A Multilevel, Control-Theoretic Framework for Integration of Planning, Scheduling, and Rescheduling

S. A. Munawar and R. D. Gudi*

Department of Chemical Engineering, Indian Institute of Technology, Bombay, Powai, Mumbai 400 076, India

In this paper, an integrated multilevel, control-theoretic framework has been proposed for effectively handling integration of planning, scheduling, and rescheduling. A general resource-constrained, multistage, multiproduct plant operating as a hybrid flowshop facility has been considered. The proposed approach is based on the inherent hierarchical decomposition of the overall decision-making process that is a typical characteristic of large enterprises. The overall problem is segregated into three levels with different horizons, wherein planning over multiperiods is at the top level followed by scheduling for a single period at the middle level and a detailed inventory management schedule for the operator at the lower level. In the hierarchical decomposition, the upper levels are equipped with abstractions of the lower levels and proactiveness for reactive scheduling. The integration of reactive scheduling is motivated by some of the process control principles like cascade control and the concepts of receding horizon. Using the philosophy of decentralized decision-making, it is demonstrated that the lower levels with accurate models have the flexibility and amenability for rescheduling without upsetting the global performance. As an illustrative case study, cyclic scheduling of a simple refinery flow sheet involving continuous lube production in a resource constrained hybrid flowshop is presented to demonstrate the proposed methodology.

1. Introduction

Planning in an enterprise is usually concerned with high level decisions such as what and where to produce, and deals with longer time horizons through an economic objective function of maximization of profit. Scheduling on the other hand is concerned with lower level decisions such as sequencing, and deals with shorter time horizons through a feasibility objective function of meeting the production targets set by the planning level. In the recent past, development of methods for efficient integration of planning and scheduling has received momentous attention in the industrial sector and in the research community, largely because of the challenges and the high economic incentives involved. Currently, most large-scale enterprises attempt the integration of planning and scheduling activities, with decision-making in a centralized fashion for the production and distribution tasks. Production and scheduling targets are then specified over a multiperiod operation. Often, they employ some improvised techniques for this integration and are generally discontented with the resulting inconsistencies in decision-making.[1]

Over the last few years, though some progress has been made in this direction for development of superior frameworks for such integrations, there is a large scope for additional improvements. The challenges are in terms of the complex issues relating to handling large-scale advanced planning and scheduling problems leading to combinatorial explosion of the problem sizes, for centralized decision-making. Moreover, the horizons of interest are broadly different at the planning and scheduling levels of a general plant. The planning models must be consistent with lower level scheduling models, and the scheduling models must again be consistent with the plant level operation thus achieving vertical integration. Additionally, the upper planning level should be revised as infrequently as possible when compared to the lower scheduling levels to avoid frequent revision of the commitments made to customers. Traditionally, the decisions in an enterprise flow in a top-down manner leaving less degree of freedom at lower levels for rescheduling, thus leading to frequent revision of targets set by the upper levels. Embedding contingency measures for integration of rescheduling has been ignored in most of the works published.

In the literature there are several works on planning and scheduling. Shah[2] gives a detailed review and current status on single and multisite planning and scheduling. Development of consistent planning and scheduling models has been identified as one of challenges in the integrated hierarchical decomposition of the overall problem. Grossmann et al.[3] reviewed the classification of planning and scheduling models arising in process operations and the recent developments in their solution techniques. They proposed a general disjunctive model for integration of planning and scheduling. Shobrys and White[1] examined the key business issues, the current practices, and the incentives and barriers in the integration of planning, scheduling, and control functions in the process industries. They recounted some of the success stories in this direction and analyzed the reasons for failure of other companies that could not achieve integration despite multiple initiatives. They identified two nontechnological challenges of coping with human and organizational behavior and finally made some recommendations to overcome these barriers to integration. Jia et al.[4] presented a spatial decomposition of complete refinery scheduling and proposed a state task network based continuous time

* To whom correspondence should be addressed. Tel.: +91-22-2576-7204. Fax: +91-22-2572-6895. E-mail: ravigudi@che.iitb.ac.in.

10.1021/ie040142s CCC: $30.25
Published on Web 04/23/2005

formulation for short-term scheduling of crude oil operations. Recently, Van den Heever and Grossmann[5] presented a two-level decomposition model for integration of planning and reactive scheduling in hydrogen supply networks. They used a simplified pipeline model at the planning level and a detailed pipeline description for the scheduling level.

Oftentimes, schedules derived on the basis of known parameters (such as demands, due dates) and theoretical considerations (mass balance, yields) need to be continuously evaluated to ensure that they do not violate critical operating constraints/goals. Other decision parameters such as process yields, processing times/rates, material arrival/lifting times, and even production targets themselves could change, after the nominal schedule is deployed. Furthermore, in the dynamic environment of processing operations, unexpected events continually occur and cause deviations from the expected production targets. Broadly, the two approaches that have been proposed in the literature, to react to such uncertainties and variations, focus on either (i) complete reevaluation of the nominal schedule under the changed conditions, or (ii) partial revision or "repair" of the nominal schedule to accommodate the deviations.

In process scheduling literature, most of the approaches to reactive scheduling have been focused on batch plants. Aspects related to robustness of schedules, in the presence of uncertainties, were studied by Mignon et al.[6] Rodrigues et al.[7] proposed a reactive scheduling technique for a state task network based formulation of the scheduling problem that also incorporated a rolling horizon based representation to simplify the problem. Sanmarti et al.[8] proposed a combined robust/reactive scheduling approach to batch processes in the presence of task processing time variations. Recently, Henning and Cerda[9] proposed a framework for knowledge based predictive reactive scheduling that incorporates knowledge of the human experts with automated scheduler capabilities. Sun and Xue[10] propose a dynamic reactive scheduling approach for modifying nominal schedules that could not be completed due to changes in the production target or manufacturing resources.

The area of process control is well matured, and recently there have been increasing applications of control-oriented frameworks for supply chain management and integration of planning and scheduling. Perea et al.[11] proposed a dynamic approach to supply chain management with ideas from process dynamics and control. Vargas-Villamil and Rivera[12] proposed a model predictive control (MPC) formulation for scheduling of reentrant semiconductor manufacturing lines. Bose and Pekny[13] proposed MPC for integration of planning and scheduling for a multiperiod operation of consumer goods supply chain. In each period, they used a forecasted model to calculate target inventories (control variable) for future periods and a scheduling model to achieve these targets by scheduling tasks (manipulated variable). Perea-Lopez et al.[14] proposed an MPC strategy for supply chain optimization with a rolling horizon approach for updating the changes to the supply chain.

In this work, we consider a hierarchical decomposition for integration of planning and scheduling and embed proactiveness for reactive scheduling. The proposed framework is developed for the complex plant configuration of a resource-constrained multistage, multiproduct hybrid flowshop facility. The overall decision-making is artfully segregated into appropriate decisions at individual levels with necessary degrees of freedom to make the model amenable to rescheduling. The approach proposed here is based on two key characteristics of the planning and scheduling problem that project a potential for problem simplification. The first is that, like in process control tasks, there is a natural, temporal decomposition of decision-making in a typical plant shop floor. Hierarchically, the upper layers focus on long-term objectives of profit and demand satisfaction while the lower layers look at relatively shorter term objectives of target satisfaction in the presence of shortfall transients. The second important characteristic is that the typical constraints at the lower hierarchical plant-floor levels are relatively short-term, flexible, and often potentially relaxable human constraints. This offers a potential for decomposition and heuristic problem solving at the lower levels. In our work, these characteristics have been exploited toward developing a framework that simplifies the reactive scheduling problem. Between the two approaches to reactive scheduling i.e., "reevaluate" or "repair", the approach proposed in this paper is closer to the "repair" option so as to maintain optimal operation in the presence of the parameter uncertainties/variations. We propose a control-theoretic approach for integration of reactive scheduling into the multilevel framework for the integration of planning and scheduling. The proposed framework also uses the receding horizon formulation, which has been so elegantly exploited in advanced process control algorithms, with real time feedback of production shortfalls, to ensure optimal satisfaction of the production targets.

This paper is structured as follows. The framework for multilevel decomposition of the overall problem of planning and scheduling, with proactiveness for reactive scheduling, is discussed in section 2 for a generic hybrid flowshop facility. Then the proposed control-theoretic approach to reactive scheduling is discussed for reacting to plant disruptions during nominal operation. In section 3, we consider a simple refinery flow sheet involving continuous lube production in a hybrid flowshop as a case study and present detailed model formulations and results in line with the proposed multilevel framework. The validation results for the reactive scheduling model of the hybrid flowshop plant are presented in section 4, followed by conclusions in section 5.

2. A General Multilevel Framework for Integration of Planning, Scheduling, and Rescheduling

In this work, we consider a hierarchical decomposition of the overall problem of planning and scheduling and then discuss the means of embedding contingency measures from the viewpoint of reactive scheduling.

Motivation for the Proposed Multilevel Decomposition. The motivation for the multilevel structure stems from the hierarchical nature of information flow that is typically seen in a production environment (Figure 1).

Typically at the *corporate level*, the horizon of planning is fairly extended and is based on actual (known) and forecasted demands. Functionally, the corporate level is minimally concerned with reacting to short-term production transients; rather the mandate at this level is to react to long-term slack (or a demand change) or shortfall that is available through feedback from the *supervisory level* (see Figure 1). The objectives here are

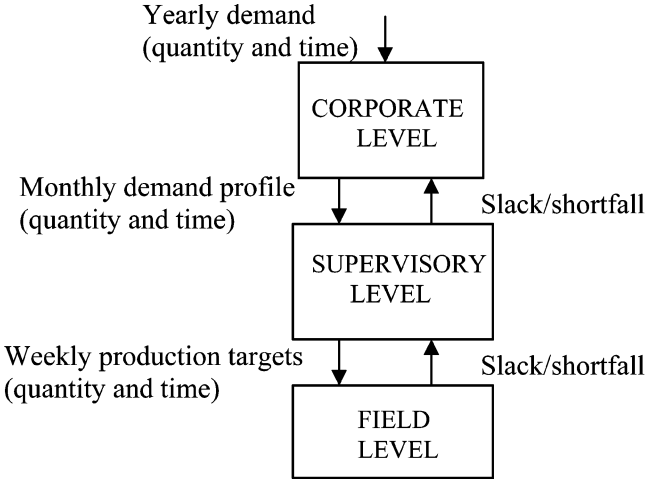

Figure 1. Information flow in a production environment.

in terms of determining production targets with a view to maximizing profits over a multiperiod operation. Additionally, we propose that the models must inherently have some means of absorbing contingency scenarios of reactive scheduling. With a view toward proactiveness for reactive scheduling, we propose that the upper levels should redistribute the demands over the multiperiods as discussed later in section 3.1, so as to have some flexibility for reactive scheduling at the lower levels.

The production targets calculated at the corporate level are passed on to the supervisory level whose objective is typically to ensure that these targets are met, keeping in mind the actual production level constraints. The horizon of focus at this level is relatively short-term; in terms of its role in reacting to short-term transients, this level could have a greater degree of involvement in decision-making. At the lower *field or operator level*, the production targets are imposed over much shorter horizons by the supervisory level. The objectives at the lower layers would be predominantly on meeting the production targets imposed by the supervisory layer.

In terms of the nature of models and constraints posed at each of the levels, the upper levels would be characterized by more abstract models, and the lower levels would be more rigorous models. For example, the upper levels may consider abstracted inventory constraints and/or assumed possible production losses in the downstream of the hybrid flowshop. The production losses could be either the trimming losses in cutting stock problems or the slopping losses during grade changeovers in refinery problems. The lower levels may consider detailed inventory constraints and/or constraints to account for the production losses mentioned above. However, at the field level, the constraints are relatively flexible and potentially relaxable; for example, the temporary nonavailability of a particular storage inventory can be accommodated at the field level, by an appropriate inventory management policy developed using operating heuristics. As seen in Figure 1, a feedback mechanism from the lower levels apprises the upper levels of the individual performance metrics so as to enable appropriate corrective planning at the upper levels.

2.1. Problem Definition. Consider a general multistage, multiproduct planning and scheduling problem with target demands specified over a multiperiod operation. In this paper, a generic hybrid flowshop configuration of various machines is assumed with different production routes for each product as shown in Figure 2. This facility can be easily simplified to any problem specific topology of series and/or parallel configuration of different stages. In the facility shown, there are a total of M stages and each can process Np grades. In each stage either there is a single unit m (as seen in stages 2, 3, 5, and so on up to M) or there is a parallel line l (as seen in stages 1, 4, and so on), with two units operating simultaneously. Additionally, there could be resource constraints on the feed side of each plant where the feed could be a continuous stream with finite storage space. Again as a problem specific situation these resource constraints may be easily simplified or dropped. In Figure 2, the continuous feed, received from an upstream plant with finite intermediate storage, enters stage 1 (S_1) with two parallel units (L_{11} and L_{12}). Stages 2 and 3 have single units (S_2 and S_3) and operate sequentially, stage 4 again has two parallel units (L_{41} and L_{42}), stage 5 has a single unit (S_5), and so on up to stage M (S_M). The global objectives are maximization of profit and timely satisfaction of the customer orders with minimal impact of plant disruptions (machine breakdowns etc.) if any. The latter objective is achieved implicitly through the proposed proactive measures and the receding horizon framework for local attenuation of the plant disruptions leading to infrequent revision of the commitments made to the customers.

With the resource constraints of feed to the plant envelope being a continuous stream with finite storage space, the plant schedule is governed by both the feed side inventories and the demand side constraints. Earlier Munawar et al.[15] proposed a generalized MINLP model for cyclic scheduling of this configuration with detailed inventory constraints. The model accounted for slopping losses during grade changeovers through a modified time slot definition. The model could handle special cases leading to empty slots (zero time duration) resulting from splitting of products into the parallel lines. However, such a model becomes intractable when we consider integration of planning and scheduling over a multiperiod operation. This is because of the presence of multiple due dates, varying demands, and longer time frames. Also from a reactive scheduling context, decisions on “repairing” or “reevaluating” the nominal schedule are not easy to make. Hence, in the next section we present a hierarchical decomposition of the overall problem.

2.2. Multilevel Decomposition. As discussed earlier, on the basis of the inherent functional decomposition of the global objectives, the overall problem can be traditionally decomposed into two major levels, an upper level for strategic (or long-term) planning over a multiperiod operation and a lower level for detailed scheduling in each time period. Additionally, we impose the requirement that the decomposition should also permit reactive scheduling at each level, although at different time frames. The upper level is revised on a less frequent basis to avoid frequent revision of the commitments made to customers. At the lower level the focus is on meeting the production targets imposed by the upper level. The schedule at this level may be revised on a relatively frequent basis but without violating the global objectives. Depending on the complexity of the problem shown in Figure 2, the lower level is further

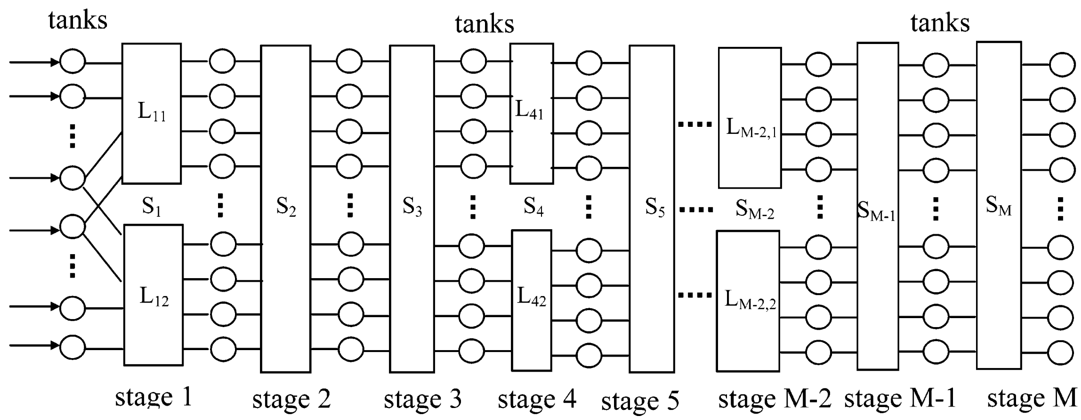

Figure 2. Schematic of a generic hybrid flowshop plant.

decomposed into a supervisory level and an operator level in the hierarchical framework of Figure 1.

The model formulation at these levels would include the mass and inventory balances and constraints related to the sequence of processing and transitions. The models at each level need to be formulated first, based on different levels of abstraction so that the schedule generated by an upper level predicts production targets that could be realistically imposed at the lower level. The model formulation at the upper level considers all demands (Q_i^t) for product i in each time period t, whereas the middle (lower) level focuses on meeting the demands of the current time period t. Toward meeting these targets at the middle level, various short-term constraints such as availability of inventory storage could be posed. These constraints can be formulated and accommodated at the third level, which considers inventory requirements on an even shorter time horizon.

The information flow in Figure 1 also indicates a feedback from the lower levels to the upper levels. From a nominal solution viewpoint this can be explained as follows. The upper levels assume an abstraction of the lower level plant constraints while the nominal plan/schedule is developed. However when the resultant targets/inventory requirements are projected onto the lower levels, these constraints need to be validated. It must be noted that the constraints may be conservatively posed during the development of the nominal schedule at each level; hence this would ensure feasibility of the constraints at the lower levels. However, any unforeseen shortfall/plant shutdowns or inventory shortage could arise. In such a case the problem formulation at each of the levels should have a provision to accommodate these changes locally so that the nominal schedule/targets projected by the upper layers are not disturbed. However if these levels cannot accommodate these unforeseen disturbances due to infeasibility, then a feedback mechanism to revise the schedule/targets at the upper level is achieved through the feedback shown in Figure 1. In the following subsection, a receding horizon based control-theoretic approach has been proposed for integrating reactive scheduling to handle unexpected machine breakdowns in a local fashion.

In the decomposition approach proposed in this paper, the presence of abstractions/relaxations of various constraints (such as those on the processing times and slopping losses) at the upper levels could cause the solution to be suboptimal. These abstractions in the proposed decomposition based approach simplify the problem at each level and enable quick solutions although the solutions at each level could be suboptimal, if the abstractions are not realistic. However, these abstractions could be chosen carefully either on the basis of a priori knowledge or in an iterative fashion based on the feedback from the lower levels, to prevent this occurrence. In comparison with a simultaneous formulation of the planning and scheduling problem using a single monolith based approach, whose size and complexity may restrict the attainment of any solution, the proposed approach could be said to be more realistic and practical.

2.3. Control-Theoretic Approach to Reactive Scheduling. During normal operation, if there are any plant disruptions, then to meet the global objectives of the overall problem, as in cascade control, we propose that any local disturbances (machine breakdowns etc.) at the lower level have to be attenuated locally before they affect the global performance.

Let us focus on how the disturbances at one level can be rejected or attenuated locally without affecting the global performance. Consider a schematic of the two-level decomposition as shown in Figure 3. Consider a shutdown time of duration T_s at time instant T_{br} (breakdown time) in time period t. Now, we consider a receding horizon window (of length T_r, as shown in Figure 3) from that time instant onward at the lower level, and meet the shortfall with increased production rates (assuming of course that the necessary degrees of freedom exist for this purpose). If necessary we may also intrude (x h) into the next time period ($t + 1$) without sending a feedback, until the next scheduled revision of the upper level. When we intrude into the next time period, it is also assumed that the corresponding targets in the next period, as decided by the upper layer in the nominal plan, get cascaded into the lower level.

The details of the reactive scheduling formulations are discussed later in section 4. In contrast, if we take the normal feedback control option, then in the same time horizon we may have to trigger the primary level thus not only forcing frequent revisions of the commitments made to customers but also spreading the effect of the disturbance over longer time periods. Thus, only if these disturbances cannot be handled locally, a feedback to the primary level is sent to seek a revision of the targets.

On the basis of the above discussion, the following generalized proposition can be made: *Disturbances can be handled better at the lower levels than at the upper levels, because of better accuracy of the models at the lower levels, over the upper levels. Local attenuation of disturbances in a cascade-control fashion would be more desirable than its counterpart of feedback control option,*

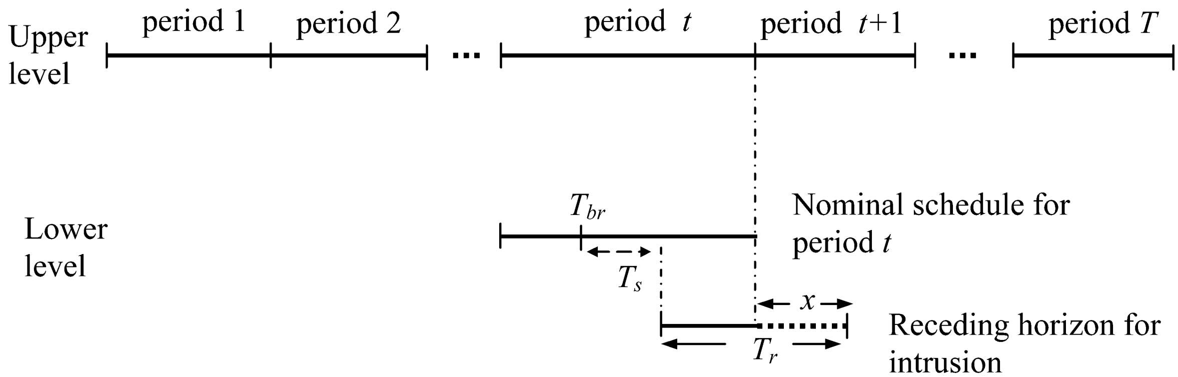

Figure 3. Proposed multilevel, control-theoretic framework with receding horizon for reactive scheduling.

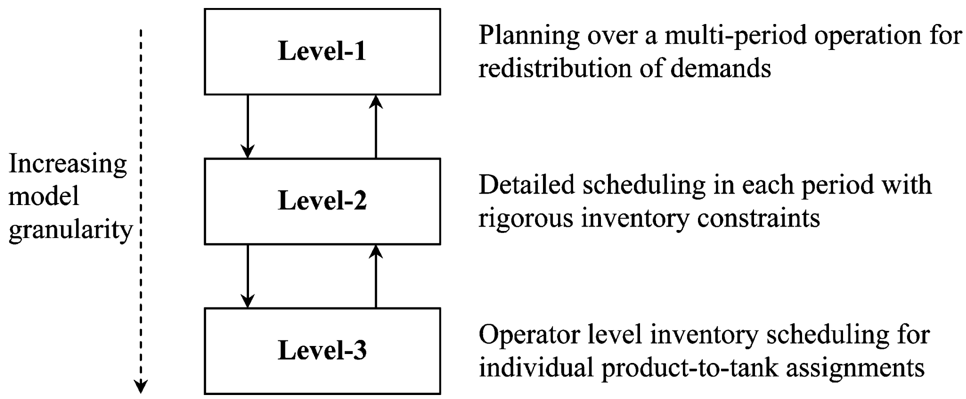

Figure 4. Proposed three-level decomposition for the hybrid flowshop facility.

as the latter would lead to frequent revision of the upper level schedules and commitments, compared to the former, which nullifies the effect of the disturbance within a shorter horizon if feasible.

The proposed methodology also has an analogy to the other well-established control theory, viz., robust process control. There is always a tradeoff between an aggressive schedule and a robust schedule. A more aggressive schedule is less amenable to react to plant disturbances. On the other hand, a less aggressive schedule can locally accommodate the disturbances better compared to a more aggressive schedule. Hence, in formulating the various models in the hierarchical decomposition of the overall problem, some proactiveness has to be incorporated at each level for enabling good reactive scheduling. For example Liao and Chen[16] proposed a heuristic rescheduling for a textile industry under machine breakdowns. As a back-off from the best solution, they used an unconventional scheduling objective of maximizing the total setup time (thus allowing some idle time of machine) to reduce the machine break down rate. Thus, a back-off from the optimum solutions on account of proactiveness to enable good reactive schedules is necessary.

The concepts of decomposition, model abstraction, and the hybrid flowshop issues that are proposed in this paper will be useful in general for all the applications of integrated planning and scheduling problems, although the details discussed in this paper appear to be problem specific. For example the assumed slopping losses at the top level can be generalized to account for the general production losses that may possibly occur at the detailed scheduling level; another example where this would be relevant is in modeling trim losses in paper manufacturing. To account for such production losses at the top levels, rigorous analysis may be required which may result in aggravation of the combinatorial aspects of the problem; hence there is the need for abstraction at these upper levels. The concept of time intrusion and proactive measures discussed in this paper would be useful for all cyclic scheduling problems with demand limited scenarios. However, if the plant has tight demands over the entire planning horizon, then the proposed concepts of proactiveness and receding horizon may not be helpful as anyway there is not adequate scope for reactive scheduling in such situations.

3. Model Formulation for the Multilevel Framework

In this section, the detailed model formulation for the proposed multilevel framework is presented followed by validation through a case study involving cyclic scheduling for lube production in a hybrid flowshop facility.

3.1. Three-Level Decomposition. Due to the complexity of the resulting problem considered in Figure 2, as discussed earlier, we propose a three-level decomposition of the overall problem as shown in Figure 4. Though the case study considered in this paper is for cyclic scheduling of continuous multiproduct plants, the proposed framework can also be readily extended to other cases of short-term scheduling. As part of an ongoing research activity, we have also extended this methodology to an application in a paper manufacturing industry.

The top level (Level-1) for long-term planning and scheduling has a 1−3 month time horizon, the production targets for which are specified over a multiperiod operation. At this level, we consider an abstract model with assumed slopping losses, and simplified inventory constraints in terms of upper bounds on processing times based on some heuristics. The objective at this level is maximization of profit and demand redistribution to have tighter schedules in the early periods on account of proactiveness from a reactive scheduling perspective as elaborated below.

For a multiperiod operation, it is evident that an effective planning would require the processing tasks to be redistributed uniformly, rather tightly in the early periods, so as to be optimal with respect to all the periods. For example, a high demand for a particular product in a given period would require its processing to be distributed over the entire time up to that period (including the previous periods), even though the demand for this product in the previous periods is low or zero. Even otherwise, from a reactive scheduling perspective we propose that the upper level must push in tighter schedules in the early periods so that there is some flexibility in the forthcoming periods to accommodate the unforeseen events. Hence, we allow overproduction in the first period of interest. As the upper level is anyway revised at the end of each period, the first period of interest rolls through in a rolling horizon fashion, and finally we have the best possible production rates in each time period. Hence the objective at Level-1 could be maximization of production in the first period of interest subject to penalties for overall costs in all the time periods. Level-1 accordingly predicts targets (Q'^{t}_{i}) for Level-2 based on tighter redistribution of demands in the early periods. Hence, the purpose of this level is primarily demand redistribution and target setting for Level-2, rather than detailed scheduling. Using these nominal targets, the detailed model at Level-2 needs to be solved for each time period to realize the production targets set by the upper Level-1. If these targets cannot be met at Level-2 due to plant/inventory constraints, then Level-1, when revised at the end of each time period, is appraised with a feedback for redistribution of demands in the remaining periods until we have realistic demand target at Level-2 in each time period.

In the next level (Level-2), we consider detailed scheduling over a short-term horizon for the target demands set by the upper level (Level-1). The model at this level has a novel modified time slot definition to account for actual slopping losses and rigorous inventory constraints as discussed in Munawar et al.[15] The time horizon at Level-2 is that of a single period (say 1 month) with the objective of maximizing total profit subject to penalties for grade change and inventory costs. If the target demands set by the Level-1 cannot be met at this level in the presence of actual slopping losses and rigorous inventory constraints, this model would yield the maximum possible demands that can be met. Then a feedback is sent to Level-1 thus seeking a redistribution of demands for the remaining periods as discussed above. The formulation for detailed product-to-tank assignments for storage of various grades is complicated at this level, and hence, we do not consider detailed inventory management in terms of tank assignments at this level, but consider an abstraction of the total inventory available for each product. The inventory management on an hourly/daily basis for individual product-to-tank assignments is done heuristically at the lowermost level (Level-3). For an efficient usage of the available tank volumes, a novel inventory slicing and tank reallocation (ISTR) algorithm is proposed at Level-3.

Initially, we solve the models for these three levels and get the detailed (nominal) production and inventory schedules for normal operation in each period. During normal operation, if there are any unexpected machine breakdowns, then we trigger the reactive scheduling model that is discussed in section 4. In the succeeding subsections, we first develop the model formulations for each of the three levels for generating the nominal schedules.

3.2. Level-1 Model Formulation. Consider the generic M-stage hybrid flowshop facility of Figure 2. Consider long-term planning and scheduling at Level-1 for a multiperiod operation involving T periods, with target demands specified in each period. As discussed earlier at this level, we consider an abstract model with assumed slopping losses, and simplified inventory constraints based on some heuristics. So, at Level-1 we consider the traditional time slot definition, the sum of transition time and the processing time. At Level-1, as an abstraction of the inventories to be handled at Level-2, upper bounds are specified on processing times for all products based on past experience or heuristics. The slopping losses are reflected by reducing the conversions or yields of grades appropriately using an apparent yield based on heuristics. The processing rates are considered as decision variables at Level-1 with specified lower and upper limits. If the model at Level-1 becomes computationally intractable, then we can use average values for the flowrates instead of defining them as variables.

All the variables and constraints defined earlier[15] for the cyclic scheduling at Level-2 are considered at Level-1 also, except that we do not consider the detailed inventory constraints and the variables corresponding to slopping. The variables here have an additional superscript t for different time periods. We have three sets of variables; the first set is common to all the sequential stages m: S_2, S_3, S_5, and so on up to S_M of Figure 2. The second set refers to all the first units of the parallel lines l: L_{11}, L_{41}, and so on. The third set corresponds to all the second units of the parallel lines l: L_{12}, L_{42}, and so on. We can lump together mathematically the grades in all the first units of the parallel lines and use a common index ($i' \in I'$); similarly for grades in all the second units of the parallel lines as well we can use a different common index ($i'' \in I''$). For the sequential stages also a common grade index is used ($i \in I$) in the formulation.

The following is the nomenclature at Level-1 for the sequential stages. For the other two sets of the parallel lines also the variables must be defined with appropriate superscripts, in a similar way as done earlier[15] for the Level-2 model.

Indices

i, j = grades
k = slots
m = stages
t = cumulative time periods ($t1$, $t2$, $t3$, ...)

Variables

y^{t}_{ikm} = binary variable denoting grade allotment to slot k of stage m in time period t

z^{t}_{ijkm} = transition from grade j to i in slot k of stage m in time period t

Nzk^{t}_{km} = variable to locate the position of zero slots in slot k of stage m in time period t

Tpp^{t}_{ikm} = processing time of grade i in slot k of stage m in time period t, h

Tsp^t_{ikm} = start time of processing of grade i in slot k of stage m in time period t, h

Tep^t_{ikm} = end time of processing of grade i in slot k of stage m in time period t, h

Rp^t_{im} = processing rate of grade i in stage m in time period t, m^3/h

T^t_c = cycle time in time period t, h

T^t_{idle} = idle time allowed in time period t

F'_i = the feed rates of the base stocks diverted to flare when there are low demands or no demand on a product, m^3/h

Sf^t_{il} = split factor of common product i denoting the fraction fed to the first line of the parallel line l in time period t

q^t_i = binary variable to denote which unit L_{11} or L_{12} starts processing first with respect to the common grade i

Parameters

α_{im} = yield or conversion of grade i in stage m

β^t_{im} = apparent yield (includes abstraction of slopping losses) of grade i of stage m in time period t

F_i = continuous feed rate of grade i into the feed inventory before stage 1, m^3/h

Ft_i = continuous total feed rates of the base stocks, m^3/h

T^t = duration of time periods (T^{t1}, T^{t2}, ...), h

Q^t_i = bulk demand of product i specified at the end of time period t, m^3

P_i = price of grade i, \$

Ctr_{ijm} = grade transition cost from grade j to grade i in stage m, \$

τ_{ijm} = grade transition time from grade j to grade i in stage m, h

The apparent yield parameter (β^t_{im}) includes the abstraction of slopping losses. For a demand limited scenario we may have to make provision for some idle time in the subsequent periods (except the first period because we are anyway allowing overproduction in the first period of interest) to avoid the case where machines are forced to run at their lower bound on processing rates; instead an idle time may be preferred for the remaining time if the demands in some periods are low. As there is no incentive for production in the subsequent periods (except the first period), the output predicted by Level-1 may show some zero productions in the later time periods. Hence, we need a formulation at this level that accommodates empty slots.

The following are the basic constraints of cyclic scheduling common to both Level-1 and Level-2:

$$\sum_k y^t_{ikm} \le 1 \qquad \forall\, i \in I \qquad \forall\, m = 2, 3, 5, ..., M \qquad \forall\, t \tag{1a}$$

$$\sum_i y^t_{ikm} \le 1 \qquad \forall\, k \qquad \forall\, m = 2, 3, 5, ..., M \qquad \forall\, t \tag{1b}$$

$$\sum_i y^t_{ikm} \le \sum_i y^t_{i(k+1)m} \qquad \forall\, k < \text{NK} \qquad \forall\, m = 2, 3, 5, ..., M \qquad \forall\, t \tag{2a}$$

$$z^t_{ijkm} = y^t_{ikm}[y^t_{j(k--1)m} + (1 - \text{Nzk}^t_{(k--1)m})y^t_{j\text{NK}m}] \qquad \forall\, i, j \in I \qquad \forall\, k \qquad \forall\, m = 2, 3, 5, ..., M \qquad \forall\, t \tag{2b}$$

$$\text{Nzk}^t_{km} = \sum_i y^t_{ikm} \qquad \forall\, k \qquad \forall\, m = 2, 3, 5, ..., M \qquad \forall\, t \tag{2c}$$

$$\left.\begin{aligned} &\text{Tsp}^t_{ikm} \le U^T y^t_{ikm} \\ &\text{Tep}^t_{ikm} \le U^T y^t_{ikm} \\ &\text{Tpp}^t_{ikm} \le U^T y^t_{ikm} \\ &\text{Tpp}^t_{ikm} = \text{Tep}^t_{ikm} - \text{Tsp}^t_{ikm} \end{aligned}\right\} \forall i \in I \forall k \;\; \forall m = 2,3,5,...,M \forall t \tag{3a}$$

$$\sum_i \text{Tsp}^t_{i1m} = \sum_i \sum_j \tau_{ijm} z^t_{ij1m} \qquad \forall\, m = 2, 3, 5, ..., M \qquad \forall\, t \tag{3b}$$

$$\sum_i \text{Tsp}^t_{i(k+1)m} = \sum_i \text{Tep}^t_{ikm} + \sum_i \sum_j \tau_{ijm} z^t_{ij(k+1)m} \qquad \forall\, k < \text{NK} \qquad \forall\, m = 2, 3, 5, ..., M \qquad \forall\, t \tag{3c}$$

$$q^t_i \sum_{k'} \text{Tep}'^t_{i'k'1} + (1 - q^t_i) \sum_{k''} \text{Tep}''^t_{i''k''1} \le q^t_i \sum_{k''} \text{Tsp}''^t_{i''k''1} + (1 - q^t_i) \sum_{k'} \text{Tsp}'^t_{i'k'1} \qquad \forall\, i \in I' \cap I'' \qquad \forall\, t \tag{4}$$

$$T^t_c \ge \sum_{k'} \sum_{i'} \text{Tpp}'^t_{i'k'l} + \sum_{i'} \sum_{j'} \tau'_{i'j'l} z'^t_{i'j'k'l} \qquad \forall\, l = 1, 4, 7, ..., M - 2 \qquad \forall\, t \tag{5a}$$

$$T^t_c \ge \sum_{k''} \sum_{i''} \text{Tpp}''^t_{i''k''l} + \sum_{i''} \sum_{j''} \tau''_{i''j''l} z''^t_{i''j''k''l} \qquad \forall\, l = 1, 4, 7, ..., M - 2 \qquad \forall\, t \tag{5b}$$

$$T^t_c \ge \sum_k \sum_i \text{Tpp}^t_{ikm} + \sum_i \sum_j \tau_{ijm} z^t_{ijkm} \qquad \forall\, m = 2, 3, 5, ..., M \qquad \forall\, t \tag{5c}$$

$$F_i + F_i' = \text{Ft}_i \qquad \forall\, i \in I \tag{6}$$

$$F_{i'} T^t_c = \text{Rp}'^t_{i'1} \sum_{k'} \text{Tpp}'^t_{i'k'1} \qquad \forall\, i' \in I' \text{ of } L_{11} \text{ only} \qquad \forall\, t \tag{7a}$$

$$\text{Sf}^t_{i'1} F_{i'} T^t_c = \text{Rp}'^t_{i'1} \sum_{k'} \text{Tpp}'^t_{i'k'1} \qquad \forall \text{ common grades } i' \in I' \text{ of } L_{11} \text{ only} \qquad \forall\, t \tag{7b}$$

$$(1 - \text{Sf}^t_{i''1}) F_{i''} T^t_c = \text{Rp}''^t_{i''1} \sum_{k''} \text{Tpp}''^t_{i''k''1} \qquad \forall \text{ common grades } i'' \in I'' \text{ of } L_{12} \text{ only} \qquad \forall\, t \tag{7c}$$

$$F_{i''} T^t_c = \text{Rp}''^t_{i''1} \sum_{k''} \text{Tpp}''^t_{i''k''1} \qquad \forall\, i'' \in I'' \text{ of } L_{12} \text{ only} \qquad \forall\, t \tag{7d}$$

$$\beta'^t_{i'l}\, \text{Rp}'^t_{i'l} \sum_{k'} \text{Tpp}'^t_{i'k'l} = \text{Rp}^t_{i'm} \sum_k \text{Tpp}^t_{i'km} \qquad \forall\, i' \in I' \text{ only} \qquad \forall\, t \tag{8a}$$

$$\beta'^t_{i'l}\, \text{Rp}'^t_{i'l} \sum_{k'} \text{Tpp}'^t_{i'k'l} + \beta''^t_{i''l}\, \text{Rp}''^t_{i''l} \sum_{k''} \text{Tpp}''^t_{i''k''l} = \text{Rp}^t_{im} \sum_k \text{Tpp}^t_{ikm} \qquad \forall \text{ common grades} \qquad \forall\, t \tag{8b}$$

$$\beta''^t_{i''l}\, \text{Rp}''^t_{i''l} \sum_{k''} \text{Tpp}''^t_{i''k''l} = \text{Rp}^t_{i''m} \sum_k \text{Tpp}^t_{i''km} \qquad \forall\, i'' \in I'' \text{ only} \qquad \forall\, t \tag{8c}$$

$$\beta^t_{im}\, \text{Rp}^t_{im} \sum_k \text{Tpp}^t_{ikm} = \text{Rp}^t_{i(m+1)} \sum_k \text{Tpp}^t_{ik(m+1)} \qquad \forall\, i \in I \qquad \forall\, m = 2, 3, 5, ..., M - 1 \qquad \forall\, t \tag{9}$$

Constraints (1) to (3) are common to all stages (m) and parallel lines (l) and hence are shown over only the set (I) for the sequential stages. For the other two sets ofparallel lines as well (I' and I'') these constraints should be written in a similar way in each time period. In each time period t, we consider as many slots as the number of grades in each cycle; however some slots may be empty. The inequalities in constraints (1) enforce unique allotment of a grade to a slot with no grade repetitions within a cycle and no allotment if the slot is empty. To ensure that the start and end times of the zero slots should be zero, we must avoid zero slots in the middle of the cycle, and hence all the zero slots are pulled toward the beginning of the cycle through constraint (2a). Here NK represents the total number of slots in a cycle. The transition variable z^t_{ijkm} is uniquely defined by constraint (2b) and is a continuous variable between 0 and 1. The operator "- -" here is a cyclic lag operator used in GAMS.[17] Nzk_{km} identifies the location of the zero slots and is defined in constraint (2c). If Nzk_{km} of slot k is zero, then that slot is a zero slot in stage m. When there are no zero slots, then the second term in eq (2b) vanishes, and the transition variable z^t_{ijkm} is calculated as the transition from grade j being produced in slot $k-1$ to grade i to be produced in the current slot k. But in the presence of zero slots, since all the zero slots are pulled toward the beginning of the cycle, the transition for the first nonzero slot of the cycle will be from the last slot of the cycle as dictated by the second term.

The non-negativity inequalities of (3a) ensure that, when a product is not assigned to a slot, the corresponding start, end, and processing times are all zero. Here U^T represents a large positive number, which may be equal to the length of the time period t. In constraint (3b) the start time of processing of the first slot is calculated, while in constraint (3c) the start time of processing of the remaining slots is calculated to be the transition time added to the end time of the previous slot. Constraint (4) ensures that the parallel lines L_{11} and L_{12} do not operate simultaneously, for the processing of common grades i fed from the common tank, from a practical perspective to avoid complicated operation as discussed in Munawar et al.[15] The cycle time is defined by constraints (5) as the maximum of summation of the lengths of all time slots across all lines and stages. In constraint (6), F'_i would be zero when there are finite demands on all products, and in case there is no demand on a product, $F'_i = \mathrm{Ft}_i$. The mass balances of feedstocks into the parallel lines of stage 1 are written in constraints (7). The mass balance between any parallel line l and the subsequent stage m are written in constraints (8). The mass balance between any stage m and the subsequent parallel line l of Figure 2 can be written in a similar way. In constraint (9), mass balance between any two sequential stages is written. The following constraints ensure redistribution of demands across the multiperiods as discussed earlier.

$$\beta^{t1}_{iM}\,\mathrm{Rp}^{t1}_{iM}(\sum_k \mathrm{Tpp}^{t1}_{ikM})\left(\frac{T^{t1}}{T^{t1}_{\mathrm{c}}}\right) \geq Q^{t1}_i \qquad \forall\, i \in I \quad (10a)$$

$$\beta^{t1}_{iM}\,\mathrm{Rp}^{t1}_{iM}(\sum_k \mathrm{Tpp}^{t1}_{ikM})\left(\frac{T^{t1}}{T^{t1}_{\mathrm{c}}}\right) + \beta^{t2}_{iM}\,\mathrm{Rp}^{t2}_{iM}(\sum_k \mathrm{Tpp}^{t2}_{ikM}) \times \left(\frac{T^{t2} - T^{t2}_{\mathrm{idle}}}{T^{t2}_{\mathrm{c}}}\right) \geq Q^{t1}_i + Q^{t2}_i \qquad \forall\, i \in I \quad (10b)$$

$$\beta^{t1}_{iM}\,\mathrm{Rp}^{t1}_{iM}(\sum_k \mathrm{Tpp}^{t1}_{ikM})\left(\frac{T^{t1}}{T^{t1}_{\mathrm{c}}}\right) + \sum_{t>1} \beta^{t}_{iM}\,\mathrm{Rp}^{t}_{iM}(\sum_k \mathrm{Tpp}^{t}_{ikM}) \times \left(\frac{T^{t} - T^{t}_{\mathrm{idle}}}{T^{t}_{\mathrm{c}}}\right) \leq \sum_t Q^{t}_i \qquad \forall\, i \in I \quad (10c)$$

Constraint (10a) allows overproduction in the first time period where there is no provision for an idle time. Even if there is a provision for idle time in the first time period, the Level-1 output predicts zero idle time because we seek a maximization of production in the first time period. Sometimes, if the specified demands in the first time period are already high, then the "≥" inequality in constraint (10a) may lead to infeasibility, in which case it needs to be relaxed. In constraint (10b), the production at the end of the second period is forced to be greater than or equal to the demands at the end of second time period. Similar constraints need to be written at the end of each time period, except for the last time period as shown by constraint (10c) where overproduction is not needed. As discussed earlier we consider the objective function to be maximization of production in the first period of interest subject to penalties for overall grade changeovers in all the time periods. For capacity limited scenarios, to find the maximum possible demands that can be met, the unsatisfied demands at the end of the last time period are penalized in the objective function in the last term in eq 11. It may be noted that in the last term of the

$$\mathrm{profit} = \sum_i P_i \beta^{t1}_{iM}\,\mathrm{Rp}^{t1}_{iM}(\sum_k \mathrm{Tpp}^{t1}_{ikM})\left(\frac{T^{t1}}{T^{t1}_{\mathrm{c}}}\right) - \sum_i\sum_j\sum_k\sum_m\sum_t \mathrm{Ctr}_{ijm}\tau_{ijm} z^t_{ijkm} - \sum_{i'}\sum_{j'}\sum_{k'}\sum_l\sum_t \mathrm{Ctr'}^t_{i'j'l}\tau'^t_{i'j'l} z'^t_{i'j'k'l} - \sum_{i''}\sum_{j''}\sum_{k''}\sum_l\sum_t \mathrm{Ctr''}^t_{i''j''l}\tau''^t_{i''j''l} z''^t_{i''j''k''l} - \mathrm{penalty}\sum_i \times \left(\sum_t Q^t_i - \sum_t \beta^t_{iM}\,\mathrm{Rp}^t_{iM}(\sum_k \mathrm{Tpp}^t_{ikM})\left(\frac{T^t - T^t_{\mathrm{idle}}}{T^t_{\mathrm{c}}}\right)\right) \quad (11)$$

objective function although the T_{idle} is used for all the time periods for readability, its value for $t = 1$ is manually assigned to be zero. About the range of the penalty, it is a large positive number which may be an order of magnitude higher than the values of the other terms in the objective function so that the last term representing the total unsatisfied demand is minimized or driven to zero at the optimum. For a demand limited scenario this term would anyway be driven to zero in the optimal solution. The schedule at Level-1 is meant to be aggressive with respect to only the first period of interest, and hence after solving Level-1, we need to solve Level-2 for the current time period (t) with the demands projected by Level-1 as the targets. As will be shown in the next subsection, Level-2 solves for the best possible demands in the presence of actual slopping losses and detailed inventory constraints. If all of the demands projected by Level-1 cannot be met at Level-2 (i.e. in the presence of ambitious targets) in the current time period, then the shortfall is sent as feedback to Level-1, thus seeking a redistribution of demands. This

Table 1. Problem Features at Levels 1 and 2

problem feature	Level-1	Level-2
time period	multiple	single
processing rates	variable	variable
abstractions:		
slopping losses	assumed	calculated
inventory	bounds on processing times	calculated
objective function	demand redistribution and profit maximization	profit maximization

communication between the levels is necessary in the event of ambitious targets, at the end of which the nominal schedule for the best achievable demands is developed at Level-2.

3.3. Level-2 Formulation. At Level-2 we consider the target demands set by Level-1 for the current time period and check if these demands can be realized in the presence of actual slopping losses and detailed inventory constraints. The model equations at this level are similar to those presented for Level-1 except that they pertain to the first time period of interest. In the continuous time domain representation the definition of a time slot is modified to account for the feed losses in slopping and additional slopping variables are defined. We consider detailed inventory constraints in terms of the inventory breakpoints that define the total inventory profile for consumption and discharge of material as discussed earlier in Munawar et al.[15] Due to brevity, we do not present the full rigorous model for Level-2 in this paper but would refer the reader to our earlier publication[15] that discusses this formulation.

The detailed product-to-tank assignments are not done here but are considered heuristically at Level-3. At Level-2 however, we consider an abstraction of the available inventory volumes for each product. As we can see later in the Level-3 model, tanks are reused for storage of different grades over the time horizon and hence we consider an overestimation for the available inventory volumes based on some heuristics or past experience and these volumes are used as upper bounds on the maximum inventory breakpoint for each grade at Level-2. The problem at Level-2 is solved sequentially for the nominal case in each period with the target demands taken from the first period of interest of Level-1 as discussed earlier. Some of the important problem features at Level-1 and Level-2 are described in Table 1.

Before going into the details of the Level-3 formulation the proposed models at Level-1 and Level-2 are demonstrated on the case study involving lube production in a hybrid flowshop facility in the following subsection.

3.4. Case Study on Lube Production in a Hybrid Flowshop Facility. Consider a case study for single-site lube production in a hybrid flowshop facility for 4 products produced in 3 stages as shown in Figure 5 (referred to as the 4P3S problem earlier in Munawar et al.[15]), wherein stage 1 has two parallel machines, viz., line 1 and line 2 that relate to the same processing task (for e.g. extraction); however line 2 is an additional parallel unit that is present to accommodate increased demands on certain products.

With the exception of component B, which can be processed in both the lines, all the products have to be processed one at a time in all the stages in the same sequence: line 1 or line 2 followed by stages 2 and 3. Additionally, the total feed rate of "C" and "D" that can be charged into line 2 is fixed by the feed rate of "E". The base stocks of A, B, and E are continuous streams with finite storage space, coming from upstream plants with the following fixed flow rates: 21, 28, and 33 m^3/h, respectively.

The total inventory space available is finite and is about 3800 m^3 overall across all stages/lines, except for the product side where unlimited inventory space (UIS) is considered. Since the feed is a continuous stream received from upstream plants, the feed inventory tanks would almost always be busy. Nevertheless, if some inventory is unused, then there is a potential for usage elsewhere in the plant from a reactive scheduling point of view. (Hence, the objective at the operator level would be to minimize the available inventory usage to the best extent possible.) The prices for the products are: 350, 500, 250, and 250 \$/m^3 respectively. The transition costs are \$3500 for all grades in all lines and stages except between grades "C" and "D" for which there is no transition cost. The inventory costs are assumed to be \$5/m^3. The yields, processing rates, and the sequence and stage dependent transition times are given in Table 2.

As an abstraction of the inventories to be handled at Level-2, upper bounds are specified on processing times, at Level-1, for all products and are assumed to be 25 h, 20 h, and 15 h for stages 1, 2, and 3 respectively (based on heuristics or past experience). The slopping losses are reflected through the assumed apparent yield of grades as shown in Table 3.

Nominal Plan for Level-1. Consider midterm planning at Level-1 for a multiperiod operation, with demands specified as in Table 3 for three periods each of 1000 h ($t = 1$), 900 h ($t = 2$), and 800 h ($t = 3$) duration, respectively. With the objective function at Level-1 being maximization of production in the first period of interest subject to penalties for overall grade changeovers in all the time periods, Table 4gives a comparison of the actual demands specified (set points) and the output from Level-1 for the three time periods considered. Here, since the problem considered corresponds to a capacity limited scenario, the total demands at the end of all time periods could not be met. The schedule is meant to be aggressive with respect to only the first period of interest, and now we solve Level-2 for the first time period ($t = 1$) with the demands projected by Level-1 as targets (set points). Level-2 predicts the best possible demands in the presence of actual slopping losses and detailed inventory constraints.

It was found that the maximum possible demands that can be met by Level-2 are 7501.41 m^3, 8953.56 m^3, 1716.5 m^3, and 1817.14 m^3 of A, B, C, and D, respectively. There is slack of about 1432.2 m^3 (3148.7 m^3 − 1716.5 m^3) of product C at Level-2 wrt the target set by Level-1. Nevertheless, compared to the demand specified of C in $t = 1$ (1470 m^3) there is still some overproduction, which means Level-1 predicted ambitious targets for C in the first time period. Now, this slack has to be sent to Level-1 as feedback for a redistribution of the demands in the remaining two periods. The second time period ($t = 2$) now at Level-1 becomes the first period of interest, and since we allow overproduction in the first period of interest, this slack from Level-2 can now be added on to the last period ($t = 3$) as it will anyway get accommodated into the second time period ($t = 2$) if feasible. Moreover, since this is a capacity limited problem, the overall slack in the

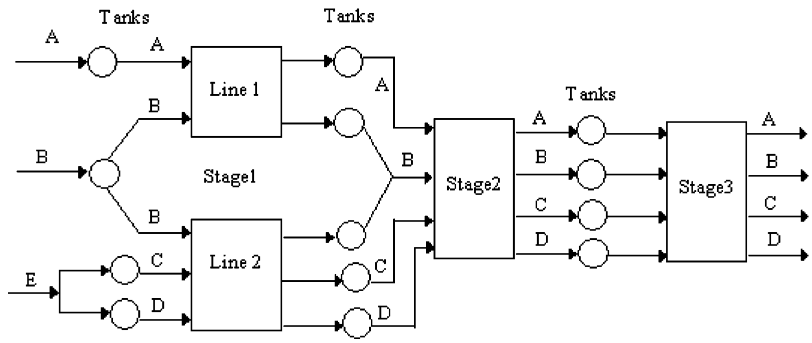

Figure 5. Schematic of lube production in a hybrid flowshop facility.

Table 2. Yields, Processing Rate Data, and Transition Time Data for the Case Studies

(A) Yields and Processing Rate Data

	yield				lower and upper bounds on processing rates (m^3/h)			
grade	line 1	line 2	stage 2	stage 3	line 1	line 2	stage 2	stage 3
A	0.73		0.73	0.91	25−60	30−72	23−57	
B	0.69	0.51	0.67	0.95	27−60	35−70	28−70	20−55
C		0.48	0.48	0.98		36−70	26−72	23−58
D		0.48	0.48	0.98		40−65	28−70	20−55

(B) Transition Time Data

	transition time (h)								
	line 1		line 2			stages 2 and 3			
grade	A	B	B	C	D	A	B	C	D
A	0	1				0	1	4	4
B	1	0	0	3	3	1	0	3	3
C			3	0	0	4	3	0	0
D			3	0	0	4	3	0	0

Table 3. Apparent Yields and Demand Data for the Lube Production Case Study

	apparent yield					demand (m^3)		
grade	line 1	line 2	stage 2	stage 3	grade	$t = 1$	$t = 2$	$t = 3$
A	0.70		0.63	0.81	A	6237	7713	8190
B	0.66	0.46	0.57	0.85	B	8414	7572	6731
C		0.45	0.38	0.88	C	1470	2780	1975
D		0.45	0.38	0.88	D	1470	2780	1975

Table 4. Demands (m^3) Specified and Projected by Level-1 for All Three Periods

	set point			output		
grade	$t = 1$	$t = 2$	$t = 3$	$t = 1$	$t = 2$	$t = 3$
A	6237	7713	8190	7501.41	6751.27	6001.13
B	8414	7572	6731	8953.56	7032.44	6731
C	1470	2780	1975	3148.70	1101.30	0
D	1470	2780	1975	1817.14	3367.96	0

Table 5. Demands (m^3) Specified and Projected by Level-1 for the Next Two Periods

	set points		output	
grade	$t = 2$	$t = 3$	$t = 2$	$t = 3$
A	6751.27	7887.32	6751.27	6001.12
B	7032.44	6731	7032.44	6731
C	1101.30	3407.22	1101.30	0
D	3367.96	1039.9	3367.96	0

demands projected by Level-1 as per Table 4 is also added on to the last time period. The set points and the output of demands at Level-1 for the remaining two periods are shown in Table 5 with the objective of maximization of production in the first period of interest, subject to penalties for the overall grade changeovers.

Now we again solve Level-2 for $t = 2$ with these demands as targets (set points) and check for feasibility.

Table 6. Demands (m^3) Specified and Projected by Level-2

	$t = 1$		$t = 2$		$t = 3$	
grade	set point	output	set point	output	set point	output
A	7501.41	7501.41	6751.27	6751.27	7887.32	6505.81
B	8953.56	8953.56	7032.44	7032.44	6731	6731
C	3148.70	1716.5	1101.30	1019.85	3488.67	1279.32
D	1817.14	1817.14	3367.96	1008.34	3399.52	2762.84

Table 7. Summary of Final output of Demands (m^3) in Each Time Period at the End of Level-2

grade	$t = 1$	$t = 2$	$t = 3$	demands met	demands specified
A	7501.41	6751.269	6505.809	20758.49	22140
B	8953.56	7032.44	6731	22717	22717
C	1716.5	1019.845	1279.319	4015.66	6225
D	1817.14	1008.337	2762.836	5588.31	6225

Finally for the last period ($t = 3$) we need not solve Level-1 as we can directly solve Level-2. All the above runs are solved for nominal schedule to find the best possible demands that can be met by Level-2 in the presence of real slopping losses and inventory constraints. Finally, the output of Level-2 is solved sequentially for each period and is shown in Table 6, and the final results are consolidated in Table 7.

The gantt chart schedule and the inventory profiles for the Level-2 solution in the first time period are shown in Figure 6 and Figure 7, respectively. The dark bands in all the gantt charts shown in this paper denote the transition times between adjacent time slots.The cycle time is 72.44 h. It is found that 74% of grade “B” is charged into line 1 ($Sf_{B1} = 0.74$).

The gantt chart schedule and the inventory profiles for the Level-2 solution in the second time period are shown in Figure 8 and Figure 9, respectively. The cycle

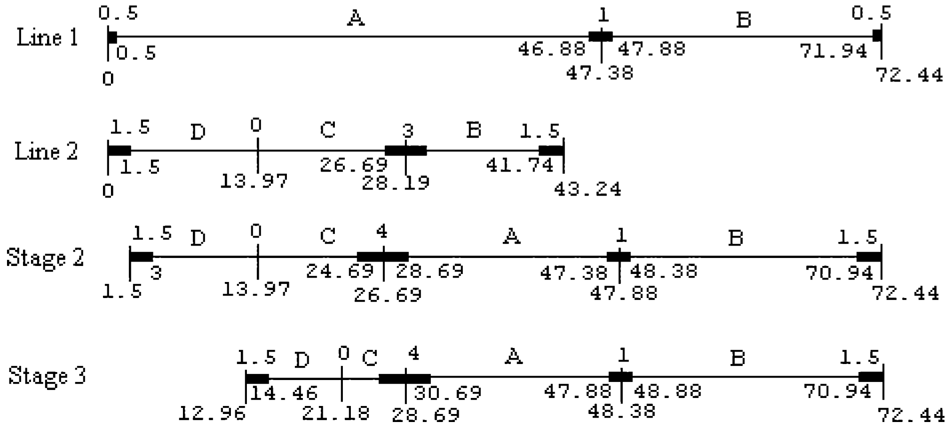

Figure 6. Gantt chart schedule at Level-2 for the first time period.

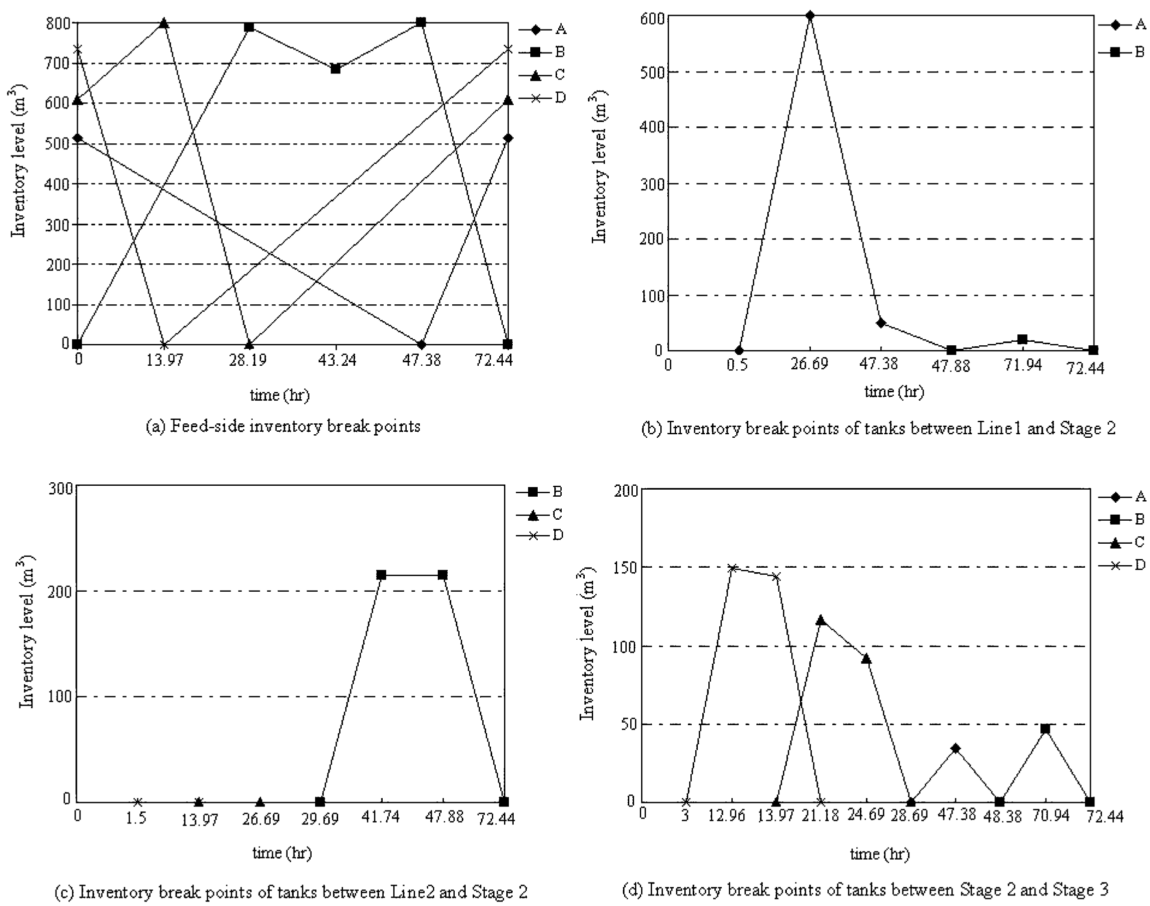

Figure 7. Inventory profiles at Level-2 in the first time period.

time is 83.42 h. It is found that 63.5% of grade "B" is charged into line 1 ($Sf_{B1} = 0.635$).

Similarly, the gantt chart schedule and the inventory profiles for the Level-2 solution in the third time period are shown in Figure 10 and Figure 11, respectively. The cycle time is 75.81 h. It is found that 58.5% of grade "B" is charged into line 1 ($Sf_{B1} = 0.585$).

3.5. Level-3 Model Formulation. Here, it is assumed that the product sequencing and the total inventory profiles are given as input from Level-2 and it is required to figure out at Level-3 if this volume can be met from the set of available tanks.

We first focus on the triangular inventory breakpoints between stage 2 and stage 3 in the first time period as shown by Figure 7d. We know that if these profiles do not overlap in the time frame, then the same tanks can be used repeatedly. For example the tanks that are used for storing either grade D or C can be used again for storing grades A and B. We exploit this feature in the proposed heuristic algorithm and make an efficient usage of the nonoverlapping profiles. However, since the feed is a continuous stream received from upstream plants, the feed inventory tanks would almost always be busy as shown in Figure 7a, thus rendering less probability for reuse of these tanks elsewhere.

Note that the inverted triangular profiles for most of the grades in Figure 7a continue to be in use as some feed is stored for use in the next cycle, unlike the rest that are traditional triangular profiles which if freed can be used elsewhere. As already mentioned earlier,

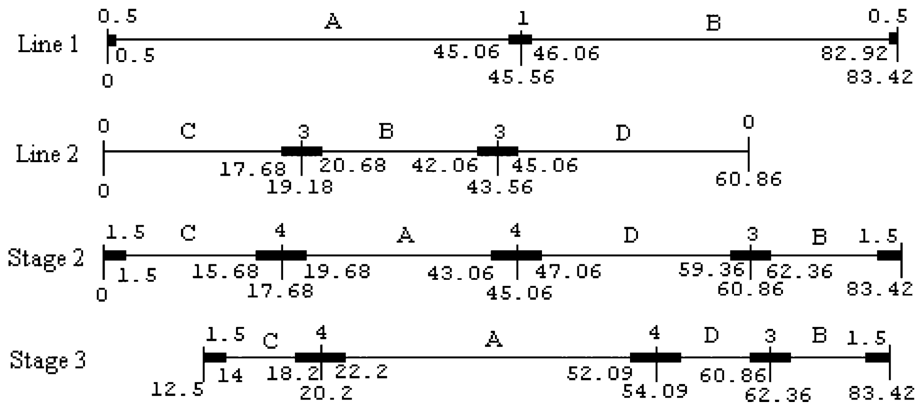

Figure 8. Gantt chart schedule at Level-2 for the second time period.

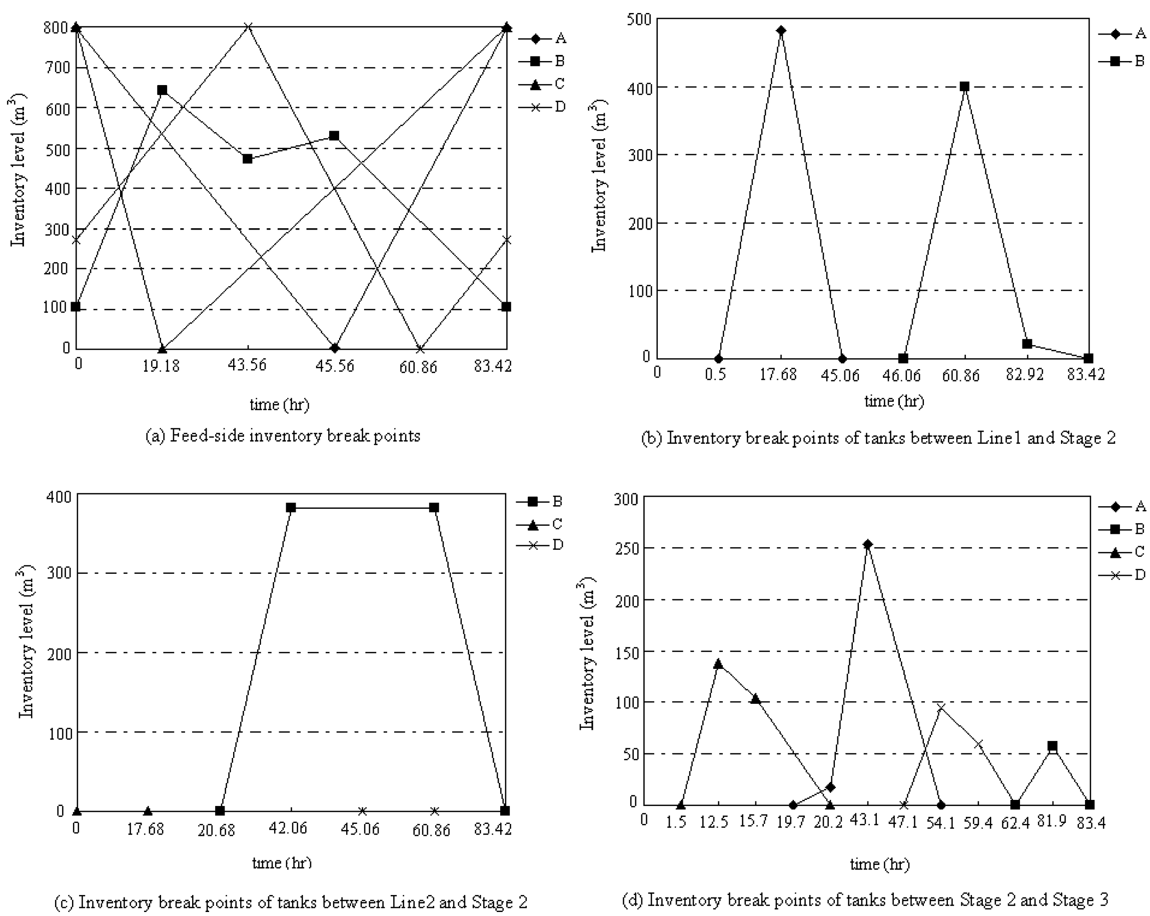

Figure 9. Inventory profiles at Level-2 in the second time period.

from the reactive scheduling point of view the objective for the operator at this level is to minimize the available inventory usage to the best extent possible.

Let us pose the problem here as one of finding the minimum number of tanks of each (say) 50 m^3 capacity required to store these grades. For each profile we first generate "subprofiles" by demarcating the 50 m^3 tank capacities as shown in Figure 7d and find the corresponding timings on the x-axis by linear interpolation. For example, for the profile of grade D, we get three subprofiles of the triangle/trapezium shape stacked on each other as seen Figure 7d. Assuming that we use three tanks T1−T3 each of volume equal to 50 m^3, then for each of these profiles, we know the exact times at which each of these tanks would be occupied and when they would be free for reuse. For example T3 would be empty at the end of 16.18 h, T2 at the end of 18.68 h, and T1 at the end of 21.18 h. Now consider the pool of all such subprofiles across all grades and look for which of these subprofiles do not overlap in the time frame and try to reuse these tanks. Using this heuristic for the inventory profiles between stage 2 and stage 3, it can be found that a minimum of four tanks of each 50 m^3 size are needed. Without reuse of tanks eight such tanks would be needed.

When we consider many stages with numerous inventory profiles, it is difficult to visualize and apply this heuristic manually. With this motive, a simple heuristic algorithm termed as inventory slicing and tank reallocation (ISTR) is proposed here which automates the generation of sliced profiles, checks for the nonoverlapping zones (subprofiles), and finds the minimum number of tanks of a given capacity (say 50 m^3 or 100 m^3)

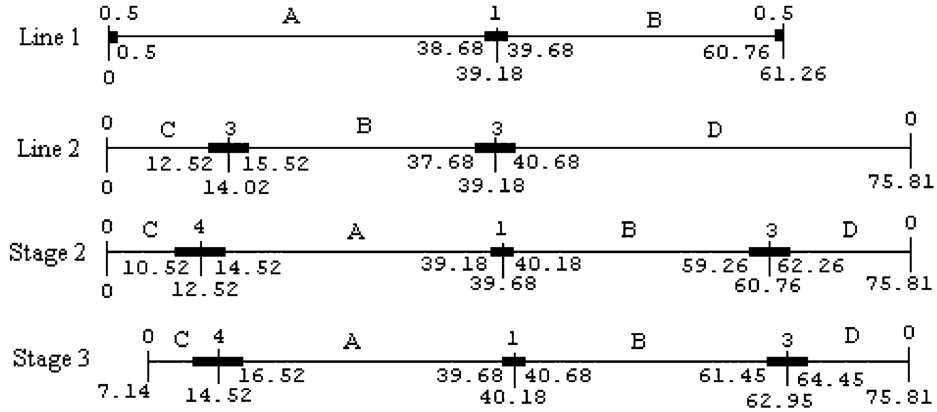

Figure 10. Gantt chart schedule at Level-2 for the third time period.

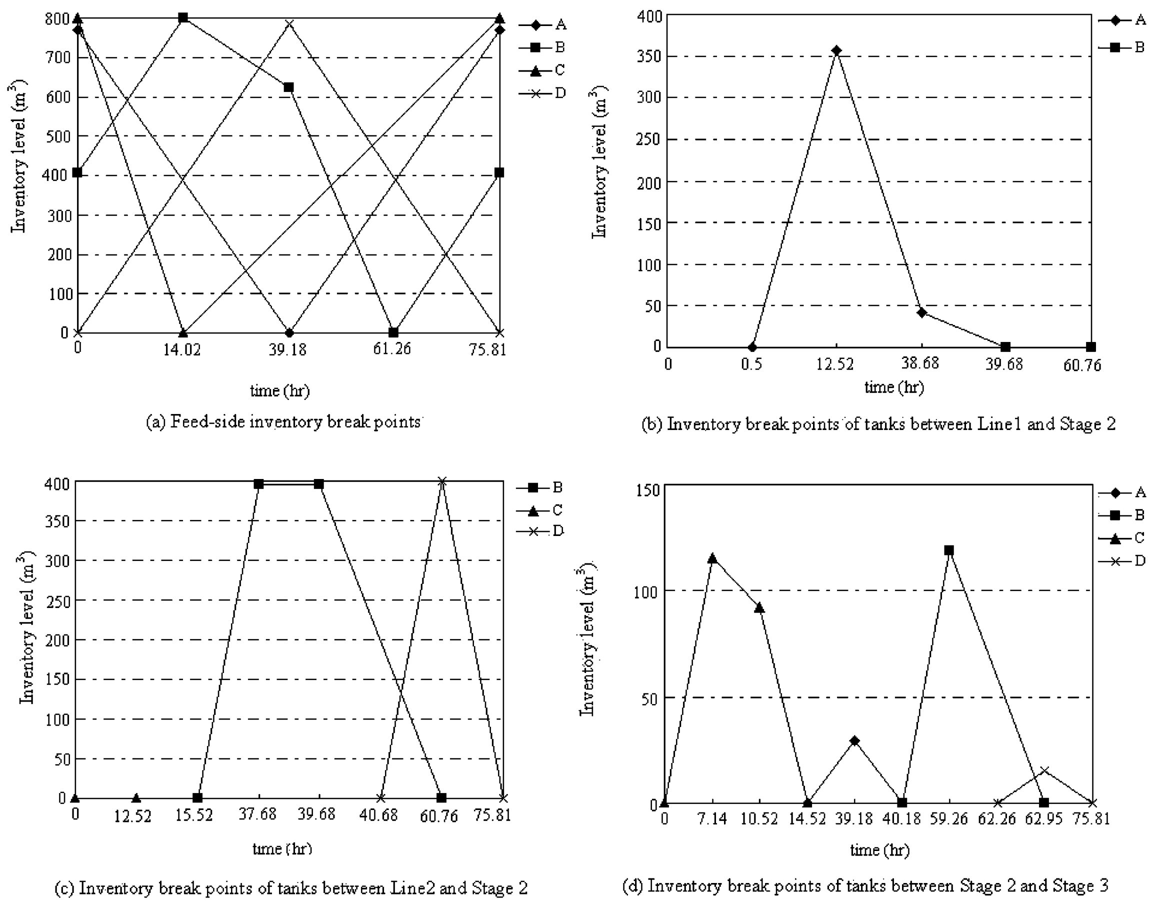

Figure 11. Inventory profiles at Level-2 in the third time period.

required to manage these inventories efficiently at Level-3.

3.5.1. ISTR Algorithm for Triangular Profiles. We first focus on the triangular inventory breakpoints between stage 2 and stage 3 of the first time period as shown in Table 8. In this algorithm for each profile we first generate subprofiles of the given tank capacity, i.e. generate time vs volume data (V_{it}) for each grade. Then slice the given profiles at every multiple of (say) 50 m^3 tank capacities and obtain the corresponding time data by linear interpolation as given in Table 9.

In Table 9 the data is modified by removing the redundant breakpoints, which are not important for tank assignments, for example entries such as 92 m^3 at 24.69 h for grade C and 144.06 m^3 at 13.97 h for grade D. The new data is shown in Table 10. It is to be noted that in Table 10 all the inventory breakpoints are now multiples of the given tank capacity 50 m^3, and get repeated gradewise, except the maximum breakpoint.

Table 8. Inventory Data (m^3) between Stage 2 and 3 from Level-2 in the First Time Period

time (h)	A	B	C	D
0				
3				0
12.96				149.15
13.97			0	144.06
21.18			116.17	0
24.69			92	
28.69	0		0	
47.38	34.75			
48.38	0	0		
70.94		46.42		
72.44		0		

Table 9. Inventory Data (m^3) between Stages 2 and 3 from Level-2 in the First Time Period after Slicing

time (h)	A	B	C	D
0				
3				0
6.34				50
9.68				100
12.96				149.15
13.97			0	144.06
16.18				100
17.07			50	
18.68				50
20.18			100	
21.18			116.17	0
23.53			100	
24.69			92	
26.52			50	
28.69	0		0	
47.38	34.75			
48.38	0	0		
70.94		46.42		
72.44		0		

Table 10. Modified Inventory Data (m^3) after Slicing

time (h)	A	B	C	D
0				
3				0
6.34				50
9.68				100
12.96				149.15
13.97			0	
16.18				100
17.07			50	
18.68				50
20.18			100	
21.18			116.17	0
23.53			100	
26.52			50	
28.69	0		0	
47.38	34.75			
48.38	0	0		
70.94		46.42		
72.44		0		

For example for grade D the entries 0, 50, and 100 all get repeated again after crossing the maximum breakpoint. We shall exploit this feature later in our algorithm. We know that the first occurrence of these entries corresponds to the time at which a tank needs to be deployed and the repeated occurrence corresponds to the time at which such tank would be freed. The remaining entries in Table 10 may be filled with some negative identifier such as −1 (not shown in Table 10 for enhanced clarity) to render those entries to be irrelevant from the algorithm viewpoint. The sliced inventory profiles after removal of the redundant breakpoints are now as shown in Figure 12.

This completes inventory slicing. Now for tank assignments and reallocations we start with Table 10. Starting at $t = 0$ and at each time instant (t), for each grade (i), we check for nonnegative entries (breakpoints, V_{it}) and start deploying new tanks (T1, T2, etc. as shown in Table 11) until the maximum breakpoint (say $V_i^{\max}$) is encountered. For the entry corresponding to a maximum breakpoint assign the same tank as was used for the previous nonnegative V_{it} of the same grade (indicating the same tank has been still in use). Now after crossing the maximum breakpoint the inventory profiles would show a decline. Therefore, before assigning a new tank at each time instant, we additionally check for each grade if the same V_{it} entry already exists in the previous time instances. This would correspond to the release of the inventory that was assigned at the latest time. After finding the first occurrence (in the backward search from the current time instant) of such an entry, we assign the same tank number as was used earlier in the previous entry at which the same V_{it} was found (indicating the same tank has been in use till now).

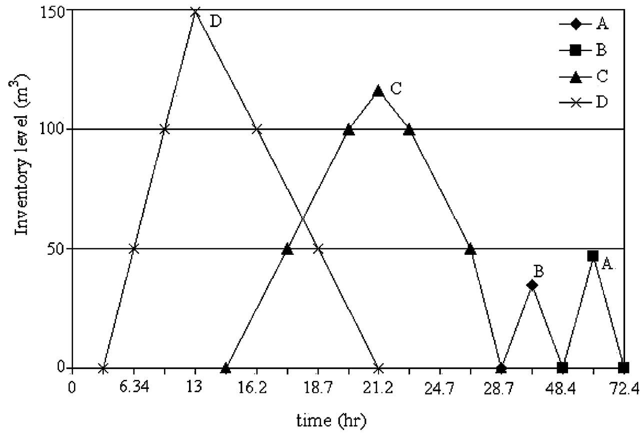

Figure 12. Sliced inventory profiles between stages 2 and 3 from Level-2.

Table 11. Tank Assignments for Four Profiles between Stages 2 and 3 Using ISTR

time (h)	A	B	C	D
0				
3				T1 (0)
6.34				T2 (50)
9.68				T3 (100)
12.96				T3 (149.15)
13.97			T4 (0)	
16.18				T3 (100) F
17.07			T3 (50)	
18.68				T2 (50) F
20.18			T2 (100)	
21.18			T2 (116.17)	T1 (0) F
23.53			T2 (100) F	
26.52			T3 (50) F	
28.69	T3 (0)		T4 (0) F	
47.38	T3 (34.75)			
48.38	T3 (0) F	T3 (0)		
70.94		T3 (46.42)		
72.44		T3 (0) F		

At this point the status of this tank number is marked as freed and available from this time instant onward. The freed tank status is marked as “F” in Table 11. Now, for the subsequent time instances, before assigning new tanks we also check if some freed tanks are available and if so we use them and remove the corresponding entry from the available tank list. If the list of freed tanks is empty, then only a new tank is deployed. In Table 11, for each grade, the time slot of a deployed tank is the time from the first occurrence of that tank number in the table to the time at which the tank is marked as “F”. For example T2 is deployed to store grade B from 6.34 h to 18.68 h. The tank assignments are depicted in Figure 13.

To find the minimum number of tanks required ($N_{\min}$), an index (initialized to zero) is kept incremented, every time a new tank is assigned. In this case $N_{\min} = 4$, totaling 200 m^3 volume required.

The following are certain remarks regarding the proposed ISTR algorithm:

Remark 1: Consider the time instant 28.69 h in Table 10. It can be seen that though T4 is also freed at the same instant we could not deploy that tank for grade A, because normally when the proposed algorithm is

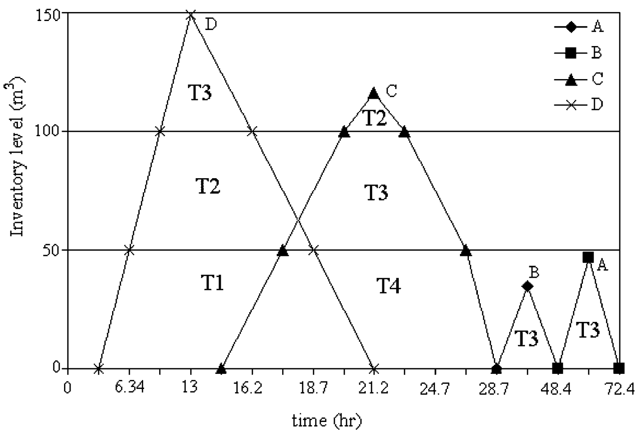

Figure 13. Tank assignments using ISTR algorithm.

implemented, at each time instant the tank assignments are made either from grade A to D or grade D to A. But depending on this sequence it may so happen that new tanks are deployed for some grades even though some tanks are getting freed at the same instant. These cases normally occur when there is more than one breakpoint at a given time instant. However, practically it may/may not be possible that a grade starts processing exactly at the same time when another grade has finished processing. These cases may be avoided by preprocessing the data before applying the ISTR algorithm. If the end breakpoint of any grade matches the timing of the first breakpoint of some other grade, we can slightly increase/decrease the timing of the latter to allow/disallow the usage of the same tank. If we want to continue the usage of the same T4, which is freed at 28.69 h for grades A and B also, then we can remove the overlap by slightly increasing the start time of grade A to 28.7 from 28.69 h. Otherwise, if we do not want to use the same tank due to practical problems, then we can make these two subprofiles overlap, by slightly decreasing the first breakpoint of grade A to 28.68 from 28.69 h. These changes can be reverted back after applying ISTR.

Remark 2: It must be noted that when more than one tank is available in the freed tank list, theoretically any tank can be deployed. For example, for grade A at 28.69 h we can deploy any of the four tanks available. While there exist several alternatives to choose the most appropriate tank, from the reactive scheduling point of view, we propose to choose the most recently freed up tank so that it continues to be in use again and again thus yielding almost a continuous usage profile for that tank. The tanks which got freed earlier would anyway have discontinuity in usage profile, and we let this discontinuity grow up thus yielding a higher probability of availability for reusage for storing other grades. As contingency measures for tank breakdowns during reactive scheduling, these discontinuities in the usage profiles would be helpful.

The proposed heuristic can be easily realized by storing the freed up tanks as a “stack” data structure, as the first to enter the stack would be the last to leave and vice versa. The standard stack operators like “push” and “pop” can be used to insert and delete tanks respectively from the stack of freed tanks. For comparison of with and without a stack arrangement for freed up tanks, consider the tank assignments in Figure 14 for the data in Table 10. If we continue to use the same T4 for grades A and B as well at 28.69 h by preprocessing the data, then the tank assignments would be as in Figure 14a for the case of stack arrangement of freed up tanks. Otherwise if we use a “queue” arrangement of the freed up tanks, where the first in would be the first out, then the tanks assignments would be as in Figure 14b. We can clearly see that the usage profile of T4 is continuous in stack arrangement while in the other case there are many discontinuities. Hence, from the reactive scheduling point of view, stack arrangement would yield a higher probability of tank availabilities, as all other tanks are free after 26.52 h, except T4. A flowchart of the ISTR algorithm is depicted in Figure 15.

Remark 3: If the start and end times of stages 2 and 3 match by chance for any grade, then the Level-2 problem may sometimes yield all inventory breakpoints to be at zero level for that grade (for example, grades C and D between line 2 and stage 2). Nevertheless, in such cases also practically we need to deploy one tank to allow the grade to pass through the tank. This case can be taken care of by deploying a tank whenever $V_i^{\max}$ hits zero for any grade. Similarly the case where $V_i^{\max}$ remains stagnant at some nonzero level also is taken care of by the proposed algorithm.

3.5.2. Amenability for Reactive Scheduling at Level-3. Consider the horizon at Level-3 as one cycle, 72.44 h (approximately 3 days) and the tank assignments as suggested by ISTR algorithm. From the viewpoint of reactive scheduling, if there are any tank breakdowns in the last 2 days of the horizon, then they can be mostly taken care of locally in a cascade-control fashion, without sending any feedback to Level-2, as most of the tanks are free in the last two-thirds of the horizon because of the way the algorithm works. If some tanks are unavailable at Level-3, then there are two ways of rejecting this disturbance. The first is a local attenuation through the proactive measure; due to the stack arrangement of the freed up tanks in the ISTR algorithm (discussed earlier under Remark 2), the resulting discontinuities in the usage profile of the tanks allow reuse of such tanks in the event of other tank breakdowns. The second way of rejecting the disturbance is through sending feedback to Level-2; if none of the tanks can be reused at Level-3, then the feedback is sent to Level-2. The proposed time-intrusion reactive scheduling model at Level-2 is now triggered considering the duration of the tank unavailability as an equivalent shutdown time, and the revised model is obtained for both Level-2 and Level-3. In the latter case above it is still a local attenuation at Level-2 without sending feedback all the way up to Level-1.

The analogy in process control applications is clear here in terms of a compromise between aggressive but nonresilient control and robust but relatively less aggressive control. Since we know that the tanks are anyway going to be reused at Level-3, we could always specify a “good” overestimate for the upper bounds on inventories at Level-2 so that it gives a feasible tank assignment at Level-3.

The output of the ISTR algorithm for the nine triangular profiles of Figure 7 (b, c, d) is shown in Table 12.

With respect to Remark 1 made earlier, no change in inventory breakpoints is made here. Grades 1 and 2 correspond to the two inventory profiles between line 1 and stage 2, grades 3−5 correspond to the three profiles between line 2 and stage 2, and grades 6−9 correspond

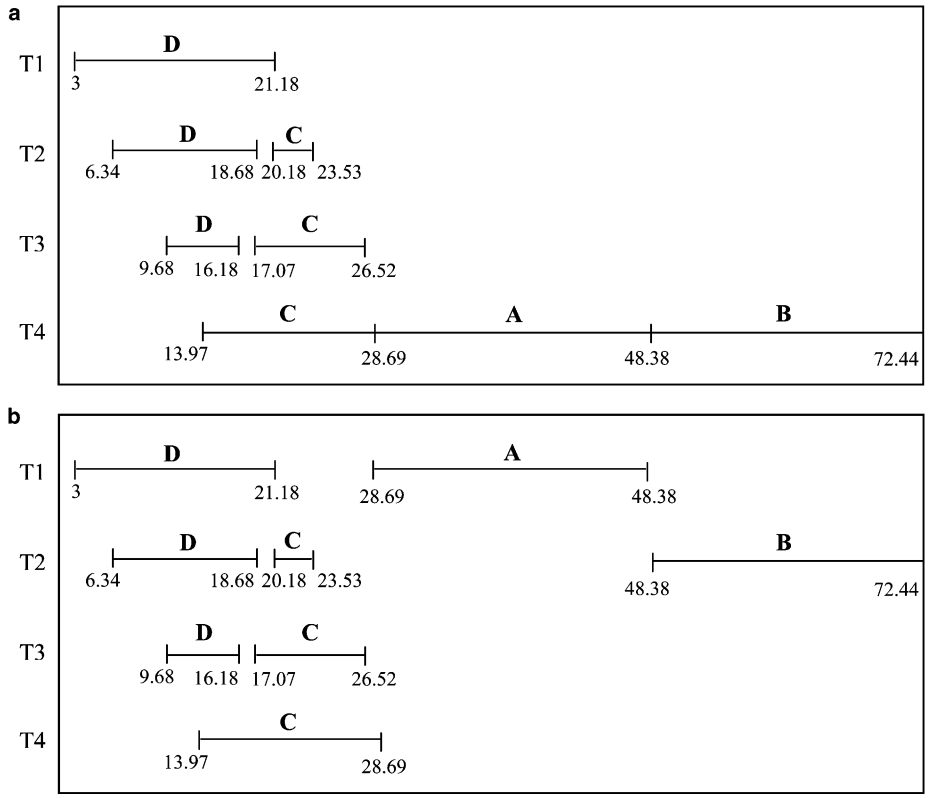

Figure 14. Comparison of tank assignments for (a) stack and (b) queue arrangements of freed up tanks.

to the four profiles between stage 2 and stage 3, respectively. For each grade at each time instant, the entry in Table 12 refers to the tank number followed by a label indicating the tank status followed by the subprofile volume in parentheses. The label S refers to a tank deployed from the stack of freed tanks, label N refers to a new tank deployed, and label F refers to the case of a tank freed at that time instant. $N_{\min}$ is found to be 15 tanks of each 50 m^3 capacity for this case totaling 750 m^3 volume requirements. The inventory upper bounds used at Level-2 for these profiles were 600 m^3, 400 m^3, and 300 m^3 respectively for each grade for the tanks after line 1, line 2, and stage 2, totaling 1500 m^3. The upper bound specifications on inventory at Level-2 were overspecified, since we anyway knew that the tanks would be reused at Level-3 and hence the actual requirements would be less. So the actual requirement here is only 750 m^3. For 100 m^3 tank capacities $N_{\min}$ is found to be 9, totaling 900 m^3 volume requirements. It is evident that the smaller the tank capacities used for slicing, the better would be the probability of reuse of tanks and hence the lower would be the total volume requirements. So the tank capacities would be an interesting parameter that can be considered from the viewpoint of integration of design and scheduling.

The Level-3 problem involves first use and reuse of tanks during different periods of operation. There are also issues related to compatibility of tanks with various intermediates and products (not presented in the paper due to brevity). The motivation behind the proposed ISTR algorithm, which is heuristic rather than numerical optimization based, stemmed from (i) the availability of such heuristics at the lower levels and (ii) relative difficulty of solving the combinatorially complex, mathematical optimization problem at the lower level.

The iterative procedure of obtaining the realistic targets that can be specified by Level-1 for the nominal plan of Level-2 (as discussed in Tables 4 to 6) also applies between Level-2 and Level-3. As discussed earlier, at Level-2 we specify an overestimate on the inventory upper bounds keeping in mind that the tanks are anyway being reused. If there are not sufficient tank volumes available at Level-3 to meet the target inventories specified by Level-2, then we appropriately reduce the assumed inventory upper bounds that were specified at Level-2 and then re-solve Level-2 until we obtain realistic targets.

The proposed ISTR algorithm also has been extended for the case of inverted triangular profiles as well, the details of which are beyond the scope of this paper. This algorithm can be explored for other applications as well that demand efficient management of *shared* resources over the scheduling horizon.

Using the ISTR algorithm for the total of thirteen profiles across all lines/stages of Figure 7, the minimum number of tanks of 50 m^3 capacity was found to be 50 (totaling 2500 m^3) in the first period and 57 and 50 tanks respectively in the second and third periods. Thus if all the tanks of the stages and lines are interconnected (13 profile case), then the inventory requirement reduces to 2500 m^3 from an otherwise volume requirement of 3200 m^3 (2150 + 600 + 250 + 200).

Though for simplicity we have presented the proposed ISTR algorithm for tanks of equal capacities, it may be noted that the algorithm can be easily extended for product assignments to tanks of unequal capacities as well. In such cases, in the ISTR algorithm the inventory slicing for generation of subprofiles needs to be done with respect to the greatest common factor (GCF) of all the available tank capacities. And while assigning tanks we additionally need to keep track of the capacities of

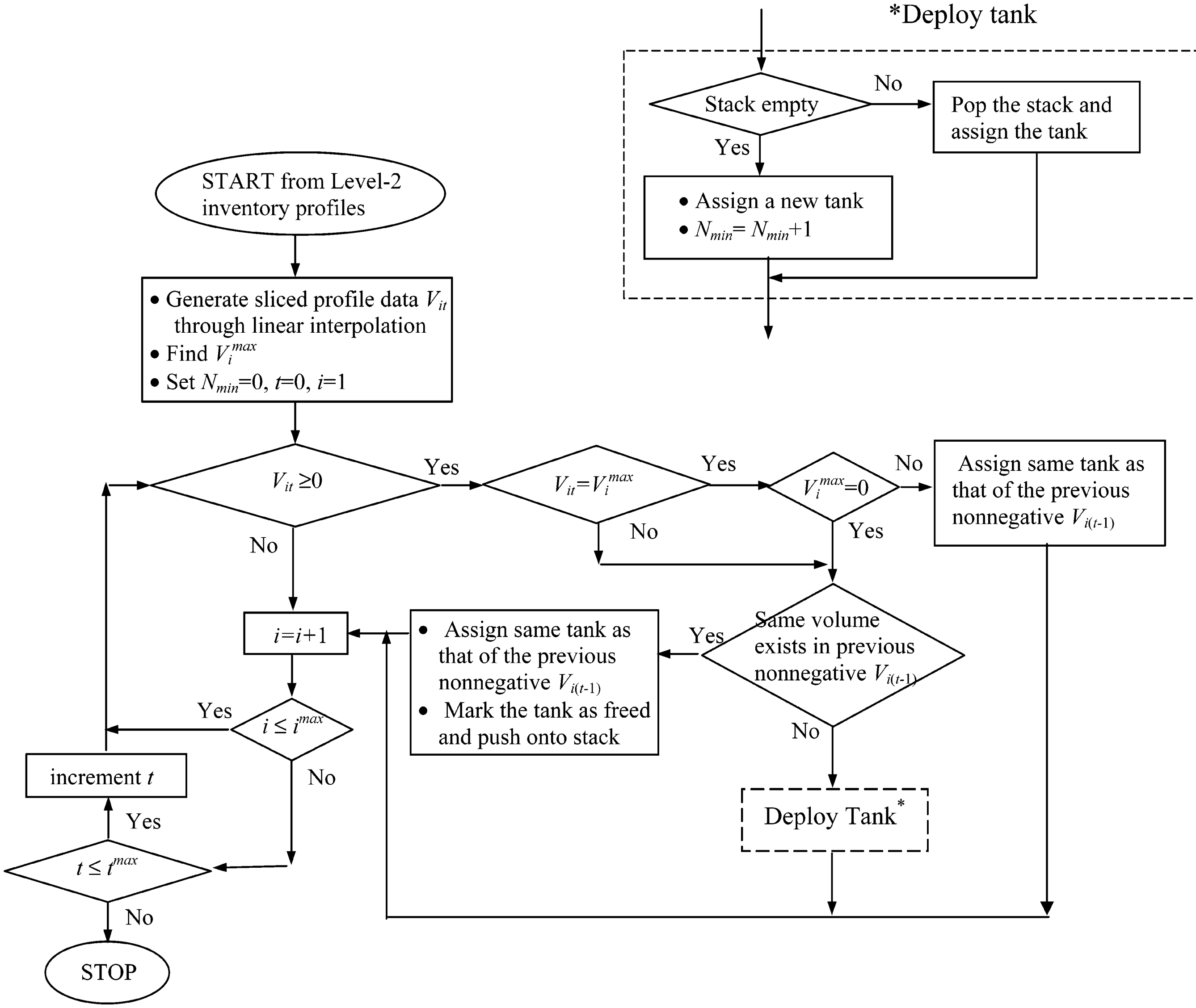

Figure 15. Flowchart of ISTR algorithm for triangular profiles.

the tanks assigned. If more than one tank is available for assignment during the execution of the algorithm, then we may choose the tank of smallest capacity available so that the tanks of larger size are kept for later usage.

The models at Level-1 and Level-2 in this paper were solved on the GAMS[17] platform. The problem at Level-1 took 108 CPU s, while the Level-2 problem in the first time period took 266.5 CPU s using the SBB solver (with SNOPT for NLP) in GAMS on a Pentium IV 1.6 GHz machine with 256 MB RAM. The C-program for the Level-3 model for all 13 inventory profiles is solved in less than 1 CPU s using the ISTR algorithm. No initial guesses were provided at any level, and also there were no convergence problems observed in obtaining the solution to these problems.

4. Reactive Scheduling Model for Lube Production Case Study

In this section, the formulation for the proposed integrated reactive scheduling model is discussed for the lube production case study described above. On the basis of the control-theoretic concepts of cascade-control and receding horizon, this approach can effectively handle reactive scheduling scenarios like equipment failure, unexpected maintenance related shutdowns, etc. in a local fashion.

4.1. Reactive Scheduling Formulation for Unexpected Breakdowns at the Lower Level. In this study we propose a control-theoretic approach to reactive scheduling with a focus on continuous multiproduct, multistage plant scheduling.

4.1.1. The Receding Horizon Framework at the Lower Level. Let us consider the model for reactive scheduling and analyze the interactions between two levels. We simulate some scenarios of machine breakdowns leading to loss of available production time in a given time period. A receding horizon time window[18] is used as discussed earlier, and we find the amount of time required to be intruded into the next time period for compensating the slack in the current period. The proposed model is more generalized as it has provision for empty slots leading to zero time duration and simplified slopping representations at the upper level.

In order to handle reactive scheduling with unforeseen machine breakdowns, at the lower level, we have a receding horizon window, which represents the time over which the overall demands are considered. The overall demands at the lower level include the unsatis-

Table 12. Tank Assignments for the Nine Triangular Profiles (for 50 m^3 Tank Capacities) in the First Time Period

time (h)	grade 1	grade 2	grade 3	grade 4	grade 5	grade 6	grade 7	grade 8	grade 9
0.5	1 N(0)								
1.5					2 N(0)				
2.683	3 N(50)								
3									4 N(0)
4.865	5 N(100)								
6.339									6 N(50)
7.048	7 N(150)								
9.23	8 N(200)								
9.678									9 N(100)
11.41	10 N(250)								
12.96									9(149.15)
13.6	11 N(300)								
13.97				12N(0)	2 F(0)			2 S(0)	
15.78	13 N(350)								
16.18									9 F(100)
17.07								9 S(50)	
17.96	14 N(400)								
18.68									6 F(50)
20.14	6 S(450)								
20.18								15 N(100)	
21.18								15(116.17)	4 F(0)
22.32	4 S(500)								
23.53								15 F(100)	
24.51	15 S(550)								
24.69									
26.52								9 F(50)	
26.69	15 (600)			12F(0)					
28.57	15 F(550)								
28.69						15 S(0)		2 F(0)	
29.69			2 S(0)						
30.45	4 F(500)								
32.33	6 F(450)								
32.49			6 S(50)						
34.21	14 F(400)								
35.29			14 S(100)						
36.09	13 F(350)								
37.98	11 F(300)								
38.1			11 S(150)						
39.86	10 F(250)								
40.9			10 S(200)						
41.74	8 F(200)		10(215.02)						
43.62	7 F(150)								
45.5	5 F(100)								
47.38	3 F(50)					15(34.75)			
47.88	1 F(0)	1 S(0)	10(215.02)						
48.38						15 F(0)	15 S(0)		
49.6			10 F(200)						
55.31			11 F(150)						
61.02			14 F(100)						
66.73			6 F(50)						
70.94							15(46.42)		
71.94		1(20.25)							
72.44		1 F(0)	2 F(0)				15 F(0)		

fied demands (foreseen and unforeseen, including those updated by real time feedback) within the current time period and some projected demands over the intruded period into the next time period as well. The sequencing and the initial inventory levels for the reactive schedule are fixed from the nominal schedule.

The framework proposed here is in line with the cascade control structure that is used in process control practice. Local disturbances/production shortfalls or transients are attenuated locally before they affect the global performance of the loop/schedule. Also, the design of the outer loop (upper level) controller (optimizer) does reflect the inherent characteristics of the inner loop (lower level); for example, bounds on the processing times at the upper level reflect the inventory constraints at the lower level. Also, the receding horizon strategy accommodates production shortfalls locally as much as possible, failing which it enables optimal extension of the due dates so as to meet the production targets. The following is the formulation used to compensate for breakdowns during reactive scheduling at the lower level.

4.1.2. Model Formulation for Reactive Scheduling. Consider a breakdown occurring in the time period t as shown earlier in Figure 3. Let T_{br} be the time at which the breakdown occurs due to which the plant remains shutdown for a period of T_s. If T_s is small, then we may still be able to meet all the demands at the end of time period t itself by operating at higher processing rates. But if T_s is large, then we may not be able to meet all the demands in the same time period t. Then we have two options: the first is to send feedback to the upper level in order to redistribute the demands from that time instant onward by incorporating the slack in the current timer period t; the second option is to intrude x h into the next time period ($t + 1$) at the lower level itself so that the sum of the slack and the corresponding demands of the next time period in the intruded time(x

Table 13. Summary of Production Rates (m^3/h) in Each Time Period at the End of Level-2

$t = 1$	$t = 2$	$t = 3$
7.501	7.501	8.132
8.954	7.814	8.414
1.717	1.133	1.599
1.817	1.120	3.454

h) is also satisfied. We prefer the latter option as we want to attenuate the local disturbances locally, with minimum effect of the disturbance on the next time period and without affecting the global performance as argued earlier. Hence the objective in the reactive formulation of the lower level is to find the minimum amount of time, x, required to be intruded into the next time period to efficiently meet the slack in the current time period. We seek a minimum value of x so that we have some flexibility in the forthcoming due dates to handle any unforeseen breakdowns in the future.

Let T_r be the length of the receding horizon window and Qr_i the overall demands in this horizon. The duration of time period t is T^t. The sequencing and initial inventory levels are fixed as parameters from the nominal schedule of Level-2. The following are the additional constraints required to handle reactive scheduling at the lower level.

$$T_r = (T^t - T_{br} - T_s) + x \qquad (12)$$

$$Qr_i = \frac{Q'^t_i(T^t - T_{br})}{T^t} + \frac{Q'^{(t+1)}_i x}{T^{t+1}} \quad \forall\, i \qquad (13)$$

$$\sum_i \left(Qr_i - \beta_{iM}\, Rp_{iM}\left(\sum_k Tpp_{ikM}\right)\left(\frac{T_r}{T_c}\right)\right) = 0 \qquad (14)$$

In eq 12, the length of the receding horizon is defined as the sum of the remaining time in time period t and the time required to be intruded into the next time period. The overall demand for each product i (Qr_i) in constraint (13) for the receding horizon (T_r) is the sum of the remaining demand in the current time period t and the corresponding demands (linearly interpolated) for the intruded time (x h) of the next time period. In constraint (14), the sum of the unsatisfied demands in the time horizon T_r is forced to be zero. And with minimization of x as the objective function we can find the minimum time required to be intruded into time period $t + 1$ in order to satisfy the overall demand. If we fail to achieve the demand requirement by intruding into the entire next time period as well, then we can include the subsequent time periods also in constraint (13). Otherwise we can choose to send the feedback to the upper level for redistributing the overall demands at the end of time period $t + 1$ in such cases.

4.2. Computational Results of Reactive Scheduling for the Lube Production Case Study. The proposed methodology has been validated on the prototype problem of the hybrid flowshop facility of Figure 5 for lube production as discussed earlier. Consider the same problem data with demands for four products specified over three time periods. The models for all the proposed three levels which need to be implemented during normal operation are solved earlier in section 3. The production rates for nominal schedules in each time period as per the solution of Level-2 are as shown in Table 13. In the following subsections the interactions

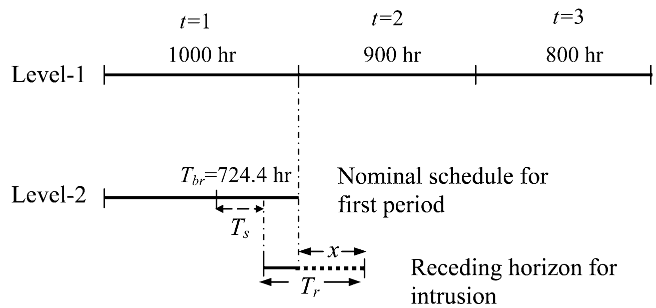

Figure 16. Simulation of shutdown times for reactive scheduling in the first time period.

Table 14. Results for Simulation of Shutdown Times in the First Time Period at Level-2

T_s (h)	x (h)	ϵ (m^3)
0	0	0
1	8.678	0
10	87.8	0
25	222.13	0
50	447.59	0
75	673.89	0
90	809.88	0
99	891.52	0
99.9	899.66	0
100	900	0.795
101	900	12.98

among different levels are discussed from a reactive scheduling perspective.

4.2.1. Simulation of Machine Breakdowns in the First Time Period at Level-2. Now, in this section some unexpected breakdowns are simulated. Consider the target demands for the Level-2 problem in each of the three time periods as shown earlier in Table 7. The demand data for the given problem in this case corresponds to a capacity limited scenario, and hence all of the specified demands could not be met in the three time periods as shown in Table 7. The gantt chart schedules and the inventory profiles for the normal operation in each of the three time periods are shown earlier. In order to simulate the unforeseen breakdown of units, a shutdown time is introduced at the end of 10 cycles (724.4 h) in the first time period ($t = 1$ of 1000 h duration) as shown in Figure 16.

The shutdown times (T_s) are varied from 0 to 100 h. The current levels of inventory at the end of 724.4 h and the processing sequence are fixed as that of the nominal schedule as shown by Figures 6 and 7. The objective function at Level-2 for reactive scheduling is posed as minimization of the amount of time required (x) to be intruded into the next time period ($t = 2$) in order to meet the overall demand (Qr_i) for each product i at end of the receding horizon window (T_r). Expectedly, for a shutdown time of zero duration ($T_s = 0$) the reactive schedule predicted is same as that of the nominal schedule for the remaining time period with zero intrusion. Then the shutdown times are gradually increased from 1 to 100 h as shown in Table 14. The last column in Table 14, denoted by ϵ, shows the unsatisfied demands as per the left-hand side of eq 14. For example, consider a shutdown time of 10 h from 724.4 to 734.4 h with the remaining time in the first time period being 265.6 h. The reactive schedule predicts the length of the receding horizon to be 353.4 h and hence an intrusion of about 87.8 h into the second time period. The production rates in the receding horizon are 7.714, 8.924, 1.62, and 1.695 m^3/h of

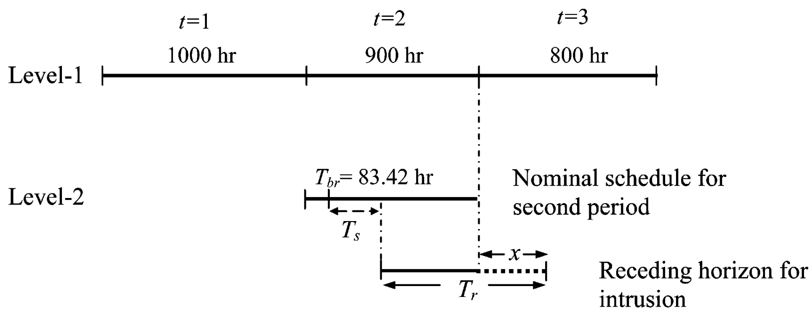

Figure 17. Simulation of shutdown times for reactive scheduling in the second time period.

products A, B, C, and D, respectively. Hence, to compensate for the slack in demands corresponding to this 10 h shutdown time, we need to implement the schedule predicted by the receding horizon with the production rates as given above until the time of 87.8 h of the second time period.

It can be observed from Table 13 that the production rates of the nominal schedule in the second time period are relatively less aggressive than that of the production rates of the reactive schedule given above. In most of the industries the increase in throughputs beyond the nominal values might be allowed only as a contingency measure and that too for small durations keeping in mind the safety and lifetime of the equipment due to operations at high throughputs. Hence, it is generally recommended that at the end of the receding horizon the production rates may be reverted back to the nominal values.

For shutdown times of 1, 10, 50, 75, and 99 h, the receding horizon approach predicted that we need to intrude 8.7, 87.8, 447.6, 673.9, and 891.5 h into the second period, respectively. For shutdown times beyond 100 h, even if all of the second period is included in the receding horizon, the bulk demands could not be met, so either the third period also has to be included in the receding horizon or the slack may be fed back to Level-1 for redistribution of demands. Since it is a capacity limited scenario, Level-1 has already done its best for redistribution of demands for the nominal schedule. Moreover, since the Level-2 model is more accurate than Level-1, it would be better if we host the receding horizon framework at Level-2 itself to minimize the propagation of the disturbance. For $T_s = 100$ h, if we include the third time period also in the receding horizon, then the reactive scheduling model predicts intrusion of 3.35 h into the third time period apart from the entire intrusion of the second time period. Similarly, for $T_s = 101$ h, the reactive scheduling model predicts intrusion of 54.66 h into the third time period. In the next subsection we simulate some shutdown times in the second time period.

4.2.2. Simulation of Machine Breakdowns in the Second Time Period at Level-2. In order to simulate the unforeseen breakdown of units, a shutdown time is introduced at the end of the first cycle (83.42 h) in the second time period ($t = 2$ of 900 h duration) as shown in Figure 17.

The current levels of inventory at the end of 83.42 h and the processing sequence are fixed as those of the nominal schedule as shown by Figures 8 and 9. The objective function at Level-2 for reactive scheduling is posed as minimization of the amount of time required (x) to be intruded into the next time period ($t = 3$) in order to meet the overall demand (Qr_i) for each product i at the end of the receding horizon window (T_r).

Table 15. Results for Simulation of Shutdown Times in the Second Time Period at Level-2

T_s (h)	x (h)	ϵ (m^3)
0	0	0
10	0	0
20	0	0
21	0	0
22	0	0
23	infeasible	0
23	0	8.32[a]
24	0	22.46[a]
27	0	64.99[a]

[a] No intrusions allowed here.

Expectedly, for a shutdown time of zero duration ($T_s = 0$) the reactive schedule predicted is same as that of the nominal schedule for the remaining time period with zero intrusion. Then the shutdown times are gradually increased as shown in Table 15. It can be observed that a maximum shutdown time of 22 h is allowed for the demands to be satisfied within second time period itself, without intruding into the next time period (i.e. $x = 0$ for $\epsilon = 0$ up to $T_s = 22$ h). The reason for this is evident from the outline of production rates in each time period as shown in Table 13. It can be observed from Table 13 that the nominal schedules in the second time period are relatively less aggressive when compared to that of the production rates in the first and last time periods. And hence, there is more flexibility for reactive scheduling in the second time period as the nominal schedule is less aggressive.

If we compare the results in Table 15 with those of Table 14 for reactive scheduling in the first time period, even for a shutdown time of 1 h we need to intrude about 8.7 h into the second time period, since the nominal schedule in the first time period is relatively more aggressive than that of the second period. In contrast, in the second time period up to 22 h of shutdown time is allowed without intruding into the third time period.

The production rates in the third period as per Table 13 are the most aggressive out of the nominal schedules across all time periods, and thus it will be less amenable to reactive scheduling. Hence expectedly, as shown in Table 15, beyond 22 h of shutdown time in the second time period the problem becomes infeasible for $\epsilon = 0$, since these require intrusion into the third time period which the schedule being most aggressive may not permit. This reasoning can be confirmed by relaxing constraint (14) and solving the reactive schedule model for minimization of ϵ subject to $x = 0$. It can be seen that, for $T_s = 23$ h, the minimum ϵ is found to be 8.32 m^3, which could not be accommodated in the third time period as the nominal schedule there is already very aggressive. In such cases when no intrusions are allowed (meaning the effect of the disturbance could not be attenuated locally), we anyway need to send feedback to the upper level thus seeking a revision of the target demands for the subsequent time periods.

5. Conclusion

An integrated multilevel, control-theoretic framework has been proposed in this work for integration of planning, scheduling, and rescheduling. In the multilevel decomposition of the overall problem of integration of planning and scheduling, proactiveness has been embedded for amenability to reactive scheduling, and it has been demonstrated that the lower levels with accurate model can attenuate the disturbances in a local fashion better than the upper levels. The proposed methodology has been demonstrated for cyclic scheduling of lube production in a hybrid flowshop facility.

Literature Cited

(1) Shobrys, D. E.; White, D. C. Planning, scheduling and control systems: why can they not work together. *Comput. Chem. Eng.* **2000**, *24*, 163.

(2) Shah, N. Single and multisite planning and scheduling: current status and future challenges. *AIChE Symp. Ser.* **1998**, *94* (320), 75.

(3) Grossmann, I. E.; Van den Heever, S. A.; Harjunkoski, I. Discrete optimization methods and their role in the integration of planning and scheduling. *AIChE Symp. Ser*. **2001**, *98* (326), 150.

(4) Jia, Z.; Ierapetritou, M.; Kelly, J. D. Refinery short-term scheduling using continuous time formulation: crude-oil operations. *Ind. Eng. Chem. Res.* **2003**, *42*, 3085.

(5) Van den Heever, S. A.; Grossmann, I. E. A strategy for the integration of production planning and reactive scheduling in the optimization of a hydrogen supply network. *Comput. Chem. Eng.* **2003**, *27*, 1813.

(6) Mignon, D. J.; Honkomp, S. J.; Reklaitis, G. V. A framework for investigating schedule robustness under uncertainty. *Comput. Chem. Eng*. **1995**, *19*, S615.

(7) Rodrigues, M. T. M.; Gimeno, L.; Passos, P. A. S.; Campos, M. D. Reactive scheduling for multipurpose chemical batch plants. *Comput. Chem. Eng*. **1996**, *20*, S1215.

(8) Sanmarti, E.; Huercio, A.; Espuna, A.; Puigjaner, L. A combined scheduling/ reactive scheduling strategy to minimize the effect of process operations uncertainty in batch plants. *Comput. Chem. Eng*. **1996**, *20*, S1263.

(9) Henning, G. P.; Cerda, J. Knowledge based predictive and reactive scheduling in industrial environments. *Comput. Chem. Eng*. **2000**, *24*, 2315.

(10) Sun, J.; Xue, D. A dynamic reactive scheduling mechanism for responding to changes of production orders and manufacturing resources. *Comput. Ind*. **2001**, *46*, 189.

(11) Perea E.; Grossmann, I. E.; Ydstie, E.; Tahmassebi, T. Dynamic modeling and classical control theory for supply chain management. *Comput. Chem. Eng.* **2000**, *24*, 1143.

(12) Vargas-Villamil, F. D.; Rivera, D. E. Multilayer optimization and scheduling using model predictive control: application to reentrant semiconductor manufacturing lines. *Comput. Chem. Eng.* **2000**, *24*, 2009.

(13) Bose, S.; Pekny, J. F. A model predictive framework for planning and scheduling problems: a case study of consumer goods supply chain. *Comput. Chem. Eng.* **2000**, *24*, 329.

(14) Perea-Lopez E.; Ydstie, B. E.; Grossmann, I. E. A model predictive control strategy for supply chain optimization. *Comput. Chem. Eng.* **2003**, *27*, 1201.

(15) Munawar, S. A.; Bhushan, M.; Gudi, R. D.; Belliappa, A. M. Cyclic scheduling of continuous multiproduct plants in a hybrid flowshop facility. *Ind. Eng. Chem. Res*. **2003**, *42*, 5861.

(16) Liao, C. J.; Chen, W. J. Scheduling under machine break down in a continuous process industry. *Comput. Oper. Res.* **2004**, 31, 415.

(17) Brooke, A.; Kendrick, D.; Meeraus, A.; Raman, R. *GAMS—A User's Guide*; GAMS Development Corporation: Washington, D.C., 1998.

(18) Munawar, S. A.; Bhushan, M.; Gudi, R. D.; Belliappa, A. M. A multi-level, control-theoretic approach to reactive scheduling in continuous plants. In *Proceedings of the conference in foundations of computer aided process operations*; Grossmann, I. E., McDonald, C. M., Eds.; CACHE Publications: Florida, 2003; p 397.

Received for review May 3, 2004
Revised manuscript received October 6, 2004
Accepted February 28, 2005

IE040142S