A Time-Decomposition Method for Sequence-Dependent Setup Scheduling under Pressing Demand Conditions

Eugene Khmelnitsky, Konstantin Kogan, and Oded Maimon

***Abstract*—This paper develops a method for continuous-time scheduling problems in flexible manufacturing systems. The objective is to find the optimal schedule subject to different production constraints: precedence constraints (bills of materials), sequence-dependent setup times, finite machine capacities, and pressing demands. Differential equations along with mixed constraints are used to model production and setup processes in a canonical form of optimal control. The proposed approach to the search for the optimal solution is based on the maximum principle analysis and time-decomposition methodology. To develop fast near-optimal solution algorithms for sizable problems, we replace the general problem with a number of subproblems so that solving them iteratively provides tight lower and upper estimates of the optimal solution.**

I. Introduction

CONSIDERABLE effort is spent on the complicated task of scheduling in flexible manufacturing systems. Scheduling remains among the hardest optimization problems, a difficulty that increases further when sequence-dependent setup times are introduced. Due to the limitations of some technological operations and machines, such setups are frequently found in different industries. The manufacture of semiconductor lines, for example, involves lithography of different layers of wafers, incurring layer-dependent changeover times. Such dependencies are found in the food industry (especially in blending and packaging operations), as well as in the consumer goods and materials fabrication industries.

Despite the well-known combinatorial explosion, some optimization techniques have been developed to cope with static models of the scheduling problem. A mixed integer linear programming model for sequence-dependent scheduling problems and a recursive technique were proposed in [1]. However, only small-scale production problems are in the range of existing mixed integer linear programming algorithms.

Deane and White in [2] developed a specialized branch-and-bound algorithm for similar problems and restrictions, in which there are parallel production lines with identical machines, given lot sizes and a work-loading constraint on the maximum makespan deviation of each line from the average.

The quadratic assignment problem approach [3], [4] and the aforementioned optimization approaches still suffer too much from the curse of dimensionality, even if the best available methods for obtaining optimal or near optimal solutions of such problems are applied. This was the case when approximation was accomplished, for instance, by a linear programming adjustment in a quadratic assignment algorithm [4]. Following a mixed integer formulation of the joint problem of lot sizing and scheduling, six initial setup cost estimators were suggested and evaluated to reduce the overall costs in various production conditions [5]. A two-phase heuristic approach, which can be used as a suboptimization method, was proposed for the simple single-stage production system with sequence-dependent setup times [6]. This approach is based on simulated annealing and a myopic rule to determine production schedules.

The other approach is to reduce the number of variables by considering dynamic continuoustime models of the scheduling problem with the aid of optimal control theory. Kimemia and Gershwin first presented a flexible manufacturing system as a continuous-time product flow passing through work stations (where setups are negligible) and buffers [7]. They described this by differential equations. The optimal flow was found from the linear problem formulated at required points of time by varying a cost functional. Since the method loses its efficiency when multilevel bills of materials and significant setup effects are added to the model, the approach commonly adopted is to decompose hierarchically the entire problem into a number of tractable ones.

At the first hierarchical level, target production rates are defined for the given demand profile (Kimemia and Gershwin's problem). The input to the next level comes from the previous one and includes production rate targets that have to be tracked as closely as possible by scheduling machine setups. However, for the approach to be useful for real-time control, tracking policies should be simple and of a distributed type [8]–[10]. When dealing with distributed tracking policies, an important issue is that of stability, i.e., whether the work-in-process in the system remains under control. Realistic manufacturing based on multi-level bills of materials and multiple machines necessitates that all machines be synchronized for stable production. Machine stability is unlikely to be achieved by such policies [11].

An alternative to the hierarchical approach is to state a generalized optimal control problem where deterministic production

Manuscript received May 30, 1997; revised July 7, 1998 and February 1, 1999. Recommended by Associate Editor, K. Rudie.

E. Khmelnitsky and K. Kogan are with the Faculty of Engineering, Department of Industrial Engineering, Tel-Aviv University, Ramat-Aviv, Tel-Aviv 69978 Israel (e-mail: xmel@eng.tau.ac.il.).

O. Maimon is with the College of Engineering, Department of Manufacturing Engineering, Boston University, Boston, MA 02215 USA (e-mail: maimon@enga.bu.edu).

Publisher Item Identifier S 0018-9286(00)03236-0.

and setup changes are considered simultaneously and synchronized at the same level. Then, stability is ensured by an optimization procedure, despite significant setup times and multi-level production structure. This direction extends Kimemia and Gershwin's approach only for the case of completely reliable machines. Although deterministic control problems do not directly yield the solution of stochastic control problems, they can serve to characterize the behavior of optimal policies and approximate the solution of a stochastic scheduling problem [12]. Therefore, the deterministic consideration could be viewed as a first step to solving complex scheduling problems in a stochastic environment. To implement the approach, a sequence-independent setup was viewed as a continuous process conflicting with the production process. Both processes were then described by differential equations with controllable production and setup rates. Analysis of the maximum principle for the problem yielded analytical properties of the optimal solution which became the basis of iterational numerical methods [13], [14].

The present paper proceeds in the latter direction to study complex dynamic scheduling problems.

- Setup and production processes are modeled and accounted for in optimization in an equal manner (in contrast to the standard decomposition approach where setups are heuristically inserted at a lower hierarchical level).
- The scheduling model incorporates now the most general, sequence-dependent setups.
- A new method is developed to capture large-scale flexible manufacturing systems.
- Lower and upper bounds are derived to estimate the quality of the obtained solutions.

Since the setups are allowed to be sequence-dependent, the numerical methods suggested so far become computationally intractable even for small-scale flexible manufacturing systems. The method developed in this paper is based on a cyclic replacement of the original problem with a number of reduced problems, which can be effectively solved by gradient time decomposition procedures [15], [16] in large-scale systems, where the straightforward time decomposition [14] and shooting methods [13] fail. Moreover, when demand is pressing, tight upper and lower bounds of the optimal solution are obtained, and the method is proven to provide the solution which converges to the optimal of the original problem.

II. Statement of the General Problem

We consider a flexible manufacturing cell operating on a number of products I (including raw materials, subproducts, assemblies, disassemblies, and end products). The cell consists of K machines, each of which may be in one of $J(k)$, $k = 1, \cdots, K$ different states at a time. Each state of a machine is a technological operation carried out by the machine. For each machine k and its state $j, j = 1, \cdots, J(k)$, the operation is modeled by vector v_{ikj} which designates the capacity of the machine when producing or consuming (in this case v_{ikj} is negative) product $i = 1, \cdots, I$. Thus, every technological operation j corresponds to a subset of the produced and consumed products according to the bill of materials selected for manufacturing. In order to switch over a machine from one state j onto another, j', a setup must be carried out. The setup time $T_{kjj'}$ is sequence-dependent.

To formalize the production process, we shall operate with production rates $w_{kj}(t)$ of machine k in state j, defined relative to the capacity v_{ikj} of the machine for every product $i, i = 1, \cdots, I$. Therefore, the production rate control variable $w_{kj}(t) \in [0, 1]$ reflects the portion of the machine capacity utilized at time t. At the same time, the setup process in the flexible cell will be controlled by the setup rate $u_{kjj'}(t)$, where $j < j'$ and $0 \le u_{kjj'} \le 1/T_{kjj'}$ if the machine is set up from state j onto j', and $0 \ge u_{kjj'} \ge -(1/T_{kj'j})$ if the machine is set up from state j' onto j. Note, by defining the first state index in $u_{kjj'}(t)$ to be always less than the second one, we decrease by half the number of setup control variables associated with machine k, which in our statement is equal to $J(k)(J(k)-1)/2$.

We model the flow of product i through its buffer of current level $X_i(t)$ by the difference between the cumulative production of the product in the cell and its demand $d_i(t)$

$$\dot{X}_i(t) = \sum_{kj} w_{kj}(t) v_{ikj} - d_i(t). \tag{1}$$

The setup process is modeled by the dynamic transformation of state variable $V_{kj}(t)$. State variable $V_{kj}(t)$ is equal to one when machine k is in state j; $V_{kj}(t)$ is equal to zero when machine k is not in state j; and it is in the range between zero and one if the machine is currently being set up either from or onto state j

$$\dot{V}_{kj}(t) = \sum_{j>j'} u_{kj'j}(t) - \sum_{j'>j} u_{kjj'}(t), \qquad V_{kj}(t) \ge 0. \tag{2}$$

The boundary scheduling conditions for (1) and (2) are

$$X_i(0) = X_i^0, \quad V_{kj}(0) = V_{kj}^0, \quad \sum_j V_{kj}^0 = 1. \tag{3}$$

Both constraints (2) and (3) ensure that each machine carries out no more than one technological operation at a time.

Natural capacity constraints are imposed on the control variables

$$w_{kj}(t) \ge 0, \qquad -\frac{1}{T_{kj'j}} \le u_{kjj'}(t) \le \frac{1}{T_{kjj'}}. \tag{4}$$

Note that we here extend the setup concept. The setup rate is commonly fixed, while we consider its fixed value as the maximal one and allow the setup to be carried out at a slower rate. Such slow setup of a machine is proved in the next section to be equivalent to the regular setup followed by idling the machine. Thus, as is the case in industry, slower setups never occur on optimal trajectories. At the same time, the new concept of setup allows the problem to be stated in the canonical form of optimal control which, in turn, allows the optimal behavior of the production system to be studied analytically.

Since production is impossible during setup process, a special restriction is imposed on the two conflicting processes

$$w_{kj}(t) \le \Theta(V_{kj}(t) - 1) \tag{5}$$

TABLE I
CAPACITIES OF THE MACHINES (PRODUCT UNITS PER DAY)

	mellow, j=1	sustain, j=2	perk, j=3	turbo, j=4
Blender A, k=1	75	75	75	75
Blender B, k=2	60	60	60	-
Blender C, k=3	75	75	-	75
Packager A, k=4	150	125	175	150
Packager B, k=5	75	60	90	75

where $\Theta(V)$ is a unit step-function ($\Theta(V) = 0$ if $V < 0$ and $\Theta(V) = 1$ if $V \geq 0$).

The performance measures most often used to judge the productivity of manufacturing lines are work-in-process (WIP) and the degree to which the line meets the demand profile. Therefore, our objective function includes penalties for both the difference between buffer levels of WIP (subproducts) and safety stock levels, and for demand violation (end products) along a given planning horizon T

$$\frac{1}{2}\int_0^T \sum_i p_i(X_i(t))\,(X_i(t) - X_i^s)^2\, dt \rightarrow \min \tag{6}$$

where X_i^s are the given levels of safety stocks and $p_i(X_i(t)) = p_i^1$ if $X_i(t) \geq X_i^s$ and $p_i(X_i(t)) = p_i^2$ otherwise. Penalty coefficients p_i^1 and p_i^2 present, respectively, buffer-carrying costs when $X_i(t) \geq X_i^s$, and penalties for shortages (stockout), when $X_i(t) < X_i^s$ [17].

Example: Prior to illustrating the stated problem, we have to emphasize here the importance of a proper choice of the planning horizon T. Specifically, we see at least two reasons not to schedule the system over very long time horizons. First, different long-term factors of the production environment, such as machine breakdowns, customer orders and costs, are unlikely to remain unchanged. Second, the computational burden increases as the planning horizon grows. When setting the time scale, it is to be related to the setup times featuring the studied manufacturing system. For example, the planning horizon can be designated at no more than 100 times minimal setup time in the system. On the other hand, when the horizon is too short, so that it is comparable with the maximal setup time in the system, the resultant scheduling might not make any practical sense. Therefore, it may be designated at no less than ten times maximal setup time.

We consider a flexible cell for coffee production which operates on three raw materials (light, medium, and dark beans); four coffee mixtures according to preset recipes (mellow mix: 80% light and 20% medium beans; sustain mix: 50% light and 50% medium beans; perk mix: 10% light, 20% medium and 70% dark beans; turbo mix: 20% medium and 80% dark beans); and four end products, one for each coffee mix (mellow, sustain, perk, and turbo brands).

There are two stages of the production process, namely blending and packaging, that are necessary to produce the end coffee products. The cell has a total of five machines: blender A for all sorts of mixtures, blender B for only mellow, sustain, and perk mixes, blender C for mellow, sustain, and turbo, and two packagers A and B for all brands.

The structure of the modeled cell is shown in Fig. 1, where blender $B(k = 2)$ has three states and currently is in state $j = 1$ (darkened area) where it mixes light $i = 9$ and medium $i = 10$ beans to produce mellow mixture $i = 5$.

According to Table I which presents machine capacities, the mellow mixture can be produced by blender B no faster than 60 units/day. Thus, the maximal capacity of mellow mixture $i = 5$ production on blender $B(k = 2)$ in state $j = 1$ is $v_{521} = 60$ units/day, while the maximal consumption rates of the light $i = 9$ and medium $i = 10$ beans are $v_{921} = -0.8 \cdot 60 = -48$ units/day and $v_{1021} = -0.2 \cdot 60 = -12$ units/day. The negative values of v_{ikj} correspond to the materials needed to produce the parent item while the positive value is assigned to the latter. All the other parameters in v_{i21} are set at zero because they are neither produced nor consumed in this blending operation.

Since the sustain mixture $i = 6$ can be produced by all blenders, the flow of this mixture through its buffer is described by the following differential equation:

$$\begin{aligned}\dot{X}_6(t) = {} & 75w_{12}(t) + 60w_{22}(t) + 75w_{32}(t) \\ & - 125w_{42}(t) - 60w_{52}(t)\end{aligned}$$

where the first three terms of the right-hand side correspond to the blenders which produce the sustain mixture and fill in its buffer, while the remaining terms are for the packagers which consume the mixture and empty the buffer. The other ten production equations are formalized similarly.

As shown in Fig. 1, blender $A(k = 1)$ is currently set up (no darkened area). The setup equation (2) and setup rate bounds presented in Table II take the following form for state $j = 2$ of blender A:

$$\begin{aligned}&\dot{V}_{12}(t) = u_{112}(t) - u_{123}(t) - u_{124}(t) \\ &-\frac{1}{0.9} \leq u_{112}(t) \leq \frac{1}{1.5}, \quad -\frac{1}{2.0} \leq u_{123}(t) \leq \frac{1}{1.5} \\ &-\frac{1}{1.5} \leq u_{124}(t) \leq \frac{1}{1.0}.\end{aligned}$$

Here, for example, $u_{112}(t)$ corresponds to the setup rate between states $j = 1$ and $j = 2$. The value of $u_{112}(t)$ is positive when the blender is set up from $j = 1$ onto $j = 2$ and is negative when the setup is carried out in the opposite direction. The

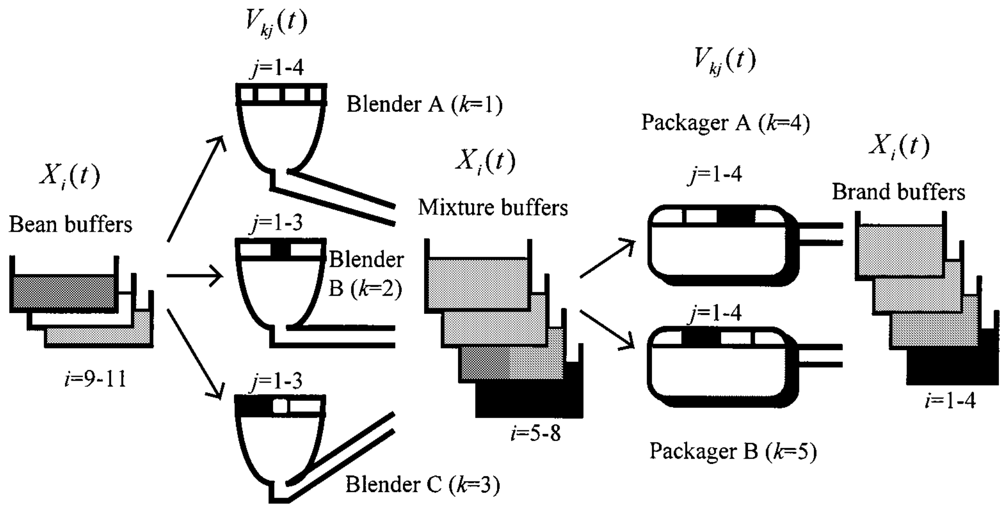

Fig. 1. Structural model of a coffee production cell.

planning horizon T is chosen to be equal to 25 days, which lies in the rational range discussed above.

III. Analysis of the System's Extremal Behavior

In order to obtain analytical properties of the optimal solution, we first approximate the step-function $\Theta(V-1)$ in (5) by a sequence of continuous functions $f_n(V)$ which will prohibit production during setups as n tends to infinity [13]. The maximum principle [18], [19] can be applied to study the problem, since we have chosen state variables (X, V) as absolutely continuous, control variables (w, u) as measurable bounded functions, and functions specifying differential equations and constraints (1)–(6) are continuous with respect to the state and control variables.

A. The Maximum Principle Formulation

The maximum principle asserts that there exist adjoint variables which are left-continuous functions of bounded variation $\psi_{kj}^V(t)$, $\mu_{kj}(t)$, piecewise absolutely continuous functions $\psi_i^X(t)$, and measurable functions $a_{kj}(t)$. The primal and adjoint variables satisfy the following conditions:

- *Nonnegativity*

$$a_{kj}(t) \geq 0, \qquad d\mu_{kj}(t) \geq 0. \tag{7}$$

- *The dual system and transversality conditions*

$$\dot{\psi}_i^X(t) = p_i(X_i(t))\,(X_i(t) - X_i^s);$$
$$\psi_i^X(T) = 0; \tag{8}$$

$$d\psi_{kj}^V(t) = -a_{kj}(t) f_n'(V_{kj}(t))\,dt - d\mu_{kj}(t);$$
$$\psi_{kj}^V(T+0) = 0; \tag{9}$$

where $a_{kj}(t) = \sum_i \psi_i^X(t) v_{ikj}$.

Note that mixed constraint (5) now takes the following form:

$$0 \leq w_{kj}(t) \leq f_n(V_{kj}(t))$$

and becomes irregular at point $V_{kj}(t) = 0$, i.e., by varying only control variable $w_{kj}(t)$ it is impossible to move this point inside the area defined by this constraint. This irregularity causes the measures $d\mu_{kj}(t)$ to appear in (9).

- *The global maximum principle*

The optimal control strategy is achieved by maximizing, for each t, the Hamiltonian

$$H = -\frac{1}{2}\sum_i p_i(X_i(t))\,(X_i(t) - X_i^s)^2 + \sum_{kj} \psi_{kj}^V(t) \left(\sum_{j>j'} u_{kj'j}(t) - \sum_{j'>j} u_{kjj'}(t) \right) + \sum_i \psi_i^X(t) \left(\sum_{kj} w_{kj}(t) v_{ikj} - d_i(t) \right) \tag{10}$$

subject to (4), (5).

- *Complementary slackness*

$$a_{kj}(t)(w_{kj}(t) - f_n(V_{kj}(t))) \equiv 0;$$
$$\int_0^T V_{kj}(t)\,d\mu_{kj}(t) = 0. \tag{11}$$

The economic interpretation of the adjoint variables is similar to the well-known "shadow price" interpretation and is discussed in [20]. Specifically, $\psi_i^X(t)\partial X$ is the change, accumulated to t, in the maximum attainable value of the objective function when $X(t)$ is increased by ∂X.

TABLE II
PRODUCT CHARACTERISTICS (BRANDS: 1-MELLOW, 2-SUSTAIN, 3-PERK, 4-TURBO; MIXTURES: 5-MELLOW, 6-SUSTAIN, 7-PERK, 8-TURBO; BEANS: 9-LIGHT, 10-MEDIUM, 11-DARK)

product, i	1	2	3	4	5	6	7	8	9	10	11
cost p_i^1 per square product unit per day, 10^{-3}	3	3	3	3	2	2	2	2	0	0	0
cost p_i^2 per square product unit per day, 10^{-1}	1	1	1	1	1	1	1	1	0	0	0
initial inventory, X_i^0	200	200	200	200	100	100	100	100	900	900	900
safety level, X_i^s	0	0	0	0	0	0	0	0	0	0	0

B. Optimal Production and Setup Regimes

The following three lemmas study the extreme behavior of the system with the aid of the maximum principle. The proofs of the lemmas are relegated to the Appendix.

Lemma 1: Given problem (1)–(6), on the optimal solution of the problem, the production rate of a machine k in state j is defined as:

i) $w_{kj}(t) = f_n(V_{kj}(t))$, when $\sum_i \psi_i^X(t)v_{ikj} > 0$ (full production regime);
ii) $w_{kj}(t) \in [0, f_n(V_{kj}(t))]$, when $\sum_i \psi_i^X(t)v_{ikj} = 0$ (underproduction regime);
iii) $w_{kj}(t) = 0$, when $\sum_i \psi_i^X(t)v_{ikj} < 0$ (idle regime). ■

Lemma 1 determines uniquely optimal production rates for the standard regimes i) and iii) and a switching surface for the singular regime ii) in terms of the dual variables. The unique optimal production rate for the singular regime is derived in the following corollary.

Corollary 1.1: Let machine k be the only machine which is on regime ii) in interval of time $[t_1, t_2]$ and

$$0 \le \sum_i b_{ikj}d_i(t) + c_{kj} \le 1, \quad \text{where } b_{ikj} = \frac{p_i v_{ikj}}{\sum_i p_i v_{ikj}^2}$$

and

$$c_{kj} = -\sum_i b_{ikj} \sum_{k'j' \neq kj} w_{k'j'}(t)v_{ik'j'}$$

then

$$w_{kj}(t) = \sum_i d_{ikj}d_i(t) + c_{kj}.$$ ■

Evidently, if several machines are on regime ii) simultaneously, then production rates for these machines are determined in the same way.

The above lemma with its corollary sustains an expected extremal behavior of the system during the production process, namely, either machine k is fully loaded [regime i)] or idle [regime iii)], or it is underloaded [regime ii)]. Below, nontrivial conditions for optimality of the setup process are derived.

Lemma 2: Given problem (1)–(6), on the optimal solution of the problem, the setup rate of a machine k, which is switched between states j and j', is defined as:

i) $u_{kjj'}(t) = 1/T_{kjj'}$, when $\psi_{kj'}^V(t) > \psi_{kj}^V(t)$ (setup from j onto j');
ii) $u_{kjj'}(t) = -(1/T_{kj'j})$, when $\psi_{kj'}^V(t) < \psi_{kj}^V(t)$ (setup from j' onto j);
iii) $u_{kjj'}(t) \in [-(1/T_{kj'j}), (1/T_{kjj'})]$, when $\psi_{kj'}^V(t) = \psi_{kj}^V(t)$ (singular setup regime between states j and j'). ■

Corollary 2.1: Given problem (1)–(6), on the optimal solution of the problem, singular setup regimes iii) can be replaced with either regime i) or regime ii) followed by no-setup regime:

iv) $u_{kjj'}(t) = 0$, when $\psi_{kj'}^V(t) = \psi_{kj}^V(t)$. ■

Lemma 2 and its corollary show that although the setup process is continuously controllable in our model, there always exists an optimal solution where the setup rates take on only their boundary values. The following lemma finally formalizes the conditions when a machine can be set up, i.e., the lemma determines a point of time and a new machine state when and where the machine can be changed over.

Lemma 3 (The Necessary Setup Conditions): On the optimal solution, if a machine k is being set up from state j onto state j' on interval $[t_s, t_s + t]$, $t \ge T_{kjj'}$, then

$$\sum_i \psi_i^X(t_s)v_{ikj} = \sum_i \psi_i^X(t_s + t)v_{ikj'}. \tag{12}$$ ■

Thus, the optimal scheduling trajectory consists of a sequence of setup and production regimes, which are changed over only at moments t_s satisfying (12).

Example (Continued): The necessary condition for a setup from state $j = 1$ onto state $j = 2$ of blender $A(k = 1)$ in our coffee example is:

$$\begin{aligned} 75\psi_5^X(t_s) - 0.8 \cdot 75\psi_9^X(t_s) - 0.2 \cdot 75\psi_{10}^X(t_s) \\ = 75\psi_6^X(t_s + t) - 0.5 \cdot 75\psi_9^X(t_s + t) \\ - 0.5 \cdot 75\psi_{10}^X(t_s + t) \end{aligned}$$

where $t \ge 1.5$. With respect to the product characteristics shown in Table III, the dual differential equations for determining, for example, $\psi_5^X(t)$ are

$$\dot{\psi}_5^X(t) = 2.10^{-3}X_5(t), \qquad \text{if } X_5(t) \ge 0,$$

and

$$\dot{\psi}_5^X(t) = 5.10^{-2}X_5(t), \qquad \text{if } X_5(t) < 0.$$

The analytical properties proven in the above lemmas and illustrated in the example allow shooting-based numerical methods to be constructed for solving the primal and dual systems (1), (2), (8), and (9) as a two-point boundary value problem. However, such methods have proved to be efficient only for small-scale production systems [17]. In fact, for

TABLE III
SETUP TIMES (DAYS)

	T_{12}	T_{13}	T_{14}	T_{21}	T_{23}	T_{24}	T_{31}	T_{32}	T_{34}	T_{41}	T_{42}	T_{43}
Blender A	0.5	0.3	1.8	0.9	0.5	1.8	0.4	2.0	0.7	1.0	1.5	1.0
Blender B	0.5	0.4	-	0.7	0.7	-	1.5	1.0	-	-	-	-
Blender C	1.0	-	0.5	1.0	-	0.3	-	-	-	0.5	0.3	-
Packager A	0.8	0.5	0.3	0.7	1.0	1.0	0.4	1.0	0.3	0.3	0.7	0.5
Packager B	1.0	0.5	0.5	1.0	1.0	0.5	0.5	1.0	0.3	1.0	0.7	0.5

scheduling problems with dynamic demands, solving the corresponding two-point boundary-value problem becomes strongly exponential in the number of control variables even for short planning horizons.

When no irregular constraint is imposed on the system, the problem of dimensionality can be overcome by applying an alternative time–decomposition approach, which is of polynomial complexity when solution accuracy is given. Since the general problem does contain irregular constraints (5), we next decompose the general problem into three subproblems which, when subsequently solved, yield the optimal solution for pressing demands.

IV. SEPARATE OPTIMIZATION OF PRODUCTION AND SETUP PROCESSES OF THE GENERAL PROBLEM

To solve the general problem, we first assume that the machines making up the system are fully loaded and then find the optimal sequence of machine changeovers (sequencing problem). Secondly, we relax the full loading requirement and find optimal loading and timing of the machines for the given setup sequence (loading and timing problems). Eventually, we prove that, under pressing demand, the solution of the resultant problems coincides with the solution of the general one.

Sequencing Problem: This problem deals with setup scheduling by varying setup rates under the full loading of the machines in the system (i.e., constraint (5) takes the form of equality: $w_{kj}(t) = \Theta(V_{kj}(t) - 1)$. Thus, the system is controlled by only setup rates $u_{kjj'}(t)$ and the problem is formulated as follows:

$$\frac{1}{2}\int_0^T \sum_i p_i(X_i(t))\left(X_i(t) - X_i^2\right)^2 \, dt \to \min$$

subject to

$$\begin{aligned}
&\dot{X}_i(t) = \sum_{kj} \Theta(V_{kj}(t) - 1)v_{ikj} - d_i(t)\\
&\dot{V}_{kj}(t) = \sum_{j>j'} u_{kj'j}(t) - \sum_{j'>j} u_{kjj'}(t), \qquad V_{kj}(t) \ge 0\\
&X_i(0) = X_i^0, \quad \sum_j V_{kj}^0 = 1, \quad -\frac{1}{T_{kj'j}} \le u_{kjj'}(t) \le \frac{1}{T_{kjj'}}.
\end{aligned}$$

Loading Problem: This problem deals with the optimization of the production process by varying machine rates under a given sequence of setups [functions $V_{kj}(t)$]. The problem formulation follows:

$$\frac{1}{2}\int_0^T \sum_i p_i(X_i(t))\left(X_i(t) - X_i^s\right)^2 \, dt \to \min$$

subject to

$$\begin{aligned}
&\dot{X}_i(t) = \sum_{kj} w_{kj}(t)v_{ikj} - d_i(t)\\
&X_i(0) = X_i^0, \qquad 0 \le w_{kj}(t) \le \Theta(V_{kj}(t) - 1).
\end{aligned}$$

Timing Problem: This problem deals with finding moments of time t_s (setup starts) which satisfy the necessary setup conditions (12) (see Lemma 3) for a given set of dual variables $\psi_i^X(t)$

$$\sum_i \psi_i^X(t_s)v_{ikj} = \sum_i \psi_i^X(t_s + T_{kjj'})v_{ikj'}.$$

The suggested decomposition leads to the problems which are computationally tractable for large-scale systems. Indeed, the general problem includes irregular mixed constraints (5) imposed on both state $V_{kj}(t)$ and control $w_{kj}(t)$ variables. These constraints prevent an efficient utilization of the time–decomposition methods. At the same time, mixed constraints do not appear in any of the three problems. Below, in Lemma 5, we also eliminate the state constraints from the sequencing problem in order to further extend the scope of practical applications.

Prior to proving that the solution of the decomposed problems is the optimal solution for the general problem, we formalize the requirement of pressing demand.

Definition: Demand is pressing if the solution of the general and sequencing problems defines the same setup sequence and durations.

Although our definition of pressing demand is based on the time–decomposition solution methodology, it implies a generalization of the well-known case when the production system is overloaded, i.e., when the demand is close to or exceeds the system's capacity along the planning horizon. An *a priori* condition sufficient for the demand to be pressing, which takes into account not only the machine capacities but also the cost relationships, technology (bill of materials), and initial buffer levels are given by the following lemma.

Lemma 4: For the demand to be pressing, it is sufficient that

$$\sum_i v_{ikj} \int_t^T \int_0^\tau d_i(y)\,dy\,d\tau > \sum_i v_{ikj}\alpha_i(t), \qquad \forall t, k, j. \tag{13}$$

where

$$\alpha_i(t) = \min_{m=1,2} \left(p_i^m \left(X_i^0 - X_i^s\right)(T-t) + \int_t^T \sum_{kj, v_{ikj}<0} v_{ikj}\tau\,d\tau \right). \quad \blacksquare$$

To formulate the next theorem, we introduce the following notation:

$(X_i(t), V_{kj}(t), u_{kj}(t), \psi_i^X(t)$, and $\psi_{kj}^V(t))$—solution of the sequencing problem;

$(X_i^*(t), V_{kj}(t), w_{kj}(t)$, and $\psi_i^{X^*}(t))$—solution of the loading problem;

$(X_i^{**}(t), V_{kj}^*(t), w_{kj}^*(t), \psi_i^{X^{**}}(t))$—solution of both loading and timing problems as described above in this section.

Since the iterations of the loading problem do not change the setup sequence, the functions $V_{kj}(t)$ are common for both sequencing and loading problem solutions. At the same time, the iterations of the timing problem do change the location of setups and loading of the machines. Therefore, an additional asterisk appears in the solution functions.

Theorem 1: Given pressing demand and the optimal solution for the sequencing problem $(X_i(t), V_{kj}(t), u_{kj}(t), \psi_i^X(t)$, and $\psi_{kj}^V(t))$, if $(X_i^*(t), V_{kj}(t), w_{kj}(t)$ and $\psi_i^{X^*}(t))$ is the optimal solution for the loading problem and $(X_i^{**}(t), V_{kj}^*(t), w_{kj}^*(t)$ and $\psi_i^{X^{**}}(t))$ is the solution of both loading and timing problems, then $(X_i^{**}(t), V_{kj}^*(t), w_{kj}^*(t)$, and $\psi_i^{X^{**}}(t))$ is the solution of the general problem.

Proof: If demand is pressing, it is sufficient to show that the solution $(X_i^{**}(t), V_{kj}^*(t), w_{kj}^*(t)u_{kj}^*(t) = \dot{V}_{kj}^*(t))$ satisfies Lemmas 1 and 3, Corollary 1.1, and all of the constraints of the general problem.

Lemma 1 and its corollary are valid because the respective (production rate) term of the Hamiltonian and the dual equation of the loading problem coincide with the Hamiltonian (10) and (8), respectively. Lemma 3 is satisfied because the solution $(X_i^{**}(t), V_{kj}^*(t), w_{kj}^*(t)$, and $\psi_i^{X^{**}}(t))$ meets the necessary setup conditions formulated explicitly in the timing problem.

With respect to the formulations of the sequencing and loading problems, one can easily observe that the primal variables $(X_i^{**}(t), V_{kj}^*(t)w_{kj}^*(t)u_{kj}^*(t) = \dot{V}_{kj}^*(t))$ satisfy all of the constraints of the general problem. Q.E.D.

Although the proven theorem provides an efficient framework in the search for the optimal solution of the general problem, presence of the state constraints in the sequencing problem reduces still further the applicability of the approach for large-scale systems. Therefore, we will relax the state constraints and obtain both upper and lower estimates of the optimal solution. To formulate the relaxation, we denote by $f_n(V)$ a sequence of functions which converges to $\Theta(V-1)$ when $V \geq 0$ and tends to the minus infinite-valued function when $V < 0$ for $n \to \infty$ (see Fig. 2). The choice of such functions is twofold: to approximate the discontinuity of the unit step-function $\Theta(V-1)$ and to penalize the violation of the state constraint $V_{kj}(t) \geq 0$.

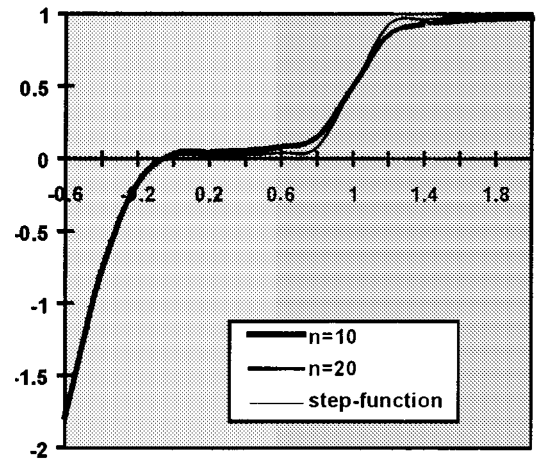

Fig. 2. Step-function $\Theta(V-1)$ and its approximation functions $f_n(V)$.

Example (Continued): For the coffee example, the following approximating functions are selected as:

$$f_n(V) = \begin{cases} \dfrac{1}{\pi}\,\mathrm{arctg}(n(V-1)) + \dfrac{1}{2} \\ \quad \text{if } V \geq 0; \\ -5V^2 + \dfrac{n}{\pi(1+n^2)}\,V - \dfrac{1}{\pi}\,\mathrm{arctg}(n) + \dfrac{1}{2} \\ \quad \text{if } V < 0, \end{cases}$$

and are depicted in Fig. 2.

Lemma 5–The Lower Bound Lemma: If the constraint (5) of the general problem is relaxed as

$$-\infty < w_{kj}(t) \leq f_n(V_{kj}(t)) \tag{14}$$

then the solution of the relaxed general problem is the lower bound of the general problem for each n. $\blacksquare$

Note, that the solution of the relaxed general problem (see Lemma 5) is not necessarily feasible for the general problem, but, when demand is pressing, it provides a close (lower) estimation for the optimal value of the objective function. Indeed, negative production rates (14) waste time and capacity on production unspecified by bills of materials. Recalling our coffee example, negative production rate of a packager would mean that it unpacks a brand and pumps it back to the buffer for the corresponding mixtures. This clearly cannot occur frequently when demands for brands are pressing.

Remark 1: Following the framework of Theorem 1, a feasible near-optimal solution for the general problem can be obtained by, first, relaxation of the equation of the production process in the sequencing problem

$$\dot{X}_i(t) = \sum_{kj} f_n(V_{kj}(t))v_{ikj} - d_i(t)$$

and exclusion of the state constraint $V_{kj}(t) \geq 0$. Second, the sequencing problem is solved by a time–decomposition method for optimal control problems with only control constraints. Due to the special form of function $f_n(V)$, only an insignificant violation of the state constraint can occur. Therefore, then the solution is approximated in a natural way to meet the excluded state constraint. Namely, on the no-setup regimes of machine k, a $V_{kj}(t)$ which is close to one is set to one while the others $V_{kj}(t)$ are set to zero. On the setup regimes, the solution is left without change. Eventually, the near-optimal solution can be found by Theorem 1. Indeed, if the solution satisfying both loading and timing problems exists, then the obtained solution is the optimal one according to Theorem 1. Otherwise, the timing problem can be used as an heuristic to bring the solution nearer to the optimal one. ■

Lemma 6: Let $X_i(t)$, $w_{kj}(t)$, and $\psi_i^X(t)$ be the solution of the loading problem for a given setup sequence $V_{kj}(t)$ and

$$\sum_i \psi_i^X(t_s)v_{ikj} \neq \sum_i \psi_i^X(t_s + T_{kjj'})v_{ikj'}$$

then there exists a small variation ζ of t_s so that the solution for the varied loading problem improves. ■

Thus, Remark 1, Lemmas 5 and 6 define the way for estimating the optimal solution of the general problem. Namely, while Lemma 5 provides the lower bound for the optimal solution, Lemma 6 specifies the upper estimate by allowing iterative improvements of admissible solutions. These results will be elaborated upon in the next section on the basis of the time–decomposition methodology.

V. A Time-Decomposition Numerical Method

The time–decomposition method is based on decomposition of an infinite-dimensional optimization problem into finite-dimensional problems of maximizing the Hamiltonian, formulated at each point of time. The time–decomposition method is of an iterative nature. On each iteration of the time–decomposition method, the only condition of the maximum principle that is not satisfied is that the control functions do not maximize the Hamiltonian everywhere on the planning horizon. Iterations of the methods will ensure that a measure of this discrepancy from the maximum principle is minimized. Thus, the central point of the time–decomposition method is the numerical construction of a descent variation of control.

Time–decomposition methods have proved their efficiency for solving complex optimal control problems with regular mixed constraints [21], [22]. However, the mixed constraint (5) is not regular and, therefore, the straightforward time–decomposition is not applicable to the original problem.

In manufacturing, numerical methods based on time–decomposition were successfully applied to scheduling large FMS with negligible setup times [14], [23]. In what follows we present two time–decomposition algorithms for both near-optimal solutions and lower bound estimates.

A. An Algorithm for Near-Optimal Solution

In order to extend the time–decomposition approach to scheduling with significant sequence-dependent setup times, an iterative method outlined in Theorem 1 is suggested. The method presented here is based on the projected gradient approach which ensures monotone decrease of the objective function on consecutive iterations. In what follows, Steps 1–7 of the algorithm are intended to solve the sequencing problem with regular constraints (see Remark 1); Steps 8–13 are for solving the loading problem; and Step 14 is for the timing problem according to Lemma 6.

Step 1: Choose a feasible solution of the sequencing problem $X_i(t)$, $V_{kj}(t)$, $u_{kjj'}(t)$, $t \in [0, T]$, (for example $u_{kjj'}(t) \equiv 0$).

Step 2: Integrate the system of dual equations $\dot{\psi}_i^X(t) = p_i(X_i(t))(X_i(t) - X_i^s)$ and

$$\dot{\psi}_{kj}^V(t) = -f_n'(V_{kj}(t)) \sum_i \psi_i^X(t)v_{ikj}$$

with the right-hand boundary conditions $\psi_i^X(T) = 0$ and $\psi_{kj}^V(T) = 0$.

Step 3: At every point of time, calculate the direction of descent $s_{kjj'}(t) = \psi_{kj'}^V(t) - \psi_{kj}^V(t)$ (see proof of Lemma 2). When $u_{kjj'}(t) + S_{kjj'}(t) > T_{kjj'}^{-1}$, assign $s_{kjj'}(t) = T_{kjj'}^{-1} - u_{kjj'}(t)$, if $u_{kjj'}(t) + s_{kjj'}(t) < -T_{kj'j}^{-1}$ then $s_{kjj'}(t) = -T_{kj'j}^{-1} - u_{kjj'}(t)$.

Step 4: Determine control variation $\delta u_{kjj'}(t) = \varepsilon s_{kjj'}(t)$ as a step ε along the calculated direction, $0 \leq \varepsilon \leq 1$.

Step 5: Integrate the primal equation

$$\dot{V}_{kj}(t) = \sum_{j>j'}(u_{kj'j}(t) + \delta u_{kj'j}(t)) - \sum_{j'>j}(u_{kjj'}(t) + \delta u_{kjj'}(t))$$

with the left-hand boundary condition $V_{kj}(0) = V_{kj}^0$.

Step 6: Integrate the primal equation

$$\dot{X}_i(t) = \sum_{kj} f_n(V_{kj}(t))v_{ikj} - d_i(t)$$

with the left-hand boundary condition $X_i(0) = X_i^0$ and calculate the value of the resultant objective function (6). If the objective function is not decreased, then decrease ε and go to Step 4; otherwise assign $u_{kjj'}(t) = u_{kjj'}(t) + \delta u_{kjj'}(t)$ and go to Step 7.

Step 7: Check a stop condition. For example, if $u_{kjj'}(t)$ cannot be described by a combination of regimes (i), (ii), and (iv) (see Lemmas 2 and its corollary) with a given accuracy

TABLE IV
CHANGE IN OBJECTIVE FOR DIFFERENT STEP-FUNCTION APPROXIMATIONS

Unit step-function approximation	n=10	n=15	n=20	n=25
Objective value, 10^4	2.221	2.206	2.195	2.190

then go to Step 2; otherwise approximate $V_{kj}(t)$ (see Remark 1) and go to Step 8.

Step 8: Choose a feasible solution of the loading problem $X_i(t)$, $w_{kj}(t)$, $t \in [0, T]$ for the found setup sequence $V_{kj}(t)$ (for example $w_{kj}(t) \equiv 0$).

Step 9: Integrate the dual equation $\dot{\psi}_i^X(t) = p_i(X_i(t))(X_i(t) - X_i^s)$ with the right-hand boundary condition $\psi_i^X(T) = 0$.

Step 10: At every point of time for which $V_{kj}(t) = 1$ with a given accuracy, calculate the direction of descent

$$s_{kj}(t) = \sum_i \psi_i^X(t) v_{ikj}.$$

When $w_{kj}(t) + s_{kj}(t) > 1$, assign $s_{kj}(t) = 1 - w_{kj}(t)$; when $w_{kj}(t) + s_{kj}(t) < 0$ then $s_{kj}(t) = -w_{kj}(t)$.

Step 11: Determine control variation $\delta w_{kj}(t) = \varepsilon s_{kj}(t)$ as a step ε along the calculated direction, $0 \le \varepsilon \le 1$.

Step 12: Integrate the primal equation

$$\dot{X}_i(t) = \sum_{kj} (w_{kj}(t) + \delta w_{kj}(t)) v_{ikj} - d_i(t)$$

with the left-hand boundary condition $X_i(0) = X_i^0$ and calculate the value of the resultant objective function (6). If the objective function is not decreased, then decrease ε, otherwise assign $w_{kj}(t) = w_{kj}(t) + \delta w_{kj}(t)$ and go to Step 13. If ε is greater than a given number then go to Step 11; otherwise decrease ζ and go to Step 8.

Step 13: Check a stop condition. For example, if $w_{kj}(t)$ cannot be described by a combination of regimes i)–iii) (see Lemma 1 and its corollary) with a given accuracy, then go to Step 9; otherwise go to Step 14.

Step 14: If all of the setups satisfy the necessary setup conditions (see Lemma 3) with a given accuracy, then exit; otherwise, shift each of the setups which do not satisfy these conditions on a step ζ in the directions defined as follows:

$$\text{if } \sum_i \psi_i^X(t_s) v_{ikj} > \sum_i \psi_i^X(t_s + T_{kjj'}) v_{ikj'}$$
$$\text{then assign } t_s = t_s + \zeta$$
$$\text{otherwise } t_s = t_s - \zeta.$$

Go to Step 8.

B. A Lower Bound Algorithm

According to Lemma 5, the relaxed general problem contains no irregularity and therefore can be efficiently solved by the time–decomposition method where both production and setup processes are integrated simultaneously.

Step 1: Choose a feasible solution of the relaxed general problem $X_i(t)$, $V_{kj}(t)$, $w_{kj}(t)$, $u_{kjj'}(t)$, $t \in [0, T]$, [for example $u_{kjj'}(t) \equiv 0$, $w_{kj}(t) \equiv 0$].

Step 2: Integrate the system of dual equations $\dot{\psi}_i^X(t) = p_i(X_i(t))(X_i(t) - X_i^s)$ and

$$\dot{\psi}_{kj}^V(t) = -f_n'(V_{kj}(t)) \sum_i \psi_i^X(t) v_{ikj}$$

if $w_{kj}(t) = f_n(V_{kj}(t))$ and $\dot{\psi}_i^V(t) = 0$; otherwise with the right-hand boundary conditions $\psi_i^X(T) = 0$ and $\psi_{kj}^V(T) = 0$.

Step 3: At every point of time, calculate the direction of setup rate descent $s_{kjj'}^u(t) = \psi_{kj'}^V(t) - \psi_{kj}^V(t)$. When $u_{kjj'}(t) + s_{kjj'}^u(t) > T_{kjj'}^{-1}$, assign $s_{kjj'}^u(t) = T_{kjj'}^{-1} - u_{kjj'}(t)$; when $u_{kjj'}(t) + s_{kjj'}^u(t) < -T_{kj'j}^{-1}$ then $s_{kjj'}^u(t) = -T_{kj'j}^{-1} - u_{kjj'}(t)$.

Step 4: At every moment of time calculate the direction of production rate descent

$$s_{kj}^w(t) = \sum_i \psi_i^X(t) v_{ikj}.$$

When $w_{kj}(t) + s_{kj}^w(t) > 1$ then $s_{kj}^w(t) = 1 - w_{kj}(t)$; if $w_{kj}(t) + s_{kj}^w(t) < 0$ then $s_{kj}^w(t) = -w_{kj}(t)$.

Step 5: Determine control variation $\delta u_{kjj'}(t) = \varepsilon s_{kjj'}^u(t)$ as a step ε along the calculated direction, $0 \le \varepsilon \le 1$.

Step 6: Determine control variation $\delta w_{kj}(t) = \varepsilon s_{kj}^w(t)$ as a step ε along the calculated direction, $0 \le \varepsilon \le 1$.

Step 7: Integrate the primal equations

$$\dot{X}_i(t) = \sum_{kj} (w_{kj}(t) + \delta w_{kj}(t)) v_{ikj} - d_i(t)$$

and

$$\dot{V}_{kj}(t) = \sum_{j > j'} (u_{kj'j}(t) + \delta u_{kj'j}(t)) - \sum_{j' > j} (u_{kjj'}(t) + \delta u_{kjj'}(t))$$

with left-hand boundary conditions $X_i(0) = X_i^0$, $V_{kj}(0) = V_{kj}^0$ and calculate the value of the resultant objective function (6). If the objective function is not decreased, then decrease ε and go to Step 5; otherwise assign $u_{kjj'}(t) = u_{kjj'}(t) + \delta u_{kjj'}(t)$ and $w_{kj}(t) = w_{kj}(t) + \delta w_{kj}(t)$.

Step 8: Check a stop condition, i.e., $\int_0^T (s_{kjj'}^u(t)^2 + s_{kj}^w(t)^2)\, dt$ is less than a given tolerance then stop; otherwise go to Step 2.

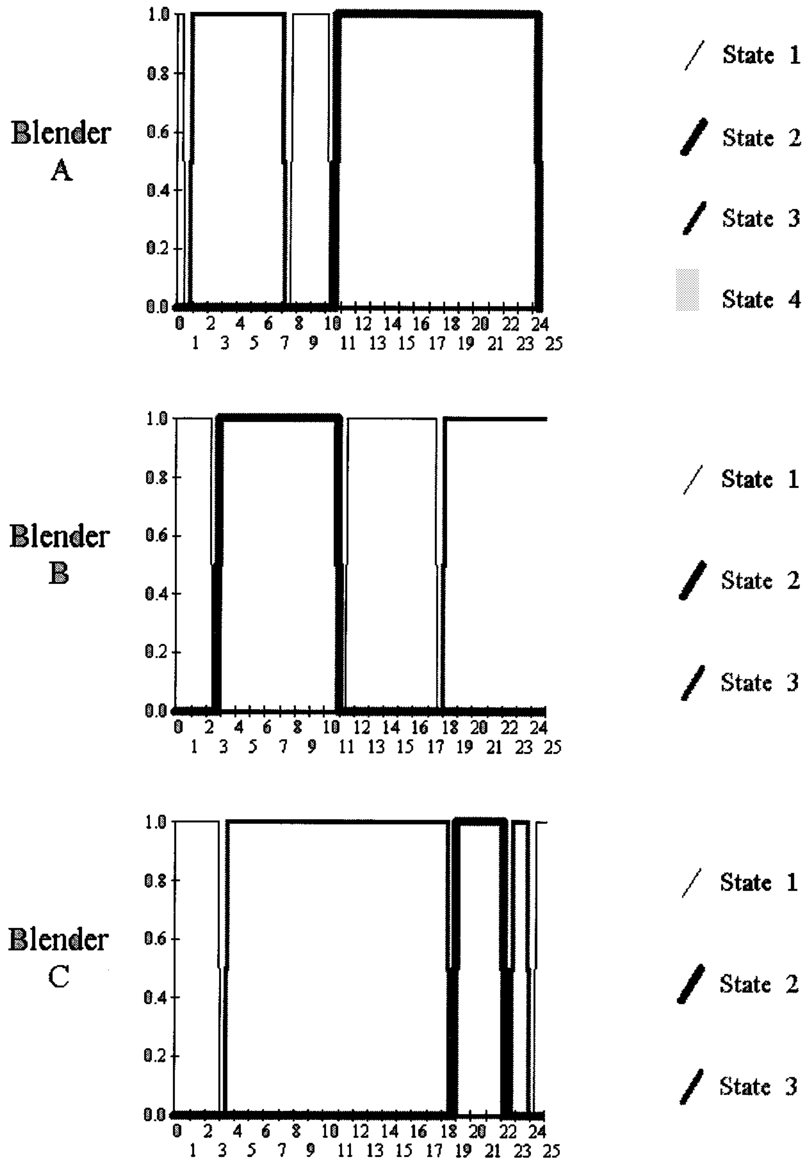

Fig. 3. Near-optimal solution for blenders.

VI. Numerical Example and Computational Results

This section presents a solution of the coffee example and summarizes computational experiments conducted for different scale manufacturing systems.

A. Application of the Method to the Coffee Production Cell

To analyze the convergence of the solution, the unit step function was approached by the sequence of approximating functions (see Fig. 2) with parameter n running from 10 up to 25.

The lower estimate of the objective function found by the lower bound algorithm for 3.2 min at IBM PC-486-66 MHz is $2.10 \cdot 10^4$ cost units, while the upper estimate provided by the near-optimal solution algorithm for 4.3 min is $2.19 \cdot 10^4$ cost units for $n = 25$. Table IV shows change in the objective of the near-optimal solution, which converges to the upper bound estimate, as n goes to 25.

The near-optimal solution $(n = 25)$ is depicted in Figs. 3 and 4 for all stages of the production process. Fig. 5 shows the demand profile and buffer behavior for the end products over the planning horizon.

B. Statistical Results

The scheduling program, realizing the developed algorithms, is written in C++, and is provided with a user-friendly interface, which is a Visual Basic project running in a Windows environment. This software has been utilized for scheduling of the avionic harness manufacturing system at Sikorsky Aircraft Corporation; 3-axis Monarch vertical machining centers at a department of Remington Arms Co.; textiles and forming machine centers in rigid panel manufacturing at Albany International HPM; final assembly at Spectra Inc.; high demand bottleneck machines in surface mount technology at Sanders, a Lockheed Martin Co.; six volt battery assembly for the Polaroid film

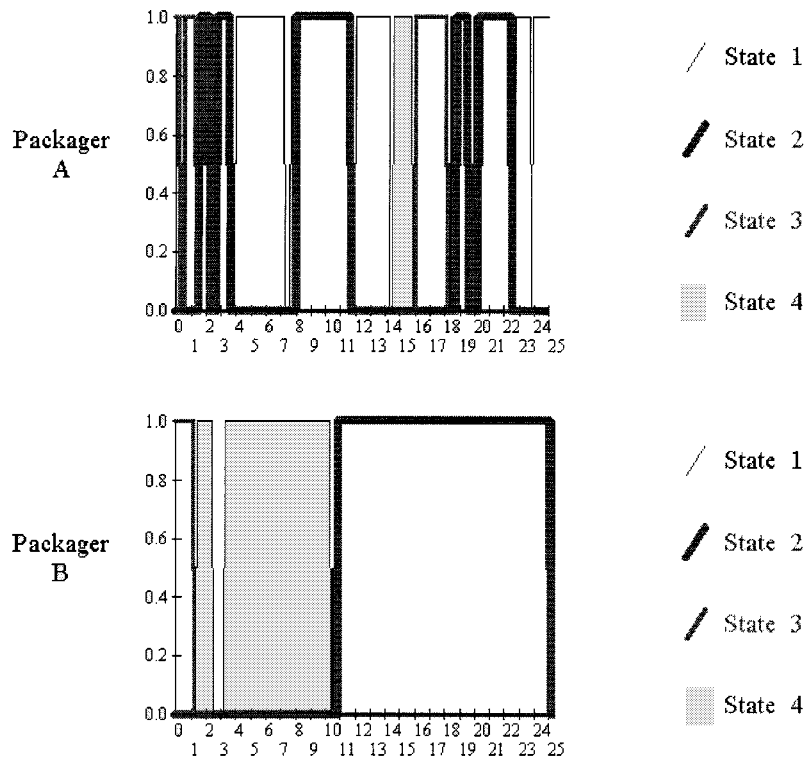

Fig. 4. Near-optimal solution for packagers.

pack at Polaroid Corp.; and heat exchanger manufacturing at Hamilton Standard.

In the experiments, the scheduling problems were solved by both the lower bound and near-optimal solution algorithms. The problems were categorized by the cumulative number of machine states $JK = \sum_k J(k)$, i.e., the sum of states for all the machines. Ten problems with different production conditions (bills of materials, demand profiles, penalty coefficients, and so on) were solved for every category. The averaged results are presented in Fig. 6 as the difference between the upper and lower estimates related to the lower-bound objective function value, ΔF. Three curves in Fig. 6 illustrate the influence of machine flexibility [$J(k)$ from 2–4] on the gap between both estimates. Although this optimality gap slowly grows with increase of machine flexibility, it meets practical requirements for the wide range of manufacturing systems. The averaged computation time for both algorithms is shown in Fig. 7.

VII. Conclusions

A new time–decomposition approach to scheduling sizable manufacturing systems with flexible machines and sequence-dependent setup times is suggested in this paper. The approach utilizes both the analytical properties of the optimal solution derived from the maximum principle and the numerical time–decomposition methodology. The theorem proves under conditions of pressing demand that the general problem can be solved to optimality by decomposing it into a number of specially constructed subproblems. At the same time, the proposed relaxation of the general problem significantly simplifies numerical procedures for the approach, while allowing quite accurate lower and upper estimates of the optimal solution. As a result, an effective algorithm for near-optimal scheduling in realistic production environment is developed. Computational tractability of the approach, observed in the conducted numerical experiments, offers the possibility of successful applications for modern flexible manufacturing systems.

Appendix

Proof of Lemma 1: The maximum principle claims that the optimal control strategy is achieved by maximizing, for each t, the Hamiltonian on the set of admissible controls. Since control variables $w_{kj}(t)$ and $u_{kj}(t)$ appear in distinct terms of the Hamiltonian (10) and since, moreover, there is no constraint which joins both controls, the maximization with respect to $w_{kj}(t)$ and $u_{kj}(t)$ can be carried out separately. Let us consider the respective (production rate) term of the Hamiltonian

$$H = \sum_i \psi_i^X(t) \left(\sum_{kj} w_{kj}(t) v_{ikj} \right). \tag{A1}$$

Because this term and constraint (5) are linear with respect to $w_{kj}(t)$, the maximum is reached in one of three cases stated in the lemma. Q.E.D.

Proof of Corollary 1.1: Let us differentiate twice the condition of regime ii). We then obtain

$$\frac{d^2}{dt^2} \sum_i \psi_i^X(t) v_{ikj} = \sum_i p_i(X_i(t)) \dot{X}_i(t) v_{ikj} = 0.$$

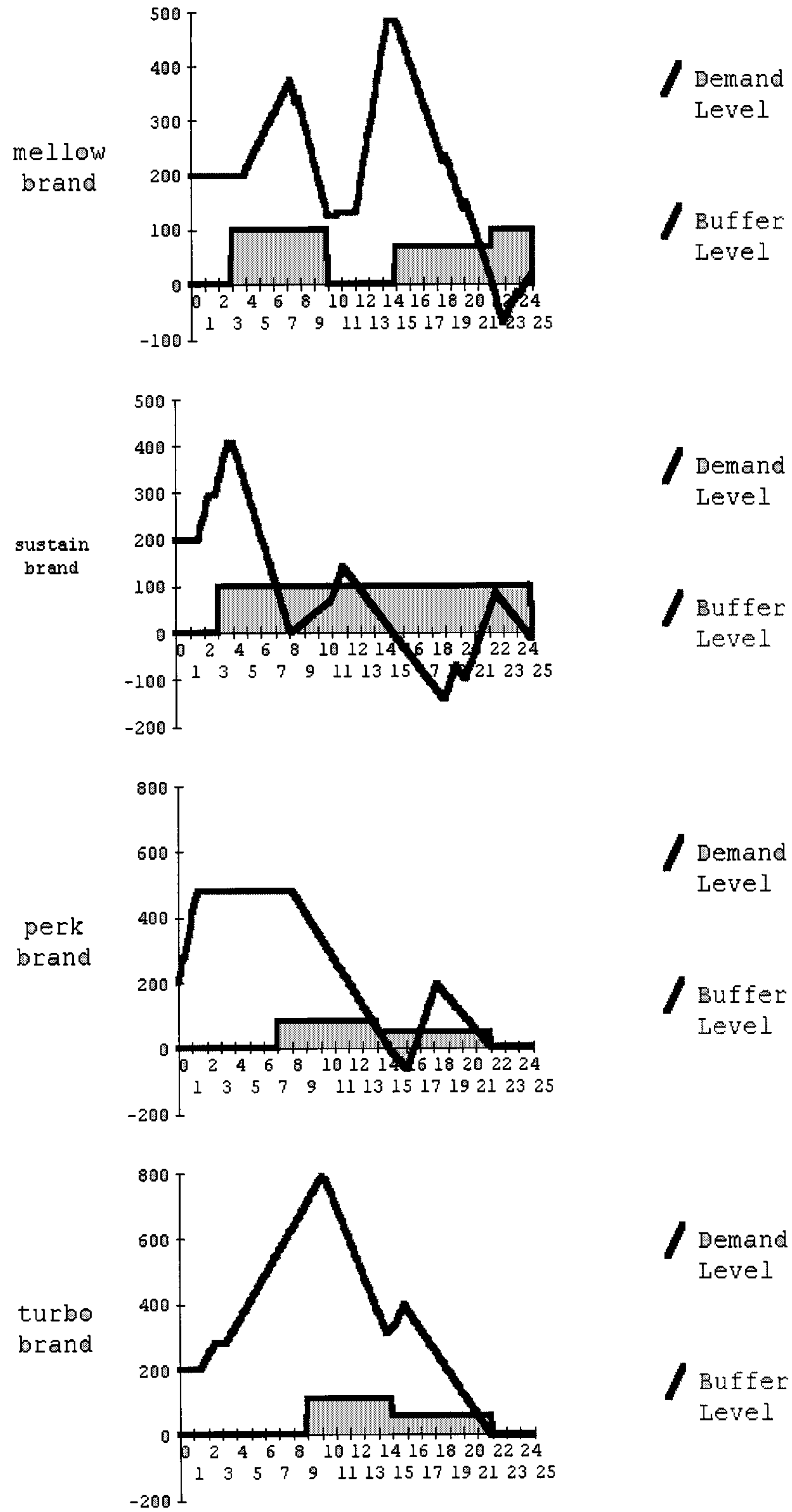

Fig. 5. Buffers behavior and demand profiles for the end products.

Replacing $\dot{X}_i(t)$ with the production process (1) and denoting

$$\frac{p_i v_{ikj}}{\sum_i p_i v_{ikj}^2} \text{ and } -\sum_i b_{ikj} \sum_{k'j' \neq kj} w_{ik'j'}(t) v_{ik'j'}$$

by b_{ikj} and c_{kj}, respectively, we immediately obtain

$$w_{kj}(t) = \sum_i b_{ikj} d_i(t) + c_{kj}.$$

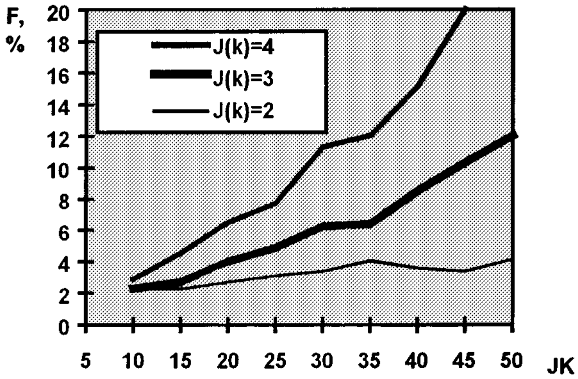

Fig. 6. Optimality gap between lower and upper estimates.

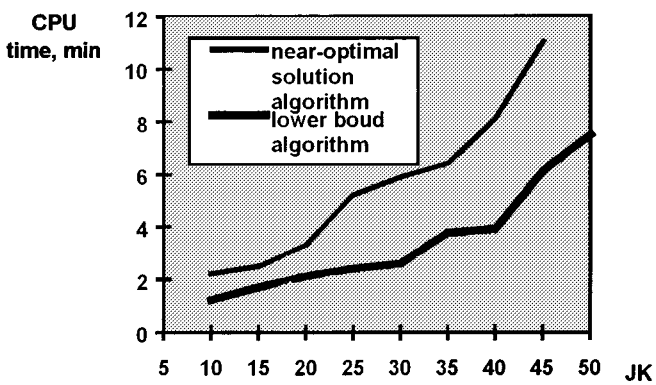

Fig. 7. Computation time for the lower bound and near-optimal solution algorithms.

If

$$\sum_i b_{ikj} d_i(t) + c_{kj}$$

is out of range [0, 1], then regime ii) cannot exist for optimal behavior of machine k, because constraints (4) and (5) are not satisfied. Q.E.D.

Proof of Lemma 2: According to the maximum principle, we now maximize the respective (setup rate) term of the Hamiltonian

$$H(u_{kjj'}(t)) = u_{kjj'}(t)\left(\psi^V_{kj'}(t) - \psi^V_{kj}(t)\right) \tag{A2}$$

on the set of admissible controls (4).

The proof immediately follows from the fact that this term and constraint (4) are linear with respect to $u_{kjj'}(T)$. Q.E.D.

Proof of Corollary 2.1: Let a singular setup regime from state j onto state j' exist on a time interval $[t_1, t_2]$ of the optimal solution, i.e.,

$$u_{kjj'}(t) \in \left[-\frac{1}{T_{kj'j}}, \frac{1}{T_{kjj'}}\right]$$

when $\psi^V_{kj'}(t) = \psi^V_{kj}(t)$, $V_{kj}(t_1) = V_{kj'}(t_2) = 0$, and $V_{kj}(t_2) = V_{kj'}(t_1) = 1$. Then from constraint (4) it follows that we can replace this regime with regime i) on the interval $[t_1, t_1 + T_{kjj'}]$ and with no-setup regime on the remaining interval $[t_1 + T_{kjj'}, t_2]$. One can observe that the objective function (6) does not change and no constraint is violated. Indeed, the objective depends only on buffer levels, and in both setup cases no product is produced on machine k during the interval $[t_1, t_2]$ which influences the buffer levels.

Likewise, a singular setup regime from state j' onto j can be replaced with regime ii) on the interval $[t_1, t_1 + T_{kj'j}]$ and with no-setup regime on the remaining interval $[t_1 + T_{kj'j}, t_2]$. Q.E.D.

Proof of Lemma 3: If setup is carried out on regime i) (see Lemma 2), then it is characterized by the setup time $t = T_{kjj'}$ and by the difference between the corresponding dual variables $(\psi^V_{kj'}(t) > \psi^V_{kj}(t))$, which are identical before and after the setup (see no-setup regime iv) in Corollary 2.1).

Thus, the necessary conditions of setting up from state j to state j' on the time interval $[t_s, t_s + T_{kjj'}]$, where t_s is an unknown moment of the setup initiation, are

$$\psi^V_{kj'}(t_s) = \psi^V_{kj}(t_s)$$
$$\psi^V_{kj'}(t_s + T_{kjj'}) = \psi^V_{kj}(t_s + T_{kjj'})$$

and

$$\psi^V_{kj'}(t) > \psi^V_{kj}(t), \qquad t \in (t_s, t_s + T_{kjj'}). \tag{A3}$$

During this setup $\dot V_{kj'}(t) = \dot V_{kj}(t) = u_{kjj'}(t) = 1/T_{kjj'}$ [see (2) and (4)], and therefore (9) takes the form

$$\dot\psi^V_{kj}(t) = -\alpha_{kj}(t) f'_n(V_{kj}(t)) \tag{A4}$$

$$\dot\psi^V_{kj'}(t) = -\alpha_{kj'}(t) f'_n(V_{kj'}(t)). \tag{A5}$$

Since $V_{kj}(t)$ and $V_{kj'}(t)$ do not equal zero during the setup, the complementary slackness condition (11) causes the measure functions $d\mu_{kj}(t)$ and $d\mu_{kj'}(t)$ to equal zero and not to enter the dual equations. From (A3)–(A5) it immediately follows that

$$\int_{t_s}^{t_s+T_{kjj'}} \alpha_{kj}(t) f'_n(V_{kj}(t))\,dt = \int_{t_s}^{t_s+T_{kjj'}} \alpha_{kj'}(t) f'_n(V_{kj'}(t))\,dt.$$

When n tends to infinity, we obtain the limit form of the necessary setup conditions

$$\sum_i \psi^X_i(t_s) v_{ikj} = \sum_i \psi^X_i(t_s + T_{kjj'}) v_{ikj'}. \tag{A6}$$

Likewise, if the setup is carried out on regime iii) (see Lemma 2), i.e., $t > T_{kjj'}$, then the necessary setup conditions take the form

$$\sum_i \psi^X_i(t_s) v_{ikj} = \sum_i \psi^X_i(t_s + t) v_{ikj'}. \tag{A7}$$

Q.E.D.

Proof of Lemma 4: One can easily observe that if the optimal production rate for the general problem takes the maximal value along the planning horizon [see Lemma 1, regime i)], then the general and the sequencing problems become equivalent. Specifically, the optimal setup sequences and durations coincide for both problems (i.e., demand is pressing). What is left to show is that (13) is sufficient for regime i) to occupy the entire planning horizon. Indeed, let the condition of regime

$$\sum_i \psi_i^X(t) v_{ikj} > 0$$

be satisfied for every state of every machine. Then, from (8) it follows that

$$-\sum_i v_{ikj} \int_t^T p_i(X_i(\tau))\,(X_i(\tau) - X_i^s)\ d\tau > 0.$$

Consequently, taking into account (1), we immediately obtain

$$-\sum_i v_{ikj} \int_t^T p_i(X_i(\tau)) \cdot \left(X_i^0 + \int_0^\tau \left(\sum_{kj} w_{kj}(y) v_{ikj} - d_i(y) \right) dy - X_i^s \right) d\tau > 0.$$

Since $w_{kj}(y)$ is always maximal, the last inequality will not be violated if:

$$-\sum_i v_{ikj} \int_t^T p_i(X_i(\tau)) \cdot \left(X_i^0 + \int_0^\tau \left(\sum_{kj, v_{ikj}<0} v_{ikj} - d_i(y) \right) dy - X_i^s \right) d\tau > 0.$$

Clearly, rearranging the terms of the inequality, we obtain

$$\sum_i v_{ikj} \int_t^T \int_0^\tau d_i(y)\,dy\,d\tau > \sum_i v_{ikj} p_i^m(X_i(t)) \left(\left(X_i^0 - X_i^s\right)(T-t) + \int_t^T \sum_{kj, v_{ikj}<0} v_{ikj} \tau\,d\tau \right).$$

Substituting the defined values of $\alpha_i(t)$ and penalty coefficients into the last inequality, we conclude with (13). Q.E.D.

Proof of Lemma 5: The proof immediately follows from the fact that the set of admissible controls for the relaxed problem includes that for the general problem. Q.E.D.

Proof of Lemma 6: Let us consider the case when

$$\sum_i \psi_i^X(t_s) v_{ikj} > \sum_i \psi_i^X(t_s + T_{kjj'}) v_{ikj'}. \tag{A8}$$

Then, assign $t_s = t_s + \zeta$, $\zeta > 0$ and find the optimal solution $X_i(t) + \delta X_i(t)$, $w_{kj}(t) + \delta w_{kj}(t)$ for the varied loading problem. Consider the variation of the objective

$$\int_0^T \sum_i p_i(X_i(t))\,(X_i(t) - X_i^s)\,\delta X_i(t)\,dt.$$

To define the sign of this variation, we replace $\delta X_i(t)$ with varied production process (1):

$$\int_0^T \sum_i p_i(X_i(t))\,(X_i(t) - X_i^s) \int_0^t \sum_{kj} \delta w_{kj}(\tau) v_{ikj}\,d\tau\,dt.$$

Next, changing the order of integration and taking into account (8) we obtain

$$-\int_0^T \sum_{ikj} \psi_i^X(t) \delta w_{kj}(t) v_{ikj}\,dt. \tag{A9}$$

From condition (A8) and Lemma 1 it immediately follows that the variation of the objective (A9) is negative, i.e., the solution of the varied loading problem improves.

Similarly, the improvement of the solution can be proved for the case when

$$\sum_i \psi_i^X(t_s) v_{ikj} < \sum_i \psi_i^X(t_s + T_{kjj'}) v_{ikj'}$$

and variation ζ is chosen negative. Q.E.D.

References

[1] T. Prabhakar, “A production scheduling problem with sequencing considerations,” *Management Sci.*, vol. 21, pp. 34–42, 1974.

[2] R. H. Deane and E. R. White, “Balancing workloads and minimizing setup costs in the parallel processing shop,” *Operational Res. Quart.*, vol. 26, pp. 45–53, 1975.

[3] E. Balas, L. T. Enos, and G. W. Graves, “Computerized scheduling of printing operations,” in *Proc. Spring Conf. Graphic Communications Computer Association*, Toronto, Canada, 1973, pp. 285–293.

[4] A. M. Geoffreon and G. W. Graves, “Scheduling parallel production lines with changeover costs: Practical application of a quadratic assignment/LP approach,” *Oper. Res.*, vol. 24, pp. 595–610, 1976.

[5] D. M. Dilst and K. D. Ramsing, “Joint lot sizing and scheduling of multiple items with sequence-dependent setup costs,” *Decision Sci.*, vol. 20, pp. 120–132, 1989.

[6] K. Morikawa and N. Nakamura, “Production panning and scheduling of multiple items with sequence-dependent setup times,” in *Proc. 12th Int. Conf. Production Res.*, Lappeenranta, Finland, 1993, pp. 335–336.

[7] J. G. Kimemia and S. B. Gershwin, “An algorithm for the computer control of a flexible manufacturing system,” *IIE Trans.*, vol. 15, no. 4, pp. 353–362, 1983.

[8] J. R. Perkins and P. R. Kumar, “Stable, distributed, real-time scheduling of flexible manufacturing/assembly/disassembly,” *IEEE Trans. Automat. Contr.*, vol. 34, pp. 139–148, 1989.

[9] P. R. Kumar and T. I. Seidman, “Dynamic instabilities and stabilization methods in distributed real-time scheduling of manufacturing systems,” *IEEE Trans. Automat. Contr.*, vol. 35, pp. 289–298, Mar. 1990.

[10] C. Humes Jr., “Some results in the Perkins-Kumar framework for distributed activity scheduling,” in *Proc. Workshop Hierarchical Control Real-Time Scheduling Manufacturing Syst.*, Lincoln, NE, 1992, pp. 35–44.

[11] A. Sharifnia, M. Caramanis, and S. B. Gershwin, “Dynamic setup scheduling and flow control in manufacturing systems,” *Discrete Event Dynamic Syst.*, vol. 1, pp. 149–175, 1991.

[12] A. Haurie and C. van Delft, "Turnpikes in flow control models of unreliable manufacturing systems," *European J. Oper. Res.*, vol. 82, pp. 359–372, 1995.
[13] E. Khmelnitsky and K. Kogan, "Necessary optimality conditions for a generalized problem of production scheduling," *Optimal Contr. Appl. Methods*, vol. 15, pp. 215–222, 1994.
[14] E. Khmelnitsky, K. Kogan, and O. Maimon, "A maximum principle based method for scheduling in a flexible manufacturing system," *Discrete Events Dynamic Syst.*, vol. 5, pp. 343–355, 1995.
[15] A. E. Bryson and J. Y.-C. Ho, *Applied Optimal Control*: Hemisphere, 1975.
[16] L. Pun, *Introduction to Optimization Practice*. New York: Wiley, 1969.
[17] R. J. Tersine, *Production/Operations Management: Concepts, Structure, and Analysis*. Amsterdam, The Netherlands: North-Holland, 1985.
[18] A. Y. Dubovitsky and A. A. Milyutin, "The theory of the maximum principle" (in Russian), in *Methods of the Theory of Extremal Problems in Economics*, V. L. Levin, Ed, Moscow: Nauka, 1981, pp. 6–47.
[19] R. F. Hartl, S. P. Sethi, and R. G. Vickson, "A survey of the maximum principles for optimal control problems with state constraints," *SIAM Review*, vol. 37, no. 2, pp. 181–218, 1995.
[20] M. W. J. Blok, *Dynamic Models of the Firm: Determining Optimal Investment, Financing and Production Policies by Computer*, Berlin: Springer-Verlag, 1996.
[21] A. Ilyutovich and E. Khmelnitsky, *A Numerical Method for an Optimal Control Problem with Constraints on State Variables, Based on the Maximum Principle* (in Russian). Moscow, Russia: VNIISI, 1991.
[22] A. A. Milyutin, N. Osmolovsky, S. Chukanov, and A. Ilyutovich, *Optimal Control in Linear Systems* (in Russia), Moscow, Russia: Nauka, 1993.
[23] J. B. Sousa and F. L. Pereira, "A hierarchical framework for the optimal flow control in manufacturing systems," in *Proc. 3rd Int. Conf. Computer Integrated Manufacturing*, Troy, NY, 1992, pp. 278–286.

Eugene Khmelnitsky received the M.S. degree in applied mathematics from Moscow Institute of Physics and Technology, 1988 and the Ph.D. degree in industrial engineering from Tel-Aviv University in 1997.

He worked as a Researcher at Boston University from 1995 to 1996 and then joined the faculty of the Deptartment of Industrial Engineering, Tel-Aviv University, Israel. His research interests include the areas of control and optimization of complex dynamic systems, production planning and scheduling.

Konstantin Kogan received the B.S. and M.S. degrees in computer science and the Ph.D. degree in industrial engineering all from Moscow Forestry-Technical Institute.

He is a Senior Lecturer of Industrial Engineering at the Department of Computerized Systems at Holon Center of Technological Education, Israel. His teaching and research interests are in the area of scheduling, material requirements and production planning.

Oded Maimon received the Ph.D. degree from Purdue University in 1984, attended at MIT, and since 1987 has attended Tel-Aviv University. He is a Professor and head of the Department of Industrial Engineering at Tel-Aviv University. He works and publishes in the area of design, production planning and control.