Scheduling a Flexible Manufacturing System with Tooling Constraints: An Actual Case Study

SETH W. FLANDERS

Department of Industrial Engineering
Lehigh University
Bethlehem, Pennsylvania 18015

WAYNE J. DAVIS

Department of General Engineering
University of Illinois at Urbana-Champaign
Urbana, Illinois 61801

Material handling and the provision of supporting resources, particularly tools, in flexible manufacturing systems (FMSs) can constrain production. To improve the productivity for an existing FMS, we first developed a simulation model that included constraints arising from material handling and tool management. Detailed simulations of the FMS showed that a simple set of heuristic rules could produce several alternative schedules of nearly optimum performance. They also showed that when the flow of tooling was considered the FMS was operating near or over capacity. The goal of the study—to generate slack production time for the manufacture of additional parts—was unachievable. Clearly companies must consider the constraints arising from material handling and the flow of supporting resources in scheduling, yet these constraints have seldom been considered in simulations of FMSs.

Scheduling flexible manufacturing systems (FMSs) has been addressed by hundreds of technical articles, yet scheduling remains a major concern for most FMS operators. Numerous approaches to solving the problem have been attempted, including mathematical programming, heuristic searches, genetic algorithms, and simulated annealing, but all methods have limitations. Currently, the goal of developing a robust solution methodology to address the wide breadth of scheduling prob-

0091-2102/95/2502/0042$01.25
This paper was refereed

SIMULATION—APPLICATIONS
INDUSTRIES—TRANSPORTATION EQUIPMENT

lems arising in a flexible manufacturing environment remains unfulfilled.

Most published scheduling approaches focus upon job flow only. That is, given a set of jobs to be done, the scheduler seeks to discover the order in which to assign each manufacturing task to the available processes. Seldom do schedulers consider the assumed secondary concerns pertaining to material handling and the provision of supporting resources, such as tooling. These constraints are nearly impossible to express mathematically, and their inclusion in a simulation model is difficult to accomplish.

In 1990, Caterpillar, Inc. came to us with a scenario that appeared to present an ideal scheduling problem. The FMS under consideration was used to manufacture a repeated ensemble of parts each day. The manufacturing engineers had optimized the processing plans for the required part types so that every part type of a given material type (either steel or aluminum) could be produced using a single set of tools positioned at each machine dedicated to processing that material type. The goal was to determine the schedule that would maximize the availability of the FMS for the production of additional parts while insuring that the ensemble of required parts would be produced each day. While we initially believed that we could achieve this goal, our efforts to generate this schedule demonstrated otherwise.

Description of the FMS

Figure 1 shows the FMS. The system consists of seven Scharmann numerically controlled milling and drilling machines, one fixturing station where parts are introduced into and removed from the system, one tooling station where tools are intro-

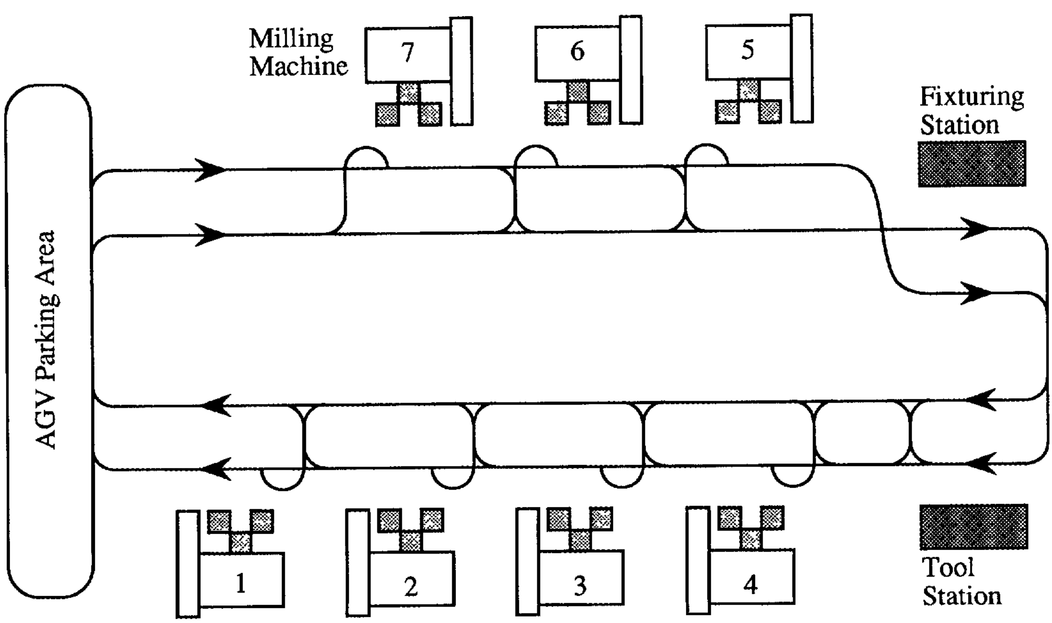

Figure 1: The modeled FMS includes seven milling machines, a fixturing station, and a tool station. Four AGVs travel along the one-way paths.

duced into and removed from the system, and four automated guided vehicles (AGVs) that deliver both jobs and tools to the stations.

The FMS produces several different transmission bell housings made of either iron or aluminum. The processing of these two material types must be segregated on different machines because different tools are used to machine the two materials and

Orders for replacement tools eventually become too frequent.

there is a desire to segregate the cuttings for recycling. At present, two of the seven milling machines are used to process aluminum parts while the rest are used for iron parts.

A typical part requires three distinct fixturing steps, each corresponding to a different orientation of the part. For each fixturing step, the milling machine will execute several machining instructions. The process engineer has designed the machining instructions so that any fixturing step for any part of a given material type (iron or aluminum) can be executed using the standard ensemble of tools provided at each machine dedicated to that material type.

The AGVs deliver tools using the same network they use for parts. To accomplish a tool delivery, the tool setter first places the replacement tool on a tool carrier that can hold up to 16 tools. After the AGV delivers the tool carrier to the milling machines, the machine's worktable positions the tool carrier so that the tool can be placed into the machine's collet. The carrier is next withdrawn, and the tool exchanger unloads the tool from the collet and places it in the machine's tool magazine. This process is repeated for each replacement tool on the carrier. The procedure is reversed to remove worn tools from the tool magazine and to place them on the tool carrier. The AGV finally returns the tool carrier with the worn tools to the tool station. The machine's tool magazine can store up to 80 tools, but an oversized tool may require several positions in the magazine.

The tool station is responsible for refurbishing and storing tools. Each tool has a different maximum life, which is defined as the total machining time the tool should be able to withstand without becoming so fatigued that it breaks or so worn that it will not cut properly.

Because the AGV tracks are one-way only, travel distances between points within the system are direction dependent. For example, the travel distance from machine 1 to machine 2 is 374 feet, while the travel distance from machine 2 to machine 1 is only 52 feet. Clearly, the scheduler must consider travel distances in determining which AGV will respond to a given transportation request. Finally, the AGVs are capable of carrying only one part pallet or tool carrier at a time.

At each machine are two on-line queuing positions for parts that are waiting for processing, parts that have finished processing, or tool carriers. Each machine also has two off-line buffer positions where parts waiting for processing can be stored. These off-line areas do not interact with the work area; the AGV must move a stored entity from an off-line queue into an on-line queue. For this reason, the off-line positions are typically used for storing only pallets that are not in use, not for storing fixtured parts or tools needed in production.

There are 24 pallets available to the system. The pallets are metal platforms that have been specifically designed to be handled by loading systems for the AGVs and the machining centers. Each pallet employs a custom fixture to hold a part during the machining process. Some pallets can carry only one specific fixture of a given part type, but others can carry multiple fixtures, allowing more than one part at a time. For the multiple fixtured pallet, usually all three fixturing steps for a given part type are processed on the same pallet.

The pallets enter the system through the fixturing station. Following a predefined production schedule, the fixturing station takes parts from bins behind this station, mounts the part(s) on the pallet, and places the pallet in a queue awaiting introduction into the system. If a processing machine cannot currently accept a pallet in its on-line queue, the AGV can take the pallet to one of the 14 off-line buffer stations within the system or the pallet can wait at the fixturing station until an on-line queue position at the assigned milling machine becomes available. After the parts have completed the processing required by a given fixturing step, they return to the fixturing station where they are either refixtured or exit the system as finished parts.

Problem Statement

Currently, the company uses the FMS to produce a standard ensemble of bell housings each day. After reviewing the performance statistics for the FMS and the previous simulation studies, the manufacturer had concluded that the FMS was being underutilized. The company asked us to define an improved production schedule that would permit it to produce the daily requirement of parts in the most efficient manner so that the resulting slack (unscheduled) production time could be reassigned to produce additional parts. Because the company wanted to produce the same ensemble of parts each day, we expected that we could define a schedule that could be repeated daily.

Our first task was to define an objective function for the scheduling problem. We chose minimizing the makespan (the total time required to produce the daily ensemble of parts) as the performance criteria for the study. From the outset, we knew that the two milling machines dedicated to producing aluminum parts were underutilized and would not affect this performance criterion. Our major concern was scheduling the five milling machines dedicated to producing iron parts. By minimizing the makespan, we thought that we could maximize the utilization of the iron machines. Furthermore, since the milling machines were completely interchangeable (every machine could process every fixturing step for every iron part), we also believed that minimizing the makespan would balance the utilizations among the milling machines dedicated to the production of iron parts to the greatest extent possible. The difference between the determined minimum daily makespan and the total available production time for each day would be the maximum possible slack time that could be assigned to the production of parts beyond those contained in the required daily ensemble.

Developing a Detailed Simulation Model

To determine the makespan required to implement a given production schedule, we developed a detailed simulation model for the FMS using SLAM [Pritsker 1986], which was requested by the company sponsor. In developing the SLAM model, we relied upon experience that we had gained developing a similar simulation model using SIMAN [Pegden, Shannon, and Sadowski 1990] to consider the tooling constraints for an FMS operated by the US Army Rock Island Arsenal [Dullum and Davis 1992; Hedlund, Davis, and Webster 1990]. In both FMS models, detailed consideration of the processing plans for each part type was essential. We also made a complete accounting of the remaining tool life for each tool residing at a given milling machine as each processing task was implemented. Whenever tools became worn, the model generated orders for replacement tools and incorporated the steps needed to ready the replacement tools at the tool station and deliver them to the machine via the AGV system. We also included the detailed mechanics of loading replacement tools from the tool carrier into the machine and removing worn tools.

Using SLAM, we considered the dynamics of the AGV system including the control logic for AGV assignment. However, we had only limited documentation for the AGV control logic because it was contained in proprietary computer code, which was not available for analysis. Perhaps our most significant deviation from real-world operating conditions in the model was omitting both machine breakdowns and the occurrence of system deadlock. Watching the operation of the FMS, we observed several modes of system deadlock. For example, an AGV with a low battery might wait to be recharged in the middle of a primary track and block passage of the other AGVs.

That there are only two on-line buffer positions at each machine also caused problems. In more than one instance, the AGV delivered a second input part to a milling machine while another part was being processed. When processing was completed, the completed part could not be unloaded from the work area because no on-line buffer positions were available. In this situation, the FMS operator had to direct the controller to bring an AGV to the deadlocked machine to remove one of the waiting parts to an off-line buffer position, allowing the finished part to be removed from the work area. During this intervention, most other material handling was delayed. In some instances, it took over 30 minutes to restore a normal operating condition. Since we did not model breakdowns or system deadlock, we realized that the detailed model would overestimate the actual performance of the system.

The detailed simulation included a dispatching mechanism that insured that parts were assigned to machines in the exact order dictated by the schedule. The development of this mechanism was complicated by the fact that sometimes two or three different parts had to be mounted upon a fixture before it was dispatched to the machine.

In addition to maintaining the scheduled processing order and machine assignments, the dispatching mechanism prevented pallets from being transferred to their assigned milling machines until it ascertained that tools with enough tool life needed to process the fixtured part(s) were available. After dispatching the pallet, the dispatching mechanism computed the life that would remain for each tool after the pallet of fixtured part(s) had been processed. Using this remaining life, it reviewed the processing plans to determine which fixturing steps would exceed the maximum tool-life of any tool currently positioned at the assigned milling machine and ordered the necessary replacement tools. The dispatching mechanism then prevented any part that violated the maximum tool-life constraint for any tool from being assigned to that machine until the essential replacement tools were replaced.

Since the tool magazines at the machines were nearly full, replacement tools could not be delivered until the currently assigned parts had been processed. Replacing tools before assigned parts were finished would waste available tool life and ultimately result in more tool deliveries. Because the tools are delivered by the AGV system and introduced into the machine through its work area using the same spindle and tool exchanger used in processing, tool deliveries decrease both the availability of a given machine for production and the availability of the AGV system to transfer parts. Veeramani, Upton, and Barash [1992] discuss the many issues associated with tooling management in the modern manufacturing environment and survey the related literature.

The FMS operator may have suspected that tool management and handling was the primary constraint.

Selection of an Optimization Approach

From the outset, it was clear that we needed a detailed simulation to accurately assess the internal dynamics associated with operating the FMS. As we constructed the simulation model, we realized that employing a mathematical programming approach for scheduling the FMS would require too many simplifying assumptions. We would have to ignore constraints pertaining to material handling, tool management, and the details of the fixturing process to use this approach because it is nearly impossible to express these constraints in the form of mathematical equalities and inequalities. Even if we could formulate the mathematical programming problem, we would still face the task of finding the optimal solution.

Since simulation models provided the only viable means to express these constraints, we sought an optimization procedure that would permit direct consideration of the simulation model. We eliminated experimental factorial design-based methodologies because we would need too many variables (factors) to define the production schedule. Davis and Stubitz [1987] have shown the difficulties that can arise in attempting to optimize a performance criteria (response) over a discrete-event decision space. Their example considered only seven factors.

We did briefly attempt to apply genetic optimization [Goldberg 1989] to define the optimal schedule. However, as we simulated the generated populations of schedules, we made a series of observations that indicated that a simple set of heuristic rules could generate a set of alternative schedules with a nearly optimal makespan. First, the simulations confirmed that machining aluminum parts did not constrain the overall makespan. Second, five machines were assigned to producing steel parts. If each machine is utilized at a 75-percent level, the probability that all five machines will be busy at any moment is approximately 24 percent. Even at 80-percent utilization, the probability of all machines being busy at a given time is less than 33 percent. Given that any machine dedicated to producing steel parts can process any fixturing of a steel part, machine availability is not the issue. Whenever any part is fixtured, a machine is likely to be available (or nearly available) to process it.

Intuitively, we could see other obvious policies that should be adopted. Pallets capable of carrying multiple fixtured parts should be filled whenever possible to minimize the number of their trips. Second, we should attempt to keep the machines dedicated to producing steel parts busy by giving priority to fixturing a steel part over fixturing an aluminum part. To minimize the overall makespan, we should keep the process utilizations across the five machines nearly equal. Using these simple heuristics, we could generate many schedules requiring a makespan of approximately 1,270 minutes (88.3 percent of the available daily production time). For this makespan, the average utilization for the iron machines was approximately 85 percent, which is near the feasible upper limit for most machines in operating FMSs.

The Real Production-Scheduling Problem

Given this discovery, we began to ask ourselves why the company was having a problem. The heuristic was so simple that the company must have discovered it during the years it operated the FMS. Still, the company was having problems in meeting the minimum daily requirement, and it often needed overtime production on the weekends to insure parts would be available for the assembly area. We had neglected several considerations in the simulation model, including breakdowns and deadlocks. Also we had not assessed lost production time during the shift changes. Nevertheless, about 12 percent of the available production time could serve as slack time to cover these omissions.

However, we had not tested one major assumption made in our simulation model. We had assumed that each simulation run would begin with a new ensemble of tools at each machine. During the simulation runs for schedules with a 1,270 minute makespan, less than 20 tools were replaced. Obviously the number of tool replacements depends upon the remaining life of the tools at each machine at the beginning of the simulation run. Since we had no basis for assigning this initial configuration, we chose to allocate a complete set of new tools for each simulation run considered during the optimization.

To test the effect of this assumption, we ran a 20-day simulation run. Figure 2 shows the slack time left each day after meeting the daily requirement. In performing the 20-day simulation study, we assumed that any slack time generated each

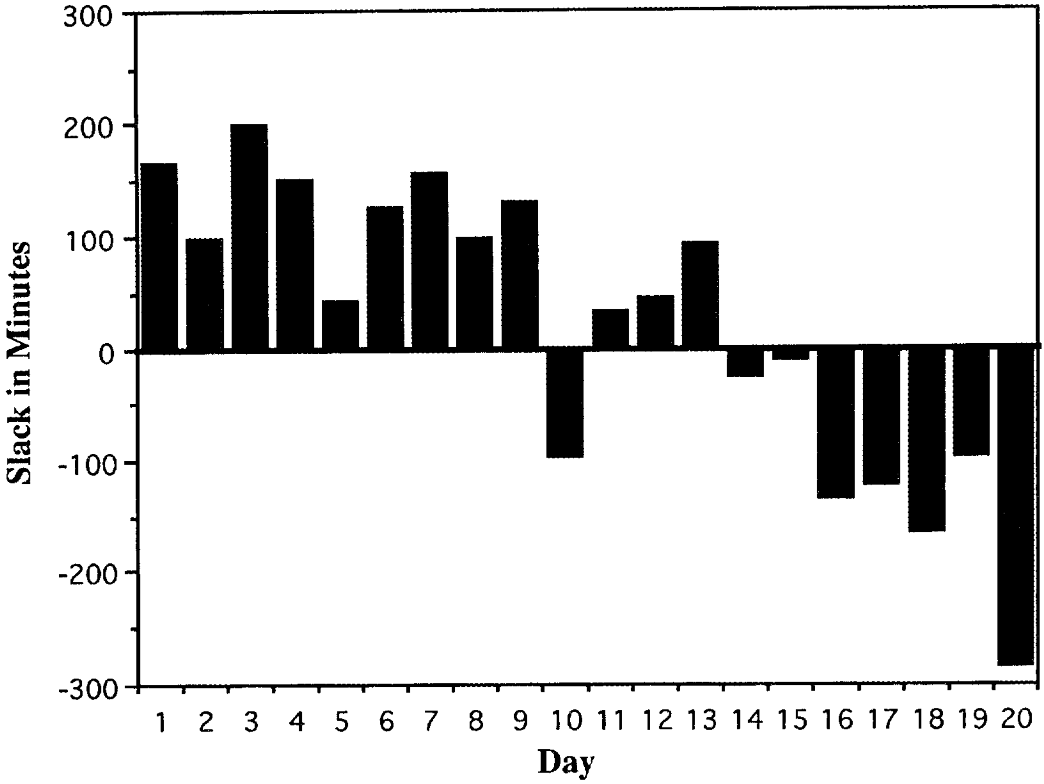

Figure 2: This figure gives the observed slack times during the 20-day simulation run. A negative slack time implies that the simulation took longer than the available time to complete the daily production. The system becomes unstable after the 14th day.

day would be used to produce additional parts Therefore, when the simulation finished a given day's production, we advanced the simulation time to the beginning of the next day. Slack time was available for reassignment during the first nine days. On the tenth day, the FMS needed nearly 1,550 minutes to produce the daily requirement. This implied that the 10th day's production spilled over into the 11th day, which fortunately generated enough slack time to complete the 10th day's requirement and do the 11th day's requirement and leave a small amount of slack for additional production. The system then returns to a situation in which there was slack time each day. By the 14th day, however, it became unstable; the system got behind and never caught up. The only source of variability in this simulated scenario is the remaining life for each tool at the beginning of the day. That is, the simulation time assumes that the processing times are deterministic and that the outcome of each processing task is a successful completion. Because of the level of detail in this simulation model, there are few (if any) random variables.

The above simulation results are conservative in their projection of the consequences arising from the consideration of tooling. First, there is no way that the current simulation study could account for the tool wear resulting from assigning slack production time to the production of other parts. Hence, using the slack production time would wear out tools and require additional orders for replacement tools beyond those observed in the 20-day simulation study. Second, we assumed that only orders requiring more than one replacement tool would use the AGV delivery system. An order for only one tool would be delivered and installed by the operator with no production downtime. The company used this replacement policy under some circumstances, and we decided to adopt it for all single tool orders since the company could not define the conditions under which it performed manual transfers. We conducted several experiments to determine the cause of the declining slack. We collected statistics on machine utility, cart utility, tool-carrier utility, and the number of tools ordered per day. After these experiments were performed, the most significant factor appeared to be the tooling replacements ordered by the machines. The number of replacement tools increases each day over the 20-day period and the number of multiple tool replacements (requiring the use of the AGV delivery system) also increases (Figure 3).

We believe that the increasing number of multiple tool orders is the main cause of the declining slack. When a machine needs a tool for processing, it will be effectively shut down until that tool arrives. Production time is lost every time a tool carrier of replacement tools are loaded into the machine magazine and worn tools are removed. As the initial condition deteriorates

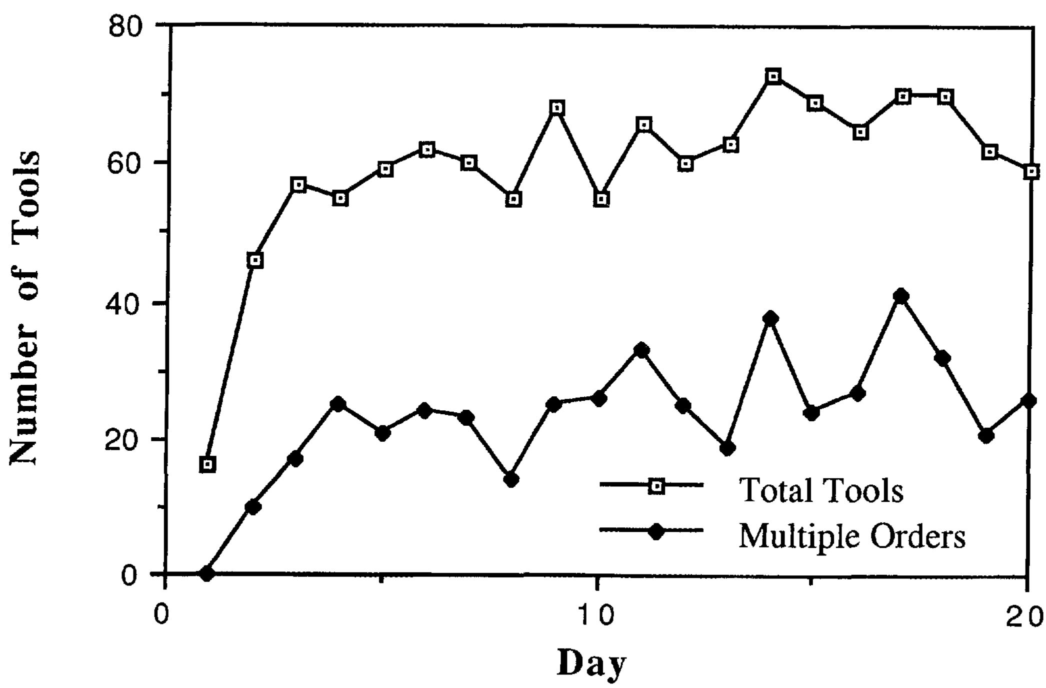

Figure 3: The total number of tools ordered and the number of orders requesting multiple tools each day increased during the 20-day simulation study.

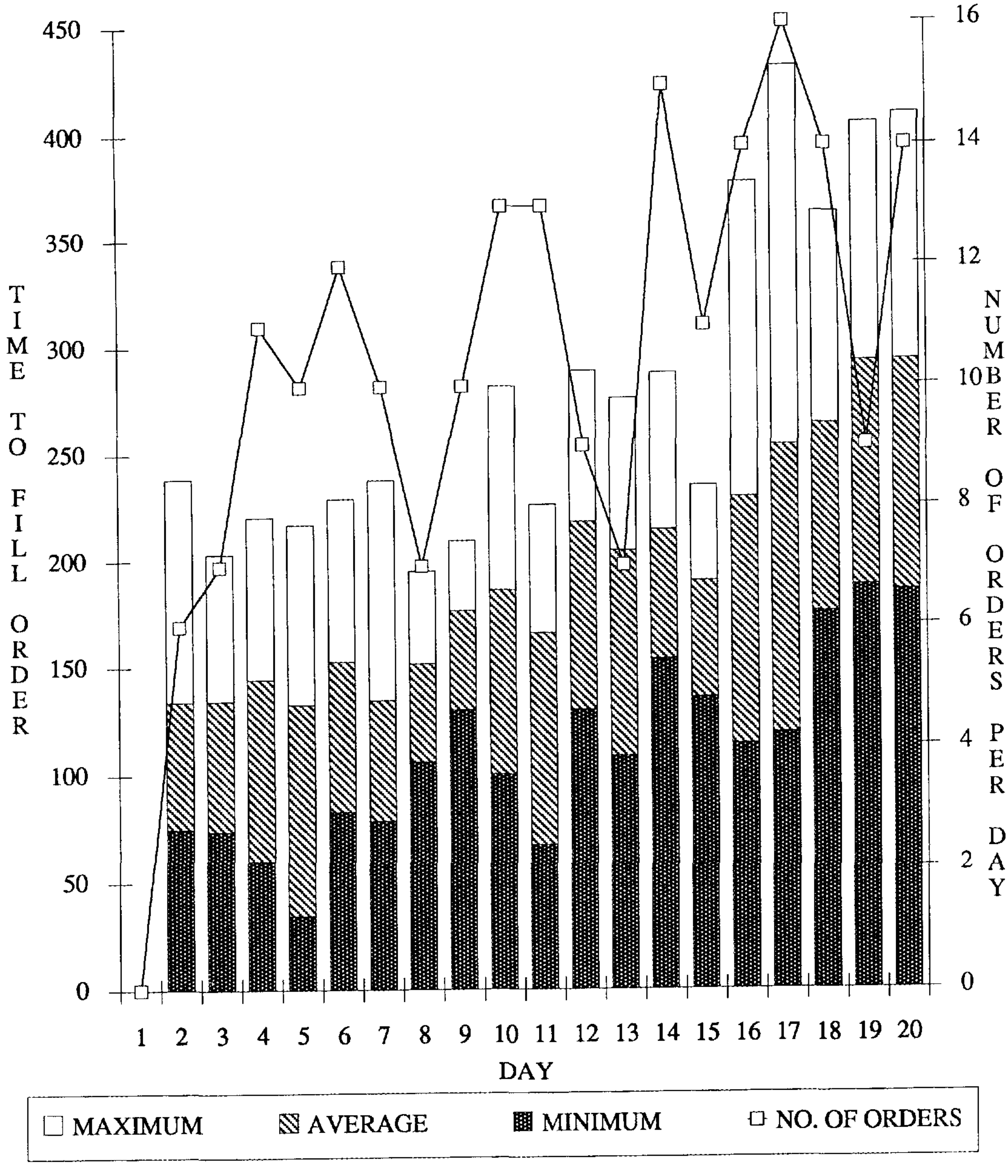

Figure 4: The line graph gives the number of orders requesting multiple tools for each day of the 20-day simulation study. The bar graphs depict the statistics pertaining to the time required to fill the tool orders on each day, including the minimum observed time, the maximum observed time, and the average time.

and tools wear, the number of requests for tools rises dramatically. Orders for replacement tools eventually become too frequent, and the time it takes to fill a tool request also rises (Figure 4). The number of multiple tool orders per day and the time required to fill tool orders are correlated. The order is generated after a part is fixtured, and the replacement order cannot be put in the machine's magazine until the part that initiated the order is finished. Also, only three tool carriers are available to the system, which becomes a production constraint when several orders for replacement tools arrive in a short period. In addition, tool orders must wait for a single tool setter to install cutters on the replacement tools and put them on the carrier for delivery.

After making these observations, we counted the replacement tool orders for each tool over the 20-day period. We found that eight tools constituted almost 60 percent of the tool orders. We then decided to place two of these tools in the tool magazines to see if this would increase slack time. We ran this configuration for a 20-day period, assuming again that the slack time would be allocated to other production.

Placing redundant tools in the magazine allowed the system to run smoothly for a longer period, but the number of tool orders still increased beyond the capacity of the system. The 30-day trial ended three minutes over the due date of 28,800 minutes. The average slack per day was 92 minutes, but four days in the period were still so overburdened with tool orders that they exceeded the time allotted for daily production (1,440 minutes). Thus, placing spare tools on the machines just delays the time when the system becomes saturated and then unstable.

The simulation also tabulated the percentage of downtime on each machine that can be attributed to tool replacement. With no duplicate tools, the five machines average 17.3 percent of the production time waiting for tools to be replaced (a maximum value of 25 percent was observed at one machine). With duplicate tools, the five machines spent an average of 8.6 percent of the production time waiting for tool replacements. This downtime is clearly a significant detriment to producing parts within the specified makespan. For the single-day simulation runs in which we assumed all tools were new, the number of requests for multiple tool orders were minimal (single tool orders incur no loss in production time), and the machines took 88 percent of the available production time to complete the daily production requirement. With average downtime due to tooling replacement of 17.3 percent, the FMS, even with duplicate tooling at each machine, operates near or over saturation. And we have not modeled breakdowns or the occurrence of deadlocks, which further diminish the available production time. It is clear why the company currently needs overtime to keep up with its daily production requirements.

Conclusions and Recommendations

We learned a great deal from our retrospective analysis of this scheduling exercise. Many troubling issues can be raised. An important question is, has this exercise helped the owner of the FMS? The answer is probably no. We did ascertain the true bottleneck to production, but we did not provide any solutions. The FMS operator may have suspected that tool management and handling was the primary production constraint before the study was authorized and hoped that we could prove otherwise.

The true scheduling concern for this FMS lies in the management of tooling. In our simulations, the tool handling constraints eventually made the system unstable, regardless of whether or not the slack production was reassigned to machine additional parts beyond the minimum daily requirement. In the real world, this system does not become significantly unstable because the company suspends production on the weekends and performs system maintenance to replace worn tools. In addition, it can use overtime production on weekends to catch up with production requirements.

It is very difficult to formulate an optimization problem to assess the optimality of a general tool management policy. We would first specify the set of alternative policies that can be considered and the planning horizon. Considering an extended planning horizon would greatly increase the simulation time required to evaluate each policy.

The tooling should not be the only area of focus, however, because the FMS is not an entity unto itself. The dynamics of part arrival and the demands of the downstream assembly areas must also be taken into consideration when this model is used. The variability of daily demand should be considered. In addition, if the company uses the slack time provided through improved scheduling to produce other parts, then these additional parts must also be specified because their production will likely affect tool replacement. All of these concerns must be addressed to develop a model capable of assessing the true system dynamics.

For the modeler of FMSs, we have an additional observation. Few simulation studies of FMSs consider the flow of such supporting resources as tooling Had we not considered tooling in this simulation study, we would have overlooked a major production constraint. This is not the first FMS we have modeled in which the flow of supporting resources is the primary bottleneck. In fact, our experience indicates that the flow of jobs is becoming a secondary concern in modeling FMSs. For every FMS we have modeled, the flow of supporting resources is much more complex and is not understood.

Most scheduling approaches focus upon job flow only.

There is a fundamental rule in optimization. Whenever a new constraint is appended to a decision, the ability to optimize the decision is at best unchanged and more likely, it is diminished. This implies that whenever operational constraints are not included in a simulation model, the simulation model will probably overstate the expected performance for the modeled system. In our collaboration with industry, we find a continued concern that FMSs cannot provide the production capability predicted by simulation studies. Invariably the cited simulation studies have ignored the constraints associated with the flow of supporting resources.

Furthermore, most FMS simulation studies assume that there is a steady state for operating the system. This is not a valid assumption. There is simply too much variability in the operation of FMSs to assume that a steady-state operating condition can be maintained over an extended horizon. Therefore, we conclude that any detailed scheduling must be performed on a near-term basis. To analyze proposed near-term schedules, one must use real-time, discrete-event simulation [Tirpak, Deligiannis, and Davis 1992]. Davis et al. [1993] state additional concerns in the modeling, scheduling, and control of FMSs and then provide an integrated solution approach to address all three tasks.

Acknowledgment

This research was supported by Caterpillar, Inc. with a grant to the University of Illinois Manufacturing Research Center.

References

Davis, W. J. and Stubitz, S. 1987, "Configuring a kanban system using a discrete optimization of multiple stochastic responses," *International Journal of Production Research*, Vol. 25, No. 5, pp. 721–740.

Davis, W. J.; Setterdahl, D.; Macro, J.; Izokaitis, V.; and Bauman, B. 1993, "Recent advances in the modeling, scheduling and control of flexible automation," *Proceedings of the 1993 Winter Simulation Conference*, eds., G. W. Evans, M. Mollaghasemi, E. C. Russell, and W. E. Biles, pp. 143–155.

Dullum, L. M. and Davis, W. J. 1992, "Expanded simulation studies to evaluate tool delivery systems in an FMC," *Proceedings of the 1992 Winter Simulation Conference*, eds., J. J. Swain, D. Goldsman, R. C. Crain, and J. R. Wilson, pp. 978–986.

Goldberg, D. E. 1989, *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley Publishing Company, Reading, Massachusetts.

Hedlund, E.; Davis, W.; and Webster, P. 1990, "Using computer simulation to compare tool delivery systems in an FMC," *Proceedings of 1990 Winter Simulation Conference*, eds., O. Balci, R. P. Sadowski, and R. E. Nance, pp. 641–648.

Pegden, C. D.; Shannon, R. E.; and Sadowski, R. P. 1990, *Introduction to Simulation Using SIMAN*, McGraw-Hill, New York.

Pritsker, A. 1986, *Introduction to Simulation and SLAM I*, John Wiley and Sons, New York.

Tirpak, T. M.; Deligiannis, S. J.; and Davis, W. J 1992, "Real-time scheduling in flexible manufacturing," *Manufacturing Review*, Vol. 5, No. 3, pp. 193–212

Veeramani, D.; Upton, D. M.; and Barash, M. M. 1992, "Cutting-tool management in computer-integrated manufacturing," *International Journal of Flexible Manufacturing Systems*, Vol. 3, No. 4, pp. 237–265.