of five equal sublots splitting policies could improve performance when the time bucket length, which is a critical parameter in the design of a good-performing MRP system, was considered.

A non-trivial issue was to use queueing theory to analyse the effectiveness of lot streaming. Benjaafar (1996) extended the work of Karmarkar (1987) to find the relationship between MLT and lot streaming. He also considered sublot transportation, and concluded that the unit-sized sublots splitting policy might not be optimal when operating the two-stage production system linked by a single discrete transportation device.

In this section, we cannot summarize the results of the previous studies due to the diversity of the problems involved. Future research can focus on the multi-objective lot streaming problems and MRP lot streaming.

7. Concluding remarks and future directions

In the past few decades, lot streaming has been shown to be an effective technique for shortening MLT and reducing WIP in manufacturing. A comprehensive survey has been presented here to make up for the lack of a literature review on this subject. We have proposed a unique structure, which includes three main dimensions, seven subdimensions, and 17 levels (see table 1), to classify lot streaming problems. In addition, a notation set containing seven elements with respect to the seven subdimensions has been defined and employed to examine each existing lot streaming problem. Based on the classification structure and the defined notation set, tables 2 through 5 summarize the single-product time-related, multi-product time-related, and single-product cost-related problems. From these summaries, readers can clearly understand the evolution of research on lot streaming. In addition, some problems that cannot be suitably classified have also been presented.

In sections 3 through 6, we have discussed the shortcomings of the previous studies and provided readers with some suggestions for future research. These suggestions are summarized as follows.

(1) For single-product time-related problems, some well-known meta-heuristic procedures (e.g., GA and TS) are recommended. Stochastic scheduling and fuzzy scheduling without lot streaming have received more attention in

recent years. Hence, stochastic and fuzzy lot streaming will be future research topics.

(2) Most of the existing studies assumed that the buffer size between any two adjacent stages was unlimited. A limited buffer size will be assumed in future research.

(3) In the past, the sublot transportation time or cost was assumed to be constant, irrespective of the sublot size. A variable sublot transportation time or cost will be assumed in future work.

(4) Arborescent shops and parallel-machine shops can be considered in multi-product time-related problems.

(5) In the case of multi-product time-related problems, the previously proposed hierarchical procedures did not guarantee that a global solution could be obtained. Developing an effective method for finding a global optimum will be of interest.

(6) If the container size and the number of kanbans attached to containers are compared with the sublot size and the number of sublots, respectively, then the existing JIT mixed-type approaches can be applied to lot streaming.

(7) Sequence-dependent setups can be assumed in multi-product time-related problems.

(8) Besides minimizing the total cost, maximizing the total revenue or the total profit can be a subject of research.

(9) The learning and forgetting effects based on sublots or units can be incorporated into single-product cost-related problems.

(10) Whereas a single performance measurement criterion was adopted in previous lot streaming models, multiple criteria can be adopted in the future. In addition, time discounting, the disposal of defective sublots after sampling and inspecting, machine maintenance and reliability, the dynamic arrivals of jobs, and discrete time-varying demand can also be investigated in future studies.

(11) Dampening nervousness in a MRP system has received some attention in the literature (e.g., Steele 1975, Blackburn *et al.* 1986, Ho *et al.* 1995, Bai *et al.* 2002). Combining MRP and lot streaming into a novel field called MRP lot streaming will be an interesting challenge. However, lot streaming can be a useful tool for dampening the nervous effect, which is similar to the bullwhip effect in a multi-echelon supply chain system.

Acknowledgements

The authors gratefully thank the anonymous referees for their helpful suggestions.

References

Adams, J., Balas, E. and Zawack, D., The shifting bottleneck procedure for job shop scheduling. *Manage. Sci.*, 1988, **34**, 391–401.

Bai, X., Davis, J.S., Kanet, J.J., Cantrell, S. and Patterson, J.W., Schedule instability, service level and cost in material requirements planning system. *Int. J. Prod. Res.*, 2002, **40**, 1725–1758.

Baker, K.R., Lot streaming in the two-machine flow shop with setup times. *Ann. Oper. Res.*, 1995, **57**, 1–11.

Baker, K.R. and Jia, D., A comparative study of lot streaming procedures. *OMEGA*, 1993, **21**, 561–566.

Baker, K.R. and Pyke, D.F., Solution procedures for the lot-streaming problem. *Decis. Sci.*, 1990, **21**, 475–491.

Benjaafar, S., On production batches, transfer batches, and lead times. *IIE Trans.*, 1996, **28**, 357–362.

Blackburn, J.D., *Time-Based Competition*, 1991 (Business One Irwin, Burr Ridge, IL).

Blackburn, J.D., Kropp, D.H. and Millen, P.A., A comparison of strategies to dampen nervousness in MRP system. *Manage. Sci.*, 1986, **32**, 413–429.

Bockerstette, J.A. and Shell, R.L., *Time Based Manufacturing*, 1993 (McGraw-Hill, New York).

Bogaschewsky, R.W., Buscher, U.D. and Lindner, G., Optimizing multi-stage production with constant lot size and varying number of unequal sized batches. *OMEGA*, 2001, **29**, 183–191.

Bukchin, J. and Masin, M., Multi-objective lot splitting for a single product *m*-machine flowshop line. *IIE Trans.*, 2004, **36**, 191–202.

Bukchin, J., Tzur, M. and Jaffe, M., Lot splitting to minimize average flow-time in a two-machine flow-shop. *IIE Trans.*, 2002, **34**, 953–970.

Cetinkaya, F.C., Lot streaming in a two-stage flow shop with set-up, processing and removal times separated. *J. Oper. Res. Soc.*, 1994, **45**, 1445–1455.

Cetinkaya, F.C. and Kayaligil, M.S., Unit sized transfer batch scheduling with setup times. *Comput. Indust. Eng.*, 1992, **22**, 177–183.

Chen, J. and Steiner, G., Lot streaming with detached setups in three-machine flow shops. *Eur. J. Oper. Res.*, 1996, **96**, 591–611.

Chen, J. and Steiner, G., Approximation methods for discrete lot streaming in flow shops. *Oper. Res. Lett.*, 1997, **21**, 139–145.

Chen, J. and Steiner, G., Lot streaming with attached setups in three-machine flow shops. *IIE Trans.*, 1998, **30**, 1075–1084.

Chen, J. and Steiner, G., Discrete lot streaming in two-machine flow shops. *INFOR*, 1999, **37**, 160–173.

Chen, J. and Steiner, G., On discrete lot streaming in no-wait flow shops. *IIE Trans.*, 2003, **35**, 91–101.

Chikan, A., *Inventory Models*, 1990 (Kluwer Academic, Boston).

Chiu, H.N., A cost saving technique for solving capacitated multi-stage lot-sizing problems. *Comput. Indust. Eng.*, 1993, **24**, 367–377.

Chiu, H.N., Discrete time-varying demand lot-sizing models with learning and forgetting effects. *Prod. Plann. Contr.*, 1997, **8**, 484–493.

Chiu, H.N. and Chang, J.H., Lot-streaming models using optimized production technology in multi-stage serial manufacturing systems. Working Paper, Department of Industrial Management, National Taiwan University of Science and Technology, Taipei, Republic of China, 1996a.

Chiu, H.N. and Chang, J.H., A study on lot-streaming models in multi-stage manufacturing systems. Technical Report NSC 85-2213-E-011-015, National Science Council of Republic of China, 1996b.

Chiu, H.N. and Chang, J.H., Optimal and heuristic transfer batch allocation models for single-product lot-streaming in a flow shop. *J. Chin. Inst. Ind. Eng.*, 1996c, **13**, 329–341.

Chiu, H.N. and Chang, J.H., Cost models for lot streaming in a multistage flow shop. *OMEGA* (forthcoming).

Chiu, H.N., Chang, J.H. and Lee, C.H., Lot streaming models with a limited number of capacitated transporters in multistage batch production systems. *Comput. Oper. Res.*, 2004, **31**, 2003–2020.

Chiu, H.N. and Chen, H.M., The effect of time-value of money on discrete time-varying demand lot-sizing models with learning and forgetting considerations. *Eng. Econ.*, 1997, **42**, 203–221.

Chiu, H.N. and Chen, H.M., An optimal production policy for the dynamic lot-sizing model with learning and forgetting in setups and production. *Int. J. Prod. Econ.* (forthcoming).

Chiu, H.N., Chen, H.M. and Weng, L.C., Deterministic time-varying demand lot-sizing models with learning and forgetting in setups and production. *Prod. Oper. Manage.*, 2003, **12**, 120–127.

Chiu, H.N. and Huang, H.L., A multi-echelon integrated JIT inventory model using the time buffer and emergency borrowing policies to deal with random delivery lead times. *Int. J. Prod. Res.*, 2003, **41**, 2911–2931.

Chiu, H.N. and Lee, C.H., Lot streaming models with limited transporter capacity in a flow shop. Working Paper, Department of Industrial Management, National Taiwan University of Science and Technology, Taipei, Republic of China, 2002.

Chiu, H.N. and Lin, T.M., An optimal lot-sizing model for multi-stage series/assembly systems. *Comput. Oper. Res.*, 1988, **15**, 405–415.

Chiu, H.N. and Xiang, S.C., A study of optimized production technology and lot-streaming in a flow shop. Working Paper, Department of Industrial Management, National Taiwan University of Science and Technology, Taipei, Republic of China, 1994.

Chiu, H.N. and Xiang, S.C., Optimal and heuristic models for single product lot streaming in a flow shop. *J. Chin. Inst. Ind. Eng.*, 1996, **13**, 73–83.

Coleman, B.J., A further analysis of variable demand lot-sizing techniques. *Prod. Invent. Manage. J.*, 1992, **33**, 19–23.

Drezner, Z., Szendrovits, A.Z. and Wesolowsky, G.O., Multi-stage production with variable lot sizes and transportation of partial lots. *Eur. J. Oper. Res.*, 1984, **17**, 227–237.

Eilon, S., *Elements of Planning and Control*, 1962 (Macmillan, New York).

Eynan, A. and Li, C.L., Lot-splitting decisions and learning effects. *IIE Trans.*, 1997, **29**, 139–146.

Fawcett, S.E. and Pearson, J.N., Understanding and applying constraint management in today's manufacturing environments. *Prod. Invent. Manage. J.*, 1991, **32**, 46–55.

Glass, C.A., Gupta, J.N.D. and Potts, C.N., Lot streaming in three-stage production processes. *Eur. J. Oper. Res.*, 1994, **75**, 378–394.

Glass, C.A. and Potts, C.N., Structural properties of lot streaming in a flow shop. *Math. Oper. Res.*, 1998, **23**, 624–639.

Goldratt, E.M., Optimized production timetable: a revolution program for industry, in *APICS 23th Int. Conf. Proc.*, 1980, pp. 172–176.

Goldratt, E.M., *What Is This Thing Called Theory of Constraints*, 1990 (North River, New York).

Goyal, S.K., Notes on 'manufacturing cycle time determination for a multi-stage economic production quantity model'. *Manage. Sci.*, 1976, **23**, 332–333.

Goyal, S.K., Determination of optimum production quantity for a two-stage production system. *Oper. Res. Quart.*, 1977, **28**, 865–870.

Goyal, S.K., Economic batch quantity in a multistage production system. *Int. J. Prod. Res.*, 1978, **16**, 267–273.

Goyal, S.K. and Szendrovits, A.Z., A constant lot size model with equal and unequal sized batch shipments between production stages. *Eng. Costs Prod. Econ.*, 1986, **10**, 203–210.

Graves, S.C. and Kostreva, M.M., Overlapping operations in material requirements planning. *J. Oper. Manage.*, 1986, **6**, 283–294.

Hancock, T.M., Effects of lot-splitting under various routing strategies. *Int. J. Oper. Prod. Manage.*, 1991, **11**, 68–75.

Harris, F.W., How many parts to make at once. *Factory, Mag. Manage.*, 1913, **10**, 135–136, 152.

Ho, C.J., Law, W.K. and Rampal, R., Uncertainty-dampening methods for reducing MRP system nervousness. *Int. J. Prod. Res.*, 1995, **33**, 483–496.

Jacobs, F.R. and Bragg, D.J., Repetitive lots: flow time reductions through sequencing and dynamic batch sizing. *Decis. Sci.*, 1988, **19**, 281–294.

Jeong, H.I., Park, J.W. and Leachman, R.C., A batch splitting method for a job shop problem in an MRP environment. *Int. J. Prod. Res.*, 1999, **37**, 3583–3598.

Jin, B., Luh, P.B. and Thakur, L.S., An effective optimization-based algorithm for job shop scheduling with fixed-size transfer lots. *J. Manuf. Syst.*, 1999, **18**, 284–300.

Johnson, S.H., Optimal two- and three-stage production schedules with setup times included. *Naval Res. Logist. Quart.*, 1954, **1**, 61–68.

Kalir, A.A. and Sarin, S.C., Evaluation of the potential benefits of lot streaming in flow-shop systems. *Int. J. Prod. Econ.*, 2000, **66**, 131–142.

Kalir, A.A. and Sarin, S.C., Optimal solutions for the single batch, flow shop, lot-streaming problem with equal sublots. *Decis. Sci.*, 2001a, **32**, 387–397.

Kalir, A.A. and Sarin, S.C., A near-optimal heuristic for the sequencing problem in multiple-batch flow-shops with small equal sublots. *OMEGA*, 2001b, **29**, 577–584.

Kannan, V.R. and Lyman, S.B., Impact of family-based scheduling on transfer batches in a job shop manufacturing cell. *Int. J. Prod. Res.*, 1994, **32**, 2777–2794.

Karmarkar, U.S., Lot sizes, lead times and in-process inventories. *Manage. Sci.*, 1987, **33**, 409–418.

Kher, H.V., Malhotra, M.K. and Steele, D.C., The effect of push and pull lot splitting approaches on lot traceability and material handling costs in stochastic flow shop environments. *Int. J. Prod. Res.*, 2000, **38**, 141–160.

Kropp, D.H. and Smunt, T.L., Optimal and heuristic models for lot splitting in a flow shop. *Decis. Sci.*, 1990, **21**, 691–709.

Kumar, S., Bagchi, T.P. and Sriskandarajah, C., Lot streaming and scheduling heuristics for *m*-machine no-wait flowshops. *Comput. Indust. Eng.*, 2000, **38**, 149–172.

Lundrigan, R., What is this called OPT? *Prod. Invent. Manage. J.*, 1986, **27**, 2–11.

McMullen, P.R., The permutation flow shop problem with just in time production considerations. *Prod. Plann. Contr.*, 2002, **13**, 307–316.

Miltenburg, J., Level schedules for mixed-model assembly lines in just-in-time production systems. *Manage. Sci.*, 1989, **35**, 192–207.

Moily, J.P., Optimal and heuristic procedures for component lot-splitting in multi-stage manufacturing systems. *Manage. Sci.*, 1986, **32**, 113–125.

Monden, Y., *Toyota Production System*, 1983 (Industrial Engineering and Management Press, Norcross, GA).

Nydick, R.L. Jr. and Weiss, H.J., An evaluation of variable-demand lot-sizing techniques. *Prod. Invent. Manage. J.*, 1989, **30**, 41–44.

Philipoom, P.R., Rees, L.P. and Taylor, B.W., Simultaneously determining the number of kanbans, container sizes, and the final-assembly sequence of products in a just-in-time shop. *Int. J. Prod. Res.*, 1996, **34**, 293–300.

Potts, C.N. and Baker, K.R., Flow shop scheduling with lot streaming. *Oper. Res. Lett.*, 1989, **8**, 297–303.

Potts, C.N. and Van Wassenhove, L.N., Integrating scheduling with batching and lot-sizing: a review of algorithms and complexity. *J. the Operational Research Society*, 1992, **43**, 395–406.

Ranga, V.R., Haizhen, F., Duncan, K.H.F. and Jack, C.H., Lot streaming in multistage production systems. *Int. J. Prod. Econ.*, 2000, **66**, 199–211.

Reiter, S., A system for managing job shop production. *J. Bus.*, 1966, **34**, 371–393.

Riezebos, J., Time bucket length and lot splitting. *Int. J. Prod. Res.*, 2004, **42**, 2325–2338.

Ronen, B. and Starr, M.K., Synchronous manufacturing as in OPT: from practice to theory. *Comput. Indust. Eng.*, 1990, **18**, 585–600.

Ruben, R.A. and Mahmoodi, F., Lot splitting in unbalanced production systems. *Decis. Sci.*, 1998, **29**, 921–949.

Sassani, F., A simulation study on performance improvement of group technology cells. *Int. J. Prod. Res.*, 1990, **28**, 293–300.

Sekhar Naidu, B.M.C., Sarma, K.V.S. and Goyal, S.K., Optimal two-stage production–inventory policy with whole lot and split lots. *Prod. Plann. Contr.*, 1995, **6**, 134–139.

Sen, A. and Benli, O.S., Lot streaming in open shops. *Oper. Res. Lett.*, 1999, **23**, 135–142.

Sen, A., Topaloblu, E. and Benli, O.S., Optimal streaming of a single job in a two-stage flow shop. *Eur. J. Oper. Res.*, 1998, **110**, 42–62.

Serin, Y. and Kayaligil, S., Lot splitting under learning effects with minimal revenue requirements and multiple lot types. *IIE Trans.*, 2003, **35**, 689–698.

Silver, E.A., Pyke, D.F. and Peterson, R., *Inventory Management and Production Planning and Scheduling*, 3rd edn, 1998 (Wiley, New York).

Smunt, T.L., Buss, A.H. and Kropp, D.H., Lot splitting in stochastic flow shop and job shop environments. *Decis. Sci.*, 1996, **27**, 215–238.

Sriskandarajah, C. and Wagneur, E., Lot streaming and scheduling multiple products in two-machine no-wait flowshops. *IIE Trans.*, 1999, **31**, 695–707.
Steele, D.C., The nervous MRP system: how to do battle. *Prod. Invent. Manage. J.*, 1975, **16**, 83–89.
Steiner, G. and Truscott, W.G., Batch scheduling to minimize cycle time, flow time, and processing cost. *IIE Trans.*, 1993, **25**, 90–97.
Stephane, D.P. and Lasserre, J.B., A modified shifting bottleneck procedure for job-shop scheduling. *Int. J. Prod. Res.*, 1993a, **31**, 923–932.
Stephane, D.P. and Lasserre, J.B., An iterative procedure for lot streaming in job-shop scheduling, in *Proceedings of the 15th Annual Conference on Computers and Industrial Engineering*, 1993b, pp. 231–234.
Stephane, D.P. and Lasserre, J.B., Lot streaming in job-shop scheduling. *Oper. Res.*, 1997, **45**, 584–595.
Szendrovits, A.Z., Manufacturing cycle time determination for a multi-stage economic production quantity model. *Manage. Sci.*, 1975, **22**, 298–308.
Szendrovits, A.Z., On the optimality of sub-batch for a multi-stage EPQ model—a rejoinder. *Manage. Sci.*, 1976, **23**, 334–338.
Szendrovits, A.Z., An inventory model for interrupted multi-stage production. *Int. J. Prod. Res.*, 1987, **25**, 129–143.
Szendrovits, A.Z. and Drezner, Z., Optimizing multi-stage production with constant lot size and varying numbers of batches. *OMEGA*, 1980, **8**, 623–629.
Szendrovits, A.Z. and Golden, M., The effect of two classes of process organisations on multi-stage production/inventory systems. *INFOR*, 1984, **22**, 40–55.
Taha, H.A. and Skeith, R.W., The economic lot sizes in multistage production systems. *AIIE Trans.*, 1970, **2**, 157–162.
Trietsch, D. and Baker, K.R., Basic techniques for lot streaming. *Oper. Res.*, 1993, **41**, 1065–1076.
Truscott, W.G., Scheduling production activities in multi-stage batch manufacturing system. *International J. Prod. Res.*, 1985, **23**, 315–328.
Truscott, W.G., Production scheduling with capacity-constrained transportation activities. *J. Oper. Manage.*, 1986, **6**, 333–348.
Ventura, J.A. and Radhakrishnan, S., Sequencing mixed model assembly lines for a Just-In-Time production system. *Prod. Plann. Contr.*, 2002, **13**, 199–210.
Vickson, R.G., Optimal lot streaming for multiple products in a two-machine flow shop. *Eur. J. Oper. Res.*, 1995, **85**, 556–575.
Vickson, R.G. and Alfredsson, B.E., Two- and three-machine flow shop scheduling problems with equal sized transfer batches. *Int. J. Prod. Res.*, 1992, **30**, 1551–1574.
Wagner, B.J. and Ragatz, G.L., The impact of lot splitting on due date performance. *J. Oper. Manage.*, 1994, **12**, 13–25.
Wagner, H.M. and Whitin, T.M., Dynamic version of the economic lot size model. *Manage. Sci.*, 1958, **5**, 89–96.
Williams, E.F., Tufekci, S. and Akansel, M., $O(m^2)$ algorithms for the two and three sublot lot streaming problem. *Prod. Oper. Manage.*, 1997, **6**, 74–96.
Yoon, S.H. and Ventura, J.A., Minimizing the mean weighted absolute deviation from due dates in lot-streaming flow shop scheduling. *Comput. Oper. Res.*, 2002a, **29**, 1301–1315.
Yoon, S.H. and Ventura, J.A., An application of genetic algorithms to lot-streaming flow shop scheduling. *IIE Trans.*, 2002b, **34**, 779–787.