ELSEVIER

#11988

Computers in Industry 1476 (2000) xxx

COMPUTERS IN INDUSTRY

www elsevier nl/locate/compind

Evaluation and comparison of production schedules

Karl Kempf [a,*], Reha Uzsoy [b], Stephen Smith [a], Kevin Gary [c]

[a] Intel, MS CH3 68 5000 W Chandler Boulevard, Chandler AZ 85226, USA
[b] School of Industrial Engineering, 1287 Grissom Hall Purdue University, West Lafayette, IN 47907-1287, USA
[c] Department of Computer Science Arizona State University Tempe, AZ 85287, USA

Abstract

The understanding of what constitutes a "good" production schedule is central to the development and evaluation of automated scheduling systems and their implementation in real-world factories In this paper, we provide a definition of a schedule and discuss potential uses for a schedule within the organization We then describe a number of different considerations that must be taken into account when assessing the quality of a schedule, and discuss their implications for the design and implementation of scheduling systems

Keywords Production schedule, Real-world factories, Design and implementation

1. Introduction

Although there has been a vast body of work on production scheduling in both the technical literature and industrial practice [36,50,58], the problem of assessing the quality of a given production schedule does not seem to have been studied extensively to date However, a clear understanding of how the quality of a schedule is assessed (i e , what constitutes a "good" schedule) is critical to the successful implementation of scheduling systems in real-world manufacturing environments Unless we can compare the schedules generated by a given system, either automated or manual, to those generated by alternative systems in some objective and quantifiable way, we will lack a systematic framework for evaluating the performance of scheduling systems and their impact on the performance of the manufacturing system as a whole Our goal in this paper is to describe some of the issues involved in assessing the quality of production schedules in order to bring them, and their implications for the development and implementation of scheduling systems, to the attention of the scheduling community and encourage research towards their effective resolution While many of these issues appear obvious when stated, many of them have not been considered in much of the scheduling work done to date Thus, there appears to be considerable benefit in presenting them within a coherent framework

We begin by discussing the nature of a production schedule — what it is, some basic characteristics of schedules and how schedules are used in a manufacturing facility We then examine the problem of assessing schedule quality from several different perspectives We conclude the paper with a discussion of the implications of the issues raised in the paper

* Corresponding author

0166-3615/00/$ - see front matter
PII S0166-3615(99)00071-8

for the development of automated scheduling systems and their implementation in practice

2 What is a schedule?

The most general definition of the *scheduling problem* is that of assigning scarce resources to competing activities over a given time horizon to obtain the best possible system performance In this paper, we will focus on the problem of *factory scheduling*, where the resources are machines and the competing activities are jobs that require processing on the machines Thus, we shall refer to a workpiece or batch of workpieces requiring processing at several different workcenters as a job The processing performed on a job at a particular machine will be referred to as an operation Hence, each job requires a number of operations, whose order is often specified by the technology in use or the part geometry A number of different machines may be capable of performing a given operation on a particular job While in a manufacturing environment, there may be many different resources such as operators and tooling to be considered, in this paper, we shall restrict ourselves to machines as the main resource to be scheduled

Over the last several years, an increasing number of authors have advocated taking a broader view of the scheduling problem There is increasing agreement that in addition to the classical scheduling decision of what operation to process on what machine when, decisions such as the timing and quantity of order release [53], due date quotation [10] and lot sizing [52] are closely related to the factory scheduling problem There is no doubt that in many environments the degree to which an effective solution to the scheduling problem is possible, or analogously, the degree to which the performance of the factory or company is affected by the quality of the solution to the scheduling problem, is significantly affected by these and other related decisions such as master production scheduling and multi-plant coordination Wein [64] and Wein and Chevalier [65] illustrate clearly the interrelation between planning decisions such as work release and due date setting and the effectiveness of scheduling policies This body of work raises the question of where to draw the boundary between the planning and scheduling problems In this paper, we shall assume, following Conway et al [12], that planning decisions specify what is to be made and when it is to be made by, specifying the mix of jobs that will be introduced into the shop and the times at which the jobs must be completed Conway et al [12] refer to these problems as "problems of pure sequence"

Given this definition of the scheduling problem, we shall consider a *schedule* to consist of a set of start times and machine assignments for each operation of each job to be scheduled In a manufacturing facility, the input to a scheduling system is generally derived from the current location of jobs in the system, the process plans describing the sequence of operations each job needs to undergo, due dates for the jobs and the state of the machines to process them (e g , up or down, current setup status), together with some estimates of uncertain events that may occur over the time period in the future covered by the scheduling decisions to be made The output from the system will be the set of job/machine/time assignments for a given time horizon

A schedule may be used on the shop floor in several different ways Generally, a schedule is intended to produce certain patterns of behavior in the manufacturing facility for which it was generated That is, the schedule is designed to induce the occurrence of a planned set of events on the shop floor We shall refer to a schedule as a *predictive schedule* when it is released to the shop floor at a certain point in time with the intention of guiding system behavior over a given time horizon The period of time into the future covered by the decisions in the current predictive schedule will be referred to as the *schedule horizon*

How strictly a predictive schedule is adhered to will vary from one manufacturing environment to another, based on factors such as manufacturing technology and organizational structure In a strict hierarchical organization, management may insist on a predictive schedule being followed to the letter, whereas in a more distributed, team-based organization, a schedule can be viewed as advice to the personnel on the manufacturing floor, who are encouraged to take advantage of opportunities for improving it that may arise In an integrated automated environment, a schedule may carry operational semantics, which directly drive the manufacturing hardware.

However, whatever the environment, there may be substantial deviations from the predictive schedule over the course of its execution due to unforeseen disruptions such as machine breakdowns or shop-floor personnel overriding the predictive schedule. The process of modifying the predictive schedule in the face of executional disruptions is generally referred to as *reactive scheduling* or *rescheduling*. The nature of the schedules developed in reaction to disruptions depends on the nature of the realized disruptions and the capabilities of the execution agent reacting to them. The reaction generally takes the form of either modifying the existing predictive schedule, or generating a completely new schedule, which is followed until the next disruption occurs. The issue of reactive scheduling will be discussed later in the paper.

The existence of executional disruptions and the necessity to react to them brings the necessity to distinguish between a predictive schedule, which is a planned schedule prior to execution, and the schedule actually executed on the shop floor. An actual set of job/machine/time assignments realized on the shop floor will be referred to as a *historical schedule*. We can think of the predictive schedule as the schedule as designed, and the historical schedule as the schedule as executed. Clearly, both schedules are of interest from the point of view of schedule quality assessment.

Another potential use for a schedule is as a yardstick by which to measure the performance of shop-floor personnel. In this situation, a predictive schedule is used to set goals, which the shop-floor personnel should achieve. The performance of shop-floor personnel is evaluated at the end of one or more planning periods using the deviations of the historical schedule from the predictive. This use of schedules is important in that it affects the way shop-floor personnel will react to unforeseen disruptions, influencing the evolution of the historical schedule as distinct from the predictive. Najmi and Lozinski [46] give an example of this use of predictive schedules.

There are a number of other potential uses for a production schedule. It may serve as a means of determining system capacity for a high-level production planning system, where a schedule is generated to determine whether in fact the production plan suggested by more aggregate planning methods is feasible. Dauzere-Peres and Lasserre [15] give an example of this approach. It may be used by sales departments to determine whether or not to accept an order with a given lead time. It may also form the input to other decisions such as the scheduling of shop-floor personnel (particularly overtime), preventive maintenance, and purchasing of raw or intermediate materials. Wu et al. [66] and Mehta and Uzsoy [42,43] discuss scheduling models addressing this issue.

The number of different potential uses for a production schedule underscores the need for a satisfactory execution of the scheduling task. On the other hand, the diversity of the groups affected by the schedule also makes the task of measuring schedule quality more difficult, since the schedule is used for different purposes by different groups, which are often trying to achieve different goals. We shall return to this point later in the paper.

3. Feasible and acceptable schedules

Whether we are dealing with a predictive or a historical schedule, the question "What makes a good schedule good?" is valid. The first condition that any schedule must satisfy is *feasibility* — it should not violate any of the constraints present in the manufacturing system in which it is to be executed. In other words, its execution over the specified scheduling horizon must be physically possible. It must put jobs through operations in the order specified by the process plans. It must assign operations to machines capable of performing them. At this point, we do not consider other types of constraints such as those that specify minimum acceptable levels of system performance ("WIP must not exceed X units at any time"), or those arising from management-induced operating policies ("Machine X must be kept busy at all times").

A second condition imposed on a schedule is *acceptability*. An acceptable schedule is one that cannot be improved by trivial changes and that is not dominated in all aspects of interest by another readily available schedule. For example, it may be feasible to run Operation A on Machine 1 or Machine 2, but it may be preferable to use Machine 1 when possible due to its being able to meet tighter tolerances A schedule that runs Operation A on Machine 2 while Machine 1 sits idle can clearly be improved by using Machine 1 The precise nature of what constitutes a "trivial change" will clearly vary from system to system, depending on the complexity of the scheduling problem in question In this paper, we shall assume a trivial change to be one that could be made by a knowledgeable person examining the schedule manually

In the rest of this paper, we shall assume that a schedule that is not both feasible and acceptable can be dismissed at once

4. Issues in schedule quality assessment

Once a feasible, acceptable schedule is available, the problem of assessing schedule quality can begin to be addressed This question can be approached from a number of different perspectives including individual schedules vs a group of schedules, absolute measurement vs relative comparison, tradeoffs between multiple metrics, static vs dynamic measurements and schedule vs state measurements These ideas are discussed briefly before addressing metrics at a more concrete level While many of these issues have been discussed in other contexts than scheduling, a discussion of schedule evaluation and comparison would not be complete without them

4 1 Measuring individual schedules vs groups of schedules

An important aspect of the schedule measurement issue is whether we are measuring an individual schedule or a group of schedules An obvious reason to measure an individual schedule is to gauge its individual performance For a predictive schedule, the result may determine whether or not it will be implemented

Probably, the most common reason for measuring groups of schedules is to evaluate the strategy or algorithm being used to develop them A strategy might be applied to a set of beginning states (scenarios), which cover the range of operating conditions found in practice to produce a group of schedules and ending states Measurement of the group of schedules can be used to draw conclusions about the strategies producing them In this situation, we are examining a number of different schedules and trying to measure the strategy producing them Hence, we need to combine the metrics for the individual schedules in some way to form meaningful metrics for the strategy A commonly used metric is the average over all schedules of the relevant schedule metrics However, it is also important to have measures of the sensitivity of the strategy to different parameters and beginning states A strategy with excellent average performance, which produces a disastrously bad schedule once every six months may not be preferable to a strategy with slightly worse but much more consistent performance Some measures of this type are the maximum and minimum values of the metrics obtained Barr et al [4] and Hooker [29,30] address a number of issues related to evaluating the performance of different scheduling algorithms using computational or simulation experiments These discussions focus on the design of effective computational experiments that allow algorithm performance to be linked to problem characteristics, as well as how to ensure fair comparisons are made between different algorithms In addition, a variety of statistical techniques exist for the analysis of the results of these experiments However, most of this work assumes that there is one performance measure of interest, with the main tradeoff being between this performance measure and the computation times of the different algorithms It relies on calculating aggregate statistics such as means and variances across a set of schedules This aggregate nature of these statistics often obscures the detailed behavior of schedules in specific circumstances at the local level, which is often the key to understanding and improving the algorithms In addition, these very local dynamics are often linked to metrics of schedule usability For example, it may not be desirable to run too many different kinds of jobs on a machine since frequently changing machine operating conditions will lead to quality problems Finally, when there are multiple metrics of interest these techniques begin to be difficult to apply, although other approaches such as data envelopment analysis [9] seem to hold promise for this area The issue of aggregation in evaluating schedules will be discussed further later in the paper

4 2 Absolute measurement vs relative comparison

An *absolute measurement* of schedule quality consists of taking a particular schedule on its own and deciding how good it is This requires some set of criteria or benchmarks against which to measure — some abstract definition of schedule quality This is difficult even if we are optimizing with respect to a single metric Ideally, we would like to compare the result of the schedule to the optimum, but since most factory scheduling problems are NP-hard [22], this is computationally impractical An alternative is to use upper and lower bounds on the metrics of interest For example, it may be possible to compute how far a job could have gotten through the shop under perfect circumstances over the time horizon and to compare this to the performance of the job in the schedule in question However, it is often difficult to compute bounds that are tight enough to provide meaningful information Another approach is the development of statistical point and interval estimators of the optimal values of large combinatorial optimization problems [14,23] In the event that there are multiple metrics of interest, the problem becomes more complex as discussed in the next section

Relative comparison assumes that two or more schedules for the same initial factory state are available, and the task is to decide which is better While this may appear to be a more tractable problem given that all the candidates are available for detailed inspection, two difficulties remain The first of these is the presence of multiple metrics of interest discussed below In a manufacturing environment, it is likely that any given schedule will do better than another on some metrics and worse on others, making it difficult to pick a clear winner without addressing the issue of how to trade the different metrics off against each other Another issue is the possibility that all the schedules under consideration may be poor by absolute standards, which may go undetected during the relative comparisons

If the scheduling system under consideration is being evaluated in the context of an operating manufacturing facility, historical information can be used in both the absolute and relative cases For absolute measurement, the historical average or best-case performance of the factory can be used, perhaps as a lower bound in conjunction with the upper bounds mentioned earlier In the case of relative measurement, average historical performance can be used as a benchmark for comparison Another possibility is to measure the performance of a schedule against trends of historical data This would allow the effects of continual improvement in the scheduling system to be observed over a period of time However, care must be taken to update historical data frequently since most manufacturing systems change continuously Similar care must be taken when interpreting the results, as it is often unclear whether particular changes in the performance of the manufacturing system are due to changes in scheduling practices or to other changes, such as shifts in process technology and product mix, that occurred over the same time horizon

4 3 Tradeoffs between multiple metrics

In most real-world scheduling applications, more than one performance measure is of interest The problem of scheduling in the face of multiple, conflicting objectives has been examined only for very simple systems in the optimization literature There have been three major approaches used in this literature, which we will briefly describe below

One body of research has addressed scheduling problems with primary and secondary criteria In this approach, the problem is to minimize the primary metric while keeping the other within some predefined range This is accomplished by constraining the secondary objective to be within the desired range, thus converting the secondary metric into a constraint Thus, the tradeoff between the two criteria is explicitly specified by the definition of primary and secondary metrics and the range for the secondary metric The work of Smith [61] on minimizing total completion time subject to no tardy jobs is the earliest work in this area This approach has been applied to a number of different problems [3,7,8,16,51,56] This research, however, has remained limited to two criteria, one primary and one secondary In addition, no system more complex than a single machine with all jobs available simultaneously has been addressed to date A survey of research on bicriterion scheduling problems can be found in Dileepan and Sen [16]

The second approach uses the ideas of dominated and efficient solutions Given a number of metrics of interest f_i, $i = 1, \ldots, n$ to be minimized, let f_{ij} denote the value of metric i obtained from schedule j A schedule j is said to dominate another schedule k if $f_{ij} < f_{ik}$ for all $i = 1, \ldots, n$ A schedule that is not dominated by any other solution is called a non-dominated or efficient schedule Given a scheduling problem with multiple conflicting objectives, the entire set of efficient schedules is generated The decision as to how different metrics are to be traded off against each other is left to the human decision-maker This approach has been followed by van Wassenhove and Gelders [63] and Nelson et al [47] among others

A third approach is to combine different metrics by using a weighted sum of the original metrics as a surrogate metric The weights capture the tradeoffs between the different metrics Sen et al [57] have followed this approach to minimize linear combinations of flow time and range of lateness on a single machine However, since the weights determine the tradeoffs between the different metrics, the determination of the weights is a non-trivial problem, depending on the decision makers involved, their attitudes to risk and so on Other examples of this approach are the development of dispatching rules where the priority index of each job is calculated as the weighted sum of a number of different attributes relating to different metrics [6]

A number of researchers have advocated using costs as a mechanism for combining different metrics Conway et al [12] point out that the main cost implications of scheduling decisions lie in the areas of capacity utilization (getting more work through the shop in less time), customer satisfaction (getting the jobs done on time) and in inventory (getting the same volume of work done with less inventory) In this approach, the costs associated with scheduling decisions are assessed for the schedule, which is then evaluated based on the sum of these costs, which may be discounted to reflect the time value of money Jones [31] provides a framework to evaluate schedules from a cost perspective, while Morton et al [45] and Scudder et al [55] give dispatching rules using cost information and evaluate their effectiveness However, the parameters required by these methods, such as tardiness costs, are often difficult to determine in practice In addition, the real costs of a scheduling decision are often the opportunity cost of making the wrong decision, which even with perfect hindsight can be difficult to evaluate

Efforts in the field of artificial intelligence (AI) have, in general, followed similar approaches to those in the optimization literature Fox and Smith [20] approach the problem of multiple objectives in job shop scheduling by setting aspiration levels for each metric of interest, thus effectively turning each metric into a constraint One then attempts to find a solution that is feasible with regard to all constraints If this is not possible, the constraint corresponding to a given metric will be relaxed until a feasible solution is finally found The difficulty here is in determining in what order the constraints should be relaxed, and by how much This problem is equivalent to that of determining the tradeoffs between metrics Other researchers, like Elleby et al [18], have addressed this problem by having the user interactively specify the relaxation of the constraints Kempf et al [34] and Smith [58] discuss a number of the issues involved in AI approaches to scheduling problems Hence, AI approaches do not seem to have brought a fundamental solution to this problem either

All in all, the problem of scheduling in the face of multiple conflicting metrics has been addressed in a variety of ways, all of which reduce to different representations of the tradeoffs between the different metrics of interest Thus, the problem of assessing schedule quality in a real-world environment becomes complex, especially when it is a group of people with different perceptions as to the relative importance of the different metrics, rather than a single decision maker, that are involved in the schedule evaluation process This is particularly the case when different scheduling metrics address different aspects of company performance, such as on-time delivery, which customers to satisfy before others and efficient use of the firm's manufacturing capacity It seems reasonable to ask whether this type of tradeoff should be addressed in the context of a shop-floor scheduling system for a given plant or department, instead of at a higher level in the planning hierarchy We shall return to this question throughout the paper

4.4. Static measurement vs. dynamic measurement

Under this heading, we need to distinguish between predictive and historical schedules, since in a historical schedule, at least in theory, the entire history of the system over the scheduling horizon is available for us to examine. However, in a predictive schedule, we are making a set of decisions based on estimates of future events, without knowing the actual realizations of these events until they actually occur. Thus, we shall define static and dynamic measurements of schedules in terms of predictive schedules alone.

Static measurement of a predictive schedule involves measuring a schedule independently of the execution environment to determine how good the result would be if the schedule was executed exactly as specified. *Dynamic measurement* of a predictive schedule is more difficult. In addition to the static quality of the schedule, we now consider how robust the schedule is to disruptions such as machine breakdowns that occur during execution. An attempt to quantify this aspect of a schedule would require a description of the different types of disruptions that might occur as well as the likelihood of their occurrence and a specification of the capabilities and goals of the execution agent (shop-floor personnel or automated system) that will react to these events. Without knowing how disruptions will be reacted to, their true impact cannot be measured.

A number of approaches to scheduling in the face of unpredictable disruptions have been taken in the literature to date. The two most prevalent are the prediction–reaction paradigm and stochastic approaches based on minimization of long-term steady-state measures. In the prediction–reaction paradigm, a predictive schedule is generated at a given point in time and then revised, either by generation of a completely new schedule or by local adjustments, as the unforeseen disruptions occur. Examples of this approach are the matchup scheduling approach of Barr et al. [4], the schedule repair approach of Smith [58] and Smith et al. [60], and the event-driven rescheduling approach of Church and Uzsoy [11]. In most stochastic approaches, one assumes that the frequency and duration of the disruptions are described in the form of probability distributions, and develops policies to minimize long-term steady-state measures of performance. Examples of this approach are the control-theoretic approaches of Bai et al. [2] and Gong and Matsuo [25], as well as the extensive literature on stochastic scheduling models [50].

However, in recent years, a number of other approaches have begun to emerge, based on different dynamic metrics of schedule performance. The robust scheduling approach suggested by Daniels and Kouvelis [13] and Wu et al. [66] aims at developing schedules that are capable of absorbing disruptions so as to minimize their effect on a primary performance measure. The implicit dynamic metric here is the difference in performance measure value between the predictive and the realized schedules. Another group of authors has developed metrics based on the difference between the predictive and historical schedules, with the idea that drastic revision of the predictive schedule during execution may have detrimental effects on shop performance. Wu et al. [66] address the problem of rescheduling in the face of disruptions, where there are costs associated with altering the predictive schedule. Mehta and Uzsoy [42,43] use a similar metric to develop predictive schedules, which can absorb disruptions without major changes in job completion times. Results from both these papers indicate that it is possible to achieve significant improvement in the dynamic metric at the cost of very minor deterioration in the static metric. Although more work is necessary in this area, these results indicate that under at least some circumstances, meaningful metrics can be devised and used constructively.

In practice, the main approach seems to be the use of dispatching rules, which by their myopic nature are completely reactive to changes in the system state, and a prediction–reaction approach where a predictive schedule is developed by an expert scheduler who, ideally, has some knowledge of the possible disruptions that may occur and hedges against them while constructing the schedule. When the disruptions actually materialize, the scheduler modifies the predictive schedule based on his or her prior experience to mitigate their effects and realize as many of the goals of the original schedule as possible. In this latter case, a great deal depends on the goals of the system as perceived by the scheduler, and on the metrics by which the scheduler is evaluated The risk here is that the scheduler will respond to short-term pressures and incentives instead of keeping the performance of the overall shop in view McKay et al [40] discuss various aspects of expert human schedulers, focusing on their knowledge of uncertainty

In recent years, a number of authors have begun to explore the benefits of incorporating information about executional uncertainties into scheduling algorithms McKay et al [41] examine the effects of including knowledge of the behavior of machines after major repair in scheduling decisions In their work on developing predictive schedules that can absorb disruptions, Mehta and Uzsoy [42,43] use statistical information on the nature and frequency of disruptions in building predictive schedules These approaches show that there can be substantial advantages to considering information about uncertainties in developing the predictive schedule, and that this can be done in a relatively straightforward manner by modifying existing deterministic scheduling models

For a historical schedule, we wish to determine how good the outcome of the execution of that schedule was This involves a number of issues — how close to the original predictive schedule motivating it the historical schedule was able to remain, what were the events that caused it to be modified, what were the capabilities of the agents making the modifications and so on One begins to see the notion of defining the quality of a given schedule relative to the "perfect possible", the best that could have been attained in the given set of circumstances, beginning to emerge Note that this requires hindsight — that the historical schedule and the history of the system affecting its execution be available for analysis, and that the analysis can be done in a reasonable period of time The issue of how to measure deviation from the predictive schedule is also an open one, which can be addressed in many different ways in addition to those outlined above

4 5 *Schedule measurements vs state measurements*

In measuring the quality of a given schedule, we are often interested in the activities that took place over the schedule horizon Typically, these questions concern the amount of work completed by a given machine or the progress towards completion made by a job over the schedule horizon These measurements related to the scheduled activities we shall refer to as *schedule measurements* Typical schedule measurements are the number of jobs of a particular type or the maximum tardiness of all jobs processed over the schedule horizon

However, schedule measurements alone are often insufficient to evaluate the quality of a schedule For example, a schedule for a given workcenter may perform very well for that workcenter over that schedule horizon, but may leave a downstream workcenter in an impossible position at the end of the horizon A common example of this in practice is the so-called "hockey-stick effect", where a workcenter measured on WIP level will produce at a tremendous rate for the last few days of the fiscal period in which it is to be measured, suddenly inundating the downstream workcenters with a vast amount of work that is not needed and cannot be processed In this type of situation, it is desirable to evaluate the end effects of the schedule at the end of the schedule horizon, i e , the state the schedule leaves the factory in One set of such measurements is that relating to the location of Work-In-Progress (WIP) inventory relative to available capacity One would expect these measurements to be more position based, rather than activity-based, and to describe the distribution of work over time or across the different work areas in the manufacturing system being scheduled These latter types of measurements we shall refer to as *state measurements* An example of a state measurement might be the variance of workload represented by the WIP in front of each workcenter Ideally, in a perfectly balanced semiconductor manufacturing plant, one would like to have roughly the same amount of WIP (in terms of workload) at each station A schedule that results in high variance will probably result in major bottlenecks developing in the next scheduling horizon

5 Selection of metrics to apply

The issue of what metrics to use in assessing the quality of a schedule is far from trivial Answering this question is equivalent to describing the kind of behavior we want the scheduling system to induce in the manufacturing system, which in turn is equivalent to deciding what the goals of factory management should be. At an abstract level, this is easy enough: the factory should be run in such a way as to maximize the value to the organization. This involves decisions such as whether the company will compete on quality or rapid delivery, which customers are more important than others, what is the priority of engineering work relative to production work, and what is an acceptable level of performance in each of these areas. These decisions are closely related to the firm's manufacturing strategy, particularly to how it chooses to use its manufacturing capabilities as a competitive weapon [28]. In many firms, particularly large, multi-division companies with multiple geographic locations, this involves a broadened definition of the scheduling problem to include consideration of due date setting, order acceptance, capacity allocation between competing product lines, blurring the distinction between production planning and shop-floor scheduling. Hence, it is immediately apparent that the choice of scheduling metrics will, and should, vary from one organization to another. Given that scheduling decisions are constrained by many other decisions, such as what to make, what orders to accept, which customers are important and how due dates are quoted, it also begs the question of whether these issues should be addressed at the level of the factory scheduling problem. We shall have more to say on this issue in the following section.

Given our definition of schedules as sets of job/machine/time assignments, it seems natural to restrict our attention to metrics that can be calculated based on the schedule itself, starting from these assignments. However, there are a number of factors that make choosing a meaningful set of metrics from among the wide range of mathematically and intuitively plausible ones difficult.

5.1. Organizational goals

A major problem is that different organizational units affected by the schedule have different, often conflicting goals and, thus, have substantially different expectations from a schedule. A sales department will often look at a schedule from the point of view of orders being delivered to the customer. A manufacturing department, on the other hand, may be under pressure to reduce costs. Thus, the sales department will look for a schedule that has good due date performance, while the manufacturing department will prefer a schedule with high machine utilization, few setups, and long production runs. An interesting discussion of this type of problem is given by Harrison et al. [27].

One can argue that this problem can be alleviated by assigning the different organizational units goals, which are coherent and directed towards the overall good of the company as a whole. However, this problem of setting compatible performance measures seems to be a long way from being solved. As one moves up the corporate hierarchy, goals tend to be expressed more and more in aggregate and financial terms, culminating in the very brief corporate mission statement. To complicate the issue further, the time horizons over which decisions are made differ vastly over different levels — from years and months at the corporate level to weeks and days at the plant level and hours and minutes on the shop floor. The short-term metrics used to evaluate the performance of shop and plant managers (such as headcount and quarterly profits) further confuse the issue. It is very difficult to reconcile a goal like "maximize profit for this quarter" with the decisions made while scheduling an 8-h shift on the shop floor. (Whether maximizing profit this quarter in a particular way contributes to the firm's long-term mission is often also a valid question.) The problem of developing coherent sets of performance measures for all levels of the corporate hierarchy to ensure that all levels are working towards the same set of corporate objectives and not adversely affecting each other seems to be an open one at present [19]. However, it would seem to suggest that some of the issues regarding the choice of scheduling metrics need to be discussed and resolved at a corporate level, rather than at the level of an individual shop or plant, which is where decisions on specific scheduling systems are often made.

The discussion, so far, has assumed that the schedules developed based on the metrics decided upon can be executed exactly as planned. However, the knowledge that unexpected disruptions will occur often leads to a set of grassroots metrics developing on the shop floor with the goal of minimizing the impact of these disruptions, such as trying to build ahead of schedule to hedge against an expected breakdown. The existence of such informal systems of metrics often adds considerably to the complexity of the metric selection problem.

5.2. Relationships between metrics

The existence of many plausible metrics for evaluating schedule quality raises the question of relationships between the metrics. Some metrics are complementary, an improvement in one bringing an improvement in the other. An example might be minimizing cycle time variance and improving on-time delivery performance. However, many metrics that are potentially conflicting are also of interest in scheduling. Maximizing machine utilization and minimizing WIP serve as an example. Maximizing the utilization of machines can be achieved by ensuring that work is always available at each machine, while minimizing the WIP might risk machines being starved at some point in time. Such interactions between metrics further confound the process of measuring schedules. In order to develop a good set of metrics for a schedule, it is important to understand the nature of these relationships as far as possible, which is a difficult task since the relations are often nonlinear in character. For example, a change in a given metric may drastically affect the significance of another. This is the case when a shop whose primary concern is on-time delivery has somehow slipped into the situation where the majority of jobs are tardy. In this case, it often makes sense for management to emphasize throughput and ignore due dates to try to get the shop back on an even keel. Thus, a change for the worse in the tardiness metric has resulted in a drastic increase in the importance of the throughput metric.

This example also illustrates the possibility that the relative importance of metrics may change over time depending on management goals and the state of the shop. Another issue is how altering the metrics over time affects the progress of the shop as a whole. While it seems obvious that as the world about the shop and management goals change, the metrics in the shop should be altered to reflect this, the cost of altering metrics too frequently is not clear. Based on the previous discussion of state and schedule metrics, it would seem that some system states are more advantageous for some metrics than for others. Hence, when metrics are changed, the system may well be in a bad state as far as the new metric goes. The strategy driven by the new metric will then try to return the system to a good state as it perceives it, which may in turn leave the system in a bad state for the succeeding metric. It is conceivable that switching metrics too often will result in the system becoming unpredictable and unstable, with ramifications for higher-level planning systems that use aggregate information on the shop-floor performance such as lead times. Kempf and Beaumariage [33] have shown that relatively simple manufacturing systems can exhibit chaotic behavior in the sense that minor changes to the system can result in major disruptions in the operation of the system. Another interesting aspect here is that if the metrics of interest are changed too frequently, it becomes impossible for shop-floor personnel to learn good strategies for the different metrics, since by the time they have accumulated some experience running the shop under the new conditions, the metric has changed.

5.3. Aggregation and segmentation of metrics

When assessing the quality of a given production schedule, the raw data used to calculate any set of metrics is that which can be calculated directly from the set of job/machine/time assignments that constitute the schedule itself. Examples of these metrics, which we shall refer to as *atomic metrics*, are the time spent in process by a particular lot at a particular machine, or the proportion of the time horizon that a certain machine was busy or down. These atomic metrics are based on the observation that a given scheduling object, i.e., a job or a machine, can enter one of several different states over the course of the execution of a schedule. The atomic metrics then record how much time each scheduling object spends in each possible state over the schedule horizon. For example, a machine may be in one of five states: busy, in setup, in maintenance, idle and down. Hence, the atomic metrics for this machine would record how long it remained in each state before a transition to a new one. In the case of a job, one might define the possible states as in process, in transport, on hold or idle, and calculate the amount of time each job spent in each state over the course of the schedule

It is apparent that we can use these atomic metrics in a number of different ways to calculate metrics relating to the overall schedule being evaluated We shall distinguish between two different operations, segmentation and aggregation, that are performed on the atomic metrics

The operation of *segmentation* consists of specifying a class of scheduling objects (jobs or machines) that form a meaningful unit in terms of schedule evaluation To illustrate this idea, consider the set of all jobs in the schedule The atomic metrics available are time spent waiting and in process at each operation Possible segmentations of these atomic metrics are by jobs of the same product, or by jobs of the same priority class (urgent/late/on time, for example) A segmentation commonly encountered in practice is by administrative area — for example, the fabrication and assembly areas of a manufacturing plant might be viewed as segments of the scheduling problem The underlying idea is that the performance of a given schedule will be evaluated in terms of its impact on the set of scheduling objects specified in the segmentation

An example of a job-based segmentation commonly encountered in semiconductor manufacturing is by the type of order Engineering jobs related to product development have the highest priority Jobs whose goal is to replenish inventories generally have low priority, while orders from specific customers may cover a wide range of priorities An example of a machine-based segmentation is separating machines into bottleneck and non-bottleneck resources as advocated by Goldratt's Theory of Constraints [24]

Developing a segmentation of a scheduling problem, which is meaningful in terms of the overall organizational objectives, is not a trivial task Any scheduling decision assigning a machine to process a job at a certain point in time may well impact the metrics of two or more different segments, potentially in conflicting ways The relationships between different segments also becomes important, in the sense that performing well in one segment may entail performing badly in another For example, consider a segmentation of the jobs into high- and low-priority classes and of the machines into bottleneck and non-bottleneck machines Suppose the goal is to maximize utilization on the bottleneck machines, and minimize the time in system of the high-priority jobs A situation may arise where a high-priority job is approaching the bottleneck machine, but will not arrive there for a while yet. The scheduler may choose to keep the bottleneck machine idle and wait for the high-priority job, thus, sacrificing the machine utilization in favor of the job's flow time The alternative is to keep the machine busy with whatever work is at hand, sacrificing the flow time of the high-priority job to the machine utilization Another example already referred to is segmentation by administrative area, where in order to make itself look good one department may pursue a strategy that leaves another in a hopeless position

An interesting application of the idea of segmentation is that of viewing each segment as an independent, intelligent agent trying to optimize its own objective function, which may differ from those of other agents A number of researchers [17,38] have advocated scheduling systems of this nature, where a bidding mechanism is used to resolve conflicts between different agents While the analogy to free-market economics is interesting, this approach still requires that the objectives of the individual agents be set in a manner that will ensure good overall system performance, which is not clear how to do In addition, most of these approaches have been tested in the context of flexible manufacturing systems with relatively few machines

Once the segmentation of the set of scheduling objects, and thus of the associated set of atomic metrics, has been specified the atomic metrics for each segment can be *aggregated* in a number of different ways The most common method of aggregation is averaging Thus, one might aggregate a set of atomic metrics by taking their average over a given set of jobs or machines specified in the segmentation, one might take the average of a set of atomic metrics associated with a given object over the length of the time horizon, or one might do both There seem to be two main dimensions, along which, one can aggregate atomic metrics the set of scheduling objects specified as a meaningful unit for schedule evaluation in the segmentation, and some specified time horizon Each of these dimensions in turn may imply a hierarchy of aggregation For example, if we segment the atomic metrics by functional departments on the shop floor, such as drilling, milling and welding, we will aggregate the atomic metrics for each department into an aggregate metric for that entire department However, the aggregate departmental metrics can, in turn, be aggregated into plant-wide metrics, which can then be aggregated into company-wide ones If we are aggregating over time, the atomic metrics may be aggregated over a time horizon such as an 8-h shift These metrics may then be aggregated into daily, weekly, monthly and yearly metrics

The segmentation and aggregation of metrics is crucial to good schedule quality evaluation It will also affect the behavior of the manufacturing system, since the way the metrics are structured will determine the criteria by which the performance of shop-floor management will be evaluated, which in turn will affect the decisions made by shop-floor management The segmentation must be over classes of objects that are meaningful not just in terms of organizational structure, but in terms of the manufacturing system being scheduled If we segment the problem based on functional departments, as is often the case in industry, we run the risk of each department optimizing its own performance independently The way in which metrics are aggregated is closely linked to the problem of establishing meaningful metrics for different levels of the corporate hierarchy discussed above Ideally, the aggregation of metrics should be accomplished in a way that will ensure the coordination and coherence, in terms of overall corporate goals, of the actions taken in the schedule The degree of aggregation is an important variable here If we aggregate too little, we face a mass of atomic data from which it is difficult to draw any conclusion about the schedule as a whole If we aggregate too much, we obtain metrics that are difficult to relate to actual events on the shop-floor or schedule characteristics These highly aggregate metrics can often be misleading when used indiscriminately A good example is looking at a plant-wide average of machine utilization Trying to maximize this metric will lead to attempts to reach 100% utilization on all machines in the plant, resulting in a dramatic increase in WIP levels and shop-floor congestion Looking at average utilization on bottleneck machines, for instance, makes a lot more sense since we are maximizing utilization where it matters, instead of indiscriminately across the whole plant. The challenge is to select a segmentation and an aggregation that provide the most meaningful information to the decision maker who is trying to evaluate the schedule Gary et al [21] illustrate a possible set of segmentations and aggregations of scheduling metrics for a semiconductor manufacturing facility

6. Implications for design and implementation of scheduling systems

The above discussion should highlight the fact that the problem of assessing the quality of a schedule is actually a complicated question that can be addressed from a number of different perspectives The first characteristic of the schedule measurement problem is its multiobjective, multiattribute nature Even when a single individual is involved in the evaluation process, the question of how to address the tradeoffs between the different metrics of interest is hard to address without keeping the human involved Another important point is that schedules mean different things to different people and are used by different organizational groups in different ways This has several implications for designing scheduling systems In this section, we shall discuss some of the implications of these difficulties for the development and implementation of scheduling systems

The difficulty of selecting a meaningful set of metrics for schedule measurement, both static and dynamic, the often conflicting nature of the metrics selected, the multiple groups involved in using and evaluating schedules and the presence of subjective factors such as attitude to risk and different tradeoffs between objectives makes a crisp definition of what a scheduling system is supposed to achieve in an organization very hard to obtain This, in turn, makes it difficult to evaluate the performance of a scheduling system When there is no clear idea shared by all users of the system on what the system is supposed to achieve, it is very difficult for the system to be deemed successful This is consistent with our (admittedly limited and anecdotal) experience that a manifest crisis in shop performance is often required for management to unite behind a scheduling system for a significant length of time Kempf [32] and Kerr and Ebsary [35] give interesting descriptions of attempts to implement scheduling systems, which were negatively affected by difficulties in evaluating the resulting schedules The difficulty in relating scheduling decisions to the "bottom line", the short-term financial performance of the plant or department, insisted on by many managers, is the most conspicuous manifestation of this problem We would conjecture that these difficulties in evaluating the effectiveness of production schedules, and, therefore, of scheduling systems, are one reason for the relatively small proportion of scheduling research transferred into industrial practice Hence, the development of effective means of schedule quality assessment, especially insight into the economic impact of scheduling systems on the performance of the company as a whole, is critical if scheduling research is to have a major impact on industrial practice While the current trend towards "balanced scorecard" approaches to evaluating performance that take into account non-financial, operational measures is encouraging, this issue will probably remain important for the foreseeable future

Given the existence of computerized inventory tracking systems, it seems reasonable to assume that atomic metrics and historical schedules will be available for examination, providing the basic data necessary for assessing schedule quality However, exactly how these data should be used for schedule evaluation is difficult to prescribe Given a good understanding of the environment in which the scheduling decisions are being made, it should be possible to develop at least reasonable segmentations of the problem, and aggregations of the different metrics If the segmentation of the problem is not effectively addressed, the different segments will be driven to optimize their own set of metrics at the expense of suboptimizing the performance of the overall system If the aggregation of metrics is not carried out in a meaningful way, the relationship between long-term corporate goals and scheduling decisions will become blurred or lost completely due to the interposition of several layers of surrogate metrics whose effects on each other and on longer-term corporate goals is at best unclear This problem is further complicated by the fact that the metrics of interest may change over time, implying the possibility of changes in the aggregation and segmentation used Resolution of these issues will probably require negotiation between the different groups using the schedules, and a close study of the interactions between the different segments of the problem and the different metrics of interest It is interesting to note that there have been relatively few in-depth studies of the interactions between different scheduling metrics Much of the current folklore on this subject is based on somewhat contrived textbook examples which show that under certain circumstances optimizing one metric will lead to very poor performance in others However, it is not obvious at all how likely these conditions are to arise in practical environments, where different job parameters such as processing times, setup requirements and due dates may be highly correlated Ovacık and Uzsoy [49] have shown in an extensive computational study of job shop scheduling algorithms that schedules that perform well with respect to maximum lateness also perform well with respect to makespan, total completion time and average tardiness This indicates that for all practical purposes, focusing on maximum lateness as a metric may be sufficient to yield satisfactory shop performance with respect to several measures, eliminating the problem of tradeoffs between metrics Further work in this area is needed, although results are likely to be somewhat situation-specific, depending on the structure and particular characteristics of the shops being considered

As discussed in Section 1, a considerable body of research over the last decade has demonstrated the close interrelation between scheduling performance and planning decisions such as due date setting, order release and lot sizing A number of these authors have gone so far as to suggest that if the planning is done right, the scheduling becomes almost a non-issue (e g , Ref [65]) In some manufacturing systems, this is probably a valid point of view If a manufacturing system produces a relatively simple product where different stages of the production process are loosely coupled, relatively simple scheduling policies can give good results and shop performance will be driven by the planning operation However, in more complex environments such as semiconductor manufacturing it has been shown that this is not the case The scheduling problem is essentially the problem of how to allocate shop capacity in the short term. Poor scheduling decisions can lead to significant loss of productive capacity, invalidating the results of even the most careful planning process.

However, the relationship between planning and scheduling may well hold the key to a significant simplification of the schedule measurement issue. Our definition of scheduling in this paper has focused on the execution of activities on the shop floor in the short term, with a typical time horizon of a day or a shift. Planning, on the other hand, involves decisions of how to allocate the firm's productive capacity among different products and customers over the longer term. Its typical output is a master production schedule specifying how much of each product to make in weekly or monthly time buckets. In particular, the planning system must consider the firm's different departments and geographical locations in an integrated way for the resulting plans to be effective. Clearly, then, the scheduling decisions in a given department or plant are tightly constrained by the planning decisions leading up to them.

Given this situation, we would suggest that the scheduling function be viewed as one of very narrow scope — that of executing the production plan as closely as possible. This simplifies greatly the issue of evaluating schedule performance, bringing it to essentially one simple question — did we meet the goals (both timing and quantity) set by the production plan or not? This is clearly a much more tractable problem than that of linking shop-floor scheduling decisions to the firm's long-term goals and strategies. It still leaves open a considerable area, in that there may be several schedules capable of achieving a given plan that differ significantly in how they do it, particularly in how efficiently resources are used. Many of the issues raised in this paper are still valid within this more limited scope. However, if we accept this viewpoint, the scheduling system's top priority becomes that of achieving the production plan, with local efficiency considerations remaining secondary to this. The major advantage of this approach is that the question of what the plant should be doing to add value to the firm — what products to push in the market, which customers to give priority over others, how to quote due dates and handle unexpected rush orders — are addressed within the production planning process with a firm-wide view of goals and resources. The need to discuss these issues, which are almost part of corporate strategy, is removed from the individual plant or shop. The issues to be resolved within the production planning process are clearly complex, and many of the complexities discussed in this paper in the scheduling context remain here also. However, these issues are now being discussed by the people with the information and the responsibility to make them, instead of having them made by default in response to local, short-term pressures.

From this perspective, it is interesting to note that many scheduling systems developed in industry, such as that developed at Intel by Kempf [32] and the ReDS system of Hadavi et al. [26] developed at Siemens, as well as several commercial products, actually contain a large planning component, addressing issues of due date quotation, material management and order release. It is also interesting to note the parallel between shop-floor scheduling and master production scheduling, both of which involve decisions as to how to allocate the firm's productive capacity over time, although in different time frames. It is indicative of the difficulty of making the trade-offs involved that neither shop-floor scheduling nor master production scheduling has been extensively automated to date in the majority of the discrete-part industries, although the intervening function of requirements planning has been automated in many companies using Material Requirements Planning (MRP) systems for the last 20 years.

Even if we treat scheduling as a problem of short-term execution of plans made at a higher level in the firm, the issue of dynamic schedule measurement remains. The issue of how to develop schedules that perform well in the face of unexpected disruptions is closely linked to the ability to define what such good performance is, creating a strong need for dynamic metrics that capture management concerns. While the study of such metrics us still in its infancy, the empirical work of McKay et al. [40] suggests that they form an important tool for human schedulers. Modelling efforts taking into account uncertainties in developing predictive schedules have merged over the last 5 years, indicating that if the set of possible disruptions can be described, as well as the actions that can be taken to remedy them, effective dynamic metrics can be developed and used to improve shop performance. There remains a strong need to characterize the disruptions that can occur in the shop in a systematic way. While there may well be some totally unexpected disruptions that can occur, it is unreasonable to expect a schedule to operate under every conceivable circumstance. If the uncertainties in the system are so prevalent that no characterization can be captured in a scheduling model, or at least some recurring types of disruptions cannot be isolated, then we would submit that the shop in question probably has bigger problems than scheduling and should focus on reducing the uncertainty to obtain a more manageable system.

In optimization terminology, the difficulty of specifying the goals a scheduling system is expected to achieve implies that defining the objective function unambiguously is difficult. This leads one to speculate whether in their general form discussed in this paper some industrial scheduling problems are mathematically ill-defined. Certainly, when a number of restrictions such as a single objective function are placed on the problem, well-defined mathematical formulations as control, optimization or heuristic search problems are available, though not always practically solvable to optimality. In recent years, considerable progress has been made in solving these problems to near-optimality, or at least developing solution procedures that improve on commonly used myopic strategies such as dispatching rules [1,2,48,54]. Thus, one is led to conjecture that under different sets of restrictions, where crisp mathematical formulations of different types are available, scheduling problems can be solved, at least in terms of developing viable computational procedures that develop usable, acceptable solutions. However, it is not always clear what these restrictions should be, and it is difficult to see the problem in all its generality described in this paper admitting of a crisp mathematical formulation. At this point, there is no substitute for in-depth knowledge of the shop being scheduled to allow the modeler to determine what to consider explicitly and what to omit. However, it may well be beneficial for management to flesh out a coherent objective function for the scheduling problem to ensure that issues are addressed with a view of the entire firm, not just the local unit trying to develop a localized scheduling solution.

One concern raised by this discussion is the nature of reasonable restrictions to place on a scheduling problem to make a mathematical formulation possible. In other words, how much of the realism of the model should be sacrificed to gain at least some mathematical resolution and tractability? The answer to this question will probably depend heavily on the environment in which the scheduling decisions are being made and used. An understanding of this issue, however, would also allow us to determine how the environment should be designed (or maybe redesigned) to make the scheduling problem more rational. This may include setting mutually compatible goals for the different organizational groups using the schedules, instituting operational protocols to simplify the scheduling task, reorganizing the shop itself to simplify the flow of work, or reducing the uncertainty in the system to reach a more predictable environment. For example, simplified scheduling is often cited as a benefit of cellular manufacturing systems.

The idea of redesigning the manufacturing environment to facilitate the scheduling task raises the question of whether the successful execution of the scheduling function is important enough to the organization to merit all this trouble. Once again, we would conjecture that the answer to this question depends on the environment. The difficulty of evaluating the performance of a schedule in itself, discussed in this paper, would seem to hint that the evaluation of the effectiveness of a scheduling system within the organization is much more difficult. This issue is of critical importance to the actual implementation of scheduling systems in practice. The economic impact of scheduling on the organization is generally not well understood or quantified, and considerable research on this more general aspect of schedule measurement is required.

Another issue raised by the ill-defined nature of scheduling problems is how to use the models and solution techniques that have been developed over the last three decades. It seems clear that it is not cost-effective, feasible or even desirable to include all constraints and considerations pertaining to the industrial scheduling problem into a model [58]. On the other hand, taking a mathematical model based on an abstraction of the problem at hand, implementing it directly in the factory and expecting it to work is unrealistic Perhaps, rather than interpret the models and techniques developed as being directed towards actual implementation on the shop floor, we should try to interpret them as providing insights into the structure of solutions to the industrial problem through rigorous analysis (theoretical or experimental) of tractable special cases and simplified models These insights can then be combined into effective heuristic procedures for the real problem An alternative interpretation is that the function of the scheduling model is to get the user into the neighborhood of a good solution for the real problem, rather than to provide the final solution This initial solution obtained via the model can then be modified by the user, perhaps via some interactive decision aid, to incorporate constraints and other aspects of the problem not easily captured by the model The development of such decision aids poses an interesting direction for future research

Finally, the issues described in this paper indicate a strong need for more empirical field work in the area of scheduling This would fill a strong need to know how scheduling problems are perceived by different functional groups in different industries, how far insights from one type of environment can be generalized to another, and what are the really pressing scheduling problems that need to be solved The majority of the scheduling problems studied in the literature are based on formulations developed in the 1960s Organizational structures, the business environment and the available set of computer and mathematical tools have changed considerably since then A study of industrial practice and perceptions would lead to interesting insights into how practitioners are addressing these problems, how existing theory can be deployed to assist them, and what new theory and tools need to be developed The work of McKay et al [39,40] is an important step in this direction, which needs to be pursued further

7. Uncited references

[5]
[37]
[44]
[59]
[62]

Acknowledgements

Part of the research of Reha Uzsoy was performed while a visiting researcher at Intel in July–August 1992 The authors would like to thank Prof K.N. McKay for his extensive, insightful comments

References

[1] J Adams, E Balas, D Zawack, The shifting bottleneck procedure for job-shop scheduling Management Science 34 (1988) 391–401
[2] X Bai, N Srivatsan, S B Gershwin, Hierarchical real-time scheduling of a semiconductor fabrication facility, in Proceedings of the Ninth IEEE International Electronics Manufacturing Technology Symposium, Washington DC,1990
[3] S P Bansal, Single machine scheduling to minimize weighted sum of completion times with secondary criterion — a branch and bound approach, European Journal of Operational Research 5 (1980) 177–181
[4] R S Barr, B L Golden J P Kelly, M G C Resende, W R Stewart, Designing and reporting on computational experiments with Heuristic methods, Journal of Heuristics 1 (1995) 9–32.
[5] J C Bean, J R Birge, J Mittenthal, C.E. Noon, Matchup scheduling with multiple resources, release dates and disruptions Operations Research 39 (1991) 470–483
[6] K Bhaskaran, M Pinedo Dispatching, in G Salvendy (Ed), Handbook of industrial engineering Wiley, New York, 1991, Chap 83
[7] S Chand, H Schneeberger, A note on the single-machine scheduling problem with minimum weighted completion time and minimum allowable tardiness, Naval Research Logistics Quarterly 33 (1986) 551–557
[8] S Chand, H Schneeberger, Single machine scheduling to minimize weighted earliness subject to no tardy jobs European Journal of Operational Research 34 (1988) 221–230
[9] Y.L. Chang, T Sueyoshi, R.S Sullivan, Ranking dispatching rules by data envelopment analysis in a job shop environment, IIE Transactions on Scheduling and Logistics 28 (1986) 631–642
[10] T C.E Cheng, M C Gupta, Survey of scheduling research involving due date determination decisions, European Journal of Operational Research 38 (1989) 156–166
[11] L.K Church, R Uzsoy, Analysis of periodic and event-driven rescheduling policies in dynamic shops International Journal of Computer-Integrated Manufacturing 5 (1992) 153–163
[12] R W Conway W.L Maxwell, L W Miller, Theory of scheduling, Addison-Wesley 1967
[13] R.L. Daniels P Kouvelis, Robust scheduling to hedge against processing time uncertainty in single-stage production, Management Science 41 (1995) 363–376
[14] D G Dannenbring, Procedures for estimating optimal solution values for large combinatorial problems Management Science 23 (1977) 1273–1283

[15] S Dauzere Peres J -B Lasserre, An integrated approach in production planning and scheduling,Lecture notes in economics and mathematical systems, Springer-Verlag, Berlin, 1994
[16] P Dileepan, T Sen Bicriterion static scheduling research for a single machine, OMEGA 16 (1988) 53–59
[17] N A Duffie, R S Piper, Non-hierarchical control of a flexible manufacturing cell, Robotics and Computer-Integrated Manufacturing 3 (1987) 175–179
[18] P Elleby, H.E Fargher T R Addis, A constraint based scheduling system for VLSI wafer fabrication, in J Browne (Ed.), Knowledge-based production management systems, Elsevier, 1989
[19] J Fisher, Use of nonfinancial performance measures, Journal of Cost Management (1992) 31–38 Spring
[20] M S Fox, S F Smith, ISIS a knowledge based system for factory scheduling Expert Systems 1 (1984) 25–49
[21] K Gary, K Kempf, S Smith R Uzsoy, Assessing the quality of production schedules, in W Scherer, D Brown (Eds.) Intelligent scheduling systems, Kluwer Academic Publishing 1995, pp 129–154
[22] M R Garey, D S Johnson Computers and intractability a guide to the theory of NP-completeness, Freeman, San Francisco, 1979
[23] B L Golden, F B Alt, Interval estimation of a global optimum for large combinatorial problems, Naval Research Logistics Quarterly 26 (1979) 69–77
[24] E.M Goldratt, The goal a process of ongoing improvement North River Press, Croton on Hudson, 1986
[25] Gong L., Matsuo H A control policy for a manufacturing system with random yield and rework The Graduate School of Business University of Texas at Austin 1990
[26] K Hadavi, M S Shahraray K Voigt, ReDS a dynamic planning scheduling and control system for manufacturing, Journal of Manufacturing Systems 9 (1989) 332–344
[27] J M Harrison, C.A Holloway, J M Patell, Measuring delivery performance a case study from the semiconductor industry in R.S Kaplan (Ed.), Measures for manufacturing excellence, Harvard Business School Press 1990
[28] R H Hayes, S C Wheelwright, Restoring our competitive edge computing through manufacturing Wiley New York, 1984
[29] J Hooker, Needed an empirical science of algorithms, Operations Research 42 (1994) 201–212
[30] J Hooker, Testing Heuristics we have it all wrong Journal of Heuristics 1 (1995) 33–42
[31] C H Jones, An economic evaluation of job shop dispatching rules, Management Science 20 (1973) 293–307
[32] K G Kempf, Intelligently scheduling semiconductor wafer fabrication in M Zweben M Fox (Eds), Knowledge based scheduling, Morgan Kaufman, 1994
[33] K G Kempf T Beaumariage, The nature and origin of chaos in manufacturing systems in Proceedings of the IEEE/SEMI Advanced Semiconductor Manufacturing Conference, 1994 pp 169–174
[34] K Kempf B Russell, S Sidhu S Barrett Artificially intelligent schedulers in manufacturing practice, AI Magazine 11 (5) (1991) 46–56
[35] R M Kerr, R V Ebsary, Implementation of an expert system for production scheduling, European Journal of Operational Research 33 (1988) 17–29
[36] E.L. Lawler, J K Lenstra, A.H G Rinnooy Kan, D B Shmoys, Sequencing and scheduling algorithms and complexity, in Handbooks in operations research and management science, S C Graves, A H G Rinnooy Kan, P Zipkin (Eds.), Logistics of production and inventory Vol 4 North-Holland, 1993
[37] Leon V.J , Wu S D, Storer R.H Robustness measures and robust scheduling for job shop Department of Industrial Engineering, Lehigh University, 1990
[38] G Y.J Lin, J J Solberg Integrated shop-floor control using autonomous agents IIE Transactions 24 (1992) 57–71
[39] K.N McKay F R Safayeni, J.A Buzacott, Job-shop scheduling theory what is relevant?, Interfaces 18 (1988) 84–90
[40] K.N McKay J A Buzacott, F R Safayeni, The scheduler's knowledge of uncertainty the missing link, in J Browne (Ed), Knowledge based production management systems, Elsevier, 1989, pp 171–189
[41] McKay K N , Morton T.E., Ramnath P Aversion dynamics, Research Report Memorial University of Newfoundland, St Johns, Newfoundland Canada, 1996
[42] S V Mehta, R Uzsoy, Predictable scheduling of a single machine subject to breakdowns International Journal of Computer-Integrated Manufacturing 12 (1999) 15–38
[43] S V Mehta, R Uzsoy, Predictable scheduling of a job shop subject to breakdowns, IEEE Transactions on Robotics and Automation 14 (1998) 365–378
[44] S Miyazaki One machine scheduling problem with dual criteria, Journal of the Operations Research Society of Japan 24 (1981) 37–50
[45] T.E Morton S R Lawrence, S Rajagopalan, S Kekre, SCHED-STAR a price-based shop scheduling module, Journal of Manufacturing and Operations Management 1 (1988) 131–181
[46] A Najmi C. Lozinski Managing factory productivity using object-oriented simulation for setting shift production targets in VLSI manufacturing, in Proceedings of the Autofact Conference Society of Manufacturing Engineers, 1989 pp 3 1–3 14
[47] R T Nelson R K. Sarin, R.L. Daniels, Scheduling with multiple performance measures the one-machine case, Management Science 32 (1986) 464–479
[48] I M Ovacik R Uzsoy A shifting bottleneck algorithm for scheduling semiconductor testing operations, Journal of Electronics Manufacturing 2 (1992) 119–134
[49] I.M Ovacik R Uzsoy Decomposition methods for complex factory scheduling problems, Kluwer Academic Publishers, Norwell MA 1997
[50] M Pinedo, Scheduling theory, algorithms and systems Prentice-Hall, NJ 1995
[51] C N Potts L.N Van Wassenhove An algorithm for single-machine sequencing with deadlines to minimize total weighted completion time European Journal of Operational Research 12 (1983) 379–387
[52] C N Potts L.N Van Wassenhove, Integrating scheduling

with batching and lot-sizing a review of algorithms and complexity, Journal of the Operational Research Society 43 (1992) 395–406

[53] G L. Ragatz, V.A Mabert, An evaluation of order release mechanisms in a job shop environment, Decision Sciences 19 (1988) 167–189

[54] Sadeh N Look-ahead techniques for micro-opportunistic job shop scheduling, PhD Thesis School of Computer Science, Carnegie-Mellon University, 1991

[55] G D Scudder, D.E. Smith-Daniels, T Rohleder, Use of the net present value criterion in a random job shop where early shipments are forbidden, Journal of Operations Management 9 (1990) 526–547

[56] T Sen, S K Gupta, A branch and bound procedure to solve a bicriterion scheduling problem, IIE Transactions 15 (1983) 84–88

[57] T Sen, F M E. Raiszadeh, P Dileepan, A branch and bound approach to the bicriterion scheduling problem involving total flow time and range of lateness Management Science 34 (1988) 254–260

[58] Smith SF Knowledge-based production management approaches, results and prospects Center for Integrated Manufacturing Decision Systems, The Robotics Institute, Carnegie-Mellon University, Pittsburgh PA 15213, 1991

[59] S F Smith, A methodology and architecture for reactive scheduling, in W T Scherer, D E Brown (Eds.), Proceedings of the Intelligent Scheduling Systems Symposium San Francisco, CA, 1992, pp 66–81

[60] S F Smith N Keng K Kempf Exploiting local flexibility during execution of pre-computed schedules, in A Famili, D S Nau, S H Kim (Eds), Artificial intelligence applications in manufacturing, AAAI Press/MIT Press, 1992

[61] W E. Smith, Various optimizers for single-stage production, Naval Research Logistics Quarterly 3 (1956) 59–66

[62] G Sullivan K Fordyce, IBM burlington's logistics management system, Interfaces 20 (1) (1990) 43–64

[63] L.N Van Wassenhove, L.F Gelders, Solving a bicriterion scheduling problem, European Journal of Operational Research 4 (1980) 42–48

[64] L.M Wein, Due-date setting and priority sequencing in a multiclass M/G/1 queue, Management Science 37 (7) (1991) 834–850

[65] L.M Wein P B Chevalier A broader view of the job-shop scheduling problem, Management Science 38 (1992) 1018–1033

[66] S D Wu, R H Storer, P C Chang, One machine rescheduling heuristics with efficiency and stability as criteria, Computers and Operations Research 20 (1993) 1–14