Enterprise-wide Optimization: A New Frontier in Process Systems Engineering

Ignacio Grossmann
Dept. of Chemical Engineering, Carnegie Mellon University, Pittsburgh, PA 15213

DOI 10.1002/aic.10617
Published online June 14, 2005 in Wiley InterScience (www.interscience.wiley.com).

Enterprise-wide optimization (EWO) is a new emerging area that lies at the interface of chemical engineering and operations research, and has become a major goal in the process industries due to the increasing pressures for remaining competitive in the global marketplace. EWO involves optimizing the operations of supply, manufacturing and distribution activities of a company to reduce costs and inventories. A major focus in EWO is the optimal operation of manufacturing facilities, which often requires the use of nonlinear process models. Major operational items include planning, scheduling, real-time optimization and inventory control. One of the key features of EWO is integration of the information and the decision-making among the various functions that comprise the supply chain of the company. This can be achieved with modern IT tools, which together with the internet, have promoted e-commerce. However, as will be discussed, to fully realize the potential of transactional IT tools, the development of sophisticated deterministic and stochastic linear/nonlinear optimization models and algorithms (analytical IT tools) is needed to explore and analyze alternatives of the supply chain to yield overall optimum economic performance, as well as high levels of customer satisfaction. An additional challenge is the integrated and coordinated decision-making across the various functions in a company (purchasing, manufacturing, distribution, sales), across various geographically distributed organizations (vendors, facilities and markets), and across various levels of decision-making (strategic, tactical and operational). © 2005 American Institute of Chemical Engineers AIChe J, 51: 1846–1857, 2005

Introduction

The process industry is a key industrial sector in the U.S. In particular, the chemical industry is the major producer in the world (24% of world production) with shipments reaching $459 billion (2% of the total U.S. GDP) and $91 billion in exports in 2003 (see http://www.eere.energy.gov/industry/about/pdfs/chemicals_fy2004.pdf). However, due to the increasing pressure for reducing costs and inventories, in order to remain competitive in the global marketplace, enterprise-wide optimization (EWO) has become the “holy grail” in process industries. For instance at the conference Foundations of Computer-Aided Process Operations that took place in Coral Springs in January 2003, under the theme “A View to the Future Integration of R&D, Manufacturing and the Global Supply Chain,” it became clear that there is great interest among a variety of process industries, such as petroleum, chemical, pharmaceutical, consumer products, to achieve the goal of EWO (see http://wwwcheme.cmu.edu/focapo, Lasschuit and Thijssen, 2004; Neiro and Pinto, 2004; Shah, 2004). As shown in Figure 1, the supply chain in the petroleum industry comprises many intermediate steps starting from the exploration phase at the wellhead, going through trading and transportation, before reaching the refinery, and finally the distribution and delivery of its products, some at the retail level (e.g., gasoline). In this case it is clear that the effective coordination of the various stages is essential to accomplish the goal of EWO. Figure 2 shows the R&D phase for the testing of new drugs in the pharmaceutical industry, which can be regarded as the initial component in the supply chain of that industry, and that is the major bottleneck. The goal of achieving enterprise-wide optimization in the two examples is clearly still elusive, and motivates the research challenges outlined in the article.

Enterprise-wide optimization is an area that lies at the interface of chemical engineering (process systems engineering) and operations research. It involves optimizing the operations of supply, manufacturing (batch or continuous) and distribution in a company. The major operational activities include plan-
ning, scheduling, real-time optimization and inventory control. Supply chain management might be considered an equivalent term for describing EWO (see Shapiro, 2001). While there is a significant overlap between the two terms, an important distinction is that supply chain management is aimed at a broader set of real-world applications with an emphasis on logistics and distribution, which usually involve linear models, traditionally the domain of operations research. In contrast in enterprise-wide optimization, the emphasis is on the manufacturing facilities with a major focus being their planning, scheduling and control which often requires the use of nonlinear process models, and, hence, knowledge of chemical engineering. We should also note that many process companies are adopting the term enterprise-wide optimization to reflect both the importance of manufacturing within their supply chain, as well as the drive to reduce costs through optimization.

One of the key features in EWO is integration of the information and decision making among the various functions that comprise the supply chain of the company. Integration of information is being achieved with modern IT tools, such as SAP and Oracle that allow the sharing and instantaneous flow of information along the various organizations in a company. The development of the internet and fast speed communication has also helped to promote through e-commerce the implementation and deployment of these tools. While these systems still require further developments to fully realize the vision of creating an agile platform for EWO (i.e., transactional information), it is clear that we are not too far from it.

While software vendors provide IT tools that in principle allow many groups in an enterprise to access the same information, these tools do not generally provide comprehensive decision making capabilities that account for complex trade-offs and interactions across the various functions, subsystems and levels of decision making. This means that companies are faced with the problem of deciding as to whether to develop their own in-house tools for integration, or else make use of commercial software from vendors.

Some commercial tools are becoming increasingly capable of addressing some parts of the enterprise-wide optimization in the process industry (e.g., Aspen Technology, Mahalec, 2001). As a specific example of a strategic planning study, BASF performed a corporate network optimization of packaged finished goods in North America. There were 17 operating divisions with multiple, heterogeneous systems: 25,000 SKU’s (stock keeping units), 134 Shipping Points, 15,000 Ship-to locations, 956 million pounds shipped direct to customers, 696 million pounds shipped to customers through distribution centers. By using optimization tools from Aspen Technology, BASF reduced transportation and facility costs by 10%, next-day volume delivery increased from 77 to 96%, the number of distribution centers was reduced from 86 to 15, generating $10 million/year savings in operating costs.

From the earlier example it can be seen that there is great economic potential in EWO, and that some progress has been made toward the goal of developing some of the basic building blocks. However, major barriers are the lack of computational optimization models and tools that will allow the full and comprehensive application of EWO throughout the process industry. This will require a new generation of tools that allow the full integration and large-scale solution of the optimization models, as well as the incorporation of accurate models for the manufacturing facilities. Given the strong tradition that chemical engineers have in process systems engineering and in the optimization area (see Biegler and Grossmann (2004) for a recent review), they are ideally positioned to make significant contributions in EWO.

**Challenges in Enterprise-wide Optimization**

In order to realize the full potential of transactional IT tools, the development of sophisticated optimization and decision-support tools (analytical IT tools) is needed to help explore and analyze alternatives, and predict actions for the operation of the supply chain so as to yield overall optimum economic performance, as well as high levels of customer satisfaction. A major challenge that is involved in EWO of process industries is the integrated and coordinated decision-making across the various functions in a company (purchasing, manufacturing, distribution, sales), across various geographically distributed organizations (vendors, facilities and markets), and across various levels of decision-making (strategic, tactical and operational), as seen in Figure 3 (Shapiro, 2001). The first two items conceptually deal with issues related to *spatial integration* in that they involve coordinating the activities of the various subsystems of an enterprise. The third item deals with issues related to *temporal integration* in that they involve coordinating decisions across different timescales. Addressing these spatial and temporal integration problems is important because they provide a basis to optimize the decision-making in an enterprise through the IT infrastructure.

In order to achieve EWO throughout the process industry, this goal will require a new generation of computational tools for which the following major challenges must be addressed:
Figure 3. Transactional and analytical IT (Tayur et al., 1999).

(a) The modeling challenge: What type of production planning and scheduling models should be developed for the various components of the supply chain, including nonlinear manufacturing processes, that through integration can ultimately achieve enterprise-wide optimization? Major issues here are the development of novel mathematical programming and logic-based models that can be effectively integrated to capture the complexity of the various operations.

(b) The multiscale optimization challenge: How to coordinate the optimization of these models over a given time horizon (from weeks to years), and how to coordinate the long-term strategic decisions (years) related to sourcing and investment, with the medium-term decisions (months) related to tactical decisions of production planning and material flow, and with the short-term operational decisions (weeks, days) related to scheduling and control? Major issues here involve novel decomposition procedures that can effectively work across large spatial and temporal scales.

(c) The uncertainty challenge: How to account for stochastic variations in order to effectively handle the effect of uncertainties (e.g., demands, equipment breakdown)? Major issues here are the development of novel, meaningful and effective stochastic programming tools.

(d) The algorithmic and computational challenge: Given the three points earlier, how to effectively solve the various models in terms of efficient algorithms, and in terms of modern computer architectures? Major issues here include novel computational algorithms and their implementation through distributed or grid computing.

Although progress has been made in some of the areas cited above, significant research effort is still required to overcome the four challenges above. The following sections briefly discuss the technical issues involved in each of the challenges. The long term goal is to produce new analytical IT tools that will help to realize the full potential of EWO in conjunction with the transactional IT tools.

The modeling challenge

While the area of planning and scheduling has seen the development of many models in operations research (OR) (e.g., Pinedo, 2001), over the last decade a significant number of planning and scheduling models have been proposed specifically for process applications (for a recent review see Pinto and Grossmann, 1998; Shah, 1998; Pekny and Reklaitis, 1998). In contrast to general OR scheduling models, the process-oriented models tend to require the use of material flows, and very often network topologies that are quite different from the more traditional serial and multistage systems. Furthermore, they address both batch and continuous processes, and may require the use of detailed nonlinear process models.

The most general batch scheduling model that has been proposed for short-term scheduling for processing applications is the State-Task Network by Kondili et al. (1993). This model has the feature that it does not preassign equipment to tasks, the batches are of variable size and can be combined and split. The original model relied on a discrete time representation which led to a mixed-integer linear programming (MILP) formulation. Pantelides (1994) proposed the resource-task network as an alternative representation that leads to a more compact MILP model. Recent efforts on this problem have extended the model to continuous time which greatly complicates the underlying MILP model (e.g., see Schilling and Pantelides, 1996; Zhang and Sargent, 1996; Ierapetritou and Floudas, 1998; Mockus and Reklaitis, 1999; Maravelias and Grossmann, 2003). On the other hand there are a good number of specific process scheduling models that have been developed to better exploit the special structure of some problems (e.g., continuous multistage with parallel units; see Jia et al., 2003), and to incorporate process performance models and explicit handling of changeovers (e.g., see Jain and Grossmann, 1998). Another common occurrence is in long term cyclic scheduling models in which at the very least the objective function must be expressed in nonlinear form (e.g., see Pinto and Grossmann, 1994). It should be noted, however, that despite the progress that has been made, the availability of a general purpose scheduling and planning model for the process industries, particularly for continuous processes, is still elusive. This is not only because of the great variety of problems that arise in practice, but also because of a number of major computational issues, namely difficulties in solving large-scale discrete and continuous optimization problems, handling of nonlinear process models, and treatment of uncertainties.

The general form of the deterministic problems in EWO problems corresponds to the following multiperiod mixed-integer programming problem

$$
\min \sum_{i \in T} f_i(x, w_i, w_{ij}, \theta_i) \\
\text{s.t. } g_i(x, w_i, w_{ij}, \theta_i) \leq 0 \quad t \in T \\
x \in \mathbb{R}^n, x_i = 0, 1 \quad i = 1, m_s \\
w_i \in \mathbb{R}^{m_{ai}}, w_{ij} = 0, 1 \quad j = 1, m_{ai}, t \in T
$$

where $f$, $g$, are scalar and vector functions (linear/nonlinear), respectively, and $T$ is a set of fixed or variable time periods. The variables $x$ represent decisions independent of the time periods, while the variable $w_i$ represent decisions at each time period $t$, where $m_s \leq n_s$, $m_a \leq n_a$, $\theta_i$ are exogenous or endogenous parameters that have fixed values for deterministic problems. When considering uncertainties, these parameters are treated as random variables and problem (P) is extended as a stochastic programming problem. Since EWO will require
the formulation and solution of problems of type (P), both linear and nonlinear that are one or several orders of magnitude larger than current planning and scheduling models, research is needed in order to produce effective computational models.

The multiscale optimization challenge

Integration and coordination are key components in EWO (Shapiro, 2004; Song and Yao, 2001). The areas outlined in the following sections correspond to major unresolved problem areas.

Integration of Production Planning, Scheduling and Real-time Optimization. The fundamental issue in this area is the integration of models across very different timescales (Shah, 1998). Typically, the planning model is a linear and simplified representation that is used to predict production targets and material flow over several months (up to one year). Also at this level effects of changeovers and daily inventories are neglected, which tends to produce optimistic estimates that cannot be realized at the scheduling level. Scheduling models on the other hand tend to be more detailed in nature, but assume that key decisions have been taken (e.g., production targets, due dates). Two major approaches that have been investigated for integrating planning and scheduling are the following:

1. Simultaneous planning and scheduling over a common time grid. The idea here is to effectively “elevate” the scheduling model to the planning level, which leads to a very large-scale multiperiod optimization problem, since it is defined over long time horizons with a fine time discretization (e.g., intervals of one day). A good example is the use of the State-Task-Network for multisite planning (e.g., Wilkinson et al., 1996). To overcome the problem of having to solve a very large scale problem, strategies based on aggregation and decomposition can be considered (see Bas et al., 1996; Birewar and Grossmann, 1990; Wilkinson, 1996). The former typically involve aggregating later time periods within the specified time horizon in order to reduce the dimensionality of the problem.

2. Decomposition techniques for integrating planning and scheduling are usually based on a two-level decomposition procedure where the upper level problem (planning problem) is an aggregation of the lower level problem (scheduling). The challenge lies in developing an aggregated planning model that yields tight bounds to reduce the number of upper and lower level problems (Papageorgiou and Pantelides, 1996a, 1996b; Bok et al., 2000). Another solution approach relies on using a rolling horizon approach where the planning problem is solved by treating the first few periods in detail, while the later periods are aggregated recursively (Dimitriadis et al., 1997).

Finally, real-time optimization (RTO) models are nonlinear and are defined over short time intervals; integration of RTO with planning and scheduling is a topic that has received virtually no attention in the literature.

Optimization of Supply Chains. When considering a specific decision level (strategic, tactical or operational), it is often desired to consider the entire supply chain of a given enterprise (e.g., Equi et al., 1997; Erenç, 1999; Neiro and Pinto, 2003). Here again problem size can become a major issue as we have to handle models across many length scales. Two major approaches are to either consider a simultaneous large-scale optimization model, or else to use decomposition either in spatial or in temporal forms (Kulkarni and Mohanty, 1996), usually using Lagrangean decomposition (Graves, 1982; Gupta and Maranas, 1999). In the case of spatial decomposition the idea is to severe the links between subsystems (e.g., manufacturing, distribution and retail) by dualizing the corresponding interconnection constraints, which then requires the multiperiod optimization of each system. In the case of temporal decomposition the idea is to dualize the inventory constraints in order to decouple the problem by time periods. The advantage of this decomposition scheme is that consistency is maintained over every time period (Jackson and Grossmann, 2003). See also Daskin et al. (2002) for combining location and inventory models.

Simultaneous optimization approaches for the integration of entire supply chains naturally lead to the definition of centralized systems. In practice, however, the operation tends to take place as if the supply chain were a decentralized system. What is needed are coordination procedures that can maintain a certain degree of independence of subsystems (Nishi et al., 2002), while at the same time aiming at objectives that are aimed at the integrated optimization of the overall system (see Pereia et al., 2001).

Uncertainty challenge

Uncertainty is a critical issue in supply chain operations. Furthermore, it is complicated by the fact that the nature of the uncertainties can be quite different in the various levels of the decision making (e.g., strategic planning vs. short term scheduling). Most of the research thus far has focused on operational uncertainty, such as quality, inventory management and handling uncertain processing time (e.g., Zipkin, 2000, Montgomery, 2000, Balasubramanian and Grossmann, 2002). Much less work has focused on uncertainty at the tactical level, for instance, production planning with uncertain demand (Gupta and Maranas, 2003; Balasubramanian and Grossmann, 2004). The reason for this is that the resulting optimization problems are extremely difficult to solve since they give rise to stochastic programming problems (Birge and Louveaux, 1997). In a stochastic program, mathematical programs are solved over a number of stages. Between each stage, some uncertainty is resolved, and the decision maker must choose an action that optimizes the current objective plus the expectation of the future objectives. The most common stochastic programs are two-stage models that are solved using a variant of Benders’ decomposition. When the second-stage (or recourse) problem is a linear program these problems are straightforward to solve, but the more general case is where the recourse is a MILP or a MINLP. Such problems are extremely difficult to solve since the expected recourse function is discontinuous and nonconvex (Sahinidis, 2004).

As an example, consider the problem of production planning across a supply chain with uncertain demands. The first-stage problem is to create a production plan. After the demand is realized, a plant process optimization problem must be solved for each plant. The challenge is to find a production plan that minimizes the expected production cost. The hierarchical nature of supply chains lends itself naturally to stochastic programming models, and in particular the decomposition principles that are used to solve them.
Algorithmic and Computational Challenges

Realizing the vision of EWO will require the development of advanced algorithms and computational architectures in order to effectively and reliably solve the large-scale optimization models. In this section, we briefly outline some of the more technical aspects that are involved in this endeavor. We should note that collaboration between researchers in process systems engineering and operations research should be most fruitful in this area.

Mixed-integer linear programming. When detailed process performance models are not used, planning and scheduling problems for EWO commonly give rise to mixed-integer linear programming problems (MILPs). These optimization problems can be computationally expensive to solve since in the worst case they exhibit exponential complexity with problem size (NP-hard). However, in the last 10 years great progress has been made in algorithms and hardware, which has resulted in an impressive improvement of our ability to solve mixed-integer programming problems (MILPs) (Bixby, 2002; Johnson et al., 2000) through codes such as CPLEX and XPRESS. Capitalizing on theory developed during the last 20 years, it is now possible, using off-the-shelf LP-based branch-and-bound commercial software, to solve in a few seconds MILP instances that were unsolvable just five years ago. This improvement has been particularly dramatic for certain classes of problems, such as the traveling salesman problem, and certain industries, such as the commercial airlines. In contrast, for the type of problems that arise in process industries, the available LP-based branch-and-bound software is not always capable of solving industrial-size MILP models. One reason is that, nonconvex functions, such as piecewise linear functions, and combinatorial constraints, such as multiple-choice, semicontinuous, fixed-charge, and job sequencing disjunctions (i.e., either job i precedes job j or vice-versa), abound in optimization problems related to process industries. For such functions and constraints, the “textbook” approach implemented in the current software is often not practical.

In the current methods, nonlinearities are often modeled by introducing a large number of auxiliary binary variables and additional constraints, which typically doubles the number of variables and increases the number of constraints by the same order of magnitude. Also, with this approach, the combinatorial structure is obscured and it is not possible to take advantage of the structure. In the case of EWO, where many of these constraints appear at the same time and the sizes of the instances are considerably larger, these issues are even more serious. Recently, an alternative method, branch-and-cut without auxiliary binary variables, inspired by the seminal work of Beale and Tomlin (1970) on special ordered sets, has proved to be promising in dealing with such constraints (de Farias, 2004). It consists of enforcing the combinatorial constraints algorithmically, directly in the branch-and-bound scheme, through specialized branching and the use of cutting planes that are valid for the set of feasible solutions in the space of the original decision variables. The encouraging computational results yielded by the method on some of the aforementioned constraints provide a serious indication that it may be of great impact on EWO problems for the process industries. The use of cutting planes in an LP-based branch-and-bound approach has also proven to be of significant importance in obtaining strong bounds to reduce the required amount of enumeration (see for example, Marchand et al., 2002).

Constraint Programming. The relatively new field of constraint programming has recently become the state of the art for some important kinds of scheduling problems, particularly resource-constrained scheduling problems, which occur frequently in supply-chain contexts. Constraint programming (CP) can bring advantages on both the modeling and solution sides. The models tend to be more concise and easier to debug, since logical and combinatorial conditions are much more naturally expressed in a CP than in an MILP framework (e.g., Milano 2003). The solvers take advantage of logical inference (constraint propagation) methods that are well suited to the combinatorial constraints that characterize scheduling problems. In particular, the sequencing aspect of many scheduling problems—the task of determining in what order to schedule activities—can present difficulties to MILP because it is difficult to model and gives rise to weak continuous relaxations. By contrast, a CP model readily formulates sequencing problems and offers specialized propagation algorithms that exploit their structure. Furthermore, heuristics can readily be accommodated in CP.

The greatest promise, however, lies in the integration of CP and MILP methods, which is currently a very active area of research (Hooker, 2000). Several recent systems take some steps toward integration, such as ECLiPSe (Wallace et al. 1997), OPL Studio (Van Hentenryck 1999), and the Mosel language (Columbani and Heipcke, 2002). Integration allows one to attack problems in which some of the constraints are better suited to an MILP-like approach (perhaps because they have good continuous relaxations) and others are better suited for a CP approach (because they “propagate well”). This is particularly true of supply-chain problems, in which constraints relating to resource allocation, lot sizing, routing and inventory management may relax well, while constraints related to sequencing, scheduling and other logical or combinatorial conditions may propagate well. In the context of scheduling problems, these models perform the assignment of jobs to machines with mixed-integer programming constraints, while the sequencing of jobs is performed with constraint programming. The motivation behind the former is to remove “big-M” constraints and exploit the optimization capability of mixed-integer programming. The motivation behind using the latter is to exploit the capability of constraint programming for effectively handling feasibility subproblems, as well as sequencing constraints. Hybrid methods have shown in some problems outstanding synergies that lead to order magnitude reductions in computation (Jain and Grossmann, 2001; Maravelias and Grossmann, 2004; Hooker 2003; Hooker et al. 1999; Hooker and Ottosson 2003).

Nonlinear programming. In order to develop real-time optimization models as part of the enterprise-wide optimization models for process industries (energy, chemicals, and materials) high fidelity simulation models are required that provide accurate descriptions of the manufacturing process. Most of these models consist of large sets of nonlinear equality and inequality constraints, which relate manufacturing performance to designed equipment capacities, plant operating conditions, product quality constraints, and operating costs. The sensitivity of these degrees of freedom to higher level decisions can also be exploited by an integrated optimization formulation. The
development and application of optimization tools for many of these nonlinear programming (NLP) models (Nocedal and Wright, 1999) has only recently been considered (see Biegler et al., 2002).

An important goal in EWO is the integration of these nonlinear performance models to determine optimal results from IT tools. This research task is essential because these performance models for real-time optimization ensure the feasibility of higher level decisions (e.g., logistics and planning) for manufacturing operations. Also, these models accurately represent operating degrees of freedom and capacity expansions in the manufacturing process. As a result, incorporation of these models leads to significantly superior results than typical linear approximations to these models. Several studies have demonstrated the importance of including NLP and MINLP optimization capabilities (Bhatia and Biegler, 1997, Jackson and Grossmann, 2003; Jain and Grossmann, 1998), and the significant gains that can be made in planning and scheduling operations. On the other hand, the research challenge is that nonlinear models are more difficult to incorporate and to handle as nonlinear optimization problems because they introduce non-monotonic behavior, nonconvexities and local solutions. In addition, the treatment of local degeneracies and ill-conditioning is more difficult and more computationally intensive optimization algorithms are required. The recent introduction of interior point (or barrier) methods for NLP (Byrd et al., 2000; Vanderbei and Shanno, 1999; Waechter and Biegler, 2003) have shown significant improvements over conventional algorithms with active set strategies. Also, more recent convergence criteria have been improved with the introduction of filter methods (Fletcher et al., 2003; Waechter and Biegler, 2005), which rapidly eliminate undesirable search regions and promote convergence from arbitrary starting points.

Mixed-integer Nonlinear Programming and Disjunctive Optimization. Developing the full range of models for EWO as given by problem (P) requires that nonlinear process models be developed for planning and scheduling of manufacturing facilities. This gives rise to mixed-integer nonlinear programming (MINLP) problems since they involve discrete variables to model assignment and sequencing decisions, and continuous variables to model flows and, amounts to be produced and operating conditions (e.g., temperatures, yields). While MINLP optimization is still largely a rather specialized capability, it has been receiving increasing attention over the last decade. A recent review can be found in Grossmann (2002). A number of methods, such as outer-approximation, extended cutting planes, and branch and bound have proved to be effective, but are still largely limited to moderate-sized problems. In addition, there are several difficulties that must be faced in solving these problems. For instance in NLP subproblems with fixed values of the binary variables, the problems contain a significant number of redundant equations and variables that are often set to zero, which in turn often lead to singularities and poor numerical performance. There is also the possibility of getting trapped in suboptimal solutions when nonconvex functions are involved. Finally, there is the added complication when the number of 0-1 variables is large, which is quite common in planning and scheduling problems.

To circumvent some of these difficulties, the modeling and global optimization of generalized disjunctive programs (GDP) seems to hold good promise for EWO problems. The GDP problem is expressed in terms of Boolean and continuous variables that are involved in constraints in the form of equations, disjunctions and logic propositions (Raman and Grossmann, 1994). One motivation for investigating these problems is that they correspond to a special case of hybrid models in which all the equations and symbolic relations are given in explicit form. An important challenge is related to the development of cutting planes that provide similar quality in the relaxations as the convex hull formulation without the need of explicitly including the corresponding equations (Sawaya and Grossmann, 2005). The other challenge is that global optimization algorithms (Floudas, 2000; Sahinidis, 1996) can in principle be decomposed into discrete and continuous parts, which is advantageous as the latter often represents the major bottleneck in the computations (e.g., through spatial branch-and-bound schemes; see Lee and Grossmann, 2001). Finally, the extension to dynamics of these models (e.g., Barton and Lee, 2004) should provide computational capabilities that are required to model real-time problems.

Computational Grid. Solving the large-scale EWO models will require significant computational effort. To achieve the goal of integrating planning across the enterprise, advances in algorithms and modeling must go hand-in-hand with advances in toolkits that enable algorithms to harness computational resources. One promising approach that has emerged over the last decade is to deliver computational resources in the form of a computational grid, which is a collection of loosely-coupled, (potentially) geographically distributed, heterogeneous computing resources. The idle CPU time on these collections is an inexpensive platform that can provide significant computing power over long time periods. For example, consider the project SETI@home (http://setiathome.ssl.berkeley.edu/), which since its inception in the mid 1990s has delivered over 18,000 centuries of CPU time to a signal processing effort. A
computational grid is similar to a power grid in that the provided resource is ubiquitous and grid users need not know the source of the provided resource. An introduction to computational grids is given by Foster and Kesselman (1999). An advantage of computational grids over traditional parallel processing architectures is that a grid is the most natural and cost-effective manner for users of models and algorithms to obtain the required computational resource to solve EWO problems.

To allow a larger community of engineers and scientists to use computational grids, a number of different programming efforts have sought to provide the base services that grid-enabled applications require (e.g., Foster and Kesselman, 1997 and Livny et al., 1997). A promising approach would seem to use and augment the master-worker grid library MW (Goux et al. 2001). The MW library is an abstraction of the master-worker paradigm for parallel computation. MW defines a simple application programming interface, through which the user can define the core tasks making up this computation, and the actions that the master takes upon completion of a task. Once the tasks and actions are defined by the user, MW performs the necessary actions to enable the application to run on a computational grid (such as resource discovery and acquisition, task scheduling, fault-recovery, and interprocess communication).

MW was developed by the NSF-funded metaNEOS project and used to solve numerical optimization problems of unprecedented complexity (e.g., Anstreicher et al. 2002, Linderoth and Wright, 2003). A major research direction here would be the development and testing of decomposition-based and branch-and-bound based algorithms for EWO models. The MW toolkit has already been used with great success to parallelize both decomposition-based algorithms (e.g., Linderoth and Wright, 2003), and also spatial branch-and-bound algorithms (e.g., Goux and Leyffer, 2003, Chen, Ferris, and Linderoth, 2001). However, for EWO the current functionality in the MW toolkit is not sufficient. The simple master-worker paradigm must be augmented with features that improve its scalability and information sharing capabilities to be able to solve the EWO models.

Illustrative Examples

In this section, we present four examples that illustrate the four challenges cited in this article on problems encountered in the area of Enterprise-wide Optimization. Example 1 deals with a multisite planning and distribution problem that incorporates nonlinear process models, illustrating the modeling challenge. Example 2 describes the simultaneous optimization of the scheduling of testing for new product development and the design of batch manufacturing facilities. This example illustrates the challenge of multi-scale modeling given the dissimilar nature of the activities and the need of combining a detailed scheduling model with a high level design model. Example 3 illustrates the third challenge with the design and planning of off-shore gas field facilities under uncertainty. Finally, Example 4 deals with a short term scheduling problem that makes use of a hybrid model that combines mixed-integer linear programming and constraint programming. This example illustrates the challenge for developing new algorithms. We
should note that while the examples presented are rather modest in size compared to what ideally one would like to strive for in EWO, examples 1 and 3 correspond to real world industrial problems.

Example 1.

This example deals with the production planning of a multisite production facility that must serve global markets (see Figure 4). The sites can produce 25 grades of different polymers. Given forecasts of demands over a 6 to 12 month horizon the problem consists of determining for each week of operation what grades to produce in each site and the transportation to satisfy demands in the various markets. An important feature of this problem is that nonlinear process models are required to predict the process and product performance at each site.

Neglecting effects of changeovers, the problem of optimizing the total profit can be formulated as a multiperiod NLP problem. The difficulty is that the size of the problem can become very large. For instance a 12 month problem involves 34,381 variables and 28,317 constraints. To circumvent this problem, Jackson and Grossmann (2003) developed a temporal decomposition scheme, based on Lagrangean relaxation. The authors showed that much better results could be obtained compared to a spatial decomposition (see Figure 5), and that the CPU times could be reduced by one or two orders of magnitude for optimality tolerances of 2-3%. The reason CPU times are important for this model is that this allows one to use it in real time for demand management when deciding what orders to accept and their deadlines.

Example 2.

This problem deals with the case where a biotechnology firm produces recombinant proteins in a multipurpose protein production plant. Products A, B, D, and E are currently sold while products C and F are still in the company’s R&D pipeline. Both potential products must pass successfully 10 tests before they can gain FDA approval (see Figure 6). These tests can either be performed in-house or else outsourced at double the cost. When performed in-house, they can be conducted in only one specific laboratory. Products A-C are extracellular, while D-F are intracellular. All proteins are produced in the fermentor P1 (see Figure 7). Intracellular proteins are then sent to the homogenizer P2 for cell suspension, then to extractor P3, and last to the chromatographic column P4 where selective binding is used to further separate the product of interest from other proteins. Extracellular proteins after the fermentor P1 are sent directly to the extractor P3 and then to the chromatograph P4.

The problem consists of determining simultaneously the optimal schedule of tests and their allocation to labs, while at the same time deciding on the batch plant design to accommodate the new proteins. Here, two major options are considered. One is to build a new plant for products C and D (assuming both pass the tests), the other is to expand the capacity of the existing plant. This problem was formulated as an MILP problem (Maravelias and Grossmann, 2001), involving 612 0 – 1 variables, 32184 continuous variables, and 30903 constraints. Here again one option is to solve simultaneously the full-size MILP, while the other is to decompose the problem into the scheduling and design functions using a Lagrangean relaxation technique similar to the one in Example 1. The schedule predicted for the tests is shown in Figure 8a. In Figure 8b, it can be seen that the model selects to expand the capacity of the various units rather than building a new plant. Also, since the model accounts for the various scenarios of fall/pass for C and F, it predicts that products D and E be phase-out in the case that the two new proteins obtain FDA approval.

Example 3

This problem deals with the design and planning of an off-shore facility for gas production. It is assumed that a superstructure consisting of a production platform, well platforms (one for each field) and pipelines is given (see Figure 9). It is also assumed that for some of the fields there is significant uncertainty in the size and initial deliverability (production rate) of the fields. The problem consists of determining over a given time horizon (typically 10 – 15 years) decisions regard-
ing the selection and timing of the investment for the installation of the platforms, their capacities and production profiles.

Goel and Grossmann (2003) developed a mixed-integer optimization model assuming discrete probability distribution functions for the sizes and initial deliverabilities. Also, the model was simplified with a linear performance model to avoid the direct use of a reservoir simulation model. The optimization problem gives rise to a very difficult stochastic optimization problem, which has the unique feature that the scenario trees are a function of the timing of the investment decisions. Goel and Grossmann (2003, 2005) have developed two solution methods, a heuristic and a rigorous branch and bound search method for solving this problem. The example in Figure 10 involves 6 fields over 15 years, with two of the fields having uncertain sizes and deliverabilities. If one were to solve directly the deterministic equivalent problem in which all scenarios are anticipated this would give rise to a very large multiperiod MILP model with about 16281 0 –1 variables, and 2.4 million constraints, which is impossible to solve directly with current solution methods. Fortunately, the methods by Goel and Grossmann circumvent the solution of such a large problem. The solution is shown in Table 1, which postpones the investment in the uncertain fields to years five and seven, with an expected NPV of $146 million, and a risk of less than 1% that the NPV be negative. Interestingly, if one simply uses mean values for the uncertain parameters, the platforms at the uncertain fields are installed in period 1 and the financial risk increases to 8%. Obviously, in practice models like these would be solved periodically by updating them with new information on the fields.

Table 1. Solution of Stochastic Model for Figure 10

<table>
<thead>
<tr>
<th>Year</th>
<th>Proposed Solution</th>
<th>ENPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PP, A, B, C, F</td>
<td>$146.32 Million</td>
</tr>
<tr>
<td>5</td>
<td>E</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>D</td>
<td></td>
</tr>
</tbody>
</table>

Example 4

This example deals with the scheduling of a batch process shown in Figure 11, using the state-task network representation in which circles represent material nodes with various storage options (finite, unlimited, zero-wait, no storage), and the rectangles represent operational tasks that must be performed (e.g., mixing, reaction, separation). This batch process produces four different products, P1, P2, P3 and P4. Note that 8 units are assumed to be available for performing the operations of the various tasks. Of course not all units can perform all tasks, but only a subset of them. Given data on processing times for each task, as well as on the mass balance, the problem consists of determining a schedule that can produce five tons of the four products, and that minimizes the makespan (completion time). If this problem is formulated with a continuous time approach, such as the one by Maravelias and Grossmann (1993) in order to accommodate arbitrary processing times, the corresponding MILP cannot be solved after 10 h of CPU-time. This can be qualitatively explained by the fact that scheduling and MILP problems are NP-hard. To address this difficulty, however, Maravelias and Grossmann (2004) developed a novel hybrid solution method that combines MILP with constraint programming. Using such a technique the problem was solved to rigorous optimality in only 5 s. This example then shows the importance of special solution methods that effectively exploit the structure of scheduling problems.

Concluding Remarks

This article has provided an overview of the emerging area of enterprise-wide optimization, that is driven by needs of the process industries for reducing costs and remaining competitive in the global marketplace. Some of the major challenges have been highlighted, and several examples presented to illustrate the nature of the applications and the problems that are faced.

It is hoped that this article has shown that EWO offers new and exciting opportunities for research to chemical engineers. While EWO lies at the interface of chemical engineering (process systems engineering), and operations research, it is clear that chemical engineers can play a major role not only in the modeling part, but also in the algorithmic part given the strong and rich tradition that chemical engineers have built in mathematical programming. Thus, in collaboration with operations researchers, chemical engineers should be well positioned for developing novel computational models and algorithms that are to be integrated with coordination and decomposition techniques through advanced computing tools. This effort should help to expand the scope and nature of EWO models that can be effectively solved in real-world industrial problems. These models and methods have the potential of providing a new generation of analytical IT tools that can significantly increase profits and reduce costs, thereby strengthening the economic performance and competitiveness of the process industries.
Acknowledgment

The author would like to acknowledge contributions to this article by Larry Biegler (Chemical Engineering, CMU), Ismael de Farias (SUNY-Buffalo, Industrial Engineering), John Hooker (Tepper Business School, CMU), Jeff Linderoth (Industrial and Systems Engineering, Lehigh), and Andrew Schaefer (Industrial Engineering, U. Pittsburgh).

Literature Cited


Jain, V. and Grossmann, I.E., “Cyclic Scheduling and Mainte-


Song, J. S., and D. D. Yao, eds., *Supply Chain Structures: Coordination, Information, and Optimization*, Kluwer, Inter-


