General Solution for the Self-Organizing, Distributed, Real-Time Scheduling of FMS-Automatic Lot-Streaming Using Hybrid Dynamical Systems

János Somló, Imre J. Rudas

Óbuda University
Bécsi út 96/B, H-1034 Budapest, Hungary
somlo@uni-obuda.hu
rudas@uni-obuda.hu

Abstract: The use of hybrid dynamical systems opens a new horizon for flexible manufacturing systems scheduling. It even makes possible directly connect scheduling and MRP. ***In the present paper the most important new result is the proposed demand rates determination method for multi-section scheduling problems.*** *Some other important achievements making possible the application of this approach are discussed,too. These are, for example:*

- ***feedback control law*** *resulting in stable (implementable with finite buffers) and regular (converging to periodic) processes is described;*
- ***optimal demand rates determination*** *for single- section problems is discussed.*

In this paper the "buffer principle" of planning is used and a bottleneck scheduling approach is applied. As the result, production times close to the global minimum of net manufacturing time, determined by the loading characteristics of bottleneck machine-groups, may be realized.

*The proposed control is totally self-organizing. No outside control commands are necessary. Every buffer is only connected (in the signal level) with the previous and the next buffer. The actions are real-time controlled. The most important feature of this control is that it significantly improves the efficiency of utilization of system resources.****The generalization for multi-section problems makes it possible to solve the most common application tasks. Ev, the solution of dynamical input problems becomes possible.***

Keywords: Flexible manufacturing systems; Scheduling; Hybrid dynamical systems; Stability; Periodic regimes; Optimal demand rates; Buffers, Bottleneck, Automatic lot-streaming; Overlapping production; Single-sections; Multi-sections; Self-organizing; Distributed; Real-time control; Scheduling and MRP

1 Introduction

In the present paper we extend the development in paper [1] for single-sections hybrid dynamical systems based FMS scheduling method for multi-section scheduling problems. This last is the most common formulation (coming from MRP level) of production tasks. The results may even be used when the tasks (inputs) have a dynamical character.

The good quality of the solution of scheduling problems is a key factor in the effective utilization of flexible manufacturing systems. The value of production of these systems is very high. This is why every improvement in processes quality has a significant economic effect. Classic methods of scheduling manufacturing processes may be used for Flexible Manufacturing Systems (FMS), too. In general, the approaches used for the solution of manufacturing scheduling problems belonged to applied operation research problems. Perkins and Kumar in [4] proposed to use for this goal Hybrid Dynamical Systems (HDS) theory based methods. By this, the scheduling problem became a control theory problem for which the approaches of dynamical systems investigation are suitable.

Concerning manufacturing production planning,problems arise when we want to exploit the advantages and opportunities of these systems, where not only the manufacturing but also the handling processes are automatically realized. As an example of the difficulties,we mention the application of lot-streaming and overlapping production. In systems with high-level computerized control it is trivial to use these methods. But it is not trivial how to realize it. Even with very the powerful technology of computations, similar tasks lead to problems with very high dimensions with no chance of effective solution. Lot-streaming and overlapping production methods can be effectively used in flexible manufacturing systems due to the small values of set-up times. But to plan the processes is extremely difficult because of the increase in dimensions. Exactly this is the field where the application of the methods of hybrid dynamical systems may result in a breakthrough.

The manufacturing scheduling problems are in the centre of attention of the scientific literature. A high number of publications are available. A survey is in [13] and others. From classical works we recall [14, 15].

Hybrid dynamical systems have attracted considerableattention in recent years (see e.g. [2, 3] and references therein). In general, HDS are those that combine continuous and discrete behavior and involve, thereby, both continuous and discrete statevariables. In many cases, such systems operate as follows. Whilethe discrete state remains constant, the continuous one obeys adefinite dynamical law. Transition to another discrete stateimplies a change of this law. In its turn, the discrete state evolvesas soon as a certain event occurs with both the evolution and theevent depending on the continuous state.

The class of HDS we are dealing with in the present paper consists of complex switched server queuing networks. This class of HDS was introduced in [4, 5] to model flexible manufacturing systems. Different aspects of the investigation of the processes in these systems were outlined in [2÷12]. A flexible manufacturing system considered in this paper produces several part-types on a network of machines. Raw parts are the inputs to the network. Parts arriving to a machine are waiting in buffers and are supplied to the machines when required. Each unit of a given part-type requires a predetermined processing time at each of several machines, in a given order. A set-up time is required whenever a machine switches from processing one part-type to another.

The investigation of the fluid analogy of similar systems is a popular research field. Especially the periodic processes in these systems have attracted significant attention (see: [16, 17, 18, 19]).

It should be emphasized here that in FMS the set-up times have small value compared with the manufacturing times. Nevertheless, as is well known from practice, and as is proved by theoretical investigations, their values may not be neglected without serious consequences.

In classic manufacturing scheduling problems, the parts are supposed to be delivered to machines in batches. The batches are properly sized. When hybrid dynamical methods are used, the basic difference in part delivery policy is that the parts are delivered to buffers serving the machines in a continuous flow. More precisely, the part demand is (equally) distributed in time. In the early works (see e.g.: [2÷9]) on HDS theories used for FMS scheduling, the inputs representing the production tasks were introduced as infinite flows without start and end. This representation was suitable for stability problem formulation and for the investigation of periodic motions in such systems. If the systems are stable, the practically interesting regimes of their motion are the periodic ones. Results regarding periodic motions in these systems were published, for example, in [3÷9].

In [10] a new aspect of input flows determination was proposed which reflected the practical requirements of scheduling. Namely, the part demand (part arrival) was determined in a way that it should result in the production of the given part-types in the given number during the given (scheduling) time. Clearly, one of the most important points is that the production time for an order be as low as possible.

The above mentioned method was developed for systems where the tasks (production order) were given for one single (common) scheduling section. **In the present paper we extend the results for cases when multi-section problems are formulated; that is, every series of part-types has its own(individual) scheduling section (interval) but these sections overlap.**This second case, of course, contains the single (common) scheduling section case, too.

In the present paper,as theoretical basis, the results of paper [1] are used. These results concern the solution of manufacturing scheduling problems by the use of hybrid dynamical methods. We remark that throughout the present paper the continuous representation is used. Furthermore, it is supposed that the number of parts is suitably large, and the set-up times are suitably small.

The content of the present paper covers the following:

After this Introduction, in **Section 2** we describe the problem statement for manufacturing scheduling.

In **Section 3** we give a simplified discussion of the theoretical investigation of the FMS scheduling problem. The main emphasis is on the aspects of practical use. So, the formulation of the feedback control law is simpler (but equivalent) than in [1]. Because the goal is to extend the results to multi-section scheduling problems the results are formulated to serve this goal.

In **Section 4** we describe the method developed for single section scheduling. We propose an approach for optimal determination of demandrates. This is important because the basis for the effective solution of the multi-section problem is the effective solution of single-section problems.

The main results of the present paper are given in Section 5 where we generalize the demand rates determination method for systems with multi-section scheduling intervals.

In **Section 6** some idea is given how the proposed in the paper make possible to **contact FMS scheduling and MRP.**

Conclusions are formulated in **Section 7**.

2 Problem Statement

Flexible manufacturing systems for scheduling by the use of HDS theories may be modeled as follows:
(i) There are P part-types labeled1; 2; . . . ; P, and a set $\boldsymbol{M}=\{1,2,\ldots,M\}$
of machine-groups which we will also call simply machines, in the following.
(ii) Parts of type p require processing at the machines $\mu_{p,1}, \mu_{p,2}, \ldots\ldots \mu_{p,n_p}$ in that order where $\mu_{p,j}\in 1,2,\ldots,M$. Here n_p is the index of the machine which processes the last operation of the given part (index of final machine-group). So, for example, when $\mu_{p,1}=3$, $\mu_{p,2}=1$, $\mu_{p,3}=4$ the part with index p is processed on machine identified with indices 3,1,4, in this order. Because the machine with index 3 is the last $n_p=3$.)
(iii) Raw parts of type p arrive to the system at the machine $\mu_{p,1}$ at a constant rate $r_p>0$.

(iv) At the j-th machine they visit, parts of type p enter the buffer labelled $b_{p,j}$ from which they are eventually processed by thismachineat a given constant rate $R_{p,j}>0$. The dimension of $R_{p,j}$ is [part/time unit]. We will use also the value$\tau_{p,j}=\frac{1}{R_{p,j}}$This is the time of processing one part on the given machine.
(v) We also assume that parts of type p incur a fixed transportation delay $l_{p,j}\geq 0$ when moving from the machine j to themachine j + 1.
(vi) The machine m is served from the buffers $B_m:=\{b_{p,j}: \mu_{p,j}=m\}$
A minimal set-up time $\delta_m^0>0$ is required when the machine with index m switches from processing parts of type p in the buffer b in B_m to processing parts of another type p' in B_m. The machine does not work during such a set-up time. Theset-up time can be artificially increased to achieve our controlgoal. In other words, set-up time $\delta_m(t)$ of the machine m is acontrol variable. However, condition

$$\delta_m(t) \geq \delta_m^0 \tag{2.1}$$

should be satisfied.
Now we face the task of scheduling the operations to perform the production orders. For that we should know the production capacities which we have assigned to perform the tasks. These capacities are given as the time intervals of available machine-groups devoted to the given operations.

2.1 Scheduling Sections

Single-section case

There is given the scheduling section as

$$0\leq\ t\ \leq T_{sch} \tag{2.2}$$

During this the given numbers of part-types should be produced in the number of items:

$$N_1, N_2\ N_3 \dots \dots \dots N_p$$

Multi-section case

The scheduling sections are given as

$re_j \leq t \leq dd_j$(2.3)

j=1,2,3…….J

Where:

$$\boldsymbol{re_j} - is\ the\ release\ time;\ \boldsymbol{dd_j} - is\ the\ due\ data\ for\ the\ given\ section$$

During the individual scheduling sections the following number of parts should be produced:

$N_{j(P_{j-1}+1)}$) , $N_{j(P_{j-1}+2)}$, ……..$N_{j(P_j)}$.

As can be recognized, the indexing of the part-types is continuous from 1, (1,2,3……). In a given section the part-types indices are from $P_{j-1}+1$ to P_j

3 Scheduling FMS using Hybrid Dynamical Methods

As was mentioned, in the present paper we use the results outlined in [1]. Here only the basic definitions and results are described to give a background for the formulation of the new results concerning the determination of demand rates for multi-section scheduling.

The structure of one layer of the manufacturing system in consideration is given in Fig. 3.1.

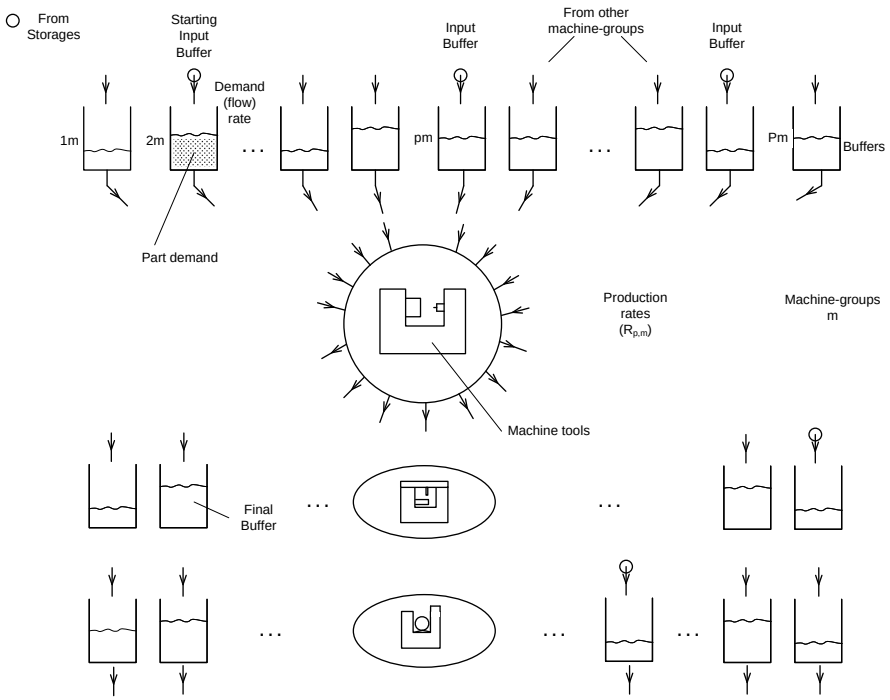


Figure 3.1

One layer of the flexible manufacturing system

This layer represents a machine-group together with the buffers serving the machines in supplying some given parts. Every part-type has its dedicated buffer. Here we remark that the buffers are understood as virtual ones because in reality these buffers may have different physical implementation. The buffers may be individual devices but may be parts of common storages or pallets, etc. Another side of this virtuality is that the numbers of part items are treated as real numbers

and not as integers, as is in real production. The buffers are filled-up from central (or local) storages, or from the output of other machine-groups. We suppose that the delivery among machine-groups is a continuous process. Among the buffers,the most common types are those which are served from some other buffers. But there are initial buffers served from storages. The momentary role of the initial buffers may be different. Some of them may supply parts for immediate processing. We name those as the starting initial buffer. Other buffers are only filled up with no output flow because the given machine (machine-group) is engaged to produce other parts. In the Figure, the final buffers are indicated, too. But they do not have any role in the systems actions. They serve only for the registration of the end of some production action.

We remark that the buffers here have symbolic meaning. In reality these indicate the buffer-machine-part-type items conglomerate. Their actions are equivalent with the actions of the conglomerate. One more remark: there are systems where the systems processes are continuous in reality. An example of these is: chemical systems processing fluid components. Of course, all that are proposed here are fully applicable to those systems.

As was already mentioned, we consider switching feedback control laws. Now, according to that the system, processes may be characterized as follows. At some particular time instant

- A given buffer is filled-up from storage (with r_p flow rate), but the machine-group which is served from this buffer is engaged in making other parts (or set-up is performed). In this case the buffer content is not reduced and its content permanently grows according to the input flow rate (demand rate).
- The same as above but the machine-group is engaged to produce the given parts so the buffer content is reduced according to the production rates.
- The buffer content is supplied from the machine before the given one (from the previous buffer) in the sequence of the production, but its content is not reduced because the given buffer for the time being is not engaged in the production.
- The same but the buffer content is reduced because the part-type items are produced.

In describing the systems work we will use the terminology: **active buffer.** An active buffer is the one which is engaged in production. That is; it receives a command for production from the previous buffer but the conditions to finish this production are not met, yet. (Where to relate the set-up times does not have any importance from the systems actions viewpoint.)

As we mentioned, the systems activities to produce parts are characterizedfrom the point of view of buffers. Accordingly, we will use for the description of processes the vector of buffer levels

$$\mathbf{x}(t)=\{x_{p,m}(t)\} \tag{3.1}$$

p∈1,2,……P

m∈1,2,…..M

Finally let $y_{p,i}$ (t) denotethecumulativeoutputofpart-type p from the buffer $b_{p,i}$ overthetimeinterval [0,t] i.e.,theamountof part-type items of type p processedbythemachine $\mu_{p,i}$ over [0,t]. Then, $y_{p,i}$ (t) is described by initialcondition $y_{p,i}(0)=0$ andequations:

if a buffer is active then $\dot{y}_{p,i}(t)=R_{\mu_{p,i}}$ (3.2)

if a buffer is not active then $\dot{y}_{p,i}(t)=0$

The cumulative output may be characterized by vector

y(t)={$y_{p,i}$(t)}

3.1 Feedback Control Law for Self-Organizing, Distributed, Real-Time Scheduling of FMS

In the system described above the continuous parts are characterized by the relations

$\dot{x}_{p,m}=u_{p,m}(t)$ (3.3)

p=∈1,2,……P

m=∈1,2,…..M

where$u_{p,m}$(t) are the input flows of the buffers.

In [1] the processes of the given system are given in algorithmic form.

This formulation gives, for example, for a buffer filled-up from the storage and reduced at the same time

$\dot{x}_{p,1}=r_p-R_{\mu_{p,1}}$ (3.4)

Another example is when a buffer is filled-up from a machine-group before and at the same timeits content is reduced according to the present production

$\dot{x}_{p,i}=R_{\mu_{p,i-1}}(t-l_{p,i-1})-R_{\mu_{p,i}}(t)$ (3.5)

For the relations describing the other processes in the system, see [1].

In [1] the following goal was formulated:

Let $d_1>0$; $d_2>0,\ldots\ldots d_P>0$ be given constants. These constants are calledproduction levels. Let$T>0$ be a given time value.The goal is to determine the part arrival rates r_1, $r_2,\ldots.r_P$ and feedback control policy such that for all p =1,2,…..P; i= 1,2,…..n_p, the value $y_{p,i}(k+1)T)-y_{p,i}(kT)$ (that is the amount of parts of type p processed by the machine $\mu_{p,i}$over the interval [kT; (k + 1)T]) is

close, in some sense, to d_p, where k =0,1,2,….. Furthermore, we wish to find the minimal time T for which this will be possible. Also, the closed-loop system should be stable. (The definition of stability is given below.)

For future use the following will be introduced:

Definition 2.1: (See [4], [5].) The closed-loop system constructed according to Fig. 3.1 and working with the use of switching feedback control is said to be stable if for any solution [$\mathbf{x}(t)$] with initial conditions $\mathbf{x}(0)=\mathbf{x}_0$ (where $\mathbf{x}_0$ is a vector with non-negative components)the vector function$\mathbf{x}$(t) is bounded on [0,∞), i.e. there exists a constant

$L(\mathbf{x_0})>0$

such that

$$x_{p,i}(t) \leq L(\mathbf{x}_0) \qquad \forall p, i, t. \tag{3.6}$$

Definition 2.2: The closed-loop system discussed above issaid to be regular with the production levels d_1, d_2,……d_P and the scheduling period T if it is stable and the following condition holds

$$\lim_{k\to\infty}\left(y_{p,i}\big((k+1)T\big)-y_{p,i}(kT)\right)=d_p \tag{3.7}$$

$$p \in 1,2,\dots P\,;\;\; i \in 1,2\,\dots\dots n_p$$

As was mentioned above, regularity with the production levels d_1,d_2.....d_P and the scheduling period T means that for any p the amount of part-type p processed over time intervals $[kT;\ (k+1)T]$ converges to d_p as k tends to infinity.

Definition 2.3: Assume that $d_1>0$, $d_2>0$,.....$d_P>0$ are given. The minimal time T_0 for which there exist constants $r_1>0$, $r_2>0$,....$r_P>0$ and a feedback policysuch that the closed-loop system is regularwith the production levels d_1, d_2,.....d_P and the schedulingperiod T_0, is called the minimal scheduling period of thesystem.

For the sake of the simplicity of future discussion, we introduce some more definitions and terminology.

The quantity

$$\mathrm{Lt_m}=\sum_{p=1}^{p=P} N_p\tau_{p,m} \tag{3.8}$$

m∈1,2,….M

is the net manufacturing time needed for any machine-group to produce the parts.

The quantity

$$\mathrm{D_m}=\sum_{p=1}^{p=P} d_p\tau_{p,m} \tag{3.9}$$

p∈ 1,2 ……P

m∈1,2,….M

is the net manufacturing time when d_p(p=1,2,…P) number of parts is produced from a part-type items in a period of periodic motions.

The quantity

$$ELt_m=[k_m\delta_m^0 + Lt_m] \tag{3.10}$$

m∈1,2,….M;

is named extended load time, where:k_m is the number of set-up events on a machine-group when performing a production order.

We will use also the following quantities

$$Mlt_m=MaxLt_m \tag{3.11}$$

m∈1,2,….M;

and

$$MELt_m=Max\ ELt_m \tag{3.12}$$

m∈1,2,….M;

These quantities belong to the bottleneck machine groups. It very frequently happens that the machine-groups for which the maximums of $Lt_m$and ELt_moccur are the same. This is because the set-up times usually are that of suitably small value. In the following, we will suppose that the bottleneck machine-group is the same as regards the net and extended load times. In the opposite case, it is very easy to modify the results.

It was proved in [1] that the minimal time T_0 for the closed loop system to be regular is

$$T_0=Max[k_m\delta_m^0+\sum_{p=1}^{p=P} d_p\tau_{p,m}]=Max[k_m\delta_m^0+D_m] \tag{3.13}$$

m∈1,2,……M

We introduce for any time value $T \geq T_0$ a quantity, which we name synchronization coefficient, as follows

$$\delta_m^M = \frac{T-D_m}{k_m} \tag{3.14}$$

The machine*m* works with $k_m$buffers. Denote the corresponding buffersb_1,b_2,…… b_{k_m} in an arbitrary order. Let us form the followingcyclic sequence of these buffers

$$b_1 \to b_2 \to \cdots \to b_{k_m} \to b_1 \tag{3.15}$$

Let bϵB_m. Then **next**[b] is the next buffer from B_m that is the next to b in the cyclic sequence (3.15).

We consider an active buffer. For the activity period of a buffer the following feedback policy was proposed. Let τ_m be the time instant when after the set-up time the processing of the parts begins. Furthermore, we introduce

$$\Delta_{p,m} = \frac{d_p}{R_{p,m}} = d_p \tau_{p,m} \tag{3.16}$$

This is the net manufacturing time to produce sub-parts of part-type p in number of items d_p in machine-group with index m.

The b_a value indicates which buffer is active at a given time instant. That is just after τ_m; $b=b_a$.

The following feedback policy is proposed

$$\textbf{if}\{x_{p,m} = 0 \ \boldsymbol{or}\ \Delta_t = \Delta_{p,m} - (t - \tau_m) = 0\}$$

$$\textbf{then}\{\delta_m = \delta_m^M + \Delta_t \quad \boldsymbol{and} \quad b(t + \delta_m + 0) = \boldsymbol{next}[b]\} \tag{3.17}$$

As was mentioned, in Equation (3.17) τ_m is the time instant when the given active buffer began to produce the given part-type items.

Finally, we introduce the demand rates as follow:

$$r_p = \frac{d_p}{T} \qquad \forall p = 1,2, \dots P \tag{3.18}$$

Now, we are can formulate the basic results concerning the problem outlined above:

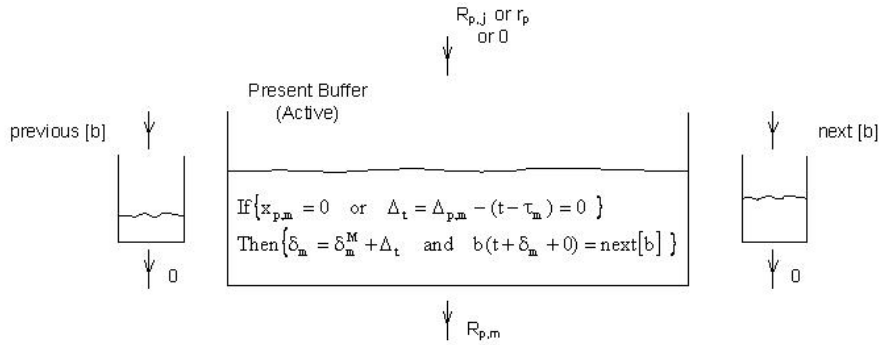
1) The minimal scheduling period T_0 of this system with the production levels d_1; $d_2, \dots\dots d_P$is defined by (3.13).

2) For any $T \geq T_0$ the closed-loop system with the part demand (3.18) and the feedback policy defined by relation (3.17) is regular with the production levelsd_1, $d_2, \dots\dots d_P$ and the scheduling period *T*.

The above results were formulated and proved in [1] (See: Theorem 1).

The given control solution is named **Enforced-Period** switching law.

In Figure 3.2 we demonstrate the above switching law. In the centre there is the present buffer which is active, that is, which produces the sub-part items p by rate $R_{p,m}$. The buffer is supplied from the previous machine-group by rate $R_{p,j}$, or is not supplied because the previous machine-group is not engaged by the production of this sub-part, or is supplied from storage (if the buffer is an input one). At the time corresponding to the situation in Figure 3.2, only $b=b_a$ is an active buffer in machine-group m. When according to the switching law the conditions for a switch are satisfied, the "activity" is transferred to the next buffer.



3.2
The switching law

4 Determination of the Demand Rates for Single-Sections

The orders for production are determined on the MRP (Material Requirement Planning) level of systems. As was mentioned, there are two cases for this:

1. single section scheduling
2. multi-section scheduling

First we deal with the single section case.

The task is to produce during the scheduling time:

$$0 \leq t \leq T_{sch} \tag{4.1}$$

the given number ofitems of the part-types

$$N=N_1, N_2,\ldots\ldots N_p,\ldots\ldots N_P \tag{4.2}$$

As was mentioned, in the classic formulation of tasks, the part numbers are given for the whole scheduling period. It is not specified when exactly in this time window the part items should be produced. The only important point is to have them at the final time instant (due date). Using hybrid dynamical approaches there is a change in this respect. We distribute the part requirement equally along the time axis. We name **demand rate** the rate at which the parts are required by the system. The part **demands** are the integral in time of the demand rates. We also use for this the terminology **arrival rate.** This is the rate the parts arrive into the buffers. Clearly, the demand rate and the arrival rate represent different ideas but are characterized by the same quantities.

Now, let us consider the manufacturing capacities necessary to perform an order. As was mentioned, the machines loads (necessary net manufacturing capacities) are:

$$Lt_m=\sum_{p=1}^{p=P} N_p \tau_{p,m} \quad (4.3)$$

$m \in 1,2,\ldots M$

The global minimum of net production time is determined by the maximum of this quantity. The machine-group for which we have this maximum is the bottleneck machine-group considering the net manufacturing times. As we mentioned, we suppose that the given machine is the bottleneck, taking into consideration the set-up times, too. We identify this machine group by index

$$m_{bt} \epsilon 1,2,\ldots M \quad (4.4)$$

In the following, for the maximum of loading time we will use (see: (3.11))

$$Mtl=MaxLt_m \quad (4.5)$$

$m=1,2,\ldots\ldots M$

In [10] it was proposed to determine the demand rates as

$$r_p = \frac{N_p}{Mtl} k \quad (4.6)$$

$p=1,2,\ldots\ldots P$

The coefficient k was named **demand rate coefficient**.

It was proposed to choose the k coefficient having a value slightly less than one. Now we discuss the reasons of this proposal.According to the description of system processes above, the part demands appear as contents in the buffers at the first machine-groups processing given part-types. So, there is an input buffer, the content of which characterizes the momentary for a part-type requirement. The demand rate choice according to Equation (4.6) is illustrated in Figure 4.1.Clearly, because Mtl is the maximum of net manufacturing time and at the same time the global minimum of production time, it is impossible to finish the overall task,for any part-type, for less time than Mtl. So, if we use equal distribution of demands along the time axis the $\frac{N_p}{Mtl}$ value, will give the slope of the upper border of the "demand sector". Similarly, the $\frac{N_p}{T_{sch}}$ value will provide the lower border because for any value less, the due date requirement may not be satisfied. (There may exist some technological restrictions (see: [10, 11, 12]) which in most of the practical cases do not affect the results.) It is expedient to choose the demand rates as big as possible for decreasing the production time. A k value slightly less than one, which provides some reserve for set-up times, can give a suitable solution for the above goal. The upper and lower borders determine the so-called "demand sector".The mentioned will be discussed later in more detail.

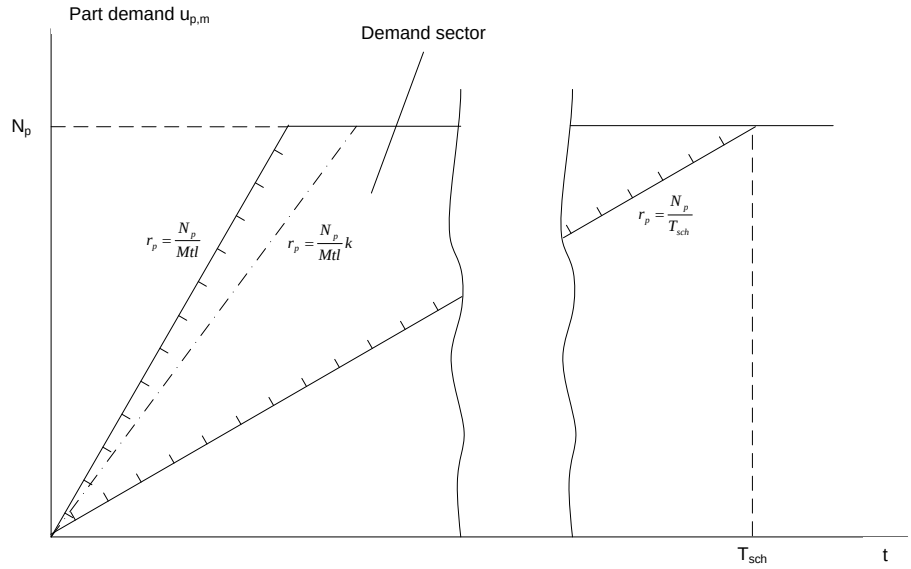


Figure 4.1
The demand sector

Returning to analyze the system processes, the demands appear at the first machine-group processing a given part-type. As was proved,the processes in the systems converge to regular. The processes at the output ofthis first machine-group will converge to periodic ones, too. At the beginning of the system actions, some starting procedure should be realized. (We will discusslater how to start the systems processes.) If the system actions are properly planned, the processes in every buffer will have some starting, transient, then periodic parts, and at the end, some final sections. Accordingly, the production time can be described as

$t_{pr} = t_{start} + t_{periodic} + t_{end}$ (4.7)

We are most interested in the $t_{periodic}$part of the motions.

It has been shown by a number of simulation studies (see: [10, 11, 12]) that the starting transient parts may be made very short. The final section has no significant effect on system performance. So, the system goodness may be characterized quite well considering the processes of the periodic part of the motions.

So, let us suppose that the planning of system processes may be performed based on periodic motion.

We assume that the scheduling task is performed during “W” period of periodic motions where W is a properly chosen integer value.

Then

$$d_p = \frac{N_p}{W} \quad p = 1,2,\ldots..P \tag{4.10}$$

According to Equations (3.13) and (3.18) the arrival rate can be determined as

$$r_p = \frac{N_p}{W\left(k_m \delta_m^o + \sum_{b(p,i)\in B_m} \frac{N_p}{W} \tau_{p,i}\right)} \tag{4.11}$$

Considering Relation (4.5) and (4.6) we get

$$k = \frac{Mtl}{Wk_m \delta_m^0 + Mtl} \tag{4.12}$$

The only free parameter in the above expression is the W value.

We remark that the low values of W exclude the use of hybrid dynamical methods, because, for example, W=1 corresponds to the classic scheduling problem. Its solution has well-known results and difficulties. The other small values of W would indicate the necessity to use classic (or modern) lot- streaming technologies (see e.g.: [20]) which, as far as we know, do not have significant, general results. The development of computational technology makes it possible to solve very sophisticated problems. But the computation difficulties and, what is even more important, the complicated realization makes the use of these approaches not very attractive. The proposed self-organizing approach is free from these difficulties.

Now, we will try to analyze the problem of the proper selection of the W (or k) value.

4.1 Optimal Demand Rate Selection

The production time (t_{pr}) when using HDS methods can be represented as

$$t_{pr} = Mtl + W * k_m \delta_m^0 + \Delta\Gamma \tag{4.13}$$

where $\Delta\Gamma$ is the time of finishing all of the operations on other than the bottleneck machine-group. The task is to choose the W value minimizing the last two terms of (4.13). It depends on the task and may be solved by simulation. To gain some general idea we may have some supposition. In Savkin, Somlo (2009) it was supposed that

$$\Delta\Gamma = T_0 \tag{4.14}$$

By that, the optimal W is (see: Savkin, Somlo (2009)):

$$W_{opt} = \sqrt{\frac{Mtl}{k_m \delta_m^0}} \tag{4.15}$$

According to some new idea about the effect of lot-streaming (see: [21, 22) it seems that a better estimation of production time may be provided by

$$\Delta\Gamma = \frac{H * Mtl}{W} \tag{4.16}$$

where the H coefficient value is

H=0,2÷0,5

The above supposition was concerned with the investigation of the solution of the scheduling problems with full load of bottleneck machines. It was supposed that for "full load" problems, scheduling the production times may be characterized as

$$t_{pr}=Mtl(1+H) + k_m\delta_m^0 \tag{4.17}$$

where H is about the above given values.

We remark that, depending on the task, H may have, in some cases, less value than the above. But these cases are trivial from scheduling points of views. The obtained schedules should be realized in the most usual way. (Lot-streaming should not be used.)

In general, substituting (4.16) into (4.17) we get:

$$W_{opt}=\sqrt{H}\sqrt{\frac{Mtl}{k_m\delta_m^0}} \tag{4.17}$$

The proper H value depends on the tasks. A rather defensive choice is H=0,5. By this

$$W_{opt}=\frac{\sqrt{2}}{2}\sqrt{\frac{Mtl}{k_m\delta_m^0}} \tag{4.18}$$

As the simulation experiments show, the system performance is not very sensitive to W value. So, wide variety "close" to the optimal value may be used.

So,

$$W_{opt}=(0,5\div 1,0)\sqrt{\frac{Mtl}{k_m\delta_m^0}} \tag{4.19}$$

valueseemsa reasonable choice.

Buffer size aspects

Because (see: Equation. (4.10))

$$d_p=\frac{N_p}{W} \qquad p=1,2,\ldots\ldots P \tag{4.20}$$

if at the chosen W, at all of the machine groups, the d_p values are below the physical buffer sizes, then the successful working regimes can be realized. In the

other case W should be increased (or buffer sizes increased). The planning issues are straightforward from the mentioned.

4.2 How to Start the System

Now, let us deal with the problem of how to start the system work. The rough parts come into the system from storages. We have given the way how the part demands are formulated. Clearly, at every machine-group it should be determined which part-type processing should begin first because the feedback control algorithm do not give any information about that at time instant t=0. (There is no previous machine which would give the command "next".) The work starting strategy also affects how the demands are fulfilled because they determine the transient processes. It is very difficult to say anything about the selection of starting input buffers from among the input buffers because the transient processes are highly nonlinear, dynamic ones, and so their parameters are very difficult to estimate. But, according to our simulation experiments (see: [11, 12]), it is not necessary. The transient processes, usually, are very short. Another point in this line is that (as the simulations have shown) the transient parts of the motions can be used for automatic scheduling, as well, without losing anything in quality. Different rules for starting buffers selection may be developed. For example, they may be chosen in decreasing order of demand rates.

Now, let us suppose we have chosen the starting input buffers. We propose the following starting strategy. Let us introduce some starting waiting time value $t_{p,start}$. The parts arrivals begins at time t=0. Then, at

$$t = t_{p,start} + \delta_m^0 \tag{4.21}$$

we propose to begin the processing of the chosen part-type item on all of the machine-groups where it is possible. It is a strategically important decision how to determine the starting waiting time value. With the proper determination of these values, the time of the transient processes may be decreased. It is easy to recognize that if we want to process part-type items in number d_p in the first sub-lot (in the bottleneck machine-group, machining the part-type with label p), the following waiting time value should be applied (see: [10, 11, 12])

$$t_{p,start} = \frac{d_p(1-\tau_{p,m}r_p)}{r_p} - \delta_m^0 \tag{4.22}$$

This is obtained from the relation

$$(t_{p,start} + \delta_m^0 + d_p\tau_{p,m})r_p = d_p \tag{4.23}$$

Then, the output of the first machine-group producing the given part-type will be exactly as at the periodic regime. But because of the dynamic processes in the system, in the following the situation will change. Hopefully, the processes will converge quickly to periodic ones. It seems to us that slightly smaller values than obtained according to (4.22) would result in favorable performance. Concerning

other machine-groups than the bottleneck, the same strategy can be used. A similar Equation to (4.22) can be used but actualized for the given machine-group.

5 Demand Rates Determination in Multi-Section Case

Now, let us consider the general formulation of the FMS scheduling tasks. The order for the production is formulated at the MRP (Material Requirement Planning) level. There are given: the time sections (time windows) of production and, for every section, the types of parts and the corresponding numbers of items to be produced.

That is:

$$re_j \leq t \leq dd_j \tag{5.1}$$

$$j = 1,2, \ldots . J$$

where:

re_j– is the release time from which the production may begin,

dd_j – is the due date,

J– is the number of scheduling sections.

We will identify the part-types as follows:

for j=1 we have the identification index p=1,2,……P_1

for j=2 we have p=P_1+1, P_1+2,……..P_2 (5.2)

•

for j=J we have p=P_{J-1}+1, P_{J-1}+2,……..P_J.

For all of the part-types, the number of part-items to be produced is given.

They are:

$N_1, N_2, N_3, \ldots\ldots\ldots\ldots N_P$ (5.3)

The scheduling sections are overlapping. This is because otherwise the tasks could be solved as outlined above for single (common) scheduling sections problems.

Let us first deal with a simple example, when there are only two sections. In the first section let be produced 2, in the second 3 part-types.

Let $re_1=0$; $re_2<dd_1<dd_2$

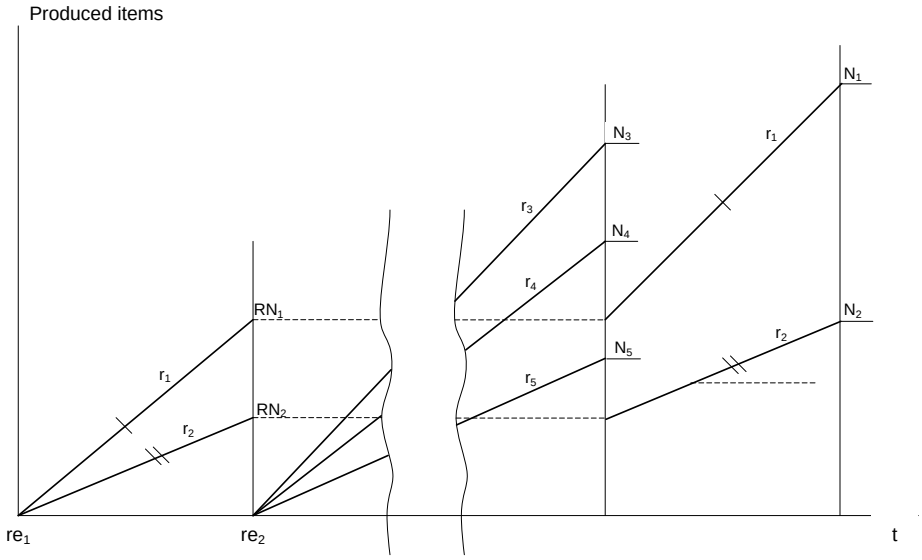


Figure 5.1
Demand rates for the example

We apply the following heuristic approach for demand rate determination. At any production period which begins at time instant re_j we produce first the part-type items scheduled for $re_j \leq t \leq dd_j$. When we arrive to a new release time instant (re_{j+1}), we interrupt to produce the items of the j-th period and begin to produce the new ones. The not finished part-type items of the j-th period will be produced later.

Now, let us consider the simple example. Let us perform for the first section all the planning steps proposed for single-sections.

We use the following symbols for the planning results:

- W_1 - is the number of sub-lots
- Mtl_1-is the global minimum of net manufacturing time for the first section
- r_1, r_2 – are the demand rates
- T_1- is the time-period of the periodic motions

In the case of the example, Mtl_1 is determined as

$$Mtl_1=\text{Max}\{N_1\tau_{1,m} + N_2\tau_{2,m}\} \quad (5.4)$$

m=1,2,…..M

Let us begin the production in a self-organizing, decentralized, real-time controlled manner. When reaching the time instant re_2, we interrupt producing the part-type items of the first period and begin to produce the parts of the second section. For the second section, we determine the quantities outlined above exactly

in the same manner as we did for the first section. The expression, we interrupt producing the part-type items of the first period, means that at $t=re_2$ the input flows are not introduced anymore for the input buffers and the buffer contents of part-types produced in the first period are freeze-in. The reason why we acted in this way is that the value of re_2 shows that having the parts of the second section is certainly more important than finishing the production of the part-types of the first section.

Furthermore, we have, certainly, some reserve in the due date of the task for the first section to be able for this kind of interruption. Below we will give some details about how the due dates effect the above.

So, for the first section we had the planning results:

$$W_1, Mtl_1, r_1, r_2, T_1. \tag{5.5}$$

After the interruption we can determine the planning parameters for the second section

$$W_2, Mtl_2, r_3, r_4, r_5, T_2. \tag{5.6}$$

The system actions we begin in self-organizing manner (see: Figure 3.1) for the first section. Then, at $t=re_2$ the processes of the second section begin. When these terminate the first section items are finished. For these finishing operations the best is to use the demand rates determined before ($re_1$and re_2). It is advisable, inspite of the fact that the freeze-in values are not exactly those which would exactly match the planning conditions. But it is possible, of course, to recalculate the demand rates having the actual parameters (numbers of produced already sub-part-type items).

Due date aspects

Now, let us consider the opportunity to begin the production at a time instant other than re_2. In this case we should use some estimation of the production time value for the second section. The simplest is:

$$t_{pr2}=k_2Mtl_2 \tag{5.7}$$

This value can be slightly less than the real.

Or we can use

$$t_{pr2}=(W_2+1)T_2 \tag{5.8}$$

It might be slightly more than the real production time.

A compromise is also possible (like using Equation (4.17)).

Anyway, a not very bad estimation is possible. Then, if we introduce some starting reserve time (indicated as Res_2) we have:

$$Res_2=dd_2-tpr_2 \tag{5.9}$$

we can begin the second production time section at any time instant (t_{re2})in the interval

$re_2 \leq t_{re2} \leq re_2+Res_2$ (5.10)

In other respects, everything stays the same. That is, the determination of the values r_1, r_2, r_3, r_4, r_5 and other parameters is unchanged.

It is an interesting opportunity that if

$dd_2-tpr_2 \geq re_1+tpr_1$…… …(5.11)

the problem can be separated in to two single-section ones.

Second method of demand rates determination

It is clear that not only the method proposed above but also many other equivalent solutions exist for demand rates determination in multi-section case. One is the following. For the second section of production we do not cancel the production of the series of the part-types items of the first section fully but apply decreased demand rates for both the first, and the second section as well. Indeed, we can consider the production of part-type itemsN_3, N_4, N_5 together with the production of part-type items not finished in the first section LN_1, LN_2. These values can be estimated as:

$LN_1=N_1-r_1(re_2-re_1)$ (5.12)

and

$LN_2=N_2-r_2(re_2-re_1)$ (5.13)

Now, we have a new scheduling task for $t>re_2$ time which can be solved as outlined for single scheduling section problems. We get the results:

$W_2',\ Mtl_2',\ r_1',\ r_{2,}',\ r_{3,}',\ r_{4,}',\ r_{5,}',\ T_2'$ (5.14)

What is outlined above in the case of the example is demonstrated in Figure 5.2.

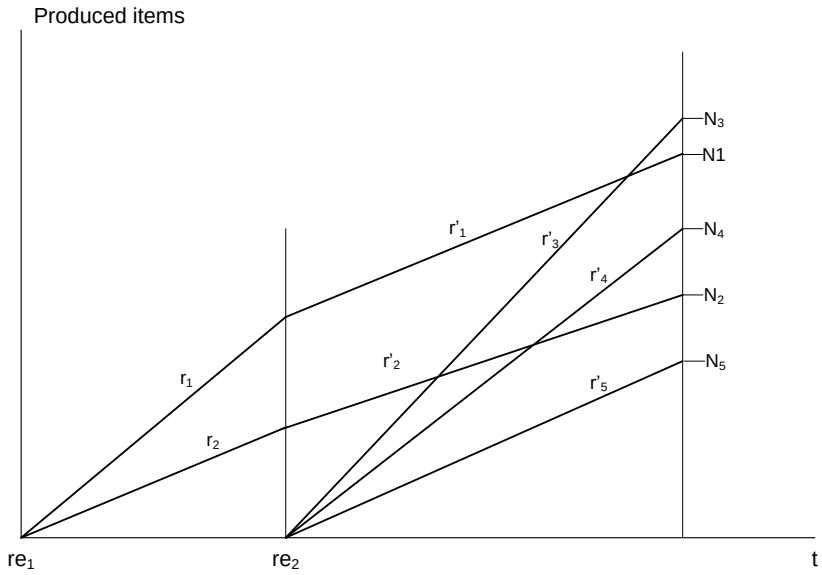


Figure 5.4
Second method of demand rates determination

General solution for demand rates determination

Now let us consider the general case as determined by Relations (5.2), (5.3).

For multi-section scheduling we propose applying the first method described above for demand rates determination.

For the first production section let us take $t=re_1=0$. By this choice we determine:

W_1, Mtl_1, r_1, r_2,……..r_{P_1}, T_1,

Then, for $t=re_2$ we determine

W_2, Mtl_2, r_{P_1+1}, r_{P_1+2}…….. r_{P_2}, T_2,

Going on we may have

……

……

W_J, Mtl_J, $r_{P_{(J-1)+1}}$, …$r_{P_{(J-1)+2}}$,…..r_{P_J}, T_J, (5.15)

Having finished the production in every section, we can estimate the number of any part-type itemswhich were not completed andleft for future production.

$LN_i=N_i-r_i\Delta t_i$ (5.16)

$i\in 1,2,\ldots\ldots\ldots P_j$

where Δt_j is the time interval a part-type was processed during any section.

These values give the simplest estimations but donot differ too much from real numbers.

If a part-type production is interrupted several times, (5.16) is applied several times in the proper sense.

This rather simple approach described above can be used for planning of the processes. In fact, as self-organizing, distributed, real-time control is used, the processes will be determined by the nonlinear dynamics of the systems. The inputs are the demand rates and the starting time values at the starting input buffers. At system level, of course, the data should be up-dated for the different scheduling sections. To understand the nature of the problems, the inter-connection of MRP and scheduling should be analyzed.

6 MRP and Scheduling Interconnections

The production orders come from the MRP systems. MRP is product oriented. By the time a product is assembled, all of the components should be available. MRP allocates the time intervals for production and at the same time checks whether the production capacities are available or not. Every time a new part-type series comes into consideration, MRP assigns the necessary production times to all of the homogeneous capacities (machine-groups). If any production capacity is overloaded (it happens at the bottleneck machine-group), the given task is rejected. Scheduling is production oriented. It allocates the loads necessary to perform the tasks (corresponding to the given orders) to the production capacities.

There is a contradiction among MRP and scheduling. The practical scheduling problems (frequently) may not have an exact solution. So, it may happen that the capacities estimated by MRP are not enough. (Practical scheduling problems (usually) may only be solved exactly with full enumeration (see, for example, French [13]) which is in most of the cases impossible).

To eliminate the above difficulty, unnecessarily big reserves should be provided at MRP level. All this constrains significantly the MRP-scheduling system efficiency.

This difficulty is eliminated when the approach proposed in the present paper is used. The production control provides that the production time is close to the global minimum. This is caused by the automatic lot streaming and overlapping production. This means that a full load strategy may be applied at MRP level.

In Section 5 of the present paper, we analyzed the question of the estimation of the production time. The global minimum of the net manufacturing time is a good basis for this estimation. If a scheduling may be produced resulting close to the

global minimum production times, it fulfils all the expectations. The proposed control solution gives close to the above goal results. Inthe classical approaches, there is no direct contact between scheduling and MRP level. So, on the MRP level, the production times should be highly overestimated. This leads to the law of effective utilization of devices. The hybrid dynamical approach may totally improve this situation.

It is possible to give a formal description of the proposed direct connection of scheduling and MRP, but because the lack of place we will not give it here.

Conclusions

In the paper we outlined **a self-organizing, distributed, real-time scheduling method for Flexible Manufacturing Systems.** This method provides production times very close to the global minimum as the result of automatic lot-streaming and overlapping production. Our earlier investigations have shown that the condition of usability is to have a suitably large number of parts (as minimum $300 \div 400$) in the series, and small set-up times. It seems to us that the sum of the maximum set-up times should be $300 \div 400$ times less than the global mimimum of net manufacturing time to produce all of the part-types in the given number. The above, is based on analytical investigations and simulation studies. **The proposed Enforced-Period-Switching-Law** and by that the **hybrid dynamical feedback control provides stability and regularity**. The first means that for every task it is possible to find buffers with given capacity which will be able to serve the stable work of machine tools (will not overflow). The second means that the processes converge to periodic ones which automatically realize lot streaming and overlapping production. The above results and the proposed control law make it possible to realize self-organizing, distributed, real-time control of flexible manufacturing systems. This is a significant achievement, not only in the respect of quality improvement but also in bringing dramatic simplification in the organization of processes control, too.

The most important achievement of the paper is the proposal for multi-section problems demand rates determination method. For the practical application, only a single planning parameter for every scheduling section should be properly chosen (the number of sub-lots or the demand rates coefficient). The paper details, also, the demand rates determination method for single-section case which is the basis for solving the multi-section problem.

The outlined makes it possible to contact directly FMS scheduling and MRP.

Acknowledgement

The research reported in the present paper was stimulated by the outstanding theoretical results of professor A. V. Savkin (UNSW, Sydney). Authors express sincere thanks for the opportunity of cooperation with him. Authors are grateful to Olesya Ogorodnikova and Akos Antal for the help in formulation of the results and the preparation of the paper.

References

[1] Savkin A. V., Somlo J. Optimal Distributed Real-Time Scheduling of Flexible Manufacturing Networks Modeled as Hybrid Dynamical Systems. Robotics and Computer- Integrated Manufacturing. V. 25, Issue 3, June 2009, pp. 597-609

[2] VanderSchaftA,SchumacherH.An Introduction to Hybrid Dynamical Systems. London: Springer; 1999

[3] Matveev A. S., Savkin A. V. Qualitative Theoryof Hybrid Dynamical Systems. Boston: Birkhauser; 2000

[4] Perkins J. R. Kumar P. R. Stable, Distributed, Real-Time Scheduling of Flexible Manufacturing/Assembly/Disassembly Systems. IEEE Trans Automat Control 1989; 34(2):139-48

[5] PerkinsJR,Humes C. J. , Kumar P. R. Distributed Scheduling of Flexible Manufacturing Systems: Stability and Performance. IEEE Trans Robotics Automat1994;10(4):133-41

[6] Savkin A. V. Regularizability of Complex Switched Server Queueing Networks Modelled as Hybrid Dynamical Systems. Systems Control Letters 1998;35(5):291-300

[7] Savkin A. V., Matveev A. S. Cyclic Linear Differential Automata: A Simple Class of Hybrid Dynamical Systems. Automatica2000; 36(5):727-34

[8] Savkin A. V. , Matveev A. S. A Switched Single Server System of Order n with All Its Trajectories Converging to (n-1)! Limit Cycles. Automatica 2001; 37(2):303-6

[9] Savkin A. V. Evans R. J. Hybrid Dynamical Systems. Controller and Sensor Switching Problems. Boston. Birkhauser, 2002

[10] Somlo J. Hybrid Dynamical Approach Makes FMS Scheduling More Effective. PeriodicaPolitechnica Ser. Mech. Eng. 2001;45(2):175-200

[11] Somlo J., Savkin A. V., Anufriev A., Koncz T. Pragmatic Aspects of the Solution of FMS Scheduling Problems Using Hybrid Dynamical Approach. Robotics and Computer Integrated Manufacturing 2004;20:35-47

[12] Somlo J. Savkin A. V. Periodic and Transient Switched Server Schedules for FMS. Robotics and Computer Integrated Manufacturing. 2006. 22;93-112

[13] Castillo I., Smith J. S. Formal Modeling Methodologies for Control of Manufacturing Cells: Survey and Comparison. J. Manuf. Systems 2002;21(1):40-57

[14] .Coffman E. G. Computer and Job-Shop Scheduling Theory. NewYork:Wiley, 1976

[15] French S. Sequencing and Scheduling: an Introduction to the Mathematics of the Job-Shop. NewYork:Wiley;1982

[16] Dai J. G., Weiss G. A Fluid Heuristic for Minimizing Make Job Shops. Oper. Res. 2002;50(4):692-707

[17] Kumar P. R., Seidman T. I. Dynamic Instabilities and Stabilization Methods in Distributed Real-Time Scheduling of Manufacturing Systems. IEEE Trans. Automat. Control.1990;35(3):289-98

[18] Rybko A. N. ,Stolyar A.L. Ergodicity of Stochastic Processes Describing the Operations of Open Queueing Networks. ProblemyPeredachi Inf. 1992;28:2-26

[19] Dai J. G., Hasenbein J. J., VandeVate J. H. Stability of a Three-Station Fluid Network. Queueing Systems.1999;33:293-325

[20] Dauzere-Peres S, Lasserre J. B. An Iterative Procedure for Lot Streaming in Job-Shop scheduling.Comput. Ind .Eng. 1993;25(1-4):231-4

[21] Somlo J, Kodeekha E. Optimal Lot Streaming for a Class of FMS Scheduling Problems – Joinable Schedule Approach. Proc. IFAC CEFIS'07 Symposium. Oct. 9-11, 2007, Istambul, Turkey

[22] Somlo J., Kodeekha E. Improving FMS Scheduling by Lot Streaming. Journal PeriodicaPolytechnika. Ser. Mechanical Engineering. 2007, 51/1, 3-14, Budapest, Hungary