IS THERE A TRADE OFF BETWEEN LEAN AND AGILE MANUFACTURING? A SUPPLY CHAIN INVESTIGATION

Michael Caramanis, Haidong Pan, and Osman Anli

Boston University, Dept. of Manufacturing Engineering, 617-353-3247, mcaraman@bu.edu

Abstract: *Lean and agile manufacturing concepts are often used under the underlying assumption that they are interchangeable and always compatible concepts. However, they are often in conflict. Lean cell design, for example, may result in a cell that can meet a certain production mix and level with the least work in process (WIP), but WIP minimization may happen at the expense of agility to reconfigure in the event production requirements change. In today's environment where production requirements change frequently due to market conditions and increasingly common outsourcing practices, agility is tantamount to the manufacturer's ability to adjust to changing conditions rapidly and efficiently. This note motivates and summarizes research that has addressed cell coordination and production planning of cellular manufacturing facilities by addressing explicitly the variable, load-dependent delays (or lead-times) of WIP in coupled cells. In particular, time scale decomposition is used to coordinate cell-supply-chains. A facility-wide master problem interacts iteratively with multiple cell-level sub problems that engage in cell specific planning, performance evaluation and sensitivity analysis under uncertainty. The interaction involves exchange of information that results in superior production planning and rapid cell reconfiguration ability that is capable of significant efficiency gains relative to current industry practice.*

Extended Summary

Motivation

Modern manufacturing productivity and competitiveness is undoubtedly time-based [2]. As a result, the design of production facilities (Product Cell) and their operation (Just In Time, Zero In Process, Lean etc. [3,4,5,6]) aim at reducing lead-times and Inventories of either work in process or finished goods. Moreover, the design of manufacturing processes and in general the choice of manufacturing technology (automation, flexible computer controlled machines for single part as opposed to batch processing; selection of type of material, raw material, and intermediate goods used in production for delayed differentiation) have lead-time reduction as a significant, if not the only, motivation. Finally, information technology is looked upon as a means to reduce product design time and also to reduce inventory levels needed to provide a desired service level through supplier-consumer information sharing (for example knowledge of a retailer's inventory replenishment policy can help its supplier to manage its own inventory better.

Significant productivity gains in general and in lead-time reduction in particular have indeed been made so far by adopting the Japanese Model (simplifying the production process, introducing visual factory standards [5]), by utilizing advanced automation and information technology, and, last but not least, by exploiting the increased set of intermediate goods choices made possible through the expansion of trade. However, for all the advances in decreasing lead-times at individual links of the supply chain, potential improvements in global supply chain lead-times have not been captured. In particular, when an individual link of the manufacturing supply chain is flexible enough to respond to production requirement fluctuations modulating its own lead-time accordingly, the rest of the supply chain cannot benefit fully from this flexibility as long as it has no information of the individual link's lead-time dynamics. In fact, present practice is for the manufacturing supply chain to plan using a worst case or maximal lead-time assumption about each individual link. The assumption that lead-times are constant – at best revised periodically to a different constant value – is the basic premise of widely used Material Requirement Planning (MRP) production scheduling algorithms [18]. The MRP practice is a serious impediment to further productivity gains. Indeed, significant gains are possible when lead-time dynamics of individual supply chain links are accounted for in the overall manufacturing supply chain coordination through synergistic decentralized production planning. Better production planning algorithms that exploit information sharing are capable of taking the industry to the next big step in productivity gains. This paper argues that *time scale decomposition* for *enhanced information sharing* and better *deterministic fluid model approximations capturing the essence of the underlying stochastic dynamics* can indeed provide these productivity gains.

Production Planning and Control Issues

A series of production planning and control decisions and actions take place routinely in cellular manufacturing facilities. Arranged in order of decreasing magnitude of the average characteristic time between consecutive decision/action implementations, they can be grouped in four categories:

1. Investment and strategic decisions (alliances, service rate targets) [time scale, years]
2. Allocation of resources (machines, tools, manpower by skills, floor space) and production responsibilities (part types and associated rout segment operations) among cells; cell layout; production process; maintenance policy, and the like [time scale, months]
3. Plant-wide production and delivery promise schedule with cell-specific weekly production targets, cell-specific manpower and shift/overtime planning, contingency planning (safety stock, kanban size, and other contingency planning policies) [time scale, week]
4. Cell-specific day to day and hour to hour work center production dispatch, detailed operations scheduling, response to contingencies such as absenteeism, quality control, yield, delivery requirements/demand variations, failures and repairs, and short term horizontal coordination between interacting cells (e.g. sequencing of runs in the presence of significant setup delays) [time scale, hour to shift]

Due to the complexity of the information involved and the underlying dynamics, the decisions listed above are made in a decentralized manner based on objectives assigned to each decision maker that are designed to be compatible with overall enterprise goals. The various decision makers, however, operate in a less than integrated enterprise and are unable to capture the efficiencies that are possible when information accessible to each decision maker is transformed to knowledge and communicated to other decision makers. The proposed algorithm aims at (*i*) decreasing significantly the cost of knowledge, and (*ii*) using this knowledge for enterprise integration and higher manufacturing productivity realization. Results of numerical investigation are reported to elaborate the significant consequence on productivity of the lack of certain knowledge and the imputed cost of this knowledge in the context of today's state of the art production planning technology.

The main function of production planning is to convert the schedule of, say, weekly multiple final product deliveries (whether deterministic or stochastic) to a schedule of weekly intermediate product production and raw material procurement. The objective of production planning is to minimize inventory holding and backlog costs (i.e. to minimize lead-time and maximize service rate) subject to production system capabilities. Key knowledge required to achieve the goals of production planning is:
1. the capacity of each cell or link in the relevant production supply chain, and
2. the lead-time of each part type's/component's operations assigned to each cell or link in the relevant production supply chain.

Capacity, and in particular lead-time at each cell, are often complex functions of decisions made by production planning (i.e. loading and production mix of each cell, and part specific lot size/run in each cell during each week in the planning horizon) as well as by cell planning and cell detailed scheduling and operational policies (i.e. shift structure and overtime, kanban size or complex production constraints involving kitting requirements and or grouping of parts for system testing, time delay constraints, etc). As a result, knowledge of these functional relationships is very hard to obtain and use. In fact:
1. It is difficult to convert detailed production system information to the required knowledge. First the production system must be converted into a model and then the model must be used in either analytic or simulation mode to extract the requisite knowledge. Second, it is not only hard to create cell models – highly skilled analysts would have to be employed – but it is even harder to maintain them in the presence of frequent changes, to make them consistent, and to integrate them through a common data structure.
2. Furthermore, even when that knowledge is available, its non-linear nature still makes it hard to use in the production planning process.

As a result, state of the art production planning models such as i2's Rhythm and mathematical programming models documented in the literature, resort to the following approaches:
1. They use fixed lead-time estimates usually from simulation [23,37] or from historical observations [26]. Adoption of a fixed lead-time requires that a "worst case" estimate is used to avoid backlogs that are costlier than inventory holding. This results in higher than needed work in process and decreases the production system's responsiveness by increasing its overall lead-time. To actually achieve constant lead-times, many Japanese manufacturers operated in the 1980s routinely at low capacity utilization. This was made possible by excessive capital investment induced by the low cost of capital in Japan during that period.
2. They approximate capacity constraints by one (or a few) bottleneck nodes of the production system network and rely on robust theory of constraints algorithms to schedule production in the presence of complex production constraints and rules at each bottleneck

point [36,35,34]. Although often successful in modeling complex constraints, this solution is not only subject to excess work in process resulting from the constant lead-time assumption, but it also requires continuous extensive expert human involvement to identify and model frequently moving bottleneck nodes.

The proposed algorithm addresses both the knowledge acquisition and non-linear relationship utilization difficulties. It employs tractable decentralized cell level sub problems to address the knowledge acquisition problem, while a novel fluid model cell coordination algorithm utilizes a subset of targeted information about the complex nonlinear relationships robustly and efficiently to achieve significant productivity gains.

Architecture of Proposed Algorithmic Structure

A synergistic decentralized computation structure is proposed for enterprise integration. It builds on tractable decentralized cell sub problems and a fluid based master problem that operate at different time scales to optimize (or continuously improve) plant wide production planning, cell planning, and finally short term production scheduling that drives manufacturing execution.

Two layers of the architecture are shown in the table below. The yearly and monthly time scale investment and cell design layers mentioned earlier are not considered here for simplicity of exposition and because they are not a tightly coupled with the planning and execution layers. Production and cell planning are carried out on a weekly basis (Planning Layer) by a single master and multiple sub problems respectively. Detailed hourly production scheduling, execution, and stochastic disturbances (Execution Layer) are modeled in the sub problems.

Layer	**Functions**	**Time Scale of Dynamics**
Planning	<u>1. Plant Production Scheduling</u> (weekly cell production targets) and <u>2. Cell Planning</u> (shift structure, labor & overtime, safety stock and contingency planning policies)	<u>Weekly</u> production Schedule by master problem. <u>Weekly</u> Planning and time averaged performance from <u>hourly</u> sub problem dynamics.
Execution	<u>Short Term Cell Production Schedule Execution</u>. (weekly targets tracking, interacting cell coordination).	<u>Hour</u>

Significant literature [27,28,31] dedicated to the investigation of the stability of actual stochastic production networks driven by fluid model generated production targets, assures us of the viability of the execution layer. However, whereas larger (monthly or yearly) time scale decisions such as investment and machine allocation are loosely coupled with the execution layer, the planning and execution layers are tightly coupled. In fact, although the planning layer makes decisions that are implemented weekly, it needs to capture the highly non-linear impact on lead-time of production load, production mix, and lot sizes. This need introduces a strong interaction with the hourly dynamics of production execution, machine failures, and other contingencies, which must be modeled to provide the requisite weekly average lead-time performance. In other words, it is not sufficient for the hourly dynamics module to merely provide averages of linear relationships. Instead, it must provide non-linear or second order information on the decreasing returns to scale as desired production approaches full capacity. Since the implementation of an efficient and effective (tractable) planning layer constitutes the major proposed innovation, this paper focuses on its description and evaluation.

Novel Fluid Model Based Production Planning Algorithm

As discussed already, the effectiveness of enterprise level production planning depends crucially on the quality with which cell (or manufacturing component) capacity and cell dynamics are modeled. Whereas capacity is ignored and dynamics are modeled by constant lead-times in the vanilla version of MRP models [18], advanced planning and scheduling (APS) approaches include adequate representations of the dynamics and detailed representations of effective (or expected) capacity. In particular APS models rely on mathematical programming techniques and hierarchical decomposition [16,17] to overcome combinatorial complexity explosion barriers while capturing the details of capacity restrictions. This task is particularly onerous in the face of integrality of discrete part production, and complex production rules and constraints, which, together with uncertainty, render stochastic integer programming formulations computationally intractable.

Two main approaches have been employed to bypass these hurdles: theory of constraints and fluid model approximations. The theory of constraints [34,35,36] approximates the model of the production system by a small number of bottleneck components that are modeled in great detail; production is scheduled around those components through constraint propagation over time. A variety of scheduling algorithms ranging from mathematical programming and Lagrangian relaxation to genetic algorithms have been used, often effectively [10,11,12,13,14,20,19,24,30]. Two main shortcoming of the theory of constraints approach is first the difficulty in identifying and modeling bottleneck components and second the fact that delays or lead-time dynamics along part routs are nonlinear and difficult to model. The systematic modeling of cells could be possibly used to alleviate the first shortcoming. However, it is very difficult to accommodate the manufacturing dynamics modeling shortcoming. The fluid model approximation venue has been also used extensively and with considerable success [1,15,22,25,29,32,37]. It has been shown that deterministic fluid model approximations of stochastic discrete production networks can be employed to predict the qualitative nature of optimal scheduling rules [11,12,15,25] and to determine the stability and robustness of the approximated stochastic discrete networks [27,28,31]. It is this line of research that is exploited in the proposed algorithm.

The major objective of the proposed fluid model algorithm is to capture second order effects of the expected cell dynamics in order to model them effectively at the courser (weekly) production planning dynamics level where they are less variable because of law of large numbers effects and can be effectively modeled as deterministic. Furthermore, detailed information on machine specific queue and setup states are not globally available, hence it is practical to share state information which is (*i*) time averaged to the coarse time scale, and (*ii*) grouped by cell. To this end, machines and production requirements are grouped and associated with cells. Production planning dynamics at the coarse weekly time scale are then characterized almost fully by expected lead-times at each cell, for each part type produced at the cell. Lead-time, and its dependence on production planning decisions (loading, mix, set-up change frequency), constitutes the second order information which we have shown that when modeled it significantly improves the ability to decrease inventory and backlog costs. We next introduce notation used in the proposed decentralized fluid model approximation algorithm and relate it to the centralized stochastic problem of section 2. The proposed algorithm employs an iterative interaction of a single production planning master problem with multiple cell planning and execution sub problems. The production system is a network of cells much like the centralized model was a network of machines. Each cell, has its own microstructure and contains a specific layout of machines.

Planning and Execution Time Scales.

-The master problem (plant-wide production planning) time scale is a week, and the planning horizon consists of K weeks shown below, where t_j denotes week j.

I **I** **I**… **I** **I**… **I** **I**

t_1 t_2 t_k t_K

-The sub problem time scale is in hours (or minutes) where $t_{k,j}$ denotes hour j of a week k which contains J hours as shown below. Cell specific production execution takes place at the hourly time scale. Cell planning at the weekly time scale is also possible at each cell level sub problem.

I| | | |… | | … |**I**… | | … **I**… **I**| … | | … |**I**… **I**… | | … **I**

$t_{1,1}$ $t_{1,j}$ $t_{1,J}$ $t_{2,j}$ $t_{k,1}$ $t_{k,j}$ $t_{k,J}$ $t_{K,j}$ $t_{K,J}$

The discrete time variable $t_{k,j}$, is an approximation of the continuous time variable t used in the centralized stochastic model of section 2. Each sub problem is able to simulate hourly execution and return to the master problem weekly time averages of the cell's hourly performance, including capacity and lead-time constraints.

Fluid model implementations of Advanced Planning and Scheduling (APS) support tools have so far omitted modeling the highly non-linear second order effects of stochasticity on lead-time requirements. At best, lead-times have been modeled as constant [26] or imported exogenously from simulations [23,37], while the only endogenous constraint on production is the effective (or average) capacity. We depart from this limitation and model lead-time constraints fully. This is done through a convenient and very powerful transformation of the lead-time requirement to a constraint on average production which is generally restricted by an upper bound that depends on the average work in progress. This transformation is based on Little's law which relates the average weekly production rate of part r in cell c_i during week t_k, $X_r^{c_i}(t_k)$, to the corresponding average weekly work in process, $\overline{Q}_r^{c_i}(t_k)$ and the average Lead-time, $LT_r^{c_i}(t_k)$, required to produce at a particular loading level and mix during week t_k. Little's Law relationship is $LT_r^{c_i}(t_k) = \overline{Q}_r^{c_i}(t_k) / X_r^{c_i}(t_k)$. Two equivalent constraints denoted here as the *explicit* and *implicit* lead-time constraints are:

- *Explicit Constraint*: Delivery of raw material and components to cell c needed to produce $X_r^{c_i}(t_k)$ parts of type r during week t_k must be made at or prior to time $t_k - LT_r^{c_i}(t_k)$.
- *Implicit Constraint*: The average Work In Process during week t_k, must equal or exceed the level required to sustain average production $X_r^{c_i}(t_k)$ during week t_k, namely, $\overline{Q}_r^{c_i}(t_k) \geq LT_r^{c_i}(t_k) X_r^{c_i}(t_k)$ or $X_r^{c_i}(t_k) \leq \overline{Q}_r^{c_i}(t_k) / LT_r^{c_i}(t_k)$

The implicit constraint was used by Graves [26] with lead-times set to a fixed value -- presumably a worst case or maximal value -- for each cell. Noting, however, that $LT_r^{c_i}(t_k)$ is in fact a non-linear function of the production of all part types in cell c_i, and the transfer lot size for each part type, and recalling that $L_r^{c_i}(t_k) = \dfrac{X_r^{c_i}(t_k)}{f_r^{c_i}(t_k)}$, i.e. the transfer lot size depends on the production requirements and the set-up change frequencies of each part type, $f_r^{c_i}(t_k)$, and the production rates, $X_r^{c_i}(t_k)$, the implicit form of the constraint becomes:

$$X_r^{c_i}(t_k) \leq \text{E NonLinF}(\overline{Q}_r^{c_i}(t_k), X_1^{c_i}(t_k), X_2^{c_i}(t_k), \ldots, f_1^{c_i}(t_k), f_2^{c_i}(t_k), \ldots) \qquad (1)$$

The nonlinear constraint (1), must be modeled effectively for every week, every cell, and every part type processed in each cell. Its nonlinear character and the cost of the knowledge required to model it effectively, is a good explanation of the fact that it has not been implemented in existing APS software. Although its significance has been recognized and efforts have been made to capture its impact through successive cycles of simulation and adjustment of the constant lead-time input to the production scheduling algorithm [23], no complete and robust modeling of the lead-time related dynamics has been implemented to our knowledge in a closed loop setting.

The Master Problem is formulated as a Linear Program which is a deterministic approximation to the stochastic centralized (DP1) problem of section 2. Continuous time is discretized to weekly intervals, integer part production is approximated by the weekly production rate modeled as a real number, and stochastic delays due to starvation, machine failures and repairs, detailed production scheduling, demand variability, etc. are approximated by the average lead-time constraint (1) and the minimum safety stock constraints $\bar{I}_r^{c_i}(t_k) \geq \bar{I}_{r,\min}^{c_i}(t_k)$. The resulting problem solved by the iterative master sub problem algorithm is in essence *an optimal open loop control* solution of the original stochastic DP problem, and should not be mistaken as a certainty equivalent approximation of the original problem. Certainty equivalent approximations merely replace random variables by their expectations, while the optimal open loop problem considers the expectation of functions that depend on the random variables, the case in point here being the non-linear lead-time constraints. In fact, fluid model approximations employed in the literature and in practice are in essence limited to certainty equivalent approximations.

Algorithmic Implementation and Numerical Evaluation of Benefits

We have so far investigated appropriate algorithms and used them to study the effectiveness of lead-time dynamics modeling [7,8,9]. The algorithms investigated so far have relied on an iterative use of a plant-wide master problem and cell-specific sub problems with the following function:

1. The Master Problem calculates tentative production schedules and frequency of set up changes, and conveys them to cell sub problems along with the corresponding WIP values.
2. Each cell sub problem performs cell planning conditional on the production and set up change frequency or equivalently transfer lot sizes obtained from the master problem. Cell planning determine labor availability and minimum safety stock that must be maintained at inventory positions between cells. It also evaluates its performance under feasible production and set up change frequency targets that are as close as possible to those provided by the master problem. It finally returns to the master problem capacity constraint, a linear or piece wise linear local approximation of the lead-time constraint, and a linear or piecewise linear local approximation of the minimum safety stock constraint. This is done for each week and each part type processed by the cell.

Two broad algorithmic implementations have been implemented to deal with the representation of the non-linear lead-time and safety stock constraints at the master problem:

1. Strategic cumulative learning of the surface of nonlinear constraints and through generalized Benders' decomposition. Cutting planes or outer linearizations are introduced progressively until they become sufficiently dense in the neighborhood of the optimum and induce convergence to the optimum solution. This algorithm involves iterative interaction of an LP based master problem with performance evaluating cell sub problems that provide support planes by determining lead-times, minimum safety stocks and their sensitivity. It is effective when (*i*) cell design and production dispatch rules are reasonable and thus yield convex feasible sets described by constraints, and (*ii*) when no cell re-planning takes place during production planning updates. The computational complexity of the algorithm increases linearly in the size of the problem and its convergence rate is insensitive to problem size.
2. Local approximations of the nonlinear constraint surface. The approximations are not cumulative as in the Benders' algorithm. They are instead regenerated after each iteration. This method's computational complexity is also linear and converges to the optimum in about one and a half to twice the number of iterations needed by the previous method. However, the local approximation method is much more robust. It can handle non-convex feasible sets, and, more importantly, re-planning at the cell level. Re-planning is important as it indeed takes place. Re-planning actions at the cell level sub-problems result in shifts of the surface of the lead time constraints rendering the cumulative learning of the Benders' method inappropriate. Since the non-linear constraint local approximations are regenerated each time under this second method, surface shifting can be accommodated.

The value of information of the non-linear lead-time constraint has been investigated [7,8,9] for various systems with multiple unreliable work stations processing multiple part types in multiple cells. No setups and safety stock constraints were considered in the above references, but additional unpublished studies have shown that their inclusion accentuate the value of modeling this additional information. Results were compared to the state of the art APS algorithms with constant lead-times. To this end, the exact same problem formulation was used with the lead-time constraint replaced by $X_r^{c_i}(t_k) \le \overline{Q}_r^{c_i}(t_k) / \overline{LT}_r^{c_i}$ where $\overline{LT}_r^{c_i}$ are constants selected to correspond to the lead-time required for production to reach the maximum capacity utilization of 90% allowed in the capacity constraint. The comparison (see [7] for details) summarized below, is for a three-cell system (two subassembly and one final assembly cell) producing three different part types. Each cell consists of three machines. It is typical of results obtained on many different systems and reveals that the value of information is very significant and can amount to substantial decrease of lead-times and work in process.

Case Description	Nonlinear Lead-time Alg.	Fixed Lead-time Alg.
Level loading, utilization of 89.9%. Individual Product Mix Varies	73	99
Level Loading, utilization of 80%. Individual Product Mix Varies	16	48

Summary of Contribution

The main contribution of the proposed approach is that it overcomes the intractable nature of an exact centralized decision making problem formulation while capturing the essence of the stochastic manufacturing supply chain dynamics. Since the exact problem cannot be solved, the paper compares the performance of the proposed algorithm to current practice algorithms and demonstrates its ability to lead to significant cost reduction and productivity gains. In conclusion, agile response to production requirements contributes significantly to the effectiveness of cellular manufacturing. Given this capability, agility does not have to be sacrificed in the interest of narrowly construed lean manufacturing.

References

1. K. Egilmez, A. Sharifnia and M. Caramanis, (1996) Scheduling of a Circuit Pack Testing Facility with Complex Constraints: An Enhanced Fluid Model Approximation with Flow Nets, 1996 Japan-USA Symposium on Flexible Automation.
2. National Research Council, (1998) *Visionary Manufacturing Challenges for 2000*, National Academy Press, Washington DC.
3. Society of Manufacturing Engineers, L.R. Nyman, Editor, (1992) *Making Manufacturing Cells Work*, McGraw-Hill, New York, NY.
4. R. Suri, (1998) *Quick Response Manufacturing: A Company Wide Approach to Reducing Lead-times*, Productivity Press, Portland, OR.
5. K. Suzaki, (1987) *The New Manufacturing Challenge: Techniques for Continuous Improvement*, The Free Press, New York, NY.
6. J.P. Womack and D.T. Jones, (1996) *Lean Thinking: Banish Waste and Create Wealth in your Corporation*, Simon & Schuster, New York, NY.
7. M. Caramanis and O. Anli (1999) *Dynamic Lead-time Modeling for JIT Production Planning*, Proceedings of the IEEE Robotics & Automation Conference, Detroit, MI., May 10-15, 1999,Vol 2, p. 1450-1455.
8. M. Caramanis and O. Anli, (1999) *Modeling Work In Process versus Production Constraints for Efficient Supply Chain Planning: Convexity Issues,* Proceedings of the IEEE CDC, Phoenix, AZ, p. 900-906.
9. M. Caramanis and I. Paschalidis, and O. Anli, (1999) *A Framework for Decentralized Control of Manufacturing Enterprizes,* Proc. DARPA-JFACC Symposium on Advances in Enterprise Control, San Diego, CA, p. 99-109.
10. A. Sharifnia, M. Caramanis, S. B. Gershwin, (1991) *Dynamic Set-up Scheduling and Flow Control in Manufacturing Systems* Discrete Event Dynamic Systems: Theory and Applications, Vol. 1, No. 2, p. 149-175.
11. M. Caramanis and G. Liberopoulos, (1992) *Perturbation Analysis for the Design of Flexible Manufacturing System Flow Controllers,* Operations Research, Vol. 40, No. 6, p. 1107-1125.

12. G. Liberopoulos and M. Caramanis, (1995) *Dynamics and Design of a Class of Parameterized Manufacturing Flow Controllers,* IEEE Transactions on Automatic Control, Vol. 40, No. 6, p. 1018-1028.
13. C. Kaskavelis and M. Caramanis (1998) *Efficient Lagrangian Relaxation Algorithms for Real-Life-Size Job-Shop Scheduling Problems* IIE Transactions on Scheduling and Logistics, Vol. 30, No 11, pp. 1085-1097.
14. E. Khmelnitsky and M. Caramanis, (1990) *One-Machine n-Part-Type Optimal Set-up Scheduling: Analytical Characterization of Switching Surfaces,* IEEE Transactions on AC, Vol. 43, No. 11, p. 1584-1588.
15. M. Veatch and M. Caramanis, (1999) *Optimal Average Cost Manufacturing Flow Controllers: Convexity and Differentiability,* IEEE Transactions on AC, Vol. 44, No. 4, pp. 779-783.
16. A. C. Hax and H. C. Meal. (1999) "Hierarchical Integration of Production Planning and Scheduling" in S. A. Irani Ed., Handbook of Cellular Manufacturing Systems, Wiley, New York.
17. G. R. Bitran, and D. Tirupati, (1993). *Hierarchical Production Planning*. In S. C. Graves, A. H. G. Rinnooy Kan, and P. H. Zipkin (eds), Logistics of Production and Inventory, Vol 4, chapter 10. Amsterdam, N. Holland.
18. H. C. Meal, M. H. Wachter, and D. C. Whybark, (1987) *Material Requirements Planning in Hierarchical Production Planning Systems*, International Journal of Production Research, 25(7), p. 947-56.
19. S. Jain, M. Johnson, and F. Safai, (1996) *Implementing setup optimization on the shop floor* OR, 43(6), 843-51
20. J. Deleersnyder, T. Hodgson, R. King, P. O'Grady, and A. Savva, (1992) *Integrating Kanban Type Pull Systems and MRP Type Push Systems: Insights from a Markovian Model*, IIE Transactions, 24(3), 43-56.
21. D. Trietsch, and K. R. Baker, (1993) *Basic techniques for lot streaming*, OR, 41, 1065-74.
22. H. Chen, and D. D. Yao, (1993) *Dynamic Scheduling of a Multiclass Fluid Network*, OR, 41(6), 1104-15
23. M. R. Lambrecht, and P. L. Ivens, N. J. Vandaele, (1998) *ACLIPS: A Capacity and Lead-time Integrated Procedure for Scheduling*, Management Science, 44(11), 1548-61.
24. Goncalves, J. F., R. C. Leachman, A. Gascon, Z. Xiong, (1994) *A Heuristic Scheduling Policy for Multi-item, Multi-machine Production Systems with Time-varying, Stochastic Demands,* Management Science, V40 (11), p. 1455-68.
25. D. Bertsimas, I Paschalidis, and J. Tsitsiklis, (1994) *Optimization of Multiclass Queueing Networks: Polyhedral and nonlinear characterizations of achievable performance*, The Annals of Applied Probability, 4, No 1, p. 43-75.
26. S. C. Graves, (1996) *A tactical planning model for a job shop,* OR, Vol. 34, No. 4, p. 522-533.
27. A. Sharifnia, (1994) *Stability and Performance of Distributed Production Control Methods base on Continuous-Flow Models,* IEEE Transations on AC, Vol. 39, No. 4, p. 725-737.
28. P.R. Kumar and S. P. Meyn, (1995) Stability of queueing networks and scheduling policies, IEEE Trans. on AC, V. 40 No 2, p. 251-260.
29. H. Chen and D. Yao, (1992) *A fluid model for systems with random disruptions*, OR, Vol. 40, No 2, p. 239-247.
30. P Brandimarte and A Alfieri, (1998) *LP-based Heuristics for the capacitated lot sizing problem,* Anals of the CIRP, Vol. 47, No 1 p.423-426.
31. P. R. Kumar and T I Seidman, (1990) *Dynamic instabilities and stabilization methods in distributed real-time scheduling of manufacturing systems* IEEE Trans. on AC, Vol 35, No 3, p. 289-298.
32. D. Conors, G. Feigin, and D. Yao, (1994) *Scheduling semiconductor lines using a fluid network model* IEEE Trans. on Robotics and Automation, Vol 10, No 2, p. 88-98.
33. F. Murphy, (1994) *Comments on "The limits of Japanese production theory" by W. I. Zangwill*, Interfaces, Vol 24, No 5., p 74-94.
34. H Chen and A. Mandelbaum, (1991) *Discrete flow networks: bottleneck analysis and fluid approximations* Operations Research, p 408-447.
35. M. Lambrecht and L Decaluwe, (1988) *JIT and constraint theory: the issue of bottleneck management* Production and Inventory Management Journal p. 61-65.
36. E. Goldratt and J. Cox, (1984) The Goal: Excellence in Manufacturing, North River Press, 262 pp.
37. Y. Feng and R Leachman, (1996) A production planning methodology for semiconductor manufacturing based on iterative simulation and linear programming calculations, IEEE Trans. on Semiconductor Manufacturing, V. 9(3), p. 257-269.