Primary production scheduling at steelmaking industries

by H. S. Lee
S. S. Murthy
S. W. Haider
D. V. Morse

This paper describes primary production scheduling in the steel industry—the problem and the approaches to the solution. The scheduling problem in steel plants is known to be among the most difficult of several industrial scheduling problems. We first describe the main steelmaking processes and show how scheduling affects the effectiveness of plant operations. We characterize the problems associated with scheduling steelmaking activities to achieve business objectives of delivering quality steel on time to customers, while minimizing operating costs. We then describe several different approaches for computerized scheduling solutions. They include application of techniques in operations research, artificial intelligence, and a hybrid of these two. We conclude by describing advanced techniques for integrated scheduling of steel plants.

1. Introduction

For more than two decades, the steel industry has been a powerful symbol of an increasingly global market economy. This industry, once indisputably dominated by the United States, has now been extended to most of the developed and developing countries of the world. The transition marked a crisis of major proportion in American industry, as hundreds of outdated steelmaking facilities were closed and tens of thousands of steelmaking jobs either "went overseas" or were eliminated entirely. Faulted for policies that neglected investment and engendered contentious employee/management relations, steel-industry management in the United States responded with major infusions of capital and with the adoption of team concepts in the workforce. Today, American steelmakers enjoy renewed prosperity and share a growing market with Japan, South Korea, Germany, and others.

While the latter half of the twentieth century has witnessed the evolution of myriad new manufacturing and construction materials, steel has retained its importance as the primary component in our automobiles, ships, and appliances. The advent of new alloys has expanded its range of physical properties to make possible new applications, from food packaging to electric-light filaments. In the construction industry, where steel has long been the material of choice for commercial buildings, bridges, and industrial facilities, it has recently received attention as a cost-effective and flexible alternative for residential construction as well.

As with any commodity, the advantage belongs to those who can produce it with economy and quality; for steel, this means obtaining raw materials and converting them

0018-8646/96/$5.00 © 1996 IBM

to marketable form at the least expense. The stakes are enormous. In a semifinished state, such as flat-rolled sheet, the value of a single metric ton of steel is about $340. In 1995 alone, worldwide production of crude steel exceeded 730 million metric tons [1].

This paper addresses one specific aspect of production efficiency: how computer approaches can be applied to optimize the complex process of primary steelmaking. The primary phase of steelmaking starts with the molten mix of raw materials and ends with stock material used to make a variety of products in subsequent finishing operations. It is an exceptional manufacturing environment in that the process is continuous at its outset, dealing with pools of molten metal, and discrete at its conclusion, dealing with individual solid components of various sizes and shapes. This is one of several conditions that make the scheduling of primary steelmaking a difficult task and preclude the use of standard packaged solutions. Despite the difficulties, however, there are two compelling reasons to address the problem. First, the return on investment for software to support improved steelmaking productivity is very high. In an industry where a single new manufacturing unit, such as a continuous caster, can cost more than $250 million, and where the annual production of the unit is measured in hundreds of millions of dollars, an expenditure in software development to improve production even by a few percent is worthwhile. Many software solutions have significantly exceeded such a modest goal. The LTV caster scheduling system discussed later is one example in which a successful scheduling system contributed to improving the production level over the nominal design tonnage of the equipment by nearly 50% [2, 3].

The second reason for focusing on software solutions to enhance scheduling efficiency is that tools capable of addressing this goal exist today, available and proven, awaiting only the innovation required to bring them to bear on the task at hand. Solution techniques from both the mathematical science and artificial intelligence communities have already been applied successfully to subsets of the overall primary scheduling problem, and this paper traces several prominent and innovative examples. While the solution approaches discussed vary in scope and focus, they are unified by the common belief that computer-based scheduling can offer a large payback and dramatic competitive advantage in the steelmaking industry.

In the sections that follow, we provide a brief discussion of the steel industry, describing the current state of technology and the overall steelmaking process. We then focus on the primary stage of the process, in which molten iron is transformed into semifinished steel products of many shapes and metallurgical chemistries. For reasons that are discussed, this stage represents the most significant challenge from a scheduling point of view. We describe the primary steelmaking process and its associated scheduling issues, and we discuss a number of solution approaches that have been developed to address aspects of the scheduling challenge. Finally, we describe some of IBM's experiences in the development of primary steelmaking scheduling solutions and discuss several new approaches that seek to address the scheduling of primary steelmaking facilities in a more comprehensive fashion.

2. Technology in the steel industry

Among people not directly involved in the industry, many retain an image of steelmaking as it was in the Industrial Revolution and essentially remained through the first half of the twentieth century: a grimy business, hostile to the environment, manpower-intensive and fraught with hazards. Others may recall the indictments of the 1970s when, in the economic shift toward global market orientation, the American steel industry was condemned as highly inefficient and hopelessly out of date. Anyone entering a modern steelmaking facility with such expectations would be happily astonished to discover that today's facilities are bright, clean, and highly automated. To be sure, the basic operation remains characterized by intense heat and smoke as the impurities of pig iron are cooked away until the purified product is attained, but the processes are contained and controlled to conform with strict environmental regulations. Mills are extensively automated. The typical mill worker one encounters today is much more likely to be in clean clothes, sitting in front of a computer screen in an air-conditioned control room, than bare-backed and toiling in the orange glow of the furnace. Advances have proceeded along two related parallel paths: improvements in the actual equipment and processes involved in making steel, and complementary improvements in the technology for directing and controlling the process.

While the equipment and processes of steelmaking are being continually tuned and improved, two "quantum leaps" in approach stand out in the latter half of the twentieth century: the conversion from an open-hearth method of steel purification to any of several closed-vessel methods, and the advance in slab-casting technology from an approach entailing multiple discrete, labor-intensive operations to a single, continuous, highly automated process.

In the area of steel purification, the open-hearth method predominated in the first half of the century. Inherent inefficiencies, however, rendered the method economically prohibitive, with rising costs of energy and manpower, expanding throughput demands, and increasing concerns for the control and treatment of emissions. No open-hearth facilities have been built in the United States since the mid-1950s. Current steelmaking practice employs closed-vessel purification processes, in which heat is generated either by injection and combustion of oxygen in the hot metal bath, in the case of the basic oxygen furnace (BOF), or by the creation of an electric arc between electrodes protruding into the vessel, in the case of the electric arc furnace (EAF). Both methods effectively contain and control the effects of the process. Additionally, like their predecessor, the open-hearth, both the BOF and EAF accommodate the addition of cold scrap steel to the molten metal in conjunction with purification.

The second major improvement in primary steelmaking has been the advent of the continuous-casting method to produce solid steel of various cross sections from purified molten steel. These "solids"—bars, slabs, billets, and the like—are the initial stock used to make finished products (e.g., wire, structural shapes) and semifinished products (e.g., steel sheet for the manufacture of automobiles or appliances). Prior to the development of continuous casting, the process of converting molten steel to solid stock involved "teeming" the hot metal into ingot molds, cooling the ingots, stripping the molds from the ingots, reheating the ingots, and rolling them into the desired solid shapes. By contrast, a continuous-casting machine accepts molten steel poured in at one end, automatically directs it through a set of adjustable molds while cooling the metal at a controlled rate, feeds the resultant solid "section" out the other end, and torch-cuts it to the desired length. Throughout the casting process, metal never stops moving and does not require handling or reheating.

Modern and sophisticated approaches to management and control of the steelmaking process have evolved in parallel with improvements in process technology. Almost universally, the control and adjustment of a modern steelmaking unit is handled by a dedicated process control computer feeding signals to actuators on the unit. Such a "level-one" computer might also support applications to monitor and analyze production data from sensors on the equipment or from human observers. Process control computers may receive instructions from human operators or from another level of computer program. "Level-two" computers direct the process from a broader perspective, considering production requirements and, potentially, production and equipment status from the process control level. Scheduling programs dedicated to a single manufacturing unit are often executed on level-two computers.

The next level in the system hierarchy generally has responsibility for activities affecting more than a single manufacturing unit in the plant. "Level-three" computers may coordinate activities over an entire plant or a major area within a plant where interdependencies among units are critical. Level three, then, is the logical place for integrated planning and scheduling systems that consider the operation of multiple manufacturing units working together.

Finally, a fourth level is generally recognized in the computer hierarchy: the "corporate" or "enterprise" level. This "level-four" system has global reach; it maintains application programs and data important to the overall organization and drives the activities of the level-three systems under its umbrella of control.

Throughout all levels, new computing technologies are gaining acceptance. Many steelmaking enterprises are embracing sophisticated data-modeling and data-management approaches as means of handling huge volumes of customer and production information efficiently. Object-oriented software technology is gaining recognition as a preferred approach for building and maintaining the complex application programs required to manage the enterprise. Knowledge-based systems have been used for years in applications ranging from diagnostics and analysis to transportation planning, business forecasting, and production planning and scheduling. Other sophisticated techniques from the artificial intelligence community, such as neural networks, have been applied to process monitoring, analysis, and response. To be sure, there are many standard business applications as well. The encouraging point is that steel-industry decision-makers have demonstrated a consistent willingness to take calculated risks and invest in leading technology when it offers a potential advantage in their highly competitive industry.

3. Overview of steelmaking and the primary steelmaking process

The overall steelmaking process can be partitioned into three distinct stages, or areas: ironmaking, primary steelmaking, and finishing. For the purposes of this paper, all three are described briefly, to provide background and context for the discussion of production-scheduling issues; then, attention is focused on primary steelmaking, since it poses the most significant challenge from the scheduling perspective [4]. Frequently used terms and abbreviations are listed in **Table 1**.

- *Ironmaking*

Ironmaking entails combining the raw ingredients of steel into a generic intermediate product known as *hot iron* or *pig iron*. The raw materials typically are preconditioned in preparation for this process. Iron ore is processed into pellets, or *sinter*, having more uniform size and consistency than the raw ore. Coal, another raw material for ironmaking, is baked in ovens to produce coke, a derivative product with higher combustion efficiency. Sinter pellets and coke are combined with limestone in a blast furnace, a tall, columnar vessel, to achieve a reduction reaction.

Table 1 Frequently used terms and abbreviations.

AMT	Applied Manufacturing Technologies (department of IBM Consulting Group)
A-team	asynchronous team approach (developed at Carnegie Mellon University)
billet	steel piece with square cross section, smaller than a bloom
bloom	steel piece with square cross section, larger than a billet
BOF	basic oxygen furnace
CAP	customizable application program
cluster	production between caster turnarounds
EAF	electric arc furnace
futures	inventory
gauge	thickness
grade	steel with a specified metallurgical composition
heat	furnace-load of steel
inventory	manufactured goods for which no order exists
linkage	coordination of production from stage to stage of steelmaking
LMF	ladle metallurgical facility
rolling group	round
round	pieces of steel processed between roll changes
slab	steel piece with elongated rectangular cross section
stock	inventory
strand	stream of steel from a caster
string	cluster
TSP	traveling-salesman problem
tundish	receptacle at top of caster
turnaround	restart of the casting process necessitated by a variety of conditions

Hot air and the coke are introduced at the bottom, creating a temperature gradient in the vessel from about 1400°C at the bottom to about 250°C at the top, where the limestone and ore pellets are introduced. As these ingredients fall through the furnace (experiencing a temperature gradient), several actions take place. The ore is smelted and reduced through combination with carbon from the coke. The molten limestone serves as a flux; i.e., it forms a liquid *slag* that carries coke ash and other impurities away from the molten metal. At the base of the furnace, slag is drawn off for disposal, and hot iron is drawn off for transport to the primary steelmaking operation. A blast furnace operates constantly, with the materials being fed continuously and the product tapped periodically. This is a necessary condition, since shutdown of the furnace could necessitate a rebuild (rehabilitating the furnace and replacing its refractory lining, a procedure that may cost 70 to 100 million dollars and require as long as a year). For this reason, hot iron produced by blast furnaces is viewed in production planning as a defined, continuous supply, and the consumption of this continuous supply is an important constraint on the planning and scheduling of the next stage: primary steelmaking.

• *Primary steelmaking*

The primary steelmaking area accepts the supply of hot iron from the blast furnace and transforms it into semifinished products in a variety of *grades* (specific metallurgical compositions of steel), shapes, and dimensions. **Figure 1** provides an overview of the steelmaking process. Hot iron arrives in insulated vessels, often via rail. At this stage, the metallurgical composition of the iron is reasonably uniform from manufacturer to manufacturer. At the beginning of the primary steelmaking process, hot iron is combined with scrap steel and further heated in a refining furnace. Modifications and refinements are performed—through selective addition of scrap steel, addition of alloy components, and controlled heating to adjust the carbon content of the "melt"—to achieve metallurgical differentiation resulting in the many, often hundreds, of grades of steel produced by large steelmaking companies. When the desired composition has been attained, each full furnace-load of steel, or *heat*, is cast into a solid shape and cooled at least enough to remain solid. At this point, the metallurgical composition of the solid product is specific, conforming to a defined grade, but the shape may be one of several. Common cast shapes include *slabs* (with elongated rectangular cross sections of various thicknesses and widths), *billets* (approximately square in cross section, from 50 mm to 125 mm square), and *blooms* (approximately square in cross section, but larger than billets). The choice of an appropriate cast shape facilitates the further processing required to achieve the desired shape of the finished or semifinished product. Flat-rolled sheet steel, for example, is derived from slabs, whereas most structural shapes are rolled from billets or blooms. Cast shapes are rolled at elevated temperatures ("hot-rolled") to complete their primary processing. Products of primary steelmaking may be shipped to customers, typically for further processing by those customers, or they may advance directly to

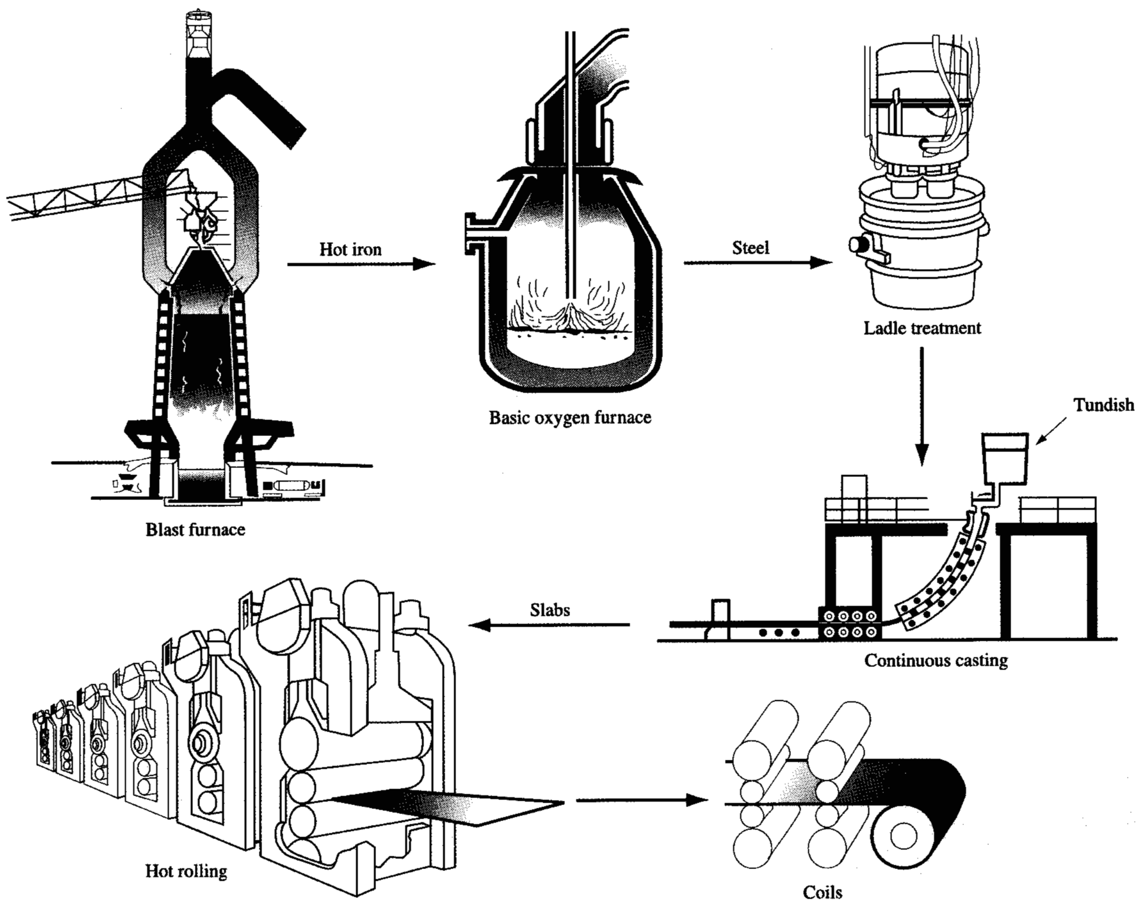

Figure 1

The primary steelmaking process.

finishing operations, in the third stage of the integrated steelmaking process.

• *Finishing*

The finishing area consists of numerous operations that can be applied selectively to the semifinished products of the primary steelmaking process, to achieve customers' specifications for final dimension, surface finish, mechanical properties, and coating. Cold-rolling (rolling the sheet or shape at ambient temperature) may be performed to reduce the product to precise dimensions, achieve a better surface finish, or modify mechanical properties. Annealing improves the material's grain size and affects its ductility. Tempering relieves internal stresses and improves ductility. A process known as pickling can be performed to clean the surface of the steel. Coatings such as galvanized zinc for corrosion resistance may be applied. Clearly, a great many options exist in the finishing stage; however, only certain subsets of the total finishing operation require close coordination, or *linkage*, from a scheduling point of view. In a coating operation, for example, it is desirable to run a continuous pickling process and the subsequent coating process in serial coordination, but many finishing operations need not be tightly linked. When the operations are linked, the combined constraints generally are less complex than the combined constraints arising from linked primary steelmaking operations, since state changes are not involved and the sizes of the batches on the various finishing units are reasonably consistent. Consequently, general-purpose discrete manufacturing-scheduling approaches have achieved respectable success in the scheduling of finishing operations, while similar approaches would fall far short of the goals in scheduling primary operations.

4. Primary steelmaking: A detailed look

Within the primary steelmaking area, several closely related process steps transform hot liquid iron into its myriad finished and semifinished forms.

When hot iron arrives in the primary area, it is poured, or *charged*, into a refining furnace, along with cold scrap steel. Further heat is then applied, to melt the combined charge into a homogeneous liquid state, remove impurities, and reduce the carbon content to a desired level. During this refining process, alloys can be introduced to achieve required metallurgical specifications for the particular grade being produced. As mentioned above, two types of refining furnaces predominate in current use: the basic oxygen furnace and the electric-arc furnace. In high-production operations, the BOF is more common. A typical production facility, or *BOF shop*, might consist of two vessels and produce about 35 heats per day, with each heat consisting of 200 to 230 metric tons of molten steel.

On the input side, refining furnaces are constrained by the requirement that they collectively must consume all hot iron arriving from the blast furnaces, a continuous supply with little available variation. On the output side, each heat of steel produced by a refining furnace is of a single specific grade, and furnaces normally are run for complete heats only. Therefore, one challenge in scheduling primary production is to make efficient use of material produced by the refining furnaces in full, grade-specific heat lots. Refining furnaces are also subject to certain constraints concerning the sequence in which different grades are made and the number of consecutive heats of certain grades they can produce, since some grades, notably low-carbon steels, require higher processing temperatures and may damage the refractory of the furnaces if too many such heats are scheduled in a row.

From the refining furnaces, molten steel is transferred via ladles, containing one heat of steel, to a ladle metallurgical facility (LMF). Strictly speaking, this step is optional, but it is becoming increasingly common. At an LMF a ladle might undergo any of several refining processes. If the desired carbon content was not achieved at the refining furnace, additional heat can be applied at the LMF to further reduce the carbon. Also, additional alloys can be added to the heat to adjust its chemical composition. Further, if there is a need to delay the steel before its next processing step, it can be held, and its temperature maintained, at the LMF. From a scheduling perspective, the LMF represents a processing step for which time must be allocated; however, it also provides some timing flexibility, since the LMF has the ability to share some duties with the refining furnace and to serve as a holding point before the material proceeds to the next step.

Molten steel from the LMF (or directly from the refining process) next moves to the casting step, where grade-specific steel first takes solid form. Continuous casting is the predominant approach in current practice. At a continuous caster, the ladles of molten steel are drained into a receptacle, called a *tundish*, at the top of the machine. Steel from the tundish flows into a long series of molds that simultaneously contain it, cool it, and move it forward. The formed steel section, or *strand*, solidifies from the outside surface inward and exits from the series of molds as a fully solid shape. Many continuous casters feature adjustable molds, enabling variation in the width of a steel strand as the casting process proceeds. Many machines also have several sets of molds operating in parallel, fed by a single tundish, so that several strands of steel (potentially of different widths) may be cast concurrently. At the tail of the casting process, gas torches cut the strands to lengths dictated by order requirements. The tundish provides a buffer of hot metal, so that an empty ladle can be removed and a new, full one positioned without interruption of the process flow. It is a refractory-lined vessel with a limited useful life: About ten heats is a typical lifespan, after which a new tundish must be "swapped in" while the casting process is momentarily suspended.

In a continuous caster, width transitions along the strand must occur gradually. For some grades of steel, width changes are limited to "narrowing in" only, since "widening out" while casting could cause the molten core of the strand to break through its surface crust, in an event known as a breakout. This restriction makes it necessary to terminate the casting process when the molds must be reset to a wider dimension, after which the casting process is begun anew. The procedure, known as a *turnaround*, is costly in terms of lost production and labor for setup.

The continuous caster imposes several constraints on primary process flow. Changes from one grade to another in the course of a casting sequence necessitate a certain amount of mixing of grades in the tundish; often, the portion of the cast product corresponding to the "mix zone" is of inferior quality. Some grades have such radically different chemistries that they cannot follow one another in a casting sequence unless the tundish is replaced. Since, ideally, casting durations should be of length sufficient to consume the full useful life of the tundish, radical grade changes should occur within a sequence only when the tundish is due to be changed in any event. Orders to be cast must be grouped in "heat lots," since molten metal is supplied to the casting process only in ladles, each containing a full heat of a specific grade. Widths of orders must be considered in the grouping process, to ensure that transitions can be accommodated within the casting sequence. Finally, the aggregate width of all strands being produced on the caster must be coordinated, over time, so that the consumption of hot metal at the caster neither outruns nor "backs up" the supply from the refining furnaces and LMF.

After steel has been cast into its initial solid shape, it is hot-rolled to the shape and size that will represent the final product of the primary steelmaking process. As the name implies, hot-rolling must be performed at high temperature. Specialized sets of rolls reduce the solid cast pieces to their final shapes. Because rolls wear down as they are used, the number of pieces that can be processed with a set of rolls is limited. Further, the order in which pieces can be processed is subject to grouping and sequencing constraints. With flat-rolled steel, for example, only certain families of grades having similar hardness can be processed together, and only a certain number of slabs can be rolled into sheet material between roll changes (the number may vary from one hundred to several hundred, depending upon the hardness of the steel). Further, for best quality, widths of the pieces to be rolled between roll changes must be sequenced from widest to narrowest, so that roll wear lines from the edge of one piece do not become embossed on the next. Additionally, each individual piece must be rolled to a particular thickness, or *gauge*. Like width changes on the caster, gauge transitions between pieces on the hot strip mill must be gradual.

Typically, some reheating is required prior to hot-rolling, even when rolling occurs directly after the casting process. Nonetheless, when compared to the handling and reheating costs associated with unlinked facilities utilizing a "stockyard" for intermediate inventory, the potential savings from linking the casting and rolling operations are significant. However, tight linkage of casting and hot-rolling steps has been an elusive goal, because of the inconsistency of the constraints between casting and rolling, and the added complexity of dealing with these units in a synchronized fashion. Consequently, the majority of primary scheduling programs in use today are focused on an individual piece of equipment. Only recently have systems evolved with the sophistication to take advantage of the benefits offered by integrating the entire primary steelmaking facility.

5. Scheduling issues in primary steel production

The primary responsibility of scheduling is to translate a higher-level plan into detailed operating sequences. Typically, scheduling is performed autonomously for each manufacturing unit or group of closely related units within a plant. A schedule generally covers a period of a week or less. Scheduling in the primary steelmaking area is concerned with the following issues:

- *Utilization of manufacturing units* As indicated in the previous section, the facilities at a steel plant are very expensive (especially the continuous caster and hot strip mill), so one of the objectives of scheduling is to produce a schedule that results in high utilization of these units.
- *Allocation of production among parallel manufacturing units* This is the issue of determining how the workload should be allocated among functionally similar units, such as two or more continuous casters within a steelmaking area. Considerations include balancing workloads, satisfying quality and metallurgical requirements, efficiently utilizing the processing units, meeting order due dates, and satisfying user preferences. In many cases, allocation may be prespecified by a higher-level plant order-processing system.
- *Specification of heats or heat groups* The process of specifying a heat entails grouping a sufficient number of planned production orders with the exact metallurgical properties and the right tonnage for a heat. Sometimes a group of heats instead of a single heat is specified. In this case, the heats within the group need not have the same metallurgical properties. These properties, however, must be compatible because of continuous-caster requirements.
- *Specification of rolling groups (rounds)* For the rolling mills, slabs, blooms, or billets must be grouped into large enough batches, called *rounds* or *rolling groups*, for efficient processing between roll changes or setups. For flat-rolled material, where slabs of various widths are to be rolled to various thicknesses, the sequence of slabs within the group is also specified as part of this task.
- *Sequencing of heats/rolling groups for manufacturing units* For the refining furnaces and continuous casters, this entails specifying the sequences in which heats of steel will be processed through the units. For rolling mills, rolling groups must be sequenced for the entire scheduling period, so that facility maintenance schedules and other operational constraints are satisfied.
- *Coordination of schedules between production stages* Balance of production and consumption rates between upstream and downstream units is considered here. As an example, schedules for BOFs and continuous casters must be coordinated to ensure that appropriate grades of hot metal are provided at the right times to keep the casters running efficiently. In addition, if tight linkage between casting and rolling operations is desired, slab sequences at casters and hot mills must be closely coordinated.
- *Rescheduling* When a disruption in production occurs, or when actual production deviates significantly from

scheduled production, rescheduling is necessary. Disruptions and deviations can occur for a variety of reasons:

- A manufacturing unit goes down.
- Excessive defects occur during an operation.
- A new, high-priority order is introduced.
- An order is canceled.

Depending on the severity of disruption or deviation, the appropriate response may be regeneration of the entire schedule, performance of minor schedule adjustments, or no formal remedial action. When rescheduling is required, determination of the scope and extent is a key issue. The best strategy might be as simple as diverting an order from an assigned manufacturing unit to another functionally similar unit with available capacity. At the other extreme, a major disruption or deviation could leave no viable alternative to regeneration of the entire schedule.

The environment in the primary production area can vary significantly from one steel plant to another because of differences in manufacturing unit capabilities, operating policies and objectives, and physical plant layout. For example, among plants producing similar flat-rolled products, casters can have single or multiple strands, adjustable or fixed-width molds, etc. Additionally, plants may operate in one or a combination of the following modes:

- *Cold-charge,* where the caster and rolling mill are totally unlinked. In the case of flat-rolled steel, slabs cool to ambient temperature in a slab yard after they are cast. Later, slabs are heated to rolling temperature in reheat furnaces prior to processing on the hot strip mill.
- *Hot-charge,* where slabs may be held for a limited time (8 to 16 hours) in insulated chambers before being rolled. This mode of operation reduces the amount of reheating that is required prior to rolling. In such a case, the caster is loosely linked to the hot strip mill.
- *Direct-hot-charge,* where slabs coming from the caster maintain the same sequence for rolling at the hot strip mill. This mode of operation requires tight linkage between caster and hot strip mill.

Further, some steel plants have more than one caster feeding a single hot strip mill that has multiple reheat furnaces. This introduces the potential to operate in a "mixed-rolling" mode; i.e., cold-charge slabs can be mixed with direct-hot-charge or hot-charge slabs in the same rolling group at the hot strip mill. This is an additional degree of flexibility not available to hot strip mills equipped with only one reheat furnace.

As these instances illustrate, a wide variety of environments can exist in the primary steel production area. Each case requires a different type of scheduling to address unique challenges such as the following:

- In slab casting, consumption of hot metal is directly proportional to the width of the slabs being cast. In situations where multiple casters exist, combined slab-width sequences can result in even greater variations in the demand for hot metal, potentially causing severe fluctuations in hot-metal inventory.
- Grouping of production orders for the casting operation must consider width-variation constraints of the casting machine as well as the sizes of the heats provided by the refining furnaces. This problem becomes particularly difficult for casters having multiple strands and adjustable-width molds.
- As described earlier, different degrees of linkage between caster and hot strip mill can have serious impacts on sequencing. In hot-charge and direct-hot-charge modes of operation, where linkage exists between the units, the following conflicts must be adequately addressed in the scheduling logic:
  - A slab caster should minimize the number of width changes in a casting sequence, yet the hot strip mill must follow a *width profile* (set of specifications regarding width changes) in rolling the slabs. Hot strip mills have several standard rolling profiles. The orders must be grouped for rolling so that the slabs in those orders conform to one of the standard profiles. Each profile has definite constraints with respect to slab width and gauge changes and other characteristics.
  - A caster should cast the same or compatible grades for as long as possible, yet the rolling mill can accept multiple grades in a single rolling profile, provided the grades are of similar hardness.
  - Gauges are irrelevant to the casting operation, yet they are a key consideration in creating rolling sequences.
  - A double-strand slab caster should cast from wide to narrow width on one strand and narrow to wide on the other in order to maintain a reasonably level hot-metal consumption rate; however, a hot strip mill should follow a profile in which widths decrease throughout the production cycle.
- Useful lives of tundishes and rolls must be considered in order to ensure that they are being utilized to the fullest extent possible.

Clearly, casters and rolling mills impose significant constraints, both on the types of items that can be grouped together for processing and on the production sequences of individual items within these groups. These constraints are most pronounced in the case of slab casting and rolling, because of the variable-width aspect of the problem. Scheduling issues in primary production of blooms or billets are similar but somewhat less complex. While bloom and billet casters typically have multiple strands, each strand has the same fixed-size mold. Further, most steel companies with bloom or billet casters operate in the cold-charge mode, while a majority of the producers of flat-rolled steel are focusing on maximizing the hot-charge mode of operation. In situations where the economical use of energy is a supreme concern, the trend is toward direct hot-charging.

The schedulers[1] in the primary area also must deal with several objectives, including on-time delivery of material to the finishing area, minimization of semifinished product inventory between caster and hot strip mill, maximization of caster and hot strip mill utilization, and maximization of tundish utilization. Some of these objectives conflict; therefore, the scheduler must deal with trade-offs while making sure that all required constraints are satisfied.

6. Scheduling for either continuous casters or hot strip mills

In this section, we discuss several approaches that have appeared in the literature to solving the problem of scheduling continuous casters and hot strip mills individually.

As discussed in Section 5, central issues in scheduling a continuous caster concern grouping orders so that they form integral heats, and sequencing heats so as to maximize capacity utilization (by reducing turnarounds) and minimize tardiness and operating costs. Tardiness measures the lateness of delivery of orders.

Most approaches to scheduling continuous casters treat them as stand-alone facilities that are coupled only loosely with further processing such as rolling. These approaches attempt to improve productivity of the continuous caster while reducing operating costs and maximizing on-time delivery.

The approaches range from a simple mixed-integer linear programming model of casting [5, 6] to complex models solved using heuristic techniques, like the models at LTV and Geneva Steel [6]. The Scheiker scheduling system [7, 8] uses a cooperative scheduling approach in which an expert system assists a scheduler in creating schedules, which the scheduler modifies using a graphical user interface.

One of the central issues in scheduling a hot strip mill is the selection of pieces that can be processed during a single setup of the mill. The group of pieces that can be processed between roll changes is referred to as a rolling group. Processing time for a single round can range from four to ten hours; a shorter round would most likely sacrifice some of the useful life of the rolls. For purposes of classifying orders and grouping, a typical operating strategy for a hot strip mill recognizes anywhere from five to twenty different round types, on the basis of parameters such as width ranges, gauge ranges, hardness ranges, and surface-quality requirements.

We discuss another approach to scheduling hot strip mills, Roll-A-Round, that uses heuristics based on the traveling-salesman problem and linear programming to build a single round at a time for a hot strip mill.

In the next few subsections we discuss details of all these approaches. We conclude by discussing merits of some of the solutions.

- *Sequencing continuous casters using mixed-integer linear programming*

Lally et al. [5] describe a mixed-integer linear programming solution [9] to the problem of caster scheduling. They consider a simple model of a steel plant in which steel is produced at an electric arc furnace, held in a ladle, and cast on a continuous caster. (They do not model the ability of continuous casters to change width.) They use a single objective cost function that is roughly the sum of the holding times in the three facilities and the cost of setting up a continuous caster after a shutdown. The program attempts to produce a schedule that minimizes this cost by coordinating the activities and minimizing turnarounds at the continuous caster.

The model can be modified to handle common constraints such as ranges in casting speed and allowable variations in time taken in the primary steelmaking area. This model produces a single solution quickly, and the scheduler can modify the inputs and obtain another solution. This model for optimizing the casting sequence has been tested on a model of the casting problem for Allegheny Ludlum. Test cases have shown a reduction of cost of the order of 4%.

- *LTV caster scheduling model*

The LTV Caster Scheduling Model [2, 3] was implemented in 1983, in support of the first continuous caster installed at LTV Cleveland Works. The model is intended to schedule caster production from customer orders while optimizing several key objectives:

- *Maximization of on-time delivery* Priority is given to orders with current (imminent) due dates and orders that are past due. Besides the obvious rationale of promoting customer satisfaction, this goal also minimizes the "inventory risk" associated with production of future orders that could be canceled before their due dates arrive.

[1] Throughout this paper we refer to human users of the scheduling software as schedulers. We also refer to the scheduling software as the scheduling engine. The scheduling engine plays a decision support role that enables the schedulers to better search the space of feasible schedules and choose one that best meets the needs of the enterprise.

- *Maximization of caster productivity* Factors that enhance productivity include casting-sequence lengths that fully utilize available tundish life, casting long sequences that minimize the required number of turnarounds and thereby minimize unproductive setup time, and casting with average widths sufficient to match hot-metal consumption in the casting process with the production capacity of the BOF shop.
- *Satisfaction of quality requirements* Overall product quality can be promoted by schedules that allow the caster to operate within its preferred ranges (e.g., casting speed, width transitions). In addition, quality requirements can be satisfied, and scrap loss minimized, by a scheduling strategy that places orders with high quality requirements early in the production sequence. This provides two degrees of flexibility. First, there is an opportunity to reassign a piece that does not satisfy the quality specifications of the order for which it was originally intended to another, less exacting order later in the sequence. Second, there is an opportunity to try again to make the piece for the original order later in the sequence.
- *Minimization of semifinished inventory* Since profit is made only when the product is sold, there is a strong motivation to construct casting sequences with minimal "stock" content (i.e., pieces included to make the sequence feasible for manufacture but not associated with a customer order). The penalties of producing stock material and future orders include inventory carrying costs and the inventory risk described above. Moreover, since facilities are booked to capacity in a strong business environment, production of inventory also represents a missed opportunity to produce a current order.

Additionally, the LTV scheduling model handles a number of constraints associated with the casting process, including

- *Grade compatibility* As a maker of engineered steels, LTV deals with customer requirements for hundreds of grades, with significant differences in metallurgy. Only certain grades can follow others within a casting sequence because of metal mixing that occurs in the tundish.
- *Width continuity* A continuous caster can change slab width only gradually within a sequence.
- *Conformance with heat-lot boundaries* Each heat of molten steel arriving from the BOF has a specific grade and will produce about ten slabs on average. Therefore, the casting schedule must recognize that each grade uses an integral number of heats.
- *Strand balancing* The LTV caster is a twin-strand machine, simultaneously casting two strands of steel from the same tundish. The slabs may be of different widths; however, the strands must be synchronized so that their lengths correspond at certain key positions in the sequence, such as heat-lot boundaries where grade changes occur and, of course, the end of the overall sequence.
- *Coordination with BOF production* The rate of molten steel production from the BOF is reasonably constant under normal conditions; however, the rate of steel consumption by the caster varies, depending on several factors, the most prominent of which is the width of slab being cast. Thus, it is necessary to coordinate the average aggregate width on both strands over the duration of the casting sequence in order to ensure that the consumption rate neither causes a backup of hot metal ahead of the casting process nor outruns the BOF.

From an operations-research perspective, the LTV caster-scheduling problem is characterized as two linked knapsack-constrained traveling salesmen, one representing each strand in the casting process. Each is knapsack-constrained in order to adhere to heat-lot boundaries. The two salesmen travel mutually exclusive routes, recognizing that any individual slab (a "city" on the route) is produced at only one point on one strand. The linkage between itineraries represents the need for both to reach certain corresponding key points at the same time (e.g., grade changes within the casting sequence, starting and ending points of the sequence). Path costs can represent such considerations as width changes between adjacent slabs, while node values can represent the relative priorities of slabs that are candidates for inclusion in the casting sequence. A heuristic algorithm was adopted, as the complexities and constraints of the problem are not amenable to a precise mathematical programming formulation.

From a metallurgical perspective, an approach is applied to reduce the number of different grades of steel that must be handled in the casting process, by creating metallurgical *cast families*, each of which can include several grades of similar chemistry. A cast family's metallurgical parameters are constrained such that the restrictions of all of its members are satisfied, thereby reducing the number of metallurgical-composition alternatives by a factor of 5 to 6 for caster-scheduling purposes.

The LTV caster-scheduling approach consists of several sequential steps. First, customer orders are evaluated by the system with respect to due dates and metallurgical similarities in order to identify candidates for inclusion in a casting sequence. At this point, the actual sequence in which orders will be produced is not determined; however, some constraints are applied during the selection process to ensure that the candidate set of selected orders can be arranged into a feasible casting sequence.

Next, the candidate set of selected orders is evaluated with respect to potential sequences that could be constructed and the corresponding starting widths on both strands associated with each sequence. At this point, the number of potential sequences is quite large, and a first-cut screening is performed to eliminate those sequences whose starting widths would not permit a consumption rate consistent with the production of the BOF shop, a sufficient number of heats between required turnarounds, or a sufficiently high average priority for the orders included.

Then, the remaining sequences are evaluated at each heat boundary according to an objective function with cost based on such production factors as due-date priorities, turnaround costs, width-change costs, and efficient tundish utilization. The evaluation also ensures that all sequencing constraints are satisfied.

The actual casting schedule is then built, for both strands in parallel, from the orders in the sequence with the least cost. The algorithm enforces satisfaction of width-transition requirements, and orders are arranged with higher-quality requirements earlier in the sequence for reasons discussed above.

Finally, the hot-metal consumption rate is computed for the composite width of both strands, and a heat sequence is constructed. The BOF shop uses this information to determine the times when it should begin production for the heats that will be consumed in the casting sequence.

The caster-scheduling model has allowed LTV to do a very effective job of generating schedules for the continuous-casting facility as a stand-alone unit, which is particularly notable considering that LTV had no prior experience with continuous-casting technology or its associated scheduling issues. The scheduling model helped LTV to achieve, in January 1989, the record monthly production for a continuous-casting machine in North America (over 232000 metric tons, well above the caster's nominal design capacity of 150000 tons). In addition, the caster scheduler has proven to be a practical and flexible tool, as evidenced by its acceptance by LTV production planners.

The model was intentionally limited in its scope to focus on the continuous-casting unit and upstream communication of requirements to the BOF shop. There is no coordination with downstream rolling requirements at the hot strip mill, because the first continuous caster at LTV was located several miles from its associated hot strip mill. Linkage is by rail car. Accumulation of intermediate inventory between casting and rolling units and resequencing of slabs prior to the rolling operation are routine aspects of the LTV manufacturing process. Communication between the human production planners responsible for developing unit schedules at the caster and at the hot strip mill has achieved a reasonable degree of coordination. By design, the caster-scheduling model takes a local view of production requirements; the system builds a single sequence without considering the balancing and coordination of production over a longer planning horizon, say one or more weeks. The more strategic view is addressed in a manual mode by production planners, using the scheduling model as a tool to support their decisions.

In 1993, a second continuous caster was commissioned at LTV Cleveland Works. While the original caster remains in operation, the new caster was conceived as part of an overall direct-hot-charge complex, wherein production flows continuously from the BOF through the casting and rolling processes. In support of this production approach, a new planning and scheduling system was developed jointly by LTV and IBM, incorporating some of the concepts of the original caster-scheduling model but adding the capabilities to schedule a linked casting and rolling operation and to schedule under the umbrella of an overall strategic production plan. This is discussed below in the section on the integrated caster–hot strip mill scheduling system for direct-hot-charge rolling.

• *Scheiker: Cooperative scheduling of steel refinement and casting*

Scheiker [7, 8] is a cooperative (man–machine) scheduling tool that helps to schedule the primary steelmaking area of a plant, including the production of slabs at a continuous caster. It has been implemented at the NKK Keihin plant, where it has been documented as saving $1 million per year.

Scheiker has been implemented as an expert system kernel with a powerful graphical user interface. In Scheiker, the scheduler, in cooperation with a scheduling engine, iteratively modifies an existing schedule through the user interface until the schedule is satisfactory.

A full-screen, visually oriented user interface allows the user to directly manipulate all features of the schedule. The scheduler typically makes global changes that increase global efficiency. The result can be a schedule in which not all local constraints are satisfied. The schedule is then passed to the scheduling engine, an expert system, which uses a rule base to recognize violations of local constraints (e.g., manufacturing unit conflict), and which makes local changes to resolve them (e.g., selects alternative units). The rule base captures experts' domain knowledge.

The scheduling engine produces the local modifications by means of a two-step process of subscheduling and merging. Subscheduling, performed from the caster backward to the BOFs, minimizes waiting time and resolves manufacturing-unit conflicts. Subschedules are merged to minimize manufacturing-unit conflicts.

The scheduling engine leaves the global structure of the schedule unchanged. The scheduler is presented with the changed schedule, along with notifications of any constraint violations. The scheduler attempts to fix these constraint violations by making global changes. This process of cooperative interactive refinement continues until the schedule satisfies all constraints and performs well with respect to the objectives of the scheduler.

This cooperative scheduling mode allows the scheduler to focus on the few changes that are globally important and the machine to focus on local improvements that eliminate constraint violations and on other local considerations.

Scheiker greatly speeds up the scheduling process while leaving the human scheduler in charge of the process; thus, it avoids the pitfalls of scheduling systems that attempt to produce schedules fully automatically but fail to take into account key constraints.

• *A caster-scheduling program at Geneva Steel*

An extensive facility-modernization project is currently under way and approaching completion at Geneva Steel. A key element in this effort is the installation of the world's widest slab caster. In order to derive maximum benefits from the capital improvements being introduced, Geneva Steel engaged consultants from the Applied Manufacturing Technologies (AMT) Department of the IBM Consulting Group to help design and develop a caster-sequencing optimization system [6].

As at LTV, radical grade changes, major width changes, an insufficient supply of hot metal, and scheduled caster maintenance cause the caster to take a turnaround. In addition, mold changes and changes in thickness cause turnarounds.

Each continuous production flow between caster turnarounds is called a *string* or a *cluster*. Since caster setup is time-consuming, longer strings are preferred in order to increase caster utilization. It is not uncommon to see a string as long as 40 heats in slab casting. However, various limitations such as order mix, facility capacity, and heat size prevent long strings often enough. On the other hand, even if ideal conditions exist at the caster, no productivity at the caster can be realized without a sufficient supply of liquid steel from upstream facilities. The way in which the caster strings are formed and sequenced directly affects the productivity of the caster and the upstream and downstream manufacturing units.

Many factors encountered specifically at Geneva Steel contribute to making caster sequencing a complicated problem:

- There can be many ways to organize the caster production, and the number of possibilities increases dramatically as the number of strings becomes larger. For example, there are 5! = 120 possible ways of sequencing five strings, but 10! = 3628800 possible ways of sequencing ten strings.
- Conflicting goals of manufacturing units often make it difficult to resolve conflicts consistently. From the point of view of the caster, long caster strings with wide slabs mean fewer turnarounds, and thus lower operating costs and more throughput; however, as at LTV, this could cause the caster to outrun the capacity of an upstream manufacturing unit such as an LMF. The Geneva scheduling system also considers the effect of caster scheduling on downstream units. While it may be to the advantage of the caster to cast a long string of the same product for a prolonged time, the capacities of the downstream facilities may not be able to accommodate the output from the caster, resulting in increased inventory and reduced profit levels.
- Tundishes and the nozzles underneath them are very expensive. Each time the caster takes a turnaround, the tundish must be changed. Even if there is no turnaround, the tundish must be changed every 10 to 12 heats, depending on the grade of steel being poured. Nozzle life is about one half of tundish life. Thus, the nozzles are changed at each tundish change and every 5 or 6 heats. A string that utilizes only partial tundish or nozzle life reduces the profit.
- Capacities of upstream and downstream units and lead times of products add to the complexity of the problem. Each mill has different operating constraints. Different products may go through different production paths. The time required for one product to go through its production path may be different from that required for some other product. When caster sequences are generated, it is necessary to ensure that capacities of downstream units are not overloaded in future weeks.
- Various time windows (scheduled operating times) for finishing units place limits on when certain orders should be cast, because crews at some finishing mills will be available to roll certain products coming directly off the caster only at certain prespecified times of each week.
- There are often more orders than can be produced because of limited capacity; therefore, selection of a subset of the orders for a specific time period is necessary. Determining which orders should be included in the plan for the specified period requires consideration of factors such as caster capacity, when the crew is available, and utilization rates at all manufacturing units. It should be noted that different crew groups are used for rolling different products, and these groups can be scheduled only during specified periods.
- While thickness changes are extremely time-consuming, other changes on the caster may be less expensive.

Different steel-grade changes cause the caster to have turnarounds of different lengths, as do mold changes, width changes, and thickness changes between strings.

Given the complexity described above and the many constraints, requirements, preferred practices, etc. that must be considered, one may ask what is the best way of planning the production at the caster for a given production period.

The Caster Sequencing Optimization program developed by IBM for Geneva Steel optimizes throughput using a group of algorithms to develop an initial schedule and using improvement algorithms to iteratively improve the initial schedule. The orders are selected and grouped into strings by applying a heuristic algorithm. The strings are then sequenced by another heuristic to develop an initial solution. With this initial sequence as the base, the string-sequencing problem is modeled as a traveling-salesman problem (TSP) and solved by means of an approximation algorithm. Two other improvement algorithms are then applied to the TSP solution to further improve the sequence. A stopping rule is put in place to monitor the progress of the improvement algorithms. If the improvement is not significant with respect to given criteria, the iteration is stopped. This is done to keep the time taken to develop the string sequence within acceptable limits.

Initially, the heuristic for grouping the orders looks at the orders, manufacturing units, capacities, maintenance schedule, operational parameters (e.g., mold sizes and types), user parameters (e.g., the week in which certain orders must be cast), etc. and develops strings. While forming the strings, the heuristic ensures that the capacities of the upstream and downstream manufacturing units are not exceeded.

The heuristic for creating initial sequences creates the strings so that the total caster downtime is minimized. During the sequence-creation process, some constraints may be violated. An example of such a violation is that some strings overlap the periods when certain manufacturing units have maintenance scheduled. In such a case, some strings may have to be broken into two or more substrings.

After the TSP approximation algorithm is used on the initial sequence of strings in order to come up with an improved sequence, two other improvement algorithms, called the k-interchange and c-interchange algorithms, are invoked. The underlying algorithms in the two are the same; they are just implemented differently. The k-interchange improvement algorithm is applied to the strings for each week, while the c-interchange improvement algorithm is applied to all of the strings for the entire scheduling period, which is a month in Geneva's case. The k-interchange improvement algorithm is applied first for all weeks in the month, after which the c-interchange improvement algorithm is invoked to further improve the sequence. The k-interchange improvement algorithm attempts to improve the solution provided by the TSP approximation algorithm by breaking the sequence of strings for each week into k parts and rearranging them, thus forming a new schedule. The schedule is broken only at the boundaries of the strings. An individual string is never broken. The c-interchange improvement algorithm breaks the entire schedule into c parts and rearranges them to form a new schedule. In both cases, the new schedule is guaranteed to be feasible and at least as good as the starting schedule. The creation of the string sequence is an iterative process. The purpose of architecting the solution in this form is to provide the scheduler control over the quality of the schedule as well as the time taken to generate the schedule. The iterative process allows the user to stop at any step when the quality of the schedule meets the needs of the enterprise.

The Caster Sequencing Optimization program also includes a heuristic for coordinating the capacities of all facilities. It also ensures that all hot metal coming out of the blast furnace is consumed.

In summary, the Caster Sequencing Optimization program incorporates state-of-the-art optimization techniques designed to maximize caster utilization, minimize caster setups, maximize tundish and nozzle utilizations, and synchronize with other upstream and downstream facilities while satisfying a wide range of constraints such as setup requirements, capacities of upstream and downstream manufacturing units, allowable buffer levels, required manufacturing-unit maintenance times, and crew-availability periods. It selects orders, builds strings, and develops an optimal sequence of strings for each week and for a month. The program provides the scheduler with the opportunity to override any of the default parameters as well as the selection decisions made by the program.

- *A hot strip mill scheduling application*

The Roll-A-Round program, developed by Balas & Martin Associates of Pittsburgh, generates unit schedules for hot strip mills [10, 11]. It schedules a single round (the pieces processed between roll changes) for a hot strip mill, according to the round type and parameters defined by the scheduler. The schedule observes the constraints of the mill and attempts optimization both by selecting the orders for inclusion in the schedule and by arranging the sequence for rolling the individual pieces. It recognizes constraints associated with inclusion of specific orders in the round under consideration, the requirement for smooth thickness (or gauge) transitions between adjacent pieces in the rolling sequence, the requirement for a generally decreasing width pattern in the rolling sequence, and the number of pieces in the round. The application program attempts to optimize the net business return on a schedule by maximizing the priority of orders included in the schedule (according to the date due to the customer), maximizing the number of pieces processed in a schedule (to exploit the full useful life of the rolls), and minimizing operating costs (such as transition costs between width groups in a sequence).

To accomplish its goal, the program breaks the problem into two subtasks. The first, preprocessing, organizes candidate orders into groups based on physical dimensions, quantity of slabs, and order priorities. Each group is evaluated on its ability to meet customer due dates and to make efficient use of the equipment. A transition cost, the production expense associated with required adjustments between the last piece in one group and the first piece in the next, is computed for each pair of groups. Preprocessing provides the data required to formulate the problem mathematically.

The second subtask is the actual solution of the mathematical problem. The problem-solver within Roll-A-Round is predicated on the observation that the distribution of orders within a candidate set has a large impact on the quality of the solution and that no single method is universally superior for creating the rolling sequence; therefore, the system provides three different routines to generate multiple candidate schedules for subsequent evaluation.

The main routine views the problem as a knapsack-constrained, prize-collecting traveling-salesman problem [12] wherein the possible linkages between orders in a round are modeled as the edges of a network and the goal is to maximize prizes (high-priority orders) collected at the nodes, while simultaneously seeking to minimize transition costs and avoid penalties incurred by failing to visit high-priority nodes. This linear-programming formulation generally yields subtours rather than complete rounds; a heuristic routine is then applied to assemble subtours into rounds.

Two other round-building routines are also applied. Both are heuristics that attempt to build a round one step at a time. One routine starts with a short path of the network that it expands to a full round by successive search and inclusion of preferred order groups. The other builds several independent paths and selectively links them to form a complete round. The two routines also differ in the selection strategies they employ for building paths from candidate order groups.

Finally, all solutions generated by the three routines are examined by an improvement algorithm that seeks to ensure local optimality by applying heuristics to exchange individual orders within a sequence. Roll-A-Round then evaluates the candidate rounds in terms of total weight of material processed and relative production cost. One or, optionally, several of the solutions are then presented to the scheduler.

Roll-A-Round is used for production scheduling of hot strip mills by several steelmakers. While its developers note that the results produced are not strictly optimal, results are believed to be of consistently high quality because of the variety of solution techniques considered. The linear programming formulation of the traveling-salesman procedure produces results that typically are within 2% of optimal, so rounds derived from this technique can reasonably claim to be nearly optimal, prior to application of the heuristic local-improvement algorithm.

By design, the application program takes a local view of production requirements at a hot strip mill. It takes the best candidates from the current order group to build a single round. In practice, a human planner considers the balancing and coordination of production requirements over a longer time period. In this context, a tool such as Roll-A-Round is useful as a decision-support tool for the planning process and as a tactical tool for building detailed, individual rounds for the hot strip mill. While Roll-A-Round is not designed to consider the tightly linked operation of a hot strip mill fed by a continuous caster in the direct-hot-charge mode, this case could be addressed through the incorporation of additional constraints to the mathematical model, to account for the sequencing requirements of a linked caster. Some components of Roll-A-Round have been incorporated into the overall direct-hot-charge solution at LTV.[2]

In this section we have discussed some of the techniques that have been used for scheduling either continuous casters or hot strip mills. The simple mixed-integer programming model of scheduling [5] does not consider all of the complexities of a real continuous caster. More complex models such as [2, 3] have been used by LTV to maximize caster productivity. The Scheiker scheduling system [8] represents the state of the art in cooperative scheduling, in which an expert system aids a human scheduler in generating a schedule. The Roll-A-Round hot strip mill scheduling program produces nearly optimal single rounds for a hot strip mill alone. The Geneva system [6] is a state-of-the-art application program that schedules a continuous caster while considering business constraints and capacity constraints of the various facilities upstream and downstream.

7. Integrated scheduling at casters and hot strip mills

Throughout the previous sections, we have described the production-scheduling problems in either casters or hot

[2] Presentation by Donald G. Herbe at the AISE Conference, Nashville, TN, February 4–6, 1996.

strip mills. None of those scheduling techniques can generate integrated schedules for both the casters and the hot strip mills. As mentioned, the operations and constraints at the casters and hot strip mills are quite different, which makes it very difficult to generate integrated production schedules. We have not found any approaches for addressing integrated production scheduling in the open literature.

In this section, we introduce techniques for integrated scheduling systems that have been developed by IBM for its customers or for exploratory purposes. The first two techniques are different approaches for the same scheduling problem at LTV Steel. The first of these uses a fast heuristic search algorithm followed by a genetic algorithm [13] that improves the solutions. The second approach uses a solution framework called asynchronous teams, or A-teams, which can use one or several different solution techniques.

Creating production schedules at steel plants operating in direct-hot-charge mode is known to be among the most difficult scheduling problems. The third approach is a new integrated scheduling technique for producing direct-hot-charge schedules for a caster and hot strip mill while optimizing the operational cost, tardiness, and inventory.

Since the first two solution approaches are based on the situation at LTV Steel that has twin-strand casters and hot strip mills, we briefly present the primary operations at LTV Steel before discussing the approaches.

• *Problem description*

The Cleveland Steel Works of LTV has two steelmaking facilities, one at the east side and the other at the west side, where each facility has a twin-strand caster and a hot strip mill. The older caster (on the east side) produces slabs, approximately 40% of which are hot-charged. The newer caster (on the west side) permits direct hot-charging. This caster takes more frequent scheduled breaks than the east caster because of the nature of direct hot-charging. As the slabs come off the caster, they go directly to the hot strip mill, in the same sequence. The hot strip mill must gradually narrow the widths of the slabs it rolls in a particular rolling group.

Casters are most productive when run continuously. The east machine is scheduled to be down for maintenance only once every three weeks; therefore, a steady supply of iron and steel is required. The caster runs at different speeds and different widths, so throughput varies. Blast furnaces also run continuously; they are not easily controlled, so one has to use the iron as it is ready. With two casters operating, scheduling the iron supply becomes extremely critical.

A cluster is an ordered set of production orders from customers. Given a set of clusters for each caster, along with planned turnaround times, the scheduling program should *quickly generate a coordinated sequence of these clusters that satisfies all constraints, meets as many objectives as possible, and, above all, manages the iron supply within control limits*. Most of the constraints and objectives apply to the casters separately, but the iron supply is a joint constraint. The program must schedule a week's operation while observing the objectives and constraints that have been described previously.

Generating such a production schedule for two twin-strand casters is difficult, with more constraints and objectives than can be handled by the existing manual methods. Moreover, even for situations in which we often cannot find solutions satisfying all of the given constraints, we still want to find a set of subsequences, each of which satisfies the constraints.

• *Production scheduling for two twin-strand casters and hot strip mills*

We describe the primary production scheduling procedure for the two twin-strand casters at Cleveland Steel Works of LTV Steel [14]. The scheduling system, developed by IBM Research and AMT, exploits a fast sequencing algorithm followed by a genetic algorithm for schedules satisfying global constraints.

Finding sequences

The scheduling process consists of three stages: The first stage, clustering, creates multi-heat clusters of single cast grades; the second stage makes sequences of these clusters; and the third stage coordinates these sequences.

Clustering begins with allocation: Each order is assigned to the east or west facility, according to strategies that take into account differences in caster and mill characteristics. Then east-side and west-side clusters are formed from current and past-due orders plus some future orders or stock, both of which are used to fill gaps in the order book and to deal with small-volume grades. For an example of a cluster, refer to the part of a typical sequence of a weekly schedule at the east side shown in **Table 2**.

The first row indicates that Monday's production begins with a cluster of grade 9001. Ten heats of this grade will be produced by the BOF. As the steel is poured into the tundish, strand 1 is set to make a width of 73 inches, while strand 2 is set at 37 inches. As the strands are formed, strand 1 is gradually narrowed, so that the tenth heat ends at 50 inches wide. Strand 2 is gradually widened until it too ends at 50 inches. Thus, the aggregate width of the two strands begins at 110 inches and ends at 100 inches, which is ideal for coordinating the caster with the BOF, since the consumption rate of the molten steel can be kept almost constant. The starting and ending widths, and how they narrow and widen, depend on the specifics of the orders.

Table 2 Part of a sequence of a weekly casting schedule. Each row represents a cluster. The last column indicates whether the old tundish is replaced with a fresh one at the end of the cluster ("fly") or the next heat uses the same tundish although its steel grade is different from that of the previous one ("mix").

Day	*Cast family*	*No. of heats*	*Strand1 width (in.)*		*Strand2 width (in.)*		*Fly or Mix*
			Start	*End*	*Start*	*End*	
Monday	9001	10	73	50	37	50	Fly
	9353	20	52	60	52	40	Fly
	9302	1	58	60	42	60	Mix
	9308	5	60	70	58	40	Fly

The sample sequence at the east side in Table 2 shows that the ending widths of each row are close to the starting widths of the next row. There is a limit to the amount of change in width allowed from one cluster to the next, since the process is continuous: The maximum change on each strand is 4 inches. The widths are sometimes chosen specifically to match the widths of adjacent clusters. For example, row 2 in Table 2 shows the grade-9353 widths both starting at 52 inches but ending at 60 inches and 40 inches. Since there are many orders for this grade in all widths, the scheduling system can deliberately choose the starting and ending widths. This permits a *bridge* between the adjacent grades, which may not be as flexible.

The last column in Table 2 indicates whether or not the tundish is replaced after the last heat in the cluster. Since the ten heats in row 1 would use up the life of the tundish, it is necessary to bring in a new one (*fly*). The second row shows 20 heats, so a tundish change is called for after the tenth heat. In the case of row 3, there is only one heat of cast family 9302. Since the tundish is still good, it is not changed; rather, the next grade, 9308, is poured into the same tundish. This is called a *mix*, because the second grade is poured into the last steel from the first grade. It is essential to make sure the two grades are compatible. Several slabs are thus a hybrid grade.

Given sets of clusters for the east and west sides, the scheduling system arranges them into a weekly sequence. As a first step, for each side a graph is built in which the clusters are nodes and any two clusters that can be scheduled in succession without violating any relevant constraints are connected by a link (usually directed).

A sequence of nodes connected by links is called a *supercluster*. For creation of superclusters, a heuristic search technique is used, which successively labels links of the graph as either "discard" or "commit." Superclusters are dynamically created by the search technique through successive labeling (merging) of nodes or other superclusters in such a manner that the number of final superclusters is minimized. By doing so, we can minimize the number of tundish changes, and at the same time maximize the use of tundishes.

In addition to the heuristics for the graph search, the search technique exploits constraint propagation to reduce the search space and the beam search technique [15], which limits the width of each search node.

At the east side, if the number of superclusters is greater than one, there are width gaps between the superclusters. A set of five superclusters A–E, for example, will require four bridging clusters to fill the gaps: A → bridge → B → bridge → · · · → E. Thus, if A ends with widths of 55 inches and 45 inches and B starts with widths of 72 inches and 38 inches, a bridging cluster would be formed to run from 55 inches to 72 inches for one strand and 45 inches to 38 inches for the other. Bridges are formed from large-volume grades for which there are orders in most widths.

Finally, sequences of superclusters are mapped onto a time line. At this point, some of the time-dependent global constraints can be accounted for. For example, it will be clear how much steel is available from one supercluster to the next. Also, planned-maintenance turnarounds can be inserted. Since the scheduling system is creating multiple candidate sequences, those with the fewest constraint violations can be chosen for further processing.

The west side is operated in the direct-hot-charge mode. Scheduling is similar to that at the east side but must account for differences in caster operation. More frequent turnarounds are allowed, but some hot-strip-mill-related constraints are added. Each cluster, for example, has a hot mill *run type*, and each type has rules governing which other types it may or may not precede. In addition, there must be opportunistic gaps created in the caster schedule to permit running the hot mill in the cold-charge mode.

The search process in the scheduling system creates one or more sets of sequenced superclusters for the west side. The sequence is provisional, since it is altered by a swapping algorithm to balance the iron supply between the east and west sides.

Coordinating east and west sequences

In balancing the iron supply, the system uses as input a profile that shows the amount of iron produced each hour. The iron can be directed to either caster as required. The goal is to schedule both casters so as to stay within the control limits. This will avoid dumping iron because it cannot be consumed fast enough, or experiencing costly turnarounds when the iron is not available.

This swapping of superclusters is performed by the *genetic algorithm* [13] which creates a population of schedules in which each member represents the sequence of superclusters. It then evaluates the merit of each member by counting violations of iron and steel constraints. (The steel constraints are not joint, since each side has its own BOF.) The population (a set of schedules) at the next generation is produced according to probability based on merit. Members with higher merit participate in this reproduction process; those with lower merit are therefore implicitly destroyed. This reproduction process continues, with occasional perturbation, until there are no violations or until time runs out. If time runs out and the schedule is unsatisfactory, the program can be rerun with a different initial pair of supercluster sets. The genetic algorithm converges toward optimal, but it may not reach it within the time limit; however, with the limited test data used so far, we have found that it is usually possible to obtain zero violations after a few reruns.

The performance of the scheduling system is quite satisfactory, requiring approximately ten seconds on an IBM 3090 mainframe to create a sequence. Approximately half of the run time is used by the last step, the genetic algorithm.

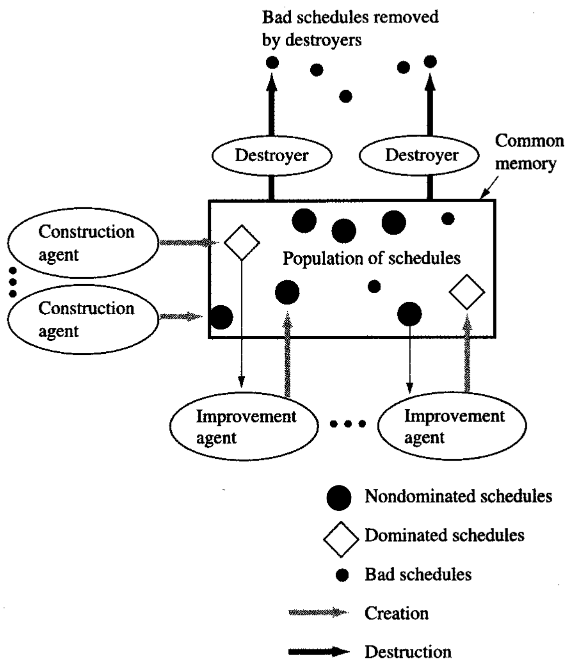

Figure 2

Creation and destruction of schedules by autonomous agents in asynchronous teams (A-teams), which are sets of independent agents that work iteratively in parallel to improve a population of solutions. Construction and improvement agents create new solutions and improve them; destruction agents delete bad solutions while keeping promising solutions. Together, they *herd* the set of solutions toward the Pareto frontier.

• *Production scheduling using the A-team framework*

The A-team approach

We discuss an approach that is well suited to addressing the unique requirements for scheduling in the primary steel production area discussed in the previous subsection. This approach, called asynchronous teams, or A-teams [16–18], was originally developed at Carnegie Mellon University by Prof. Sarosh Talukdar. An A-team is a framework that facilitates combining multiple problem-solving programs (also called agents) into a team, so that the results are better than those produced by any single agent. These multiple agents work together, but independently, on a common problem in order to produce a population of high-quality solutions. Since the primary steel production scheduling problem requires a variety of solution methods, the A-team solution architecture works well in this area.

One key feature of A-teams is that there is no central control. Individual agents have control over what they work on and when they work. Therefore, it is easy to add or delete agents. The A-team approach provides an open architecture that is very useful when building tools for solving difficult problems such as primary steel production scheduling. We can incorporate several problem-solving agents that use different strategies at the same time. Even if one agent fails to come up with a good solution, another agent may find such a solution.

In an A-team, agents work on a population of solutions that are accessible to all agents, rather than only on the results of their own work. This feature provides robustness and allows cooperation among the agents. An agent can work on a solution produced by another agent with different abilities, to produce even better results; in effect, agents build cooperatively on one another's abilities. This process of iterative improvement with multiple agents cooperating is an important feature of A-teams.

Figure 2 is a simple diagram of the A-team approach. The rectangle in the center is the common computer memory, accessible to all agents, that stores the population of schedules. We have implemented an A-team with three kinds of agents: creation agents, destruction agents (or destroyers), and evaluation agents. Creation agents can be of two kinds: construction agents that generate new solutions from the inputs, and improvement agents that select existing solutions and attempt to improve their performance on desired objectives. These agents produce new schedules, which are added to the population. Some of these schedules are good with respect to predefined objectives (minimizing order tardiness, operating costs, etc.), whereas others are bad. Evaluation

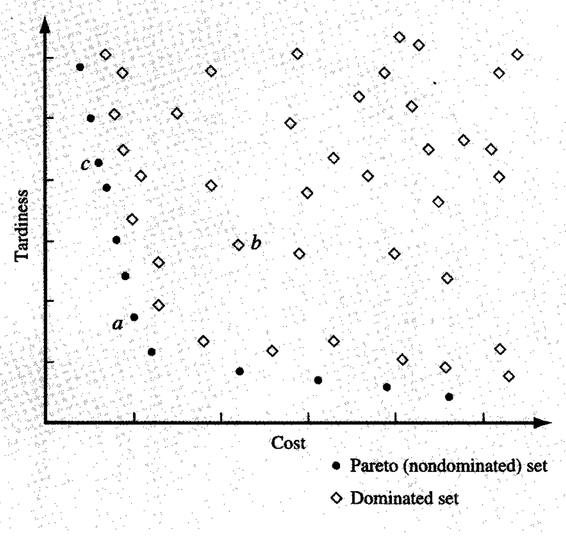

Figure 3

Domination and the Pareto frontier in multi-objective problems. For the two-objective problem (considering cost and tardiness), schedule *a* dominates schedule *b*. Neither schedule *c* nor schedule *a* dominates the other. The set of feasible schedules that are nondominated (solid circles) constitutes the Pareto frontier.

agents evaluate schedules and store the results with the schedules.

If creation processes were the only ones in A-teams, the population would reach unmanageable proportions; therefore, it is necessary to control the population. Other agents, called destroyers, perform this function. The creation agents, both construction and improvement, attempt to produce schedules that are close to the Pareto frontier (defined in the following subsection). Destruction agents destroy schedules selectively. One heuristic that may be used is to destroy schedules that are far from the Pareto frontier. Together, the improvement and destruction agents "herd" the population of schedules toward the Pareto frontier.

• *Multi-objective problems and the Pareto frontier*

The scheduling problem in the primary production area is a multi-objective problem. Single-objective optimization approaches have the premise that it is possible to combine all objectives into a single metric that is then optimized. In many types of scheduling problems, this metric is related to cost or the profit margin. The scheduling system attempts to find a least-cost or maximum-profit solution. The main drawback of such an approach is that it is difficult to specify the trade-offs between different objectives by using specific "weights." In the primary steel production scheduling problem, it would be necessary to specify weights for objectives such as operating costs and order tardiness. These weights could vary according to day-to-day operations and other business factors such as the importance of a customer order. An additional complicating factor is that the weights are not known *a priori*. Only after carefully analyzing a set of schedules can a scheduler make a guess at them.

The concept of a Pareto frontier is relevant in a multi-objective problem environment. We explain this concept with the help of **Figure 3**, which shows a number of schedules whose performances on the objectives differ. The two objectives used in the figure are cost and order tardiness. Schedule *a* dominates schedule *b*, since it has both lower tardiness and lower operating cost; however, neither schedule *a* nor *c* dominates the other, since schedule *a* has lower tardiness than schedule *c*, while schedule *c* has a lower operating cost. A schedule lies on the Pareto frontier if and only if it is not dominated by any other feasible schedule.

Because of the drawbacks of single-objective optimization discussed above, it is desirable that multi-objective scheduling problems be solved by approaches which provide schedules that are on the Pareto frontier. The scheduler can then choose the schedule on the Pareto frontier that provides acceptable trade-offs among the objectives. This changes the task of the scheduler from that of "massaging" a scheduling system to achieve the desired results to one of understanding trade-offs among multiple alternatives produced automatically by the system.

We use three main kinds of construction and improvement techniques for production scheduling:

- *Mathematical algorithms* Many of the problems in scheduling steel production correspond to well-known mathematical problems for which very efficient solution algorithms have been developed; e.g., the traveling-salesman problem, the bin-packing problem, and the shortest-path problem [19].
- *Domain-dependent heuristics* Using the characteristics of the scheduling-problem domain and the wealth of human experience in constructing schedules "from scratch" and improving such schedules, one can develop powerful problem-solving heuristics. Such heuristics may not individually be able to solve the problem; however, the A-team architecture allows one to combine these disparate algorithms and heuristics so that the results of the team are better than those of any single agent.
- *Search algorithms* Many of the scheduling problems in production industries are characterized by the fact that process and operating constraints limit the choice of feasible schedules. IBM has developed fast search

techniques that use these constraints to efficiently limit the search space and produce multiple feasible solutions.

IBM has extensive experience in using the above techniques to solve a variety of scheduling problems in primary steel production. In the next sections we discuss how we have used the A-team architecture for solving specific scheduling problems in the primary steel production area.

Application of A-teams to LTV cluster-sequencing problem
We have implemented an A-team of construction and improvement agents to solve the LTV cluster-sequencing problem. The agents in the team are as follows:

1. *Evaluation agents* These agents evaluate the tardiness, operating costs, and downtime of a complete schedule.
2. *Construction agents* We have implemented one construction agent that builds initial schedules by producing subsequences of clusters on the east and west sides and by sequencing these subsequences. The subsequences are produced using only local considerations; hence, the resulting schedules are suboptimal.
3. *Improvement agents* The A-team that solves the LTV cluster-sequencing problem has 53 built-in heuristic improvement agents; a few of these are the following:
   - *Improve the iron supply* This agent attempts to remove troughs and peaks in the iron supply by resequencing strings of compatible clusters on the east and west sides.
   - *Reduce turnarounds* This agent attempts to reduce turnarounds by appending sequences of compatible clusters. One method it employs is to use low-priority clusters that can fill holes.
   - *Reduce operating costs (tundish costs)* This agent attempts to reduce operating costs by reducing tundish costs. It does so by growing or shrinking sequences of clusters so as to almost fully utilize tundish capacity.
   - *Reduce operating costs (dissimilar round types)* This agent attempts to decrease the number of dissimilar round types in a sequence of clusters.
   - *Reduce operating costs (cold-charging)* This agent attempts to produce hot-chargeable clusters during the window that is assigned for hot-charging.

   The improvement agents choose which schedule to work on based on a modified form of *means-ends* analysis [18, 20, 21]. This ensures that improvement agents work mainly on schedules that they can improve.

   The improvement agents in this A-team are simple and very easy to implement. They are based on techniques discussed in standard textbooks on scheduling and combinatorial optimization [19, 22, 23]. We wish to stress that the A-team architecture makes it possible to keep the heuristics simple. As long as there is enough diversity of agents, if any individual heuristic fails to produce good results, other heuristics may succeed.
4. *Destruction agents* Destruction agents manage the size of the population and shape it by deleting undesirable schedules. To do so, they employ heuristics. One surprising aspect of A-teams is that the design of destruction heuristics is as important as that of construction heuristics. Following are two such agents:
   - This agent deletes schedules on the basis of an estimate of the distance of the schedules from the Pareto frontier, ensuring that the population of schedules in the A-team does not worsen with time.
   - This agent deletes schedules that do not meet *ad hoc* scheduler constraints. It refers to a file of constraints that is input by the scheduler. The agent enforces these constraints by deleting schedules that violate them. Because there is usually a lot of diversity of schedules in the population, it is generally possible to find schedules that meet the constraints input by the scheduler.

The construction agent produces an initial feasible schedule. The improvement agents work on this and other schedules produced by improvement agents and deposit their results in the common population of schedules. Improvement agents move the population of schedules toward the Pareto frontier, while the destruction agents remove schedules that are far from the frontier. The combination of the two sets of agents moves the population of schedules toward the Pareto frontier.

The A-team works by a process of continual improvement. When changes in the order book or in the plant occur, the evaluation of existing schedules changes. In this case, the A-team starts from the existing set of schedules and improves them to accommodate the new constraints. In this way, it easily and quickly handles the problem of rescheduling. The other advantage of this process of continual improvement is that the schedules get better as the A-team works. If the scheduler does not like the trade-offs represented by the current set of schedules, he can choose to wait and see whether the system can find better schedules.

We have implemented the A-team for the production-scheduling problem at LTV as described above. In about five minutes on an IBM RS/6000™ Model 320, the program generates feasible solutions that meet all constraints on the east and west sides and meet the iron-supply constraint. In about 30 minutes, it greatly improves these feasible schedules and provides roughly 50 nondominated schedules that represent trade-offs among operating cost, tardiness, and throughput. We have implemented a graphical user interface to enable the scheduler to view the set of generated schedules, display each schedule in detail, or view reports about each schedule. The scheduler can also modify schedules and return them to the A-team population.

We have found that the A-team provides a new way to implement cooperative man–machine scheduling. The system runs multiple experiments in the background and captures the results in a population of solutions. When presenting results to the scheduler, however, it "cleverly" chooses a subset of solutions available; no dominated schedules are presented. The scheduler examines the trade-offs represented by the set of nondominated schedules and chooses one that best meets the objectives of the organization. We have found that this process is intuitively easy for schedulers to understand.

- *Integrated caster–hot strip mill scheduling system for direct-hot-charge rolling*

For a number of steel mills, it is beneficial to have an integrated primary steel scheduling system supporting direct-hot-charge rolling. Direct-hot-charge rolling reduces the energy utilization and material handling caused by the resequencing of slabs. However, if the scheduling is not properly implemented, it may reduce the utilization of the continuous caster and the hot strip mill.[3] This is because it is very difficult to simultaneously consider constraints and objectives at both the casters and hot strip mills. As far as we know, no comprehensive reports of such work are to be found in the literature.

IBM Research and AMT initiated a joint project to develop a comprehensive, integrated, exploratory scheduling prototype for the primary area of steel plants, continuous casters, and hot strip mills in the direct-hot-charge mode. It is called the *Customizable Application Program*, or CAP. The problem is similar to that of the LTV west-side plant, but with the following differences:

- A twin-strand continuous caster is followed by a single reheat furnace and a hot strip mill. It is desired to have this configuration operate entirely in the direct-hot-charge mode.
- The steel mill's order book has one hundred or more cast grades and approximately ten round types. The order book typically has approximately 4000 orders, with past, current, and future due dates.

All constraints described above in the section on an overview of steelmaking are taken into account. For the caster, width-jump limits, criteria for mixing grades in a tundish, and tundish life are considered. For the hot strip mill, we consider the following constraints: slab-quantity limits for each width, width profile, width and gauge-jump limits, and roll life. In addition, a variety of scheduling objectives are included to drive the schedule-generation process. The key objectives considered are the following:

- Maximization of the percentage of time in the direct-hot-charge mode. The aim is to produce a schedule in which most of the slabs produced by the continuous caster are fed directly to the hot strip mill.
- Minimization of operating costs, including tundish costs, roll costs, reheat costs, inventory costs, and other schedule-related costs.
- Minimization of the tardiness of orders. The aim is to produce a schedule in which all current and past-due material is scheduled. However, capacity problems and conflicts with the preceding objectives of maximizing direct-hot-charge percentage and minimizing operating costs cause tardiness to increase.

The difficulty associated with the direct-hot-charge environment is twofold: First, the constraints on the caster and the hot strip mill are sometimes in conflict. For example, optimization for the caster alone results in processing the same grade and width of slabs as long as possible, whereas optimization for the hot strip mill results in following a width and gauge profile—grades are not as important. In a cold- or hot-charge mode, this problem can be managed by rolling the slabs in a different sequence from that in which they were cast. However, this flexibility is not available in a direct-hot-charge environment.

Second, this problem has multiple objectives; it is difficult to solve by using conventional problem-solving methods.

- *Solution approach*

We have completed an exploratory prototype that addresses a significant part of the overall problem. We have used some of the solution techniques described in the previous section and developed a few new ones.

The fundamental solution approach of CAP is essentially different from traditional steel-scheduling methods in that it does not create initial clusters from the order book in a preprocessing step before invoking the scheduling system. In the traditional approaches, those clusters are usually formed without consideration of many of the downstream constraints and objectives.

In CAP, on the other hand, the orders are not batched at the beginning of the scheduling process. Orders in the order book are classified according to their round type and cast grade, resulting in a two-dimensional matrix. During the scheduling, cells of the matrix are iteratively scanned, and orders corresponding to the cell are fetched

[3] A full discussion of these trade-offs is beyond the scope of this paper.

from the order book. Those orders are then sequenced to form intermediate data structures. When necessary, the order book is referenced and updated dynamically, so that orders included in the intermediate data structures are marked as already used.

Yet another outstanding aspect of the solution technique is that both the caster and hot strip mill operating constraints are considered simultaneously during the schedule-generation process.

We have implemented a near-optimal sequencing algorithm that has the ability to generate, within an acceptable time, solutions satisfying most of the constraints, even if no feasible solution exists. In addition, several heuristic methods are used that help increase the direct-hot-charge percentage while reducing the order tardiness and the operating costs.

The system is implemented in C++™, according to the object-oriented paradigm. We have been successful, with the prototype, in generating schedules having reasonable quality and achieving nearly 100% direct hot-charging.

We believe that the techniques developed for this prototype will be good for most steel-scheduling problems. With minor modifications, these techniques will provide near-optimal schedules for the difficult existing scheduling problems at steel mills. Added to the technical advantage provided by this approach is the advantage of a reduction in the development cycle of steel-scheduling projects, which represents a major saving of development cost.

In this section, we have discussed three approaches for the integrated scheduling of continuous casters and hot strip mills. The LTV scheduling system considers two parallel facilities, one operating in the direct-hot-charge mode, that draw on a common iron supply from a blast furnace [14]. The two approaches for the LTV scheduling problem, a search algorithm followed by a genetic algorithm [14], and the A-team solution, represent leading-edge integrated scheduling systems. A common aspect of those two approaches is that schedules are improved after reasonable initial solutions are found. The CAP was developed by exploiting advanced techniques, which include a very fast sequencing algorithm and several domain heuristics. It provides a sound technical foundation for difficult scheduling problems in modern steel mills with tightly coupled casters and hot strip mills.

8. Conclusions

Stiff global competition among steelmakers has necessitated reengineering the production process to meet the objectives of cutting costs while producing high-quality steel and delivering it to the customer on time. Steelmakers have increasingly been turning to computerized techniques to schedule their processes so as to efficiently meet these diverse business objectives.

This paper has addressed the scheduling problem in steelmaking industries and a variety of solution approaches for different scheduling areas, from individual casters or hot strip mills to integrated steel mill complexes with casters tightly linked to hot strip mills. It has focused on how recently developed, leading-edge scheduling approaches can be applied to optimize the complex process of primary steelmaking. These approaches are either from IBM's own experiences with customers or from the literature. Most approaches described in the literature do not consider the tight coupling of continuous casters and hot strip mills found in large steel plants.

IBM Research and the AMT have started a joint project to develop an integrated scheduling prototype for casters and hot strip mills. The prototype has demonstrated good results for scheduling in the direct-hot-charge mode. With this prototype and its associated techniques, we hope to help a number of steelmakers with similar scheduling problems.

References

1. Steel production statistics, *Iron & Steelmaker* magazine, p. 10 (March 1996).
2. R. E. Box and D. G. Herbe, "A Scheduling Model for LTV Steel's Cleveland Works' Twin-Strand Continuous Slab Caster," *Interface* **18,** No. 1, 42–56 (January–February 1988).
3. Donald G. Herbe, Jr. and Ronald T. Ivey, "Twin Strand Continuous Slab Caster Scheduling Model," *Proceedings of the Third International Oxygen Steelmaking Congress*, London, England, May 1990.
4. Richard Preston, *American Steel*, Prentice-Hall, Inc., New York, 1991.
5. B. Lally, L. Biegler, and H. Henein, "A Model for Sequencing a Continuous Casting Operation to Minimize Costs," *Iron & Steelmaker,* October 1987.
6. X. Tong, J. Silverman, and S. Clausen, "A Caster Sequencing Optimization Program for Geneva Steel's Facility Modernization Project," presented at the 1994 Annual AISE Convention, September 1994.
7. Numao Masayuki and Shin-ichi Morishita, "Scheplan—A Scheduling Expert for Steelmaking Processes," *Proceedings of the International Workshop on Artificial Intelligence for Industrial Applications*, Hitachi City, Japan, 1988, pp. 467–472.
8. Numao Masayuki and Shin-ichi Morishita, "Cooperative Scheduling and Its Application to Steelmaking Processes," *IEEE Trans. Indust. Electron.* **38,** No. 2, 150–155 (April 1991).
9. George Nemhauser and Laurence Wolsey, *Integer and Combinatorial Optimization*, John Wiley & Sons, Inc., New York, 1988.
10. Egon Balas and Clarence Martin, "Combinatorial Optimization in Steel Rolling," presented at the DIMACS/RUTCOR Workshop on Combinatorial Optimization in Science and Technology, Rutgers University, New Brunswick, NJ, April 1991.
11. *Roll-A-Round* (descriptive brochure), Balas & Martin Associates, Pittsburgh, PA.

12. Egon Balas, "The Prize Collecting Traveling Salesman Problem," *Networks* **19** (1989).
13. D. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley Publishing Co., Reading, MA, 1989.
14. David Meil and Ho Soo Lee, "Scheduling Steel Production on Two Twin-Strand Casters," *Proceedings of the International Manufacturing Productivity Symposium*, East Fishkill, NY, October 1993.
15. P. Winston, *Artificial Intelligence*, Addison-Wesley Publishing Co., Reading, MA, 1984.
16. S. Talukdar, P. DeSouza, and S. Murthy, "Organizations for Computer-Based Agents," *Engineering Intelligent Systems* **1**, No. 2, 75–87 (September 1993).
17. Pedro De Souza, "Asynchronous Organizations for Multi-Algorithm Problems," Ph.D. thesis, Electrical and Computer Engineering Department, Carnegie Mellon University, Pittsburgh, 1993.
18. S. Murthy, "Synergy in Cooperating Agents: Designing Manipulators from Task Specifications," Ph.D. thesis, Electrical and Computer Engineering Department, Carnegie Mellon University, Pittsburgh, 1992.
19. E. Lawler, J. Lenstra, A. Rinnooy, and D. Shmoys, *The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization*, John Wiley & Sons, Inc., New York, 1985.
20. *The Handbook of Artificial Intelligence*, Avron Barr and Edward A. Feigenbaum, Eds., Vol. 1, Ch. 2, William Kaufmann Inc., New York, 1981.
21. A. Newell and H. A. Simon, *Human Problem Solving*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1972.
22. Thomas E. Morton and David W. Pentico, *Heuristic Scheduling Systems*, John Wiley & Sons, Inc., New York, 1993.
23. Michael Pinedo, *Scheduling: Theory, Algorithms & Systems*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1994.

Received February 25, 1995; accepted for publication February 26, 1996

Ho Soo Lee *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (HSLEE at YKTVMV, hslee@watson.ibm.com).* Dr. Lee is a research staff member in the Production, Distribution and Transportation Research (PDTR) Department at the IBM Thomas J. Watson Research Center. He received B.S. and M.S. degrees in electrical engineering from Seoul National University and the Korea Advanced Institute of Science and Technology in South Korea in 1975 and 1977, respectively. He was with the Agency for Defense Development in Korea from 1977 to 1981 as a researcher, working in the field of image processing and pattern recognition. Dr. Lee received his Ph.D. in computer science from Northwestern University in 1985, joining the Thomas J. Watson Research Center in the Mathematical Sciences Department that same year as a research staff member. Until 1988, he worked in the development of rule-based programming languages and enabling technology in artificial intelligence. Since 1989 he has been working in constraint satisfaction problems, scheduling, optimization and search techniques, and their applications to industrial problems. Dr. Lee has conducted research and development activities relating to the production-scheduling projects at steel plants, including direct hot-charge-roll scheduling at LTV Steel, Cleveland Works. In 1993, he joined the PDTR Department at the Watson Research Center, where he has led steel-scheduling projects and has consulted a number of steelmakers in the U.S., Europe, Asia, and Latin America. Dr. Lee is a member of the IEEE, ACM, AAAI, and INFORMS.

Seshashayee S. Murthy *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (MURTHY at YKTVMV).* Dr. Murthy holds a Ph.D. from Carnegie Mellon University and has been a research staff member at the Watson Research Center since 1983. His main research interests are in the use of artificial intelligence to develop solutions for hard problems, and in architectures for combining multiple problem-solving methods. He has developed and fielded leading-edge scheduling solutions for the paper and steel industries.

S. Wali Haider *IBM Consulting Group, Applied Manufacturing Technologies, 3200 Windy Hill Road, Atlanta, Georgia 30339 (HAIDER at ATLVM10).* Dr. Haider is a senior consultant with the IBM Consulting Group's Applied Manufacturing Technologies Department. He is responsible for generating worldwide business opportunities for IBM in the areas of production planning and scheduling for the steel industry. Before moving to Atlanta, he was responsible for the Center of Competence for Manufacturing Modeling in the CIMS group in Boca Raton, Florida. Before joining IBM in January 1984, Dr. Haider participated in teaching and research activities in the areas of planning/scheduling/control of automated manufacturing systems as a faculty member in the Department of Industrial Engineering at Texas A&M University. He also worked for Bethlehem Steel Corporation, where he was involved in the applications of simulation and mathematical modeling techniques to production/facilities planning, operations analysis, and material handling. Dr. Haider received both his Ph.D. and M.S.I.E. from Purdue University in industrial engineering. He received a graduate research award sponsored by IBM in 1973, and is a senior member of the IIE.

David V. Morse *IBM Consulting Group, Applied Manufacturing Technologies, Four Allegheny Center, Pittsburgh, Pennsylvania 15212 (DMORSE at DETVMIC1, dmorse@vnet.ibm.com).* Dr. Morse is a manager and senior consultant in the Applied Manufacturing Technologies (AMT) Department in the IBM Consulting Group. AMT is focused on the development of advanced solutions for metals and other process industries in the areas of planning and scheduling, inventory management, and order routing. His primary interest is in the application of knowledge-based solution techniques to these and other complex problems in the fields of manufacturing and engineering design. Dr. Morse holds a B.S. in civil engineering from Pennsylvania State University and M.S. and Ph.D. degrees, both in civil engineering, from Carnegie Mellon University. He is a member of the Association of Iron and Steel Engineers, the Iron and Steel Society of the American Society of Mechanical Engineers, and the American Society of Civil Engineers.