Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

Real-time, Cooperative Enterprises for Customized Mass Production; Challenges and Solution Approaches

Monostori, L.*,, Váncza, J.* ,**, Kis, T.*, Erdős, G.*, Karnok, D.* ,**, Egri, P.***

**Computer and Automation Research Institute, Hungarian Academy of Sciences, Budapest, Hungary*
***Department of Manufacturing Science and Technology, Budapest University of Technology and Economics, Hungary*
(e-mail: laszlo.monostori@sztaki.hu)

Abstract: The paper underlines the main requirements of customized mass production, with special emphasis on the real-time ability and cooperativeness. Main goals of a large-scale national industry-academia R&D project aimed at improving the performance of a production network that produces consumer goods in large quantities and variability are highlighted. An integrated approach is presented for planning the behavior of the system at network-, factory- and plant level, as well as for adapting various plans to real execution conditions.

1. INTRODUCTION

The problem each manufacturer faces time and again is how to meet demand by making available the required quantities of products at proper time. Giving answer is hard because market demand is uncertain and distributed, while production processes are complex involving geographically dispersed producers of raw materials, components and end-products. Resources are finite and their performance or even availability is subject to change. Customers have a tendency to wait for meeting their needs for less and less time: typically, acceptable order lead times are much shorter than actual production lead times. Decisions are made under the pressure of time, relying also on uncertain and incomplete information.

The markets are typically served by *production networks* that consist of autonomous enterprises. Taking high *service level* as their main priority, manufacturers can hedge against demand uncertainty only by maintaining time, capacity and/or material buffers. This however, incurs extra equipment, labor, inventory and organizational costs, as well as – especially under dynamic market conditions – the risk of producing obsolete inventory. Partners are legally independent entities, with their own resources, performance objectives and internal decision mechanisms. They have to find their own trade-offs between service level and cost that are acceptable for their partners. Such a solution can only emerge from the interaction of local and asynchronous decisions. The main issues are as follows:

- There is an inevitable need to design organizations to perceive and respond to market demand by sustaining *coordination* and, if possible, *cooperation* among network members.
- Essential *production planning and scheduling* problems must be solved locally. This is a key also to predictable behavior.
- *Execution* of production plans and schedules should be supported by real-time control that is able to adapt plans and schedules to changing conditions, with minimal ramification of changes.

Our specific interest is in *customized mass production* that is aimed at satisfying volatile demand on markets of *mass products* where demand appears for a complex and ever changing variety of goods, both for small and large quantities, in hardly predictable temporal patterns. Demand must be fulfilled with mass production efficiency, but in very short times: acceptable delivery times are only fractions of the production lead times. The products are typically consumer goods like low-tech electronics, mobile phones, electric bulbs, cosmetics, etc. Customer demand is anticipated and satisfied directly by a manufacturer of end-products that works in the *focal point* of the network, while other members supply the manufacturer with necessary components including packaging materials.

The motivation of this work comes from a large-scale national *industry-academia R&D project* aimed at improving the performance of a production network that produces consumer goods in large quantities and variability (Monostori, *et al.*, 2006). Fig. 1 summarizes the main endeavours of the project: the research and development of solutions from the level of production networks through single enterprises to production lines, which can ensure the optimal / near to optimal behaviour of the whole system, and moreover, in a *real-time* fashion required by the given level of production. The importance of the *time* is illustrated by the watches in the figure, which incorporates the different levels (network, enterprise, production line) of the production expected to react on the external and internal changes and disturbances (indicated by thunderbolts) with a *reaction time* characterising the level in question.

978-1-1234-7890-2/08/$20.00 © 2008 IFAC

10.3182/20080706-5-KR-1001.3681

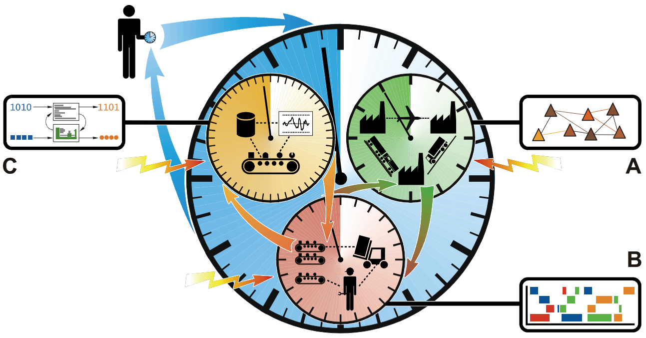

Fig. 1. General concept of the VITAL project.

The industrial partners form a complete focal network: a central assembly plant with several external and internal suppliers. The focal manufacturer produces altogether several million units/week from a mix of thousands of products. Some of the products are sold by big box retailers under their own labels; these products have to meet extra requirements and are customized variants of the manufacturer's own products. While such products typically cover 80% of the product spectrum, they give only 20% of the overall volume. The setup costs are significant throughout the whole network. Service level requirements are extremely high: keeping due dates is the main priority, even though certain retailers require shipment within 24 hours. The autonomous partners are willing to share even private business information.

We are interested in *planning and controlling the behavior* of the network, where control of future events has to be exercised on several aggregation levels, on different time horizons, but on each horizon in a *real-time* manner. Hence, the focus of our work was set to the coordination the future intentions – i.e., plans and schedules – of the partners. Since the focal manufacturer gives the heartbeat of the whole system, we put special emphasis on *scheduling* of its operations. The performance of the overall system hinges on whether the focal manufacturer really works according to schedule. Hence, methods of adaptive *executing monitoring and control* were developed to accommodate schedules to eventual changes.

In the sequel, key issues and our solutions are outlines according to this logic.

2. COOPERATIVE PRODUCTION AND SUPPLY LOGISTICS

2.1 Logistics Platform for Coordination

We assume a focal supply network of autonomous partners where there is no overlap between the channels. The network-level problem is stated as follows: the common goal of each partner is (1) to provide *high service level* towards its buyer, while, at the same time, (2) keeping overall expected production and logistics *costs at a minimum*. These requirements are conflicting:

- Due to uncertain market conditions, inventories (of components, packaging materials, products) are inevitable to provide service at the required level.
- In mass production technology, low costs can be achieved only with large lot sizes, which involve, again, higher product and component inventories as well as increased work-in-process.
- Markets of customized mass products are volatile. If the demand unexpectedly ceases for a product, then accumulated inventories become obsolete and cause significant losses.

Though there exists a number of enterprise resource planning (ERP) and supply chain management (SCM) systems that provide technology for information storing, retrieval and sharing within and between the nodes of a production network, these systems are mainly transactional: they do not really support coordinated decision making of autonomous partners (Stadler 2005; Li and Wang, 2007).

As a solution to the above dilemma, we developed a so-called *logistics platform* for coordinating the partner's decisions along individual supply channels. The key idea is to detach the two main conflicting objectives and thus make both of them manageable: while service level is tackled on the short-term, where detailed schedules provide reliable information about the close future, cost-efficient production is concerned on the medium-term. Hence, the platform consists of two levels:

1. On the *scheduling platform*, the supplier meets the exact, short-term component demand of the manufacturer. This demand is generated from the actual daily production schedule of the manufacturer in form of call-offs and should be satisfied by direct, just-in-time delivery from an inventory. Decisions are made on a daily basis, with a horizon of 1 to 2 weeks. With this short look-ahead, demand uncertainty is hedged by safety stocks.
2. On the *planning platform*, the supplier manages an inventory. As input, the supplier receives medium-term demand forecasts of components from the manufacturer, together with some information about the reliability of forecasts. On this platform, decisions are made in a weekly cycle.

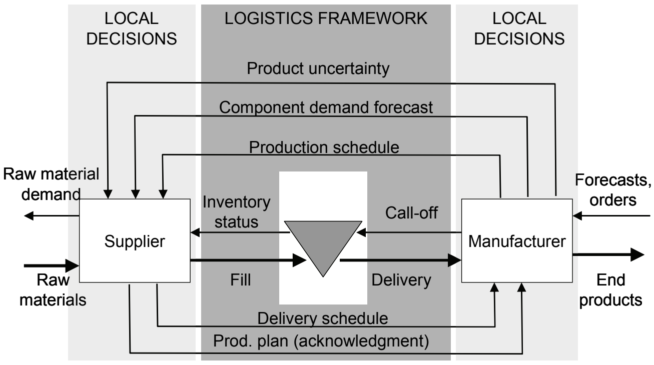

Fig. 2. Information flow through the logistics platform

Note, that the platform provides an interface between various planning functions of the manufacturer and the supplier (see also Fig. 2). Decisions have to be made locally, but the various checks (whether production schedule of the manufacturer is really served by a delivery schedule of the supplier) and inventory management policies are performed within the logistics platform (Egri and Váncza, 2006a).

2.2 Managing Inventories

For supporting *inventory planning* that links the planning and scheduling functions of autonomous partners along a supply channel, we developed a portfolio of novel lot sizing methods that take the total production and logistics costs into account, regarding also the uncertainty of demand. Decisions that *coordinate a channel* can be made on the basis of information coming partly from the manufacturer (component demand forecast and its uncertainty) and partly from the supplier (setup, production and inventory holding costs). Two different situations have been modeled:

1. *Run-out* of a product can occur with a certain probability any time in the future, but no further details are known
2. At the *end of a product's life-cycle*, the fact of the run-out and its date are known, but the amount of demand is still uncertain.

For the case of the unknown run-out date, we introduced *heuristic policies* and an extended version of the classical *Wagner-Whitin* (WW) method (Hopp and Spearman, 1996). These methods depart from a medium-term multi-period demand forecast and the probability of run-out, as well as from the cost factors of production, setup and inventory holding and determine the *optimal lot size*. As a novelty, the expected cost of obsolete inventory is also considered. The heuristic methods disregard the less trusted remote forecasts and minimize the expected average costs (Váncza and Egri, 2006). The modified WW method plans the whole horizon by minimizing the total cost; it also determines the setups (Egri and Váncza, 2006a).

For the second case we have extended the so-called *newsvendor* model. The standard one-periodic model describes the situation when stocks cannot be carried from one period to another (Cachon 2003). Demand is given as a random variable with a particular distribution, and satisfying the complete demand is a specific constraint. This may necessitate an additional setup. All in all, the expected total cost consists of the expected cost of setup, purchase, obsolete surplus and additional setup costs. The lot sizing decision is the responsibility of the supplier. Under realistic conditions the problem has a unique solution. For details and some industrial test results, see (Egri and Váncza, 2006b).

2.3 Cooperative Planning

The above framework is based on information sharing, but assumes *truthfulness* of the partners: channel coordination hinges on whether the manufacturer reveals in the logistics platform its real demand forecasts and/or product related uncertainties. However, in order to avoid component shortage, the manufacturer has an incentive to inflate its forecast or to underestimate its uncertainty. In any case, the effect will be more optimistic: it yields larger lot sizes and inventories. The manufacturer can be on the safe side, but the network will operate with higher inventories than necessary and the chance of producing obsolete inventories will also increase. Hence, selfish (rational) *distortion of information* will necessarily lead to additional costs.

We considered the problem of designing a *mechanism* that drives the partners towards disclosing and using unbiased information when trying to coordinate the channel. The supplier provides a *service* to the manufacturer by committing itself to meet all short-term demand. In return, the manufacturer pays for (1) the components delivered, (2) the flexibility of the supplier, and (3) the forecast deviation. We have defined for the newsvendor case such a *payment scheme* that drives the manufacturer to communicate its real forecast. Hence, *cooperation is self-interest* of the partners and the channel can operate at a global optimum even though decisions are made locally, using asymmetric information. In this case, the logistics platform that controls the flow of information and goods is to be augmented with the flow of financial assets. The basis of financial calculations is a fair share of the costs of operating on a risky market.

3. DAILY PRODUCTION SCHEDULING

We have developed and deployed a scheduler system for solving the short-term, daily scheduling problem of the focal manufacturer. Though the model has now a unique application, it is generic and adaptable to industries performing mass production.

The role of daily scheduling over a rolling horizon is twofold. On the one hand, it has to *schedule new production orders*, and on the other hand, it has to *guarantee feasibility of the next few days*. When scheduling new production orders (POs), first processing alternatives have to be selected from appropriate sets of alternatives, and then the POs have to be scheduled on the machines. It is a common requirement that the machine assignments and the sequence of the already scheduled tasks should be modified only if there is no other way to improve on the cost function. While scheduling algorithms mainly focus on temporal feasibility, in customized mass production where the same material can be built into several production orders, the material stock and expected shipment must cover the demand of all scheduled items in order to ensure a smooth execution. However, sufficient material supply is required in the next few days only, while on the longer term the scheduler computes only the material demand and contrasts it with the known stock levels and expected shipments. This provides information as whether additional material must be called off from the suppliers.

In the sequel we describe first the scheduling problems to be solved and then sketch the solution methodologies. Finally, we give some details of the implementation and examples of the possible functionalities of our system.

3.1 The Scheduling Problems

The problem of scheduling new production orders can be cast into the following general model: There is a set of *production orders* $\mathrm{PO}_1,\ldots,\mathrm{PO}_n$, where PO_i requests the production of a specific end-product or intermediate product in a given quantity q_i. Each PO_i has a *release date* r_i and a *due date* d_i. There is a *time horizon*, two weeks, say, and the r_i falls on the beginning of some day, while the d_i is the end of some day within the horizon. Each PO_i is divided a-priori into a finite number of *jobs*, where job $j \in \mathrm{PO}_i$ contains q_j items from PO_i and $\sum\{q_j \mid j \in \mathrm{PO}_i\} = q_i$. The production process consists of a sequence of *main production steps* and each PO requires a subsequence of this. Each PO_i has a few *routing alternatives* $R_i^1,\ldots,R_i^{a(i)}$, where each R_i^l is a sequence of stages, each *stage* being a subsequence of steps. The stages of each routing alternative must be disjoint and their union must be the set of steps required by PO_i. With each stage s, there is associated a set of machines M_i^s. The *processing time* of job $j \in \mathrm{PO}_i$ on some machine M_k is defined as $p_{j,k} = q_j / v_i^k$, where v_i^k is the *yield* of M_k (items/time unit) when processing any job of PO_i. Each machine M_k has a calendar specifying those time periods when the machine is available for processing. There are *sequence dependent setup times* between the jobs of different POs scheduled on the same machine. It is assumed that the setup times satisfy the triangle inequality.

A solution to a problem in this model selects a routing alternative for each PO, assigns a machine to each stage of each job and specifies an order of job-stages on each machine. The quality of the solution can be measured by e.g. total PO tardiness, that is, the sum of the tardiness of each PO, where the tardiness of PO_i is the time passed after d_i while its last job is completed (this can be 0 when all jobs of PO_i get finished not later than d_i).

In order to ensure schedule feasibility of the next few days, a different problem must be solved. Namely, given a solution to the above problem, and in addition material requirements for each stage of each job along with initial stock levels and expected shipment days and quantities for the different materials, a subset of job-stage pairs have to be removed from the schedule and inserted back after this period, in order to ensure that all scheduled jobs on the next t days have sufficient materials. To fill in the gaps, other jobs from the future have to be reinserted within the next t days while maintaining temporal and material feasibility in this period.

3.2 Scheduling of New Production Orders

This scheduling problem is solved in three phases. Firstly, for each new PO a routing alternative is selected and a possible distribution of the jobs of the POs on the machines is chosen by solving a relaxation of the problem. This problem can be formalized as a mixed-integer linear program (MILP) and handled by a standard solver. Then, each job of each PO is inserted into the schedule by using the stages of the selected routing alternatives and the distribution on the machines. The result is an *initial schedule* which contains also the new jobs. Finally, the initial schedule is improved by local search, the objective being to minimize the total PO tardiness as a primary objective, and to minimize the total job tardiness as a secondary objective. In fact, without the secondary objective it would be hard to decrease total PO tardiness, because the search process would have no clue how to improve on the schedule.

3.3 Rescheduling to Ensure Temporal and Material Feasibility

The rescheduling of jobs in the next t days is performed in two steps: firstly, a subset of tasks is selected for displacement into the future such that the remaining tasks have sufficient material supply. This subset is chosen by solving one multidimensional knapsack problem using the well-known MIP formulation for each of the t days. The second step consists of moving forward some jobs to fill in gaps. This task is performed by a simple heuristic procedure, which, nevertheless, ensures that no material shortage is created.

3.4 Implementation and Testing

We have designed a scheduling system with modular structure and implemented the various functionalities by combining these modules. The main modules are the “Assigner”, “Temporal Scheduler”, and “Material Scheduler”. Roughly speaking, the Assigner selects a *processing alternative* for each production order, the Temporal Scheduler builds an *initial schedule* and then *improves* it by local search, while the Material Scheduler *reschedules* the next few days to ensure temporal and material feasibility. All modules have been implemented in the programming language C++ and we also use the commercial solver ILOG CPLEX for solving the mathematical programs.

As for possible uses of these modules, we mention (1) Scheduling of new POs: first apply the Assigner, then the Temporal Scheduler, (2) Update the schedule with finished jobs: apply the Temporal Scheduler, (3) Ensure that there is no material shortage on the next 2 days: invoke the Material Scheduler.

We have thoroughly tested the scheduling of new POs both on computer generated and real-world problem instances. On the computer generated instances we also computed a lower bound which we obtain when selecting a routing alternative for each new PO. We found that the final solution of the temporal scheduler is 1.2 to 1.5 times off from the lower bound on the total PO tardiness on the hardest instances. These test instances consist of about 1,000 POs and 5,000 jobs to be scheduled on more than 100 machines. Each PO has 2 routing alternative on average. Nevertheless, a time-feasible solution of the above quality can be obtained in about 10 minutes on a PC with 3 GHz CPU.

4. REAL-TIME CONTROL OF DAILY PRODUCTION

Real-time control of the daily production is an important prerequisite of customer responsiveness. Its main function is to adapt the operations planned to the changing environment, while preserving efficiency with respect to cost, time and quality requirements.

Problems of such kind are extensively investigated world-wide. A real-time schedule monitoring and filtering approach based on statistical throughput control is described in (Wilhelm, et al., 2000), for recognizing and evaluating the impact of disturbances. The schedule repair algorithm is activated only in case of severe disturbances in order to decrease system nervousness. A deadlock-free rescheduling algorithm is introduced in (ElMaraghy, ElMekkawy, 2002). Intelligent techniques for recognizing changes and disturbances and to adapt the production rapidly to current internal and external circumstances are enumerated, e.g., in (Monostori, 2003).

In the solution described here, the reference of real-time production control is the optimized, daily schedule described in the previous section. The information about the overall factory is collected in the MES (Manufacturing Execution System) Cockpit (Fig. 3), which has a database common with the Production Monitoring system and the Scheduler.

By default, the MES Cockpit itself provides an overall view of the factory; however, the status of separate plants, cells or specific machines can also be checked. The platform also notifies the users about deviations from the production schedules together with the option to find the cause of the deviation (e.g. raw material unavailability, machine breakdown, lack of operator, etc.).

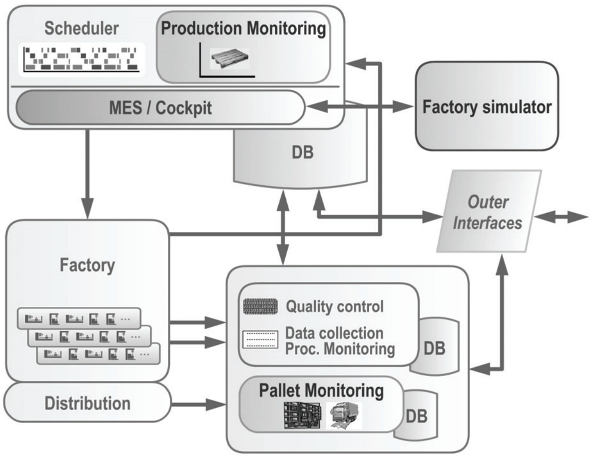

Fig. 3. Structure of the realized system

4.1 Simulation Support for Production Control

The deployed system consists also a discrete event simulation tool integrated in the MES Cockpit. Its main functions or operation modes, are as follows:

- Off-line validation, sensitivity analysis of the schedules. Evaluation of the robustness of daily schedules *prior to the execution* against uncertainties, such as machine unavailability or job slipping. By this way, it can point out those resources which can endanger the realization of the daily schedule.
- On-line, anticipatory recognition of deviations from the planned schedule by running the simulation parallel to the plant activities. By using a look ahead function (supposing of keeping the sequences as planned), support of situation recognition (*proactive operation mode*, Fig. 4).
- On-line analysis of the possible actions and minimization of the losses after a disturbance already occurred (*reactive operation mode*, Fig. 4).

The model structure in the simulator is the same for the three operation modes, however, the granulation (level of modeling detail), time horizon, applied failure models and considered outputs depend on the purpose of the assignments.

In the on-line modes the simulation models represent various virtual mirrors of the plants and run parallel to the real manufacturing environment, simulating also the future processes for a predefined short period. The performances of the predicted and the so far executed schedule are compared (highlighted as 'Performance measure of interest' in Fig. 4).

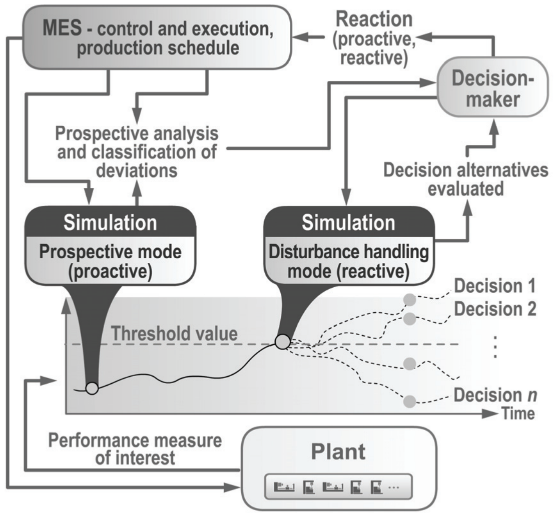

Fig. 4. Plant-level active disturbance handling by using reactive/proactive operation modes of simulation

The off-line operation mode refers to either the factory or individual plants, while in the on-line modes the work of a plant-level Decision-maker is supported (Fig. 4). The main goal of the Decision-maker is to ensure the completion of the jobs assigned by the scheduler to the given plant, and if it is not possible, to minimize the lateness of jobs. In case of intervention, a rescheduling action has to be performed with limited scope (in space and time) in correspondence to the sphere of authority of the Decision-maker. The control action made in this rescheduling point incorporates the selection of the appropriate rescheduling policy and method.

The situation detecting algorithms and the rescheduling policies have been intensively tested on industrial data and their introduction is planned for the final period of the project.

5. CONCLUSIONS

One of the most “vital” features of enterprises is their ability to cooperate as well as to give quick responses to changes and disturbances. In this paper, we presented cooperative supply planning, as well as production scheduling and execution monitoring methods that were developed to improve the overall logistic and production performance of a supply network that has to operate under uncertain market and technical conditions. Since the methods come from various, novel areas of informatics, operational research and knowledge-based systems, their integration is expected to balance the aspects of optimization, autonomy, and cooperation.

All the described solutions have been deployed at the premises of the industrial partners of the project and are in everyday use.

6. Acknowledgements

The work was supported by the NKFP Grant No. 2/010/2004 (VITAL), the OTKA Grant on “Production structures as complex Adaptive Systems”.

REFERENCES

Cachon GP (2003) Supply Chain Coordination with Contracts. In: de Kok AG, Graves SC (eds.) Supply chain management: Design, coordination and cooperation, Elsevier: 229-339

Drótos M, Erdős G, Kis T (2007) Scheduling the mass production of a large variety of products: an industrial case study. Manuscript, to be published

Egri P, Váncza J (2006a) A logistics framework for coordinating supply chains on unstable markets. In: Digital Enterprises Technology – Perspectives and Future Challenges (eds. Cunha, P.F., Maropoulos, P.G.), 2007, Springer: 59-66

Egri P, Váncza J (2006b) Incentives for cooperative planning in focal supply networks, In: Proc. of 6th International Workshop on Emergent Synthesis (IWES 2006), pp 17-24

ElMaraghy HA, ElMekkawy TY (2002) Deadlock-free rescheduling in flexible manufacturing systems. CIRP Annals – Manufacturing Technology, 51/1:371-374

Hopp WJ, Spearman ML (1996) Factory physics – Foundations of manufacturing management. Irwin

Li X, Wang Q (2007) Coordination mechanisms of supply chain systems. European Journal of Operational Research 179:1-16

Monostori L (2003) AI and machine learning techniques for managing complexity, changes and uncertainties in manufacturing. Engineering Applications of Artificial Intelligence, 16/4:277-291

Monostori L, Váncza, J, Kis T, Kádár B, Viharos, ZsJ (2006) Real-time, cooperative enterprises: requirements and solution approaches, In: Information Control Problems in Manufacturing 2006, Elsevier, Vol. I, pp. 561-566

Stadtler H (2005) Supply chain management and advanced planning – Basics, overview and challeges. European Journal of Operational Research 163:575–588

Tseng MM, Lei M, Su C (1997) A collaborative control system for mass customization manufacturing. CIRP Annals – Manufacturing Technology 46/1:373-376

Váncza J, Egri, P (2006) Coordinating supply networks in customized mass production – A contract-based approach. CIRP Annals – Manufacturing Technology 55/1:489-492

Wilhelm R., et al. (2000) Finite schedule monitoring and filtering in a computer integrated manufacturing environment. CIRP Annals – Manufacturing Technology, 49/1:335-338