Simultaneous Planning for New Product Development and Batch Manufacturing Facilities

Christos T. Maravelias and Ignacio E. Grossmann* [*]

Department of Chemical Engineering
Carnegie Mellon University
Pittsburgh, PA 15213

ABSTRACT

One of the greatest challenges in highly regulated industries, such as pharmaceuticals and agrochemicals, is the process of selecting, developing and efficiently manufacturing new products that emerge from the discovery phase. This process involves the performance of regulatory tests, such as environmental and safety tests, for the new products, and the plant design for manufacturing the products that pass all tests. In order to systematically address this problem, we consider the simultaneous optimization of resource-constrained scheduling of testing tasks in new product development, and design/planning of batch manufacturing facilities. A multiperiod mixed-integer linear programming (MILP) model that maximizes the expected net present value of multiple projects is proposed. The model takes into account multiple trade-offs and predicts which products should be tested, the detailed test schedules that satisfy resource constraints, design decisions for the process network, and production profiles for the different scenarios defined by the various testing outcomes. In order to solve larger instances of this problem with reasonable computational effort, a heuristic algorithm based on Lagrangean decomposition is proposed. The algorithm exploits the special structure of the problem and computational experience shows that it provides optimal or near optimal solutions, while being significantly faster than the full space method. The application of the model is illustrated with three example problems.

Introduction

A large number of candidate new products in the agricultural and pharmaceutical industry must undergo a set of tests related to safety, efficacy, and environmental impact, in order to obtain certification. Depending on the nature of the products, testing may last up to 5 years, and their scheduling should be made with the goal of minimizing the time to market and the cost of the testing. In order to be competitive, a company must have a number of promising products at different stages of the testing process, and at the same time must be prepared to produce a new product as soon as testing is successfully completed. Since the building of a new plant, the specification and procuring of equipment, and the validation process last more than two years, an investment decision for the manufacturing of the new product(s) must be made well before testing is completed. Furthermore, investment decisions become even more important as the pressures for reducing the costs of pharmaceutical products increase. Clearly, the timing of these decisions depends on the completion time of testing of new product(s), and the size of investment depends on the number of new products and their production levels. Thus, given these complex trade-offs, decisions on the testing of new products, and on the design and planning of manufacturing facilities should be optimized simultaneously.

Although optimization methods for addressing problems separately in new product development[1,2,3,4,5,6,7] and supply chain management[8,9,10,11,12] have been reported in the literature, the integrated problem addressed in this paper does not appear to have been reported before. Currently there are no methods and tools to explicitly address this problem, and in practice, decisions are made on an ad-hoc basis. A new optimization model for addressing this design integration problem is presented in this paper. The proposed model is a large-scale

[*] Author to whom correspondence should be addressed.
E-mail: grossmann@andrew.cmu.edu. Phone: 412-268-2230. Fax: 412-268-7139

MILP that predicts, from a portfolio of potential new products, which products should be tested, the detailed test schedules, design decisions for the process network, and production profiles of existing and new products. Since the integration leads to a very large and hard MILP problem to solve, a solution method based on a heuristic Lagrangean decomposition is also proposed.

Literature Review

Schmidt and Grossmann[3] proposed several optimization models for the optimal scheduling of testing tasks in the new product development process, and Jain and Grossmann[4] extended these models to take into account resource constraints. Honkomp et al.[5] addressed the problem of selecting process development projects from a pool of projects, and scheduling the use of limited resources to maximize the expected return from research and development operations. This problem is similar to the scheduling of testing tasks, as each process development project requires a specific sequence of tasks, each of which has a probability of failure. Subramanian et al.[6] proposed a simulation-based framework for the management of the R&D pipeline, and Blau et al.[7] developed a simulation model for risk management in the new product development process. The focus of these references, however, is the new product development process, and not the design and planning of manufacturing facilities. In most of these references it is assumed, for instance, that there are no capacity limitations, or that the production level of a new product is not affected by the production levels of other products. Furthermore, investment costs are not explicitly included in the calculation of Net Present Value (NPV) of projects. Papageorgiou et al.[10] proposed an optimization-based approach for selecting from a set of candidate products the ones to be commercialized. Their approach includes a capacity planning strategy, but does not account for the testing tasks that need to be performed for certification.

There is a large number of papers on batch design (Reklaitis[13]) that deal with detailed sizing and scheduling. Norton and Grossmann[13] proposed a simplified high level, multiperiod planning investment model for processing networks with dedicated and flexible plants. This model is a multiperiod mixed-integer linear programming model, which for given forecasts of product demands and pricing, maximizes the net present value of processing network's operations and expansion decisions over a long time horizon. In that model, it is assumed that products are known and that all products can be produced immediately. In contrast, in the problem addressed in this paper, from a given portfolio of potential products, only a few will be selected to be tested and produced if they pass all tests successfully.

In this work we develop a new integrated scheduling and planning model that uses as a basis the scheduling model of Jain and Grossmann[4], and the design/planning model of Norton and Grossmann[13]. Jain and Grossmann[4] assumed that all potential products are to be tested. However, in this work one has to decide which products should be tested, since there might not be enough testing facilities, or not all products may be profitable. Therefore, the proposed model must select which products to test. In order to account for this decision we use disjunctive programming. If a product is selected to be tested, all the constraints that describe resource-constrained scheduling are active; if it is not selected, all variables referring to testing of this product are set to zero. In most planning models (e.g. Norton and Grossmann[13]) it is assumed that the level of sales is limited by plant capacity and demand (which is either known or uncertain within a fixed range). In new product development, however, the production of a potential product depends also on whether this product has been selected and has passed all tests. This is a discrete-state type of uncertainty, which has received little attention in the literature (Straub and Grossmann[14]). It is, therefore, necessary to extend planning models to account for discrete uncertainties, which will be done in this paper with the use of scenarios.

Representation

For the representation of testing schedules, we use Gantt charts from the perspective of resources to display the use of each resource over time, and activity-on-node directed graphs to display the optimal sequence of tasks. To illustrate, consider the example of a product X that requires four tests and resources from two categories. Tests 1 and 2 require resource A and tests 3 and 4 require resource B. There are also technological precedence constraints: test 1 must finish before tests 3 and 4 begin. Using the proposed representations, a solution to this problem is presented in Figure 1. T_X is the completion time of testing of product X.

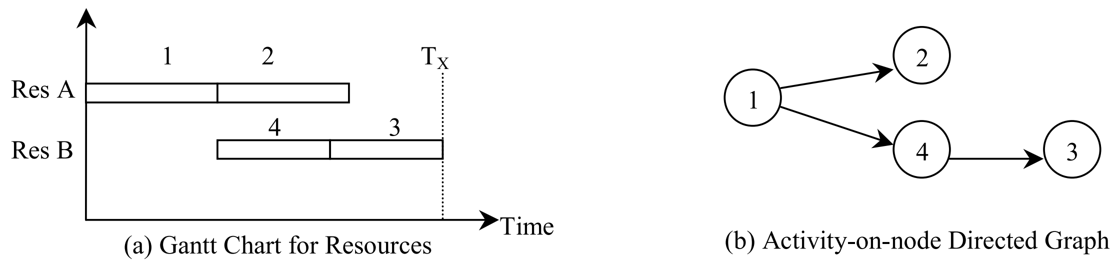


Figure 1: Schedule Representation for Tests

Motivating Example

To illustrate the issues associated with the problem in this paper, consider the example of an agrochemical company that currently sells products A and B, and has two potential products C and D in its R&D pipeline. Product A is broadly produced under constant demand forecasts, product B is gradually phased out due to competition of another firm with a similar product. Product C is a new product intended to replace B, and product D is another new product, which the company predicts will be very profitable. Forecasts of demand levels are given in Table 1. A horizon of 6 years divided into 18 four-month periods is considered.

Table 1: Demand forecasts in tons/month for Example 1

Product	1^{st} yr	2^{nd} yr	3^{rd} yr	4^{th} yr	5^{th} yr	6^{th} yr
A	8	8	8	8	8	8
B	8	8	6	6	4	4
C	-	6	8	10	12	12
D	-	4	4	6	8	8

Every new product must pass successfully two toxicology tests and three field trials. All tests can either be performed within the company (utilizing existing resources), or outsourced at a higher cost. Product C has already passed toxicology and one of the field trial tests, while product D has just been discovered, and no test has yet been performed. The process development of D is anticipated to last much longer than the process development stage of product C (18 months compared to 8 months). Process development can be treated as one test that cannot be outsourced. Cost, duration, probability of success, resource requirement and technological precedences for each test are given in Table 2. Technological precedences are also given schematically in Figure 2.

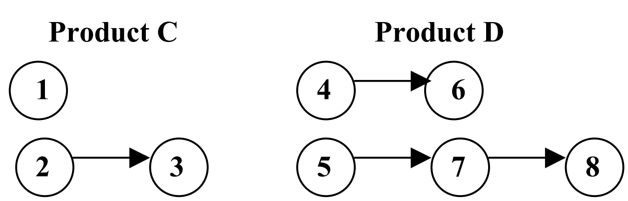


Figure 2: Technological Precedence Constraints of Example 1

Table 2: Testing Data of Example 1

Product	Test	Resource Requirements	Prob. of success	Duration (months)	Cost ($\$10^3$)	Cost of out-sourcing ($\$10^3$)	Technological Precedences
C	PDC	Process Dev Group	1	8	1000	-	-
	1	Toxic Group	1	6	600	1200	-
	2	Field Trial Group	1	4	400	800	-
	3	Field Trial Group	0.95	8	200	400	2
D	PDD	Process Dev Group	1	18	2500	-	-
	4	Toxic Group	0.9	4	800	1600	-
	5	Field Trial Group	0.85	6	600	1200	-
	6	Toxic Group	1	6	500	1000	4
	7	Field Trial Group	1	6	400	800	5
	8	Field Trial Group	0.95	8	1200	2400	5,7

The production network for all products is given in Figure 3. As can be seen for the production of A, raw material RA is converted into intermediate IA1 in unit P1, which in turn is then purified into intermediate IA2 in unit P2, and IA2 is purified into product A in unit P4. The production of products B, C and D is quite similar, the only difference being that the first purification of products B and D takes place in unit P3 instead of unit P2. Note that although process development is not yet completed, the units that will be used for the production of new products are known from preliminary studies. Processes P1 and P4 may operate under four different production schemes (S1, S2, S3, S4), while processes P2 and P3 may operate under two different production schemes (S1, S2). Each scenario involves a different mode of operation (e.g. different inputs, outputs, turnovers). For simplicity, we assume that conversions and operational costs of different production schemes of the same process are equal. Conversion factors, initial capacities, expansion and operational costs for all processes are given in Table 3. Prices of raw materials and final products are given in Table 4. Income and all costs are discounted at a rate of 9% annually.

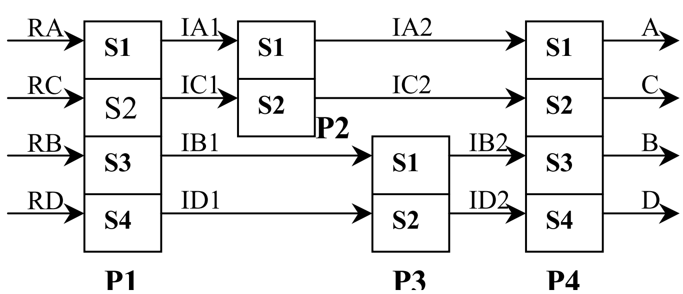


Figure 3: Process Network of Example 1

Table 3: Process Network Data Example 1

Process	Conversion (all schemes)	Initial Capacity (tons/month)	Fixed Cost ($\$10^3$)	Variable Cost ($\$10^3$·month/tons)	Operational Cost (for all schemes) ($\$10^3$/tons)
P1	1/1.4 = 0.7143	20	450	140	1.8
P2	1/1.5 = 0.6666	10	250	60	2.2
P3	1/1.3 = 0.7692	10	240	80	1.4
P4	1/1.2 = 0.8333	16	500	120	1.6

If we try to decide which products to pursue, without considering resources, we might conjecture that both products are profitable. If we now look for the optimal schedule, design and planning policy, assuming that both products C and D are to be tested, and taking resources into account, we find a solution with NPV equal to **$9,118,500**. The Gantt charts of resources for products C and D are presented in Figure 4. Design decisions (expansions) are given in Table 5, and an analysis of costs and revenues is presented in Table 6. As it will be shown later in the paper, this solution is in fact suboptimal.

Table 4: Prices of Raw Materials and Final Products for Example 1

Raw Material	Price ($\$10^3$/ton)	Final Product	Price ($\$10^3$/ton)
RA	3	A	26
RB	3.5	B	28
RC	4.5	C	36
RD	4	D	58

Table 5: Design Decisions of Example 1 (Heuristic): Expansions (ton/month)

Process	1st period	4th period
P1	17.680	-
P2	-	8.88
P3	-	-
P4	-	6.044

Table 6: Revenues and Costs ($\$10^3$) of Example 1 (Heuristic)

Sales Revenues	39,092.7
Testing Cost	6,851.0
Investment Cost	4,760.5
Operational Cost	7,990.2
Raw Material Cost	10,372.5
Expected NPV	**9,118.5**

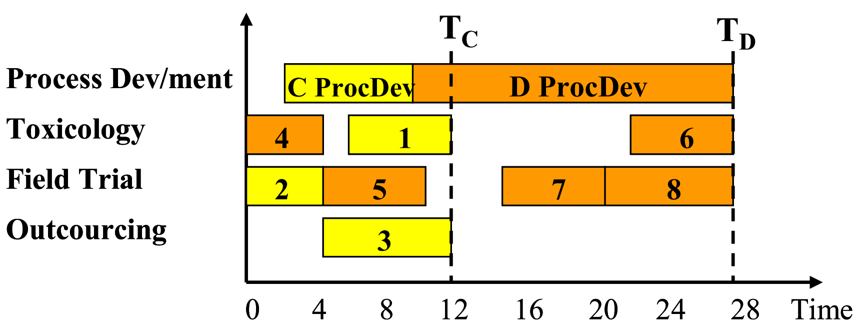


Figure 4: Gantt Chart of Resources of Example 1 - Heuristic

Problem Statement

Given are a set of existing products, and a set of potential products that are in various stages of the company's R&D pipeline. Each potential product is required to pass a series of tests. Failure to pass any of these tests implies termination of the project. Each test has a probability of success, which is assumed to be known, and an associated duration and cost which are known as well. Furthermore, only limited resources are available to complete the testing tasks and they are divided into different resource categories. If needed, a test may be outsourced at a higher cost, and in that case none of the internal resources are used.

On the manufacturing side, a time horizon divided into several time periods is given. A network consisting of existing and new potential plants is considered. Each plant (existing or potential) consists of a set of processes, that can be dedicated or flexible. Flexible processes are typically batch processes that operate under different production schemes, using different inputs and/or producing different outputs. Each process can be expanded only at the beginning of a time period. The network involves a set of chemicals (raw materials, intermediates and final products). The demand for both existing and potential final products is known. Raw materials and final products can be purchased and sold in different markets.

There are three major decisions associated with the testing process. First, the selection of the subset of potential products that are to be tested. Second, the decision about outsourcing and the assignment of resources to testing tasks, and third, the sequencing of tests. The major decisions regarding design and planning are the following: first, the selection of the new plant(s) or the expansion of the existing plant(s)/processes, as well as their timing, to accommodate the new products. Second, the production levels of existing and new products during each time period.

Since the outcome of testing is stochastic, production levels cannot be determined uniquely. If a potential product that has been selected passes all testing tasks, its production can begin. If not, it cannot be produced and thus, there is some unused capacity that can be used for another product. To anticipate the effect of testing outcomes a multiscenario approach is used. If n is the number of new products, there are $|M|=2^n$ scenarios, each corresponding to a different combination of testing outcomes. These scenarios are used in a two-stage stochastic programming[15] approach. In the first stage, here-and-now decisions are made for the testing of new products and the investments needed. Given these decisions, production levels are determined at the second stage in a wait-and-see fashion. Thus, for a specific set of testing and investment decisions, different production levels are determined for each scenario. The objective function for the optimization is to maximize the net present value over a long-range horizon. Income from sales, along with investment, testing, operating and raw material costs are taken into account.

In this paper resource constraints are enforced on exact requirements, and the option of outsourcing is used when existing resources are not sufficient. Thus, the schedule is always feasible, and rescheduling is needed only when a product fails a test. Resources are discrete in nature, and they can handle only one task at a time. Tests are assumed to be non-preemptive, and their probability of success is known a priori. The processing stages of the new products are assumed to be known, and the demands of both existing and potential products are deterministic. For any process type and production scheme, material balances are expressed linearly in terms of the "nominal" production rate of that scheme.

Model

We present the model by first discussing the equations for product selection and resource-constrained scheduling, next for the multiscenario design and planning of the process network, and finally for the integration of the two models.

1. Product Selection and Resource-Constraint Model

The major decision regarding the testing process is the selection of the potential products that will be tested. The indices, sets, parameters and variables used in the scheduling model are the following:

Indices:

k,k' Tests

r Resource Categories

q Resources

n Grid points for linearization of testing cost

j Chemicals

Sets:

J Set of chemicals

JE Set of existing products ($JE \subset J$)

JP Set of new products ($JP \subset J$)

K Set of tests

$K(q)$ Set of tests that can be scheduled on unit q

$K(j)$ Set of tests of potential product $j \in JP$ ($K = \cup_{j \in JP} K(j)$ and $\cap_{j \in JP} K(j) = \emptyset$)

KK(k) Set of tests corresponding to the product that has k as one of its tests
Q Set of resources
R Set of resource categories
QC(r) Set of resources of category r ($Q=\cup_{r\in R} QC(r)$ and $\cap_{r\in R} QC(r)=\varnothing$)
QT(k) Set of resources that can be used for test k
Parameters:
d_k Duration of test k
p_k Probability of success of test k
c_k, $\hat{c}_k$ Costs of test k when performed in-house and when outsourced
N_{rk} Resources of category r needed for test k
α_n Grid points for linear approximation of testing cost
A(k,k') Matrix: $a_{kk'}$ is 1 if test k should be finished before k' starts
ρ Discount factor
U Upper bound on the completion time of testing
Binary Variables:
z_j 1 if product j is tested (defined only for $j\in JP$)
$y_{kk'}$ 1 if test k must finish before k' starts and $k\in K(j)$ and $k'\in K(j)$ for some $j\in JP$
$\hat{y}_{kk'}$ 1 if test k must finish before k' starts and $k\in K(j)$ and $k'\notin K(j)$ for some $j\in JP$
x_k 1 if test k is outsourced
$\hat{x}_{kq}$ 1 if resource q is assigned to test k
Continuous Variables:
C_k Discounted cost of test k
T_j Completion time of testing of potential product $j\in JP$
s_k Starting time of test k
w_k Exponent for discounted cost calculation of test k
λ_{kn} Weight factors for linear approximation of discounting of cost of test k when performed in-house
Λ_{kn} Weight factors for linear approximation of discounting of cost of test k when outsourced

To simplify the presentation we first assume that all potential products are to be tested. For each test k, the most important decisions are its start time (s_k), whether the test should be outsourced or not (x_k), the assignment of resources q to that test ($\hat{x}_{kq}$), and its relative sequence to other tasks ($y_{kk'}$ and $\hat{y}_{kk'}$). The expected cost of completing test k, is determined by an outsourcing decision, and is a function of the probability of conducting the test, and its discounted cost:

$$\begin{bmatrix} \neg x_k \\ C_k = c_k\, e^{-\rho s_k} \prod_{k'\neq k} p_{k'}\, y_{k'k} \end{bmatrix} \vee \begin{bmatrix} x_k \\ C_k = \hat{c}_k\, e^{-\rho s_k} \prod_{k'\neq k} p_{k'}\, y_{k'k} \end{bmatrix} \qquad (1)$$

Disjunction (1) involves nonlinear functions that can be linearized by first using the logarithmic transformation for the product of probabilities, and then a piecewise linear approximation (Schmidt & Grosmann[3]). Thus, applying the convex hull formulation (Balas[16], Turkay and Grossmann[17]) to (1) the cost of performing test k is described by equations (2)-(6):

$$C_k = c_k \left(\sum_n e^{\alpha_n} \lambda_{kn}\right) + \hat{c}_k \left(\sum_n e^{\alpha_n} \Lambda_{kn}\right) \qquad \forall k\in K \qquad (2)$$

$$w_k = -\rho \cdot s_k + \sum_{k'\neq k} \ln(p_{k'})\, y_{k'k} \qquad \forall k\in K \qquad (3)$$

$$w_k = \sum_n \alpha_n \left(\lambda_{kn} + \Lambda_{kn}\right) \qquad \forall k\in K \qquad (4)$$

$$\sum_n \lambda_{kn} = (1 - x_k) \qquad \forall k\in K \qquad (5)$$

$$\sum_n \Lambda_{kn} = x_k \qquad \forall k\in K \qquad (6)$$

Constraints (7)-(9) are used for the sequencing and timing of tests:

$$y_{kk'} = 1,\ y_{k'k} = 0 \qquad \forall (k,k')\in A \qquad (7)$$

$$s_k + d_k \le s_{k'} + U(1 - y_{kk'}) \qquad \forall j\in JP,\ \forall k\in K(j),\ \forall k'\in K(j) | k\neq k' \qquad (8)$$

$$s_k + d_k \le T_j \qquad \forall j\in JP,\ \forall k\in K(j) \qquad (9)$$

Constraints (10)-(12) are logic cuts, which although not necessary, strengthen the LP relaxation:

$$y_{kk'} + y_{k'k} \le 1 \qquad \forall j\in JP\ \forall k,\ k'\in K(j) | k<k' \qquad (10)$$

$$y_{kk'} + y_{k'k''} + y_{k''k} \le 2 \qquad \forall j\in JP\ \forall k,\ k',\ k''\in K(j) | k<k'<k'' \qquad (11)$$

$$y_{k'k} + y_{kk''} + y_{k''k'} \le 2 \qquad \forall j\in JP\ \forall k,\ k',\ k''\in K(j) | k<k'<k'' \qquad (12)$$

Constraint (13) ensures that if a task is not outsourced, then the required number of units from each category are assigned to that test:

$$\sum_{q\in(QT(k)\cap QC(r))} \hat{x}_{kq} = N_{kr}\,(1 - x_k) \qquad \forall k\in K,\ \forall r\in R \qquad (13)$$

In order to enforce resource constraints we use the following logical condition: *if a test k is assigned to resource q, then for resource constraints to hold, any other test k' is either not assigned to resource q or there is an arc between test k and k':*

$$\hat{x}_{kq} \Rightarrow \neg\, \hat{x}_{k'q} \vee y_{kk'} \vee y_{k'k} \qquad \forall q\in Q,\ \forall k\in K(q),\ \forall k'\in (K(q)\cap KK(k)) | k<k' \qquad (14)$$

$$\hat{x}_{kq} \Rightarrow \neg\, \hat{x}_{k'q} \vee \hat{y}_{kk'} \vee \hat{y}_{k'k} \qquad \forall q\in Q,\ \forall k\in K(q),\ \forall k'\in (K(q)\backslash KK(k)) | k<k' \qquad (15)$$

Using the transformations described in Raman and Grossmann[18], we obtain the following constraints:

$$\hat{x}_{kq} + \hat{x}_{k'q} - y_{kk'} - y_{k'k} \le 1 \qquad \forall q\in Q,\ \forall k\in K(q),\ \forall k'\in (K(q)\cap KK(k)) | k<k' \qquad (16)$$

$$\hat{x}_{kq} + \hat{x}_{k'q} - \hat{y}_{kk'} - \hat{y}_{k'k} \le 1 \qquad \forall q\in Q,\ \forall k\in K(q),\ \forall k'\in (K(q)\backslash KK(k)) | k<k' \qquad (17)$$

As shown in Appendix A, constraints (16) and (17) are tighter than the ones proposed by Jain and Grossmann[3]. Finally, constraint (18) is needed to enforce the sequence between two tests that do not correspond to the same product (as constraint (8) is needed for tests of the same product):

$$s_k + d_k - s_{k'} - U(1 - \hat{y}_{kk'}) \le 0 \qquad \forall q\in Q,\ \forall k\in K(q), \forall k'\in K(q)\backslash KK(k) \qquad (18)$$

In the general case, however, one has to decide which products should be tested (z_j), since there might not be enough testing facilities or not all products may be profitable. The proposed model, therefore, should select which products to test. This is equivalent to a disjunction, whose first term corresponds to the case where product $j\in JP$ is tested (z_j=*True*), and the second term to the case where product j is not tested (z_j=*False*). In the first term, all the constraints that describe resource constrained scheduling for product j are active, while in the second term, all variables referring to the testing of product j are set to zero. Thus, constraints (2)-(13), and (16) are included in the first term, while in the second term variables s_k, λ_{kn}, Λ_{kn}, w_k, x_k, $\hat{x}_{kq}$, $y_{kk'}$ and T_j are set to zero. Note that constraints (17) and (18) are not included in this disjunction because they contain variables $\hat{y}_{kk'}$ that are used for the sequencing of tests that correspond to different products.

$$\begin{bmatrix} z_j \\ \\ (2)\text{ - }(13),\ (16) \\ \\ \end{bmatrix} \bigvee \begin{bmatrix} & \neg z_j & \\ T_j = 0, & x_k = 0 \quad \forall k\in K(j), & \hat{x}_{kq} = 0\ \ \forall k\in K(j),\ \forall q\in Q \\ s_k = 0 \quad \forall k\in K(j), & \hat{y}_{kk'} = 0 \ \ \forall k,k'\in K(j), & \lambda_{kn} = 0\ \ \forall k\in K(j),\ \forall n \\ w_k = 0 \quad \forall k\in K(j), & y_{kk'} = 0 \ \ \forall k,k'\in K(j), & \Lambda_{kn} = 0\ \ \forall k\in K(j),\ \forall n \end{bmatrix} \qquad (19)$$

By applying the convex hull formulation[18,19], and eliminating variables and redundant constraints (Appendix B), the disjunction in (19) is expressed as follows:

$$C_{TEST} = \sum_k C_k = \sum_k \{c_k (\sum_n e^{\alpha_n} \lambda_{kn}) + \hat{c}_k (\sum_n e^{\alpha_n} \Lambda_{kn})\} \quad (20)$$

$$w_k = -\rho \cdot s_k + \sum_{k' \neq k} ln(p_{k'})\, y_{k'k} \qquad \forall k \in K \quad (21)$$

$$w_k = \sum_n a_n (\lambda_{kn} + \Lambda_{kn}) \qquad \forall k \in K \quad (22)$$

$$\sum_n \lambda_{kn} = z_j - x_k \qquad \forall j \in JP,\ \forall k \in K(j) \quad (23)$$

$$\sum_n \Lambda_{kn} = x_k \qquad \forall k \in K \quad (24)$$

$$y_{kk'} = z_j,\ y_{k'k} = 0 \qquad \forall j \in JP,\ \forall k,\ k' \in K(j),\ \forall (k,k') \in A \quad (25)$$

$$s_k + d_k z_j - s_{k'} - U (z_j - y_{kk'}) \le 0 \qquad \forall j \in JP,\ \forall k,\ k' \in K(j) \quad (26)$$

$$s_k + d_k z_j \le T_j \qquad \forall j \in JP,\ \forall k \in K(j) \quad (27)$$

$$y_{kk'} + y_{k'k} \le z_j \qquad \forall j \in JP,\ \forall k,\ k' \in K(j) |\ k<k' \quad (28)$$

$$y_{kk'} + y_{k'k''} + y_{k''k} \le 2z_j \qquad \forall j \in JP,\ \forall k,\ k',\ k'' \in K(j) | k<k'<k'' \quad (29)$$

$$y_{k'k} + y_{kk''} + y_{k''k'} \le 2z_j \qquad \forall j \in JP,\ \forall k,\ k',\ k'' \in K(j) | k<k'<k'' \quad (30)$$

$$\sum_{q \in (QT(k) \cap QC(r))} x_{kq} = N_{kr} (z_j - x_k) \qquad \forall j \in JP,\ \forall k \in K(j),\ \forall r \in R \quad (31)$$

$$x_{kq} + x_{k'q} - y_{kk'} - y_{k'k} \le z_j \qquad \forall q \in Q,\ \forall j \in JP,\ \forall k \in (K(q) \cap K(j)),\ \forall k' \in (K(q) \cap KK(k) \cap K(j)) | k<k' \quad (32)$$

$$0 \le T_j \le U \cdot z_j \qquad \forall j \in JP \quad (33)$$

$$T_j,\ s_k,\ \lambda_{kn},\ \Lambda_{kn} \ge 0, \qquad w_k \le 0, \qquad y_{kk'},\ \hat{y}_{kk'},\ x_k,\ \hat{x}_{kq},\ z_j \in \{0,1\} \quad (34)$$

Note that constraints (23) and (25)-(32) are now a function of z_j, whereas in constraints (5), (7)-(13) and (16) this value was set to one. Furthermore, we can impose bounds on the binaries $\hat{y}_{kk'}$. If $k \in K(j)$, and $k' \notin KK(k)$ and product j is not tested then $\hat{y}_{kk'}$ and $\hat{y}_{k'k}$ are zero (and vice versa), which are enforced by:

$$0 \le \hat{y}_{kk'} \le z_j \qquad \forall q \in Q,\ \forall j \in JP,\ \forall k \in K(j) \cap K(q),\ \forall k' \in K(q) \backslash KK(k) \quad (35)$$

$$0 \le \hat{y}_{kk'} \le z_j \qquad \forall q \in Q,\ \forall j \in JP,\ \forall k' \in K(j) \cap K(q),\ \forall k \in K(q) \backslash KK(k) \quad (36)$$

In practice, a company may have a specified policy for some new products, or the sales policy of the company may require that a new product must be launched, for instance, to maintain market share. In other cases, when the company has two new similar products and limited resources, only one may be pursued. Such types of constraints can easily be modeled with the corresponding binary variables z_j. For example:

Product A must be tested: $\quad z_A = 1 \quad$ (37.1)

From a set $JP^* \subseteq JP$ at most N products can be tested: $\quad \sum_{j \in JP^*} z_j \le N \quad$ (37.2)

Product A cannot be tested if B is not tested: $\quad z_A \le z_B \quad$ (37.3)

Thus, the proposed MILP model that describes resource-constrained scheduling with product selection comprises of equations (17), (18), (20) to (36), and possibly some type of constraint in (37).

2. Multiscenario Design/Planning Model

As explained above, in order to address the effect of the stochastic nature of testing, we use a multiscenario design/planning model. The indices, sets, parameters and variables used in this model are the following:

Indices:

t	Time periods
p	Plants
i	Processes
s	Production schemes
j	Chemicals
l	Markets
m	Scenarios

Sets:

T	Set of time periods
P	Set of plants
PN	Set of potential new plants ($PN \subset P$)
I	Set of processes
I'	The subset of processes for which a restriction on the number of expansions applies
$I(p)$	Set of processes of plant p
J	Set of chemicals
$Ij(j)$	Set of processes that consume chemical j
$Oj(j)$	Set of processes that produce chemical j
$PS(i)$	Set of production schemes of process i
$J(i,s)$	Chemicals involved in production scheme s of process i

Parameters:

$HTOT$	Time horizon
P^m	Probability of occurrence of scenario m
H_t	Duration of period t
HT_t	Total time until the end of period t ($HT_t = \sum_{t'=1\ldots t} H_{t'}$)
$Q0_i$	Existing capacity of process i at time $t=0$
QE^L_{it}/QE^U_{it}	Lower/upper bound to the expansion of process i in period t
ρ_{is}	Dimensionless relative production rate coefficient for scheme s of process i
μ_{ijs}	Material balance coefficients of chemical j, scheme s, process i
HP_{it}	Maximum time for which process i is available during period t
aN_{pt}	Fixed cost for building new plant p, in period t
αE_{it}	Fixed cost for expansion of process i in period t
β_{it}	Variable cost for expansion of process i in period t
δ_{ist}	Unit operating cost of process i under scheme s during period t (\$·month/kg)
γ_{jlt}	Price of sale of chemical j in market l during time period t (\$/kg)
φ_{jlt}	Price of purchase of chemical j in market l during time period t (\$/kg)
a^L_{jlt} / a^U_{jlt}	Lower/upper bound of availability of raw material j in market l during period t (kg/month)
d^L_{jlt} / d^U_{jlt}	Lower/upper bound of demand of finished product j in market l during period t (kg/month)
f_{mj}	Demand factor – 1 if product j passes all tests in scenario m
CI_t	Upper bound on capital investment during period t
$NEXP_i$	Maximum allowable number of expansions for process i

Binary Variables:

yE_{it}	1 if process i is expanded at the beginning of period t
yNP_{pt}	1 if new plant p is built at the beginning of period t

Continuous Variables:

Q_{it}	Capacity of process i during period t (kg/month)
QE_{it}	Capacity expansion of process i at period t (kg/month)
θ_{istm}	“Nominal” production of scheme s of process i in period t for scenario m (kg)
W_{ijtm}	Total amount of product j produced in process i during period t for scenario m (kg)
PP_{jltm}	Purchases of product j from market l during period t for scenario m (kg)

SS_{jltm} Sales of product *j* in market *l* during period *t* for scenario *m* (kg)

First, we consider the simple case of a process network of an existing plant that involves only existing products, and thus, does not involve different scenarios. The major decisions are the type and timing of investments (yE_{it}). As a basis, we use the model proposed by Norton and Grossmann[12] (constraints (38)-(50)). Decisions referring to the timing and the size of expansions, are described by constraints (38)-(41), and the cost of investment is calculated by equation (42):

$$yE_{it}QE^{L}_{it} \le QE_{it} \le yE_{it}QE^{U}_{it} \qquad \forall i\in I,\ \forall t\in T \qquad (38)$$

$$Q_{it} = Q_{it-1} + QE_{it} \qquad \forall i\in I,\ \forall t\in T \qquad (39)$$

$$\textstyle\sum_t yE_{it} \le NEXP_i \qquad \forall i\in I' \qquad (40)$$

$$\textstyle\sum_i(\alpha E_{it}\, yE_{it}+\beta_{it}\, QE_{it}) \le CI_t \qquad \forall t\in T \qquad (41)$$

$$C_{INVEST} = \textstyle\sum_i\sum_t (\alpha E_{it}\, yE_{it} + \beta_{it}\, QE_{it}) \qquad (42)$$

The mass balance of chemical *j*, at each time period *t* is expressed by equation (43). Equation (44) determines the amount of chemical *j*, consumed/produced in process *i*, during period *t*. Constraint (45) is used to express capacity limitations, and constraints (46) and (47) to express market conditions:

$$\textstyle\sum_l PP_{jlt} + \sum_{i\in O(j)} W_{ijt} = \sum_l SS_{jlt} + \sum_{i\in I(j)} W_{ijt} \qquad \forall j\in J,\ \forall t\in T \qquad (43)$$

$$W_{ijt} = \textstyle\sum_{s\in PS(i)} \mu_{ijs}\, \rho_{is}\, \theta_{ist} \qquad \forall i\in I,\ \forall j\in J(i,s),\ \forall t\in T \qquad (44)$$

$$\textstyle\sum_{s\in PS(i)} \theta_{ist} \le Q_{it}\, HP_{it} \qquad \forall i\in I,\ \forall t\in T \qquad (45)$$

$$a^{L}_{jlt}\cdot H_t \le PP_{jlt} \le a^{U}_{jlt}\cdot H_t \qquad \forall j\in J,\ \forall l\in L,\ \forall t\in T \qquad (46)$$

$$d^{L}_{jlt}\cdot H_t \le SS_{jlt} \le d^{U}_{jlt}\cdot H_t \qquad \forall j\in J,\ \forall l\in L,\ \forall t\in T \qquad (47)$$

The income from sales, the cost of purchasing raw materials and the operational cost are calculated from equations (48)-(50):

$$Income = \textstyle\sum_j \sum_l \sum_t \gamma_{jlt}\, SS_{jlt} \qquad (48)$$

$$C_{PURCH} = \textstyle\sum_j \sum_l \sum_t \varphi_{jlt}\, PP_{jlt} \qquad (49)$$

$$C_{OPER} = \textstyle\sum_i \sum_{s\in PS(i)} \sum_t (\delta_{ist}\, \rho_{is}\, \theta_{ist}) \qquad (50)$$

Production amounts of chemicals are determined in terms of the nominal production rate θ_{ist}. θ_{ist} is not necessarily equal with the actual production (or the consumption) of a chemical *j*. Process capacity is expressed linearly in terms of the nominal production of a reference scheme of this process. Note that in this model variables θ_{ist}, W_{ijt}, PP_{jlt} *and* SS_{jlt} are not indexed by scenario *m*.

In order to model the alternative of building a new plant for the production of the new product(s), we postulate a superstructure where a set of processes have zero capacity at time $t=0$ and they are only expanded if it has previously been decided to build a new plant. The condition that a process $i\in I(p)$ *can only be expanded after plant p has been built* is enforced by:

$$yE_{it} \le \textstyle\sum_{t'\le t} yNP_{pt'} \qquad \forall p\in PN,\ \forall i\in I(p),\ \forall t\in T \qquad (51)$$

Constraints (40) and (41) are now written as follows:

$$\textstyle\sum_p \alpha N_{pt}\, yNP_{pt} + \sum_i(\alpha E_{it}\, yE_{it}+\beta_{it}\, QE_{it}) \le CI_t \qquad \forall t\in T \qquad (52)$$

$$C_{INVEST} = \textstyle\sum_p\sum_t \alpha N_{pt}\, yNP_{pt} + \sum_i\sum_t (\alpha E_{it}\, yE_{it} + \beta_{it}\, QE_{it}) \qquad (53)$$

The design/planning model consisting of constraints (38)-(39) and (42)-(53) assumes that all final products can be produced. In new product development, however, we do not know a priori whether a new product will pass all tests successfully, and thus, whether its production can begin. We only know its probability of passing all tests. This implies that the income from sales, and the corresponding costs (purchase of raw materials and operational costs) of potential products are stochastic. To account for the stochastic nature of the problem, we use scenarios based on the outcomes of testing. As explained in the Problem Statement, if we have n different potential products, there are $|M|=2^n$ different combinations (scenarios) of testing outcomes and for each combination we can calculate its probability of occurrence. The probability of a product $j \in JP$ to pass all tests is $P_j = \prod_{k \in K(j)} p_k$, where p_k is the probability of test k being successful. In general, if scenario m corresponds to the case where potential products $j_1, j_2, \ldots j_k$ pass and products $j_{k+1}, j_{k+2}, \ldots j_n$ fail, the probability of occurrence of scenario m will be $P^m = Pj_1 \cdot Pj_2 \cdot \ldots \cdot Pj_k \cdot (1-Pj_{k+1}) \cdot (1-Pj_{k+2}) \cdot \ldots \cdot (1-Pj_n)$ with $\sum_{m \in M} P^m = 1$. Hence, the corresponding variables PP_{jltm}, SS_{jltm}, W_{ijtm}, and θ_{istm} are indexed also by scenarios.

If product j passes testing in scenario m, then its sales SS_{jltm} must be bounded by d^L_{jlt} and d^U_{jlt}, but if it fails, its sales are zero, implying that its production (and the corresponding purchasing and operational costs) is also zero. To account for that, we introduce a parameter f_{mj}, which is 1 if in scenario m product j passes testing, and zero if it fails. For existing products f_{mj} is always one. Constraints (43) to (50) that describe production planning are now written as follows:

$$\sum_l PP_{jltm} + \sum_{i \in O(j)} W_{ijtm} = \sum_l SS_{jltm} + \sum_{i \in I(j)} W_{ijtm} \qquad \forall j \in J, \forall t \in T, \forall m \in M \qquad (54)$$

$$W_{ijtm} = \sum_{s \in PS(i)} \mu_{ijs}\, \rho_{is}\, \theta_{istm} \qquad \forall i \in I, \forall j \in J(i,s), \forall t \in T, \forall m \in M \qquad (55)$$

$$\sum_{s \in PS(i)} \theta_{istm} \le Q_{it}\, HP_{it} \qquad \forall i \in I, \forall t \in T, \forall m \in M \qquad (56)$$

$$a^L_{jlt}\, H_t \le PP_{jltm} \le a^U_{jlt}\, H_t \qquad \forall j \in J, \forall l \in L, \forall t \in T, \forall m \in M \qquad (57)$$

$$d^L_{jlt}\, H_t\, f_{mj} \le SS_{jltm} \le d^U_{jlt}\, H_t\, f_{mj} \qquad \forall j \in J, \forall l \in L, \forall t \in T, \forall m \in M \qquad (58)$$

$$Income = \sum_m P^m \{\sum_j \sum_l \sum_t \gamma_{jlt}\, SS_{jltm}\} \qquad (59)$$

$$C_{PURCH} = \sum_m P^m \{\sum_j \sum_l \sum_t \varphi_{jlt}\, PP_{jltm}\} \qquad (60)$$

$$C_{OPER} = \sum_m P^m \{\sum_I \sum_{s \in PS(i)} \sum_t (\delta_{ist}\, \rho_{is}\, \theta_{ist})\} \qquad (61)$$

where P^m is the probability of occurrence of scenario m.

The modified MILP model for Design and Planning consists of equations (38)-(39) and (51)-(61) with

$$Q_{it}, QE_{it}, \theta_{istm}, W_{ijtm}, PP_{jltm}, SS_{jltm} \ge 0, \; yE_{it}, yNP_{pt} \in \{0,1\} \qquad (62)$$

3. Integration of Testing with Design/Planning

The major challenge in this work is to integrate the scheduling of tests with the design and planning of the manufacturing facilities. The testing process affects production in two different ways, and it is this testing-production interaction that prevents the problem to be decomposed into two stages, making the integration of the two models necessary. On the one hand, the testing outcome determines whether a new product is ever going to be introduced into the market. On the other hand, testing completion time determines the earliest possible time of commercialization.

The goal of this integration is to obtain a set of constraints that coordinate the production of new products after the testing is completed. This is not a trivial task because scheduling is represented with a continuous time domain, while for the design/planning model a discrete time representation is used (i.e. investment decisions are considered at fixed times like every 6 months). The integration is made by introducing a boolean variable zI_{jt} that is *True* if testing of potential product j is completed within time period t. This is accomplished by the following disjunction:

$$\left[\begin{array}{c} z_j \\ \vee_t \left[\begin{array}{c} zI_{jt} \\ HT_{t-1} \leq T_j \leq HT_t \end{array}\right] \end{array}\right] \vee \left[\begin{array}{c} \neg z_j \\ T_j = 0 \end{array}\right] \qquad (63)$$

where HT_{t-1} is the beginning of time interval t.
Using the convex hull formulation[18,19] disjunction (63) is expressed in mixed integer form as (Appendix C):

$$\sum_t zI_{jt} = z_j \qquad \forall j \in JP \qquad (64)$$

$$T_j = \sum_t T_j^t \qquad \forall j \in JP \qquad (65)$$

$$HT_{t-1}\, zI_{jt} \leq T_j^t \leq HT_t\, zI_{jt} \qquad \forall j \in JP, \forall t \in T \qquad (66)$$

The condition that no sales take place before testing is completed is imposed by constraints (67) and (68):

$$S_{jltm} \leq d_{jlt}^U f_{mj}\, H_t \sum_{t' \leq t} zI_{jt'} \qquad \forall j \in JP,\ \forall l \in L,\ \forall t \in T,\ \forall m \in M \qquad (67)$$

$$S_{jltm} \leq (HT_t - T_j^t)\, d_{jlt}^U f_{mj} \qquad \forall j \in JP,\ \forall l \in L,\ \forall t \in T,\ \forall m \in M \qquad (68)$$

As is shown in Figure 5, constraint (67) does not allow sales before testing is completed, while constraint (68) is used to adjust sale levels during the time period that the introduction of the new product takes place.

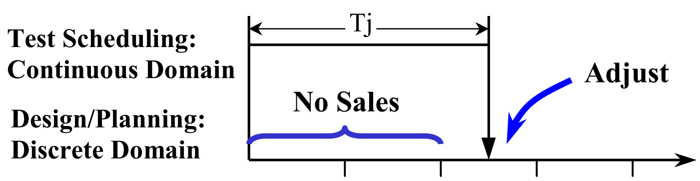


Figure 5: Integration of Testing with Design/Planning

The objective of the integrated model is the maximization of the Net Present Value of multiple projects:

$$Max\ NPV = Income - C_{TEST} - C_{INVEST} - C_{PURCH} - C_{OPER} \qquad (69)$$

The integrated MILP model then consists of (17), (18), (20)-(37), (38)-(39), (51)-(62) and (64)-(69), which in general leads to a very large scale optimization model.

Remarks

Before addressing the question of how to solve efficiently the proposed model and illustrating its application, it is worth to note the two following points. First, in order to develop a general model that can be applied in all cases of new product development, we did not include some industry-specific issues. In the pharmaceutical industry, for example, the market-production of a new drug begins only after the completion of qualification runs that they are also needed for FDA approval. These runs may turn out to be important, since the amounts produced during these runs are not sold and the whole process may be quite expensive. The process development stage that usually runs in parallel with testing is also very important. Nevertheless, the proposed model is flexible for accommodating such issues. Qualification runs, for example, can be modeled as testing tasks that should be performed after the completion of all the other tasks. The process development stage can be modeled as a testing task that requires a specific kind of resources (process development group), and where outsourcing is not allowed.
Second, constraint (68) is used for the adjustment of sales during the period of introduction, and has the general form *Sales(kg)≤Demand(kg/month)·Time(month)*, where *Time* is the time between the testing completion and the end of the time interval during which testing is completed. The production of a new drug, however, is not bounded in a similar manner (*Production≤Capacity·Time*), since this introduces the bilinear term *Capacity·Time*. From the material balance of the final products we get *Sales=Production*, but since

production time is not explicitly bounded, the production may begin earlier, i.e. we may have the equality $Demand \cdot T_1 = Capacity \cdot T_2$, where T_2 is greater that T_1. When a process is used for the production of many products, this simplification does not introduce any error, because we can assume that during each time period the process is first used for the production of the existing products, and then (after testing has been completed) for the production of the new product. Moreover, given an optimal solution, we can check whether the capacity constraints hold during the interval of new product introduction and correct accordingly. If there are processes, however, that are used only for the production of new drugs, this simplification may lead to suboptimal solutions. To overcome this difficulty we can use a stricter formulation, where we do not allow sales before the end of the time interval during which testing is completed. This is accomplished by using constraint (70) instead of constraints (67) and (68):

$$S_{jltm} \leq d_{jlt}^{U} \cdot f_{mj} \cdot H_t \cdot \sum_{t'<t} zI_{jt'} \qquad \forall j \in JP, \forall l \in L, \forall t \in T, \forall m \in M \qquad (70)$$

Note that the summation over t' for binaries zI_{jt} is made for $t' < t$, whereas in (67) is made for $t' \leq t$. This means that in (70) non-zero sales are allowed only if testing is completed within an earlier time interval.

Third, the proposed model can be easily extended to account for construction times. This can be done by (i) discounting the NPV back to the actual time of investment, and (ii) by adding a constraint that does not allow any expansions if there is not enough time. If, for instance, an investment decision for a process is for the 2nd year, and the construction time is 6 months, then the investment for the NPV is discounted at 1.5 years, and not at 2 years. In addition, a single constraint that does not allow any expansion of that process before the 6th month is added (i.e. for 6-month periods we should add $yE_{it} = 0$ for t=0). Constraint (52) must also be modified. If τi is the number of periods needed for the expansion (or construction) of process i, then constraint (52) reads:

$$\sum_p \alpha N_{p,t+\tau i}\, yNP_{p,t+\tau i} + \sum_i (\alpha E_{i,t+\tau i}\, yE_{i,t+\tau i} + \beta_{i,t+\tau i}\, QE_{i,t+\tau i}) \leq CI_t \qquad \forall t \in T \qquad (52')$$

The model can also accommodate discrete sizes in capacity, by introducing the following two constraints:

$$\sum_z yE_{itz} = yE_{it} \qquad \forall i \in I,\ \forall t \in T \qquad (71)$$

$$QE_{it} = \sum_z yE_{itz}\, QS_{iz} \qquad \forall i \in I,\ \forall t \in T \qquad (72)$$

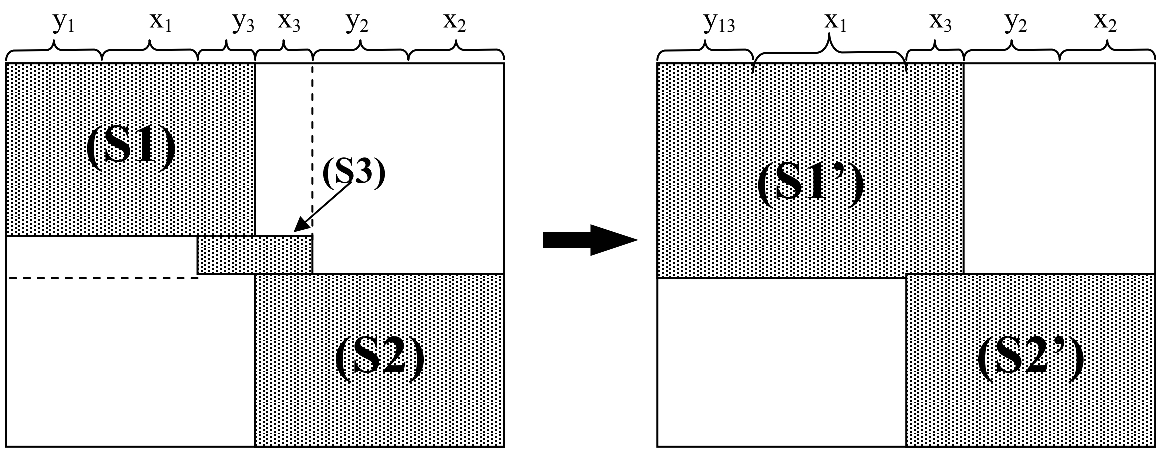
where z is the index for the discrete sizes, yE_{itz} is a binary variable that is 1 if z is the size of the expansion of process i, that takes place at the beginning of period t, and QS_{iz} are the expansion sizes of process i.

Fourth, by extending the notion of scenario, the effect of testing on sales can be taken into account. In the pharmaceutical industry, the results of clinical trials determine the group of patients for which the new drug is appropriate. It may be found, for example, that due to combined action with other drugs, the new drug is appropriate only for a subset of the initial set of potential patients. This aspect can be modeled extending the notion of scenarios. Consider, for instance, the case of only one new product with three different testing outcomes (scenarios). The first scenario has probability 0.3 and corresponds in the case where testing was successful on more types of patients and sales can be as large as the initially estimated upper bound d^U_{jlt}. The second scenario has probability 0.4 and corresponds to the case where the testing was successful on fewer types of patients and thus sales cannot be greater that 70% of d^U_{jlt}. The third scenario with probability 0.3 corresponds to testing failure. We can then define $P^1 = 0.3$, $P^2 = 0.4$, $P^3 = 0.3$, and $f_{1A} = 1$, $f_{2A} = 0.5$, and $f_{3A} = 0$, where parameter f_{mj} corresponds to the maximum percentage of sales of product j that can be materialized in scenario m. Thus, by increasing the number of scenarios, but without changing the formulation (constraints (67) and (68) remain the same), we have captured the dependence of sales on testing outcomes. The equations describing the scheduling of tests need not be changed, because scheduling is affected only by pass/failure events, and thus we are interested in the lumped probability of passing.

Finally, in the proposed model we assume that each process i, is available HP_{it} units of time, during each time period t, where $HP_{it} < H_t$. This assumption is made in order to account for setup times and plant shutdowns. The reason for this rough approximation is that in new product development it is not only the product, but also the process that is being developed. This means that exact processing times and cleaning procedures for the new products are not known accurately at the time that some first, high-level investment decisions have to be made.

Solution Method

In this section we discuss a Lagrangean decomposition[19,20,21] scheme, since the proposed MILP model is potentially very expensive to solve. As the number of new products, tests, resources, processes and time periods becomes larger, the number of binary variables increases and the problem becomes intractable with a full space method. Therefore, for the proposed models to have practical application, a more efficient solution method is needed. In this work, we propose an approach that exploits the special structure of the problem. If $y_1=\{z_j, x_k, \hat{x}_{kq}, y_{kk'}, \hat{y}_{kk'}\}$, $y_2=\{yE_{it}, yNP_{pt}\}$, $y_3=\{zI_{jt}\}$, $x_1=\{T_j, T_j^t, s_k, -w_k, \lambda_{kn}, \Lambda_{kn}\}$, $x_2=\{Q_{it}, QE_{it}, \theta_{istm}, W_{ijtm}, P_{jltm}, S_{jltm}\ \forall j \notin JP\}$, and $x_3=\{S_{jltm}\ \forall j \in JP\}$, the structure of the incidence matrix of the constraints is as the one presented in Figure 6a. Submatrix (S1) consists of the testing and linking constraints (16), (18), (20)-(37) and (64)-(66), submatrix (S2) consists of the design/planning constraints (38)-(39), and (51)-(62), and submatrix (S3) includes the linking constraints (67) and (68), or constraint (70), if the stricter formulation is used. If we treat constraints (S1) and (S3) as one group of constraints, the structure of the incidence matrix can be rearranged into the matrix of Figure 6.b, where submatrix (S1') corresponds to the union of constraints (S1) and (S3) in Figure 6a, and y_{13} corresponds to the union of y_1 and y_3. The incidence matrix is now partitioned into two submatrices. The testing and linking constraints are included in submatrix (S1'), and the design/planning constraints in submatrix (S2').



(a) (b)

Figure 6: Structure of Incidence Matrix of the Constraints

The proposed model **(P)** can then be written as follows:

(P) $\quad Max\ NPV = c_1^T x_1 + c_2^T x_2 + c_3^T x_3 + d_2^T y_2$

$$s.t. \quad A_1^{13} y_{13} + B_1^1 x_1 + B_1^3 x_3 \le b_1 \qquad (S1')$$

$$A_2^2 y_2 + B_2^2 x_2 + B_2^3 x_3 \le b_2 \qquad (S2')$$

$$x_1, x_2, x_3 \ge 0, \quad y_{13}, \ y_2 \in \{0,1\}$$

The only variables that have nonzero entries in both (S1') and (S2') are x_3 ($S_{jltm}\ \forall j \in JP$), which appear in constraints (67) and (68) (or (70)) in (S1'), and in constraints (43) and (47)-(48) in (S2'). Thus, if we duplicate x_3 we get problem **(P*)** where constraints (S1') and (S2') have no variables in common:

(P*) $\quad Max\ NPV = c_1^T x_1 + c_2^T x_2 + c_3^T x_3 + d_2^T y_2$

$$s.t. \quad A_1^{13} y_{13} + B_1^1 x_1 + B_1^3 x_{3A} \le b_1 \qquad (S1')$$

$$A_2^2 y_2 + B_2^2 x_2 + B_2^3 x_{3B} \le b_2 \qquad (S2')$$

$$x_{3A} = x_{3B} \qquad (S3')$$

$x_1, x_2, x_{3A}, x_{3B} \geq 0, \qquad y_{13}, y_2 \in \{0,1\}$

By dualizing (S3'), we add the term $\lambda^T (x_{3B} - x_{3A})$ to the objective, where λ are the Lagrange multipliers. In this way we get a problem which can be decomposed into two independent problems *(P1)* and *(P2)*:

(P1) $\quad Max\ NPV1 = c_1^T x_1 - \lambda^T x_{3A}$

$\quad s.t. \quad A_1^{13}\ y_{13} + B_1^1 x_1 + B_1^3 x_{3A} \leq b_1 \qquad (S1')$

$\quad x_1, x_{3A}, \geq 0, \qquad y_{13} \in \{0,1\}$

(P2) $\quad Max\ NPV2 = c_2^T x_2 + c_3^T x_{3B} + d_2^T y_2 + \lambda^T x_{3B}$

$\quad s.t. \quad A_2^2 y_2 + B_2^2 x_2 + B_2^3 x_{3B} \leq b_2 \qquad (S2')$

$\quad x_2, x_{3B} \geq 0, \qquad y_2 \in \{0,1\}$

Problems **(P1)** and **(P2)** are used in an iterative scheme, where they are solved independently. Their solution (NPV1+NPV2) provides an upper bound to the objective of **(P)** and a feasible set of binary values, which are then used to solve **(P)** with fixed binaries, providing a lower bound in the objective of **(P)**. The initial values for Lagrange multipliers are obtained from the solution of **(P)** as a relaxed LP. Problem **(P)** when solved with the integer requirement relaxed is called (LRP), and when solved as LP with fixed binaries is called (RP). λ_{jltm} are the multipliers of (S3') which are updated with the subgradient optimization procedure by Fischer[21].
The algorithm is as follows:

0. Initialize: Select K^{MAX}, ε, r, α^k. Set $UB=+\infty$, $LB=-\infty$. Solve (LRP) to obtain λ^1.

   For $k = 1..K^{MAX}$

1. Solve **(P1)** and **(P2)** at λ^k. Set $ZUB^k=NPV1+NPV2$. If $ZUB^k<UB$ then $UB=ZUB^k$.
2. Solve (RP) for fixed binaries determined from Step1. Set $ZLB^k=NPV$

   If $ZLB^k>LB$ then $LB=ZLB^k$.

   If $k=K^{MAX}$ or $(UB-LB)<\varepsilon$ or LB remained the same for more than 2 iterations Stop. Return current LB and corresponding solution.
3. $k=k+1$

   Update α^k if LB has remained unchanged for r iterations.

   $\lambda^{k+1}=\lambda^k + t^k (x_{3B}^k - x_{3A}^k)$, $t^k=\alpha^k (ZUB^k-LB)/||\ x_{3B}^k - x_{3A}^k||^2$. Return to Step 1.

In the above, K^{MAX} is the maximum number of iterations, t^k is a scalar stepsize that should gradually converge to zero, x_{3A}^k and x_{3B}^k are the optimal solutions of problems **(P1)** and **(P2)** at step k, α^k is a scalar chosen between 0 and 2, that is reduced by a factor of two whenever ZUB^k has failed to decrease in a specified number of r iterations, and $||\ x_{3B}^k - x_{3A}^k\ ||^2$ is the sum of squared deviations of constraints (S3'), denoted with *Dev* in the schematic diagram of the algorithm given in Figure 7. The algorithm stores the solution that corresponds to the best lower bound, but it can be modified to store all solutions found.
Due to the duality gap[22], the upper and lower bounds will not necessarily converge. Therefore, the predicted solution that corresponds to the current lower bound is not guaranteed to be optimal, although the upper bound provides a valid estimation of the maximum error. Our computational experience, however, has shown that very often the optimal solution is found in fewer than five iterations. Furthermore, as will be shown, computational times are significantly smaller than the full space method.

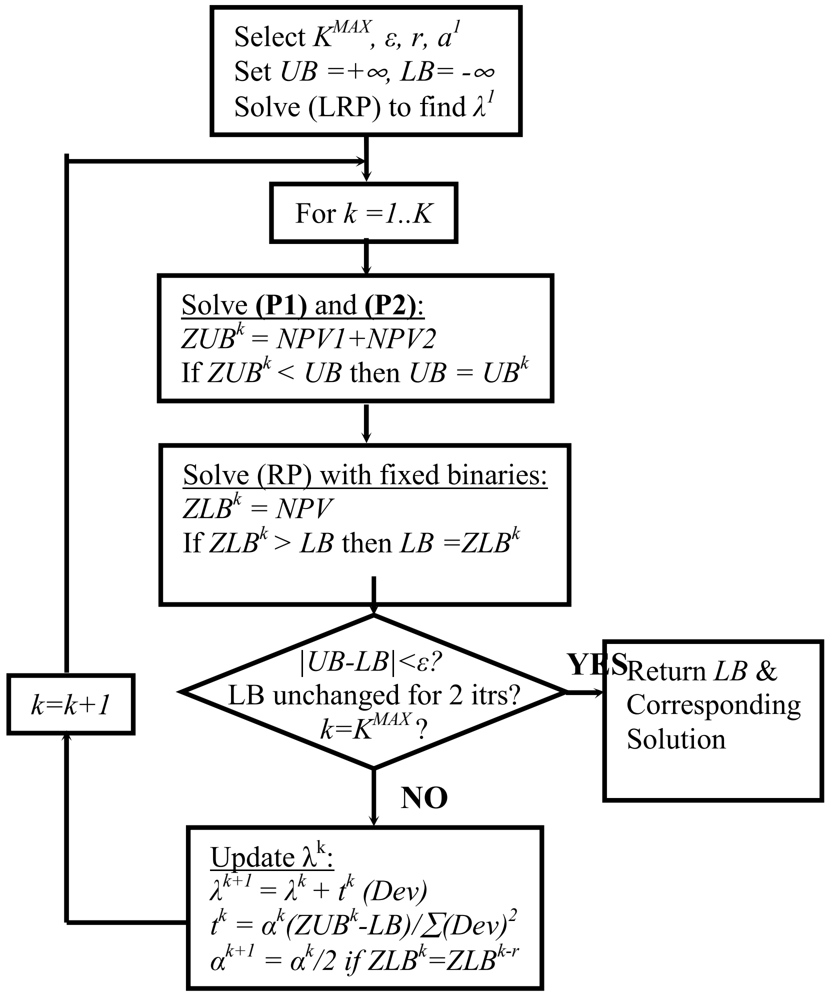


Figure 7: Flowsheet of Lagrangean Decomposition Heuristic Algorithm

Examples

In this section, we present three examples that illustrate the use of the proposed model. First, we revisit the motivating example, and next we study two examples that are closer to real-world applications.

Example 1

In the motivating example we saw that when resource constraints are not taken into account both new products C and D are selected. However, it is not only the resource constraints, but also the existing plant capacity, and various economic constraints (e.g. availability of capital) that make this problem quite difficult. Resource constraints, for instance, either lengthen testing completion times, or make testing more expensive (due to outsourcing), or both. If on the other hand, we consider unused existing capacity we may decide that no additional investment is needed for a new product. Moreover, availability of capital coupled with limited existing capacity may pose restrictions on expansion policy, and thus restrict the maximum number of new products. Applying the proposed optimization model that takes into account all these issues, we find that due to resource constraints only product C is selected. Product D is not selected because it is not profitable. The optimal NPV for this alternative is **$9,519,800**, i.e. higher than the one from the heuristic ($9,118,500). The Gantt chart, the design decisions and an analysis of costs and revenues are given in Figure 8, and Tables 7-8.

Table 7: Design Decisions of Example 1: Expansions (ton/month)

Process	1st period	2nd-18th period
P1	6.8	-
P2	-	-
P3	-	-
P4	-	-

Table 8: Revenues and Costs ($\$10^3$) of Example 1 (Model)

Sales Revenues	28,350.7
Testing Cost	4,658.1
Investment Cost	1,402.0
Operational Cost	5,945.3
Raw Material Cost	6,826.3
Expected NPV	**9,519.0**

Furthermore, let us assume that there are also restrictions on the capital available for investment during each period. In our example, if we require that no more than $1,000,000 is spent on investments during each time period, the optimal policy would be to pursue only product D, and the expected NPV would be **$9,096,900**. The solutions of the cases described above are given on Table 9.

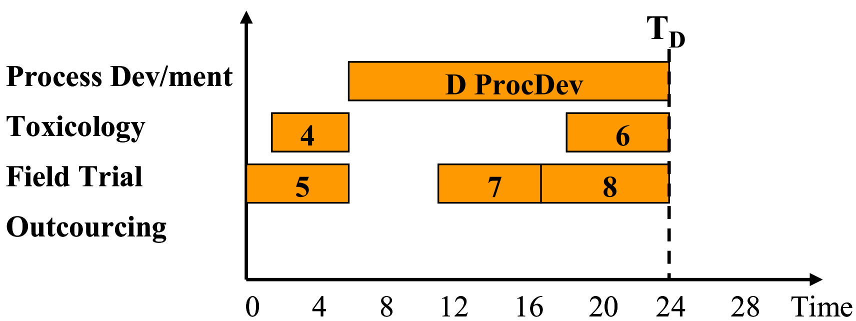


Figure 8: Gantt Chart or Resources of Example 1 – Optimal Solution

Even in this simple example, it is not straightforward to see why capital limitations lead to selection of product D instead of product C. This happens because C needs high levels of production in order to achieve high profits (note that demand for C is much higher than demand for D). In the unrestricted case, the investment needed for C can be made, but when we impose capital limitations, only limited investment is allowed and in this case product D is selected because it has larger marginal profit. In a more complex problem it would not be possible to have such insight, and even if it were possible, it would be very hard to determine the point at which product D becomes as profitable as product C.

Table 9: Optimal Solutions of Example 1 for different cases

Case	Additional Constraints	NPV ($\$10^3$)	Selected Products
1.Heuristic	-	9,118.5	C, D
2. Proposed Model	-	9,519.8	C
3. Proposed Model	Capital Restriction	9,096.9	D

Example 2

For the second example, we have used testing data similar to the one presented in example 4 of Jain and Grossmann[4], and a simplification of the production network described in Pinto et al[22]. Consider the case where a biotechnology firm produces recombinant proteins in a multipurpose protein production plant. Products A, B, D, and E are currently sold, and products C and F are in the company's R&D pipeline. Both potential products must pass successfully ten tests before they gain FDA approval. These tests can be performed either in-house, or outsourced at double cost. When performed in-house, they can be conducted in only one specific laboratory. Duration, costs, probability of success, resource requirements and technological precedences for each test are given in Table 10.

Table 10: Testing Data of Example 2

Product	Test	Cost ($\$10^4$)	Cost of out-sourcing ($\$10^4$)	Duration (months)	Prob. of success	Resource Requirement	Technological Precedences
C	1	80	160	5	1	Lab1	-
	2	80	160	4	1	Lab2	-
	3	50	100	4	1	Lab3	-
	4	10	20	1	0.84	Lab4	3,7
	5	490	980	3	0.98	Lab1	1
	6	111	222	6	1	Lab2	2
	7	60	120	1	0.95	Lab3	3
	8	1740	3480	7	1	Lab4	3,4,7
	9	620	1240	9	1	Lab1	1,5
	10	10	20	1	1	Lab2	2,6
F	11	160	320	3	1	Lab1	15,19,20
	12	1130	2260	1	0.87	Lab2	16,20
	13	10	20	1	0.91	Lab3	16,17,20
	14	130	260	3	1	Lab4	18
	15	530	1060	2	1	Lab1	19,20
	16	90	180	1	1	Lab2	20
	17	117	234	3	1	Lab3	16,20
	18	400	800	3	1	Lab4	-
	19	570	1140	5	1	Lab3	-
	20	230	460	3	1	Lab4	-

Products A, B and C are extracellular, while D, E and F are intracellular. All proteins are produced in the fermentor P1. Intracellular proteins are then sent to the homogenizer P2 for cell suspension, then to extractor P3 and lastly to the chromatographic column P4 where selective binding is used to further separate the product of interest from other proteins. Extracellular proteins after the fermentor P1 are sent directly to the extractor P3, and then to the chromatograph P4. In reality, a microfilter is used before the homogenizer and ultrafilters are used before the extractor and the chromatograph, but in order to keep the production network simple we will consider only the basic processes of such a plant. The fermentor, extractor, and chromatographic column operate under six different production schemes, while the homogenizer operates under three different production schemes (one for each protein). The production network is shown in Figure 9, and the process data for plant design is given in Table 11. For simplicity, operating costs and conversions for all operating schemes of the same process are assumed equal. The time horizon is 6 years. Four 6-month and four 12-month time periods are considered for the investment in manufacturing facilities. Capacity expansions cannot be greater than 4 ton/month, and each process cannot be expanded more than four times. Investment costs must be less than \$15,000,000 during any period. An annual discount rate of 9% has been used. All production schemes have the same relative capacity ($\rho_{is} = 1 \ \forall i,s$). The forecast of demand for existing and potential products is given in Table 12. The prices of raw materials and final products are given in Table 13. We consider the four following scenarios: in scenario 1 both C and F pass, in scenario 2 C passes and F fails, in scenario 3 C fails and F passes, and in scenario 4 both C and F fail.

Table 11: Process Design Data of Example 2

Process	Conversion (all schemes)	Initial Capacity (kg/month)	Fixed Cost ($\$10^3$)	Variable Cost ($\$10^3$·month/kg)	Operating (all schemes) ($\$10^3$/kg)
P1	1/1.3=0.7692	10	4,500	1,100	18
P2	1/1.2=0.8333	6	4,000	800	14
P3	1/1.4=0.7143	8	3,500	1,000	16
P4	1/1.1=0.9091	8	5,000	1,200	14

Table 12: Demand forecasts (kg/month) for Example 2

Product	1st yr	2nd yr	3rd yr	4th yr	5th yr	6th yr
A	3	3	3	3	3	3
B	3	3	3	4	4	4
C	4	4	4	4	4	4
D	5	5	5	5	5	5
E	4	4	2	2	0	0
F	6	6	6	6	6	6

Table 13: Prices of Chemicals ($\$10^3$/kg) for Example 2

Raw Material	Price	Product	Price
A1	40	A	300
B1	50	B	350
C1	60	C	550
D1	50	D	450
E1	60	E	400
F1	70	F	700

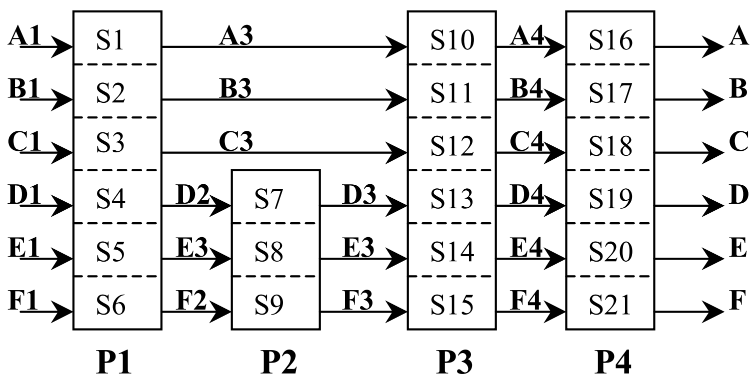


Figure 9: Process Network of Example 2

The optimal solution has a NPV of **$95,986,500**. It was found that only product F is profitable to pursue, and that process P1 should be expanded twice and processes P2, P3 and P4 should be expanded once. The Gantt chart for the resources is shown in Figure 10, and the capacity of the processes (design decisions) is given in Table 14. Production levels are presented in Figures 11 to 15. Since product C is not selected for testing, scenarios 1 and 3 are equivalent (F passes), and scenarios 2 and 4 are equivalent (F fails). Finally, a breakdown of the expenses and revenues is given in Table 15.

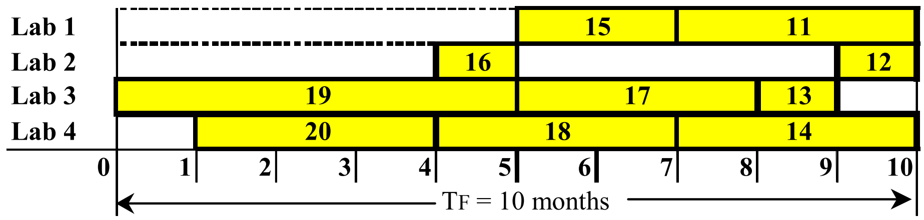


Figure 10: Gantt Chart of Resources of Example 2

Table 14: Design Decisions of Example 2 – Capacity of Process Network (kg/month)

Process	Initial	0-6 month	7-12 month	13 - 18 month	18-72 month
P1	10	14	14	18	18
P2	6	6	6	10	10
P3	8	8	11.429	11.429	11.429
P4	8	8	10.39	10.39	10.39

Table 15: Costs and Revenues of Example 2

Revenues	$276,996,000
Costs	
Testing	$31,456,000
Investment	$37,759,000
Operational	$41,351,000
Raw Materials	$71,444,000
NPV	**$94,986,000**

The model anticipates that the optimal policy is to test only product F. In this example, capacity and investment limitations play a major role. Since the initially installed capacity is insufficient, investment for the production of both existing and new products is needed. Due to the testing uncertainty, and since the price of product C is similar to the prices of the existing products, it is better not to make additional investments for product C that may not be redeemed. Product F is chosen because the expected revenues outweigh investments, due to its high price. Note also, that due to capital limitations, process expansions are made gradually: process P1 is expanded at the beginning of the first and third period, processes P3 and P4 are expanded at the beginning of the 2nd period, and process P2 is expanded at the beginning of the 3rd period.

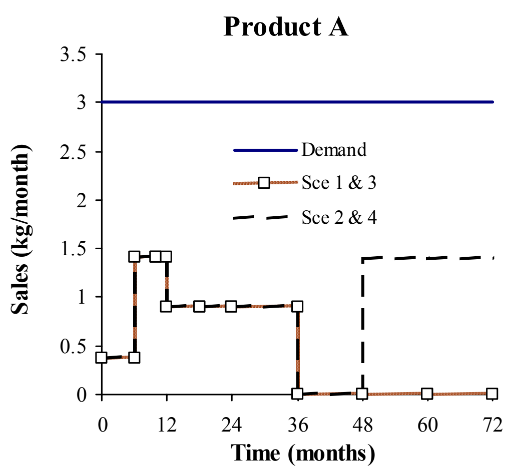


Figure 11: Production Level of Product A

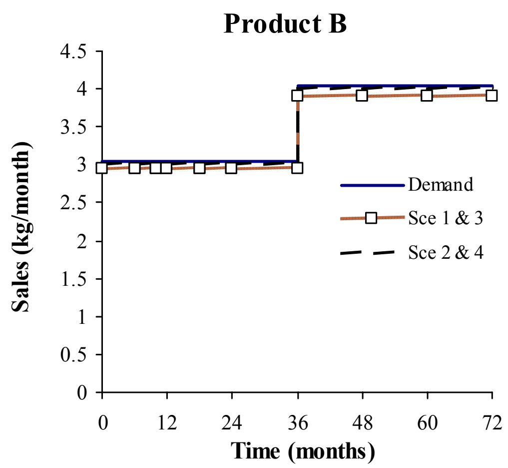


Figure 12: Production Level of Product B

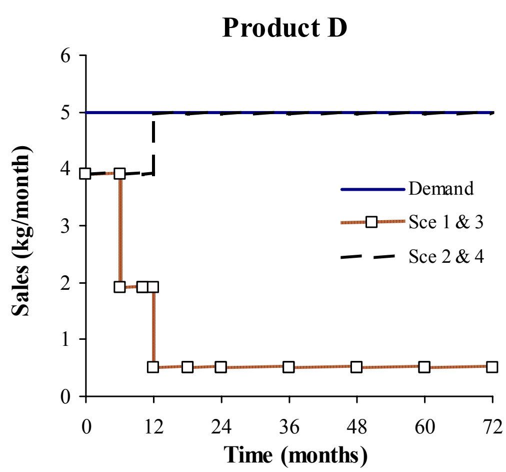


Figure 13: Production Level of Product D

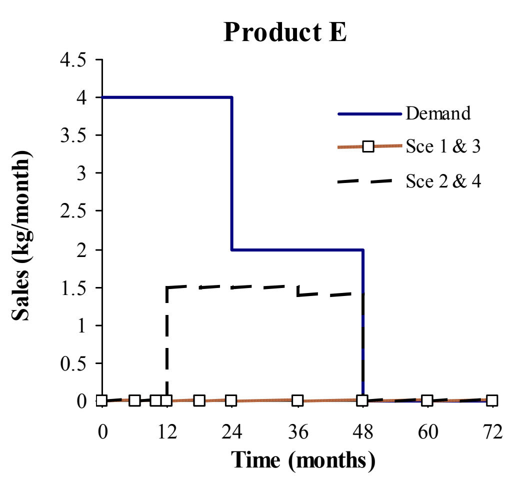


Figure 14: Production Level of Product E

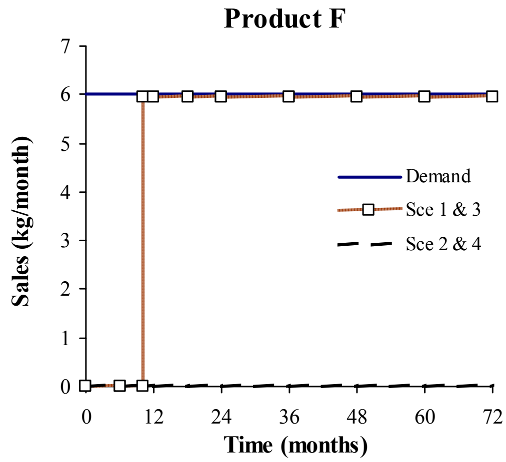


Figure 14: Production Level of Product E

As shown in Figures 11 to 15, product F is the most profitable, and thus, if it passes testing, its production meets the demand and begins as soon as testing is completed. Product B is the next more profitable product. In the scenario where F does not pass testing, its production is always as large as its demand. If F passes, its production is slightly smaller after the 3rd year (when its demand increases) due to capacity limitations. Product D is also profitable, but since its production requires the use of P2, which is a bottleneck for the production of products D, E and F, its production decreases when the production of F begins. In the scenario where F does not pass testing, its production is as large as its demand after the necessary capacity expansions are made. Finally, A is produced only when there is available capacity; i.e. after expansions and before the increase of B production. Product E is not produced at all when F passes testing, but it is produced if F fails.
Solution in full space using CPLEX 7.0 on a Pentium III at 667MHz requires 57 sec. The proposed decomposition technique required 38 sec for five iterations (the optimal was found after three iterations), and obtained the same solution.

Example 3

Consider a pharmaceutical firm that produces drugs A, C and E, and has three new drugs B, D and F in different stages of its R&D pipeline. Each new drug, in order to gain FDA approval, should pass successfully the twelve tests shown in Figure 16. Note that these tests are different for each product; e.g. test 9 for product B (denoted by B9) is different from test 9 for product D (denoted by D9). Most of the R&D work for drug B has already been conducted, drug D is a newer product, and product F has just been synthesized in the lab. Thus, product B has passed successfully tests 1 to 8, product D has passed successfully tests 1 to 4, while no testing has yet been performed for product F. The process development stage of products B, D and F is expected to last 6, 12 and 20 months respectively. The resources are four different groups of scientists that are able to perform different tests. Testing data is given in Table 16 (process development is treated as a testing task that cannot be outsourced). The cost of outsourcing is assumed twice as large as in-house cost.

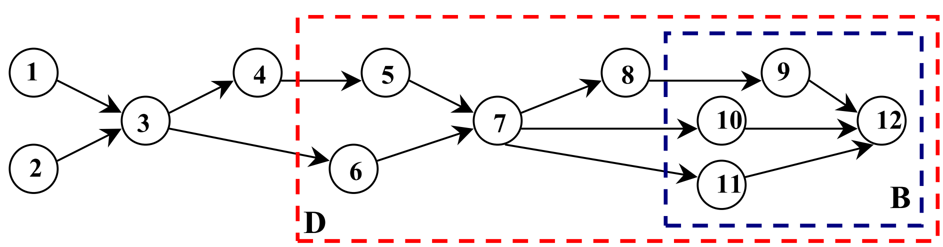


Figure 16: Technological Precedence Constraints of Example 3

Table 16: Testing Data of Example 3

	Duration			Probability of success			Cost ($\$10^3$)			Resource Requirement
Test	B	D	F	B	D	F	B	D	F	B, D & F
PrDev	6	12	20	1	0.975	0.95	4000	9000	12000	Process Devel. Group
1	3	2	3	0.97	1	1	1000	800	800	Chemical Group A
2	2	2	2	1	1	1	500	500	400	Chemical Group A
3	1	1	1	0.95	1	0.9	1200	1300	1500	Chemical Group A
4	4	5	6	1	0.98	1	1100	1200	1000	Clinical Tries Group
5	7	6	8	1	0.98	0.98	2500	2000	1800	Clinical Tries Group
6	6	6	6	1	1	1	500	800	600	Chemical Group B
7	3	2	2	0.95	0.9	0.95	1200	800	1000	Chemical Group B
8	4	4	5	0.98	1	1	2000	2400	2200	Clinical Tries Group
9	4	5	5	0.95	0.975	0.95	2000	1800	2500	Clinical Tries Group
10	3	4	6	0.99	0.99	0.88	500	700	600	Chemical Group B
11	5	3	3	1	1	1	800	1000	800	Chemical Group B
12	2	1	2	0.95	0.9	0.975	700	600	700	Chemical Group A

The production network is given in Figure 17. Note that for the production of intermediate ID3 a new plant is needed. This means that processes P3, P4 and P7 can only be built and expanded if new plant NP is decided to be built. Alternatively, intermediate ID3 can be purchased. Moreover, process P5 does not exist at t=0, but it can be built within the existing site, which means that the additional fixed investment is smaller. The process design data is given in Table 17. The fixed cost for new plant NP is $5,000,000. Demand forecasts for the final products are given in Table 18. Prices for raw materials, intermediate ID3 and final products, which for simplicity are assumed to be constant, are given in Table 19.

A time horizon of 8 years, divided into six 6-month periods (years 1-3), and five 12-month periods (years 4-8), has been used. Each process cannot be expanded more than 5 times and each expansion cannot be greater than 4 tons/month. In order to be competitive the company has decided that at least two new products should be pursued, even if this is not the optimum at the present moment. An annual discount rate of 9% is used.

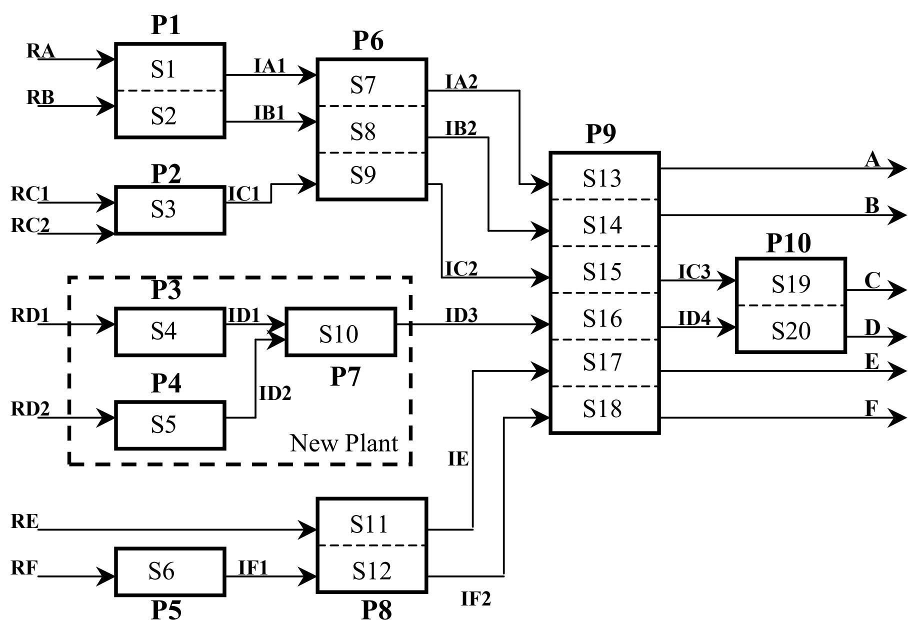


Figure 17: Process Network of Example 3

Table 17: Process Design Data of Example 3

Process	Conversion	Initial Capacity (tons/month)	Fixed Cost ($\$10^3$)	Variable Cost ($\$10^3$month/ton)	Operational Cost ($\$10^3$/ton)
P1	0.833	2	600	400	8
P2	0.8RC1+0.6RC2→IC1	2	500	250	8
P3	0.833	0	400	350	8
P4	0.909	0	300	450	8
P5	0.909	0	800	280	6
P6	0.909	5	1000	1000	S7:6, S8&S9: 4
P7	0.6ID1+0.7ID2→ID3	0	400	400	4
P8	S11: 0.833, S12:0.769	4	1000	500	3
P9	0.833	10	400	1500	6
P10	0.909	4	1200	1000	10

Table 18: Demand Forecasts (ton/month) for Example 3

Product	1st yr	2nd yr	3rd yr	4th yr	5th yr	6th yr	7th yr	8th yr
A	4	4	4	4	4	4	4	4
B	2	2	3	3	3	4	4	4
C	4	3.5	3	2.5	2.5	2	2	1.5
D	0	2.5	3	3.5	3.5	4	4	4.5
E	3	3	3.5	3.5	3.5	4.0	4.0	4.0
F	0	0	2	2.5	2.5	3	3	3.5

Table 19: Prices of raw materials, intermediate ID3, and final products ($\$10^3$/ton) for Example 3

RA	10	RD2	10	B	220
RB	15	ID3	80	C	300
RC	10	RE	18	D	400
RC	10	RF	20	E	240
RD1	20	A	180	F	260

The optimal solution yields a NPV of **$135,024,600**. Products B and D are selected for testing. Outsourcing is used only for test B11. The new plant is decided not to be built, which means that for the production of D intermediate ID3 is purchased, although at a rather high price. Investment is made on processes P1, P2, P6, P8, P9 and P10. The Gantt chart for the resources is given in Figure 18, the capacity of processes is given in Table 20 (highlighted cells correspond to expansions), and a breakdown of costs and revenues is given in Table 21.

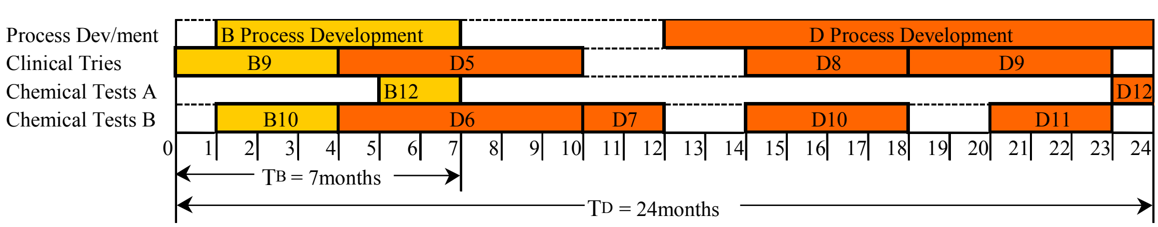


Figure 18: Gantt Chart of Resources for Example 3

The problem was solved in almost 14 min using CPLEX 7.0 on a PIII 667MHz. Using the proposed heuristic Lagrangean scheme, we found a solution with a NPV of **$134,834,900** (0.14% smaller than the optimal), in almost 4 min, and 6 major iterations.

Table 20: Design Decisions of Example 3 - Capacity of Processes Network (ton/month)

Process	Initial	0 – 6 month	7-12 month	13-24 month	25-36 month	37-96 month
P1	2	6	10	10	10	10
P2	2	5.808	5.808	5.808	5.808	5.808
P6	5	8.4	12.36	12.36	12.36	12.36
P8	4	4	4	4	4.8	4.8
P9	10	10	14	14	18	18
P10	4	4	4	4	6	6

Table 21: Costs and Revenues of Example 3

Revenues	$248,743,300
Costs	
Testing	$24,458,100
Investment	$36,651,900
Operational	$20,480,400
Raw Materials	$32,128,300
NPV	**$135,024,600**

Computational Results

In this section we summarize the computational results for Examples 1 (case 2), 2 and 3. In examples 1 and 2, the Lagrangean decomposition heuristic obtained the optimal solution after five and three iterations respectively, while in example 3 it obtained a solution that was suboptimal by only 0.14% after six iterations. In all examples the heuristic was terminated because the LB remained unchanged for 2 iterations. Figure 19 shows the duality gap as a function of the number of iterations, while Figure 20 shows the convergence of upper and lower bounds for examples 2 and 3. In order to construct these figures we deactivated the option for early termination, and run the heuristic for ten iterations. Note that as the problem size increases, the proposed heuristic outperforms the full space method.

Table 22: Computational Results

	Example 1	Example 2	Example 3
Binary Variables	236	354	612
Continuous Variables	9,372	7,456	32,184
Constraints	8,255	6,825	30,903
Full Space Method			
LP Relaxation	11,040,717	113,983,858	146,567,730
Optimal Solution	9,518,951	94,986,475	135,024,622
CPU Time (sec)	8.9	57.2	836.6
Nodes	222	1656	24627
Decomposition Heuristic			
Best Solution	9,518,951	94,986,475	134,834,914
Major iterations for best solution	5	3	6
CPU Time (sec)	20.5	37.6	252.4
Upper Bound	9,895,477	95,083,790	135,877,141
Relative Duality Gap	3.96%	0.10%	0.77%
% of CPU s for P1	17.2	20.1	36.7
% of CPU s for P2	52.0	60.9	53.1
% of CPU s for RP	30.8	10.0	10.2

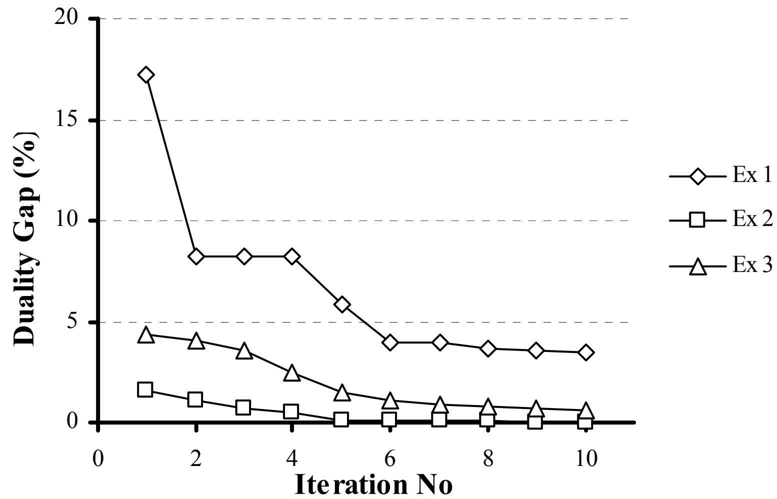


Figure 19: Lagrangean Decomposition – Duality Gap

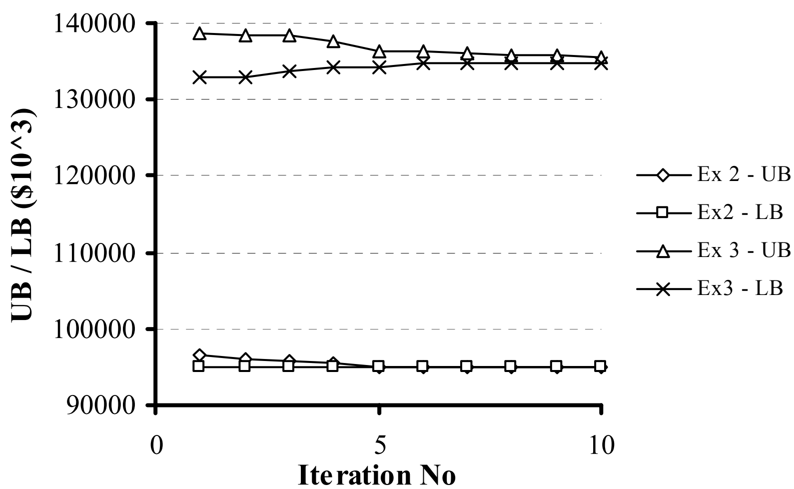


Figure 20: Lagrangean Decomposition – Convergence of Upper and Lower Bounds

Conclusions

The problem of simultaneous planning for new product development and batch manufacturing facilities has been addressed in this paper. This type of problem is important in highly regulated industries, such as pharmaceutical, biotechnology and agrochemical, where a new product is required to pass a number of tests. The proposed MILP model integrates a continuous scheduling model for testing with a discrete model for design/planning into a single framework for simultaneous optimization. A two-stage stochastic optimization approach is adopted to account for the uncertainty in the outcome of the tests. Since the resulting MILP model is rather hard to solve due to its size, an iterative heuristic based on Lagrangean decomposition was developed. Computational experience shows that the proposed algorithm provides optimal or near optimal solutions, and is considerably faster than the full space method.

Acknowledgment
The authors would like to gratefully acknowledge financial support from the National Science Foundation under Grant CTS-9810182. The financial support of Onassis Foundation to Christos Maravelias is gratefully acknowledged.

APPENDIX A: Tightness of resource allocation constraints

Jain and Grossmann[4] used the following logical condition, to derive the resource constraints (see Figure 21, with *k=3* and *k'=2*): *if a test k is not outsourced and is assigned to resource q, then for resource constraints to hold, any other test k' is either outsourced or is not assigned to resource q or there is an arc between tests k and k':*

$$\neg x_k \wedge \hat{x}_{kq} \Rightarrow x_{k'} \vee \neg \hat{x}_{k'q} \vee y_{kk'} \vee y_{k'k} \qquad \forall q \in Q,\ \forall k \in K(q),\ \forall k' \in (K(q) \cap KK(k)), k<k' \qquad (A1)$$

$$\neg x_k \wedge \hat{x}_{kq} \Rightarrow x_{k'} \vee \neg \hat{x}_{k'q} \vee \hat{y}_{kk'} \vee \hat{y}_{k'k} \qquad \forall q \in Q,\ \forall k \in K(q),\ \forall k' \in (K(q) \backslash KK(k)), k<k' \qquad (A2)$$

Using the transformations described in Raman and Grossmann[20], we obtain the following constraints:

$$\hat{x}_{kq} + \hat{x}_{k'q} - y_{kk'} - y_{k'k} - x_k - x_{k'} \le 1 \qquad \forall q \in Q,\ \forall k \in K(q),\ \forall k' \in (K(q) \cap KK(k)), k<k' \qquad (A3)$$

$$\hat{x}_{kq} + \hat{x}_{k'q} - \hat{y}_{kk'} - \hat{y}_{k'k} - x_k - x_{k'} \le 1 \qquad \forall q \in Q,\ \forall k \in K(q),\ \forall k' \in (K(q) \backslash KK(k)), k<k' \qquad (A4)$$

It should be noted, however, that if a test uses resource q, then it cannot be outsourced and, conversely, if it is outsourced it cannot use resource q, which means that the above logical condition can be simplified into the one proposed in this paper: *if a test k is assigned to resource q, then for resource constraints to hold, any other test k' is either not assigned to resource q or there is an arc between test k and k'.*
Constraints (A3) and (A4) can be rearranged as:

$$\hat{x}_{kq} + \hat{x}_{k'q} - y_{kk'} - y_{k'k} \le 1 + x_k + x_{k'} \qquad (A5)$$

$$\hat{x}_{kq} + \hat{x}_{k'q} - \hat{y}_{kk'} - \hat{y}_{k'k} \le 1 + x_k + x_{k'} \qquad (A6)$$

It is clear that (A5) is a relaxation of (16), and that (A6) is a relaxation of (17). Thus, the formulation proposed in the present paper is tighter.

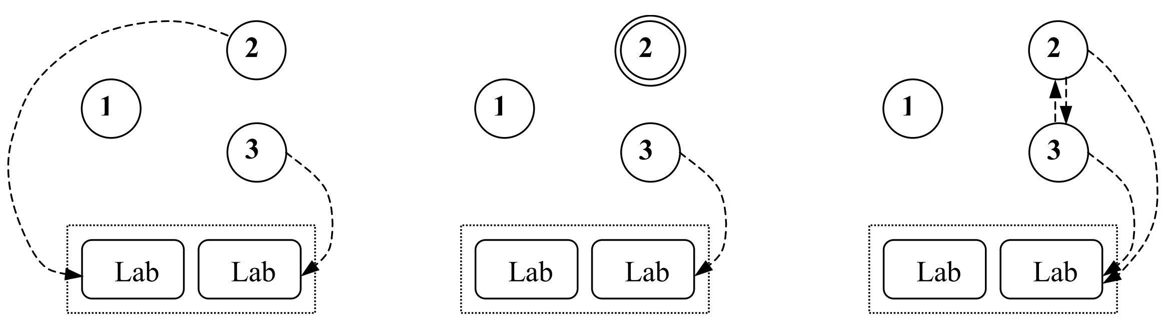


Figure 21: Graph-based Representation for Resources

APPENDIX B: Derivation of mixed-integer equations from disjunction in (19)

If we group constraints of the first term of disjunction (19) into "less or equal" inequalities and equalities, and move all variables in LHS and constants in RHS, we get the following disjunction for every potential product $j \in JP$:

$$\left(\begin{array}{ll} \multicolumn{2}{c}{z_j} \\ C_k - c_k\,(\sum_n e^{a_n}\,\lambda_{kn}) + \hat{c}_k\,(\sum_n e^{a_n}\,\Lambda_{kn}) = 0 & \forall k \in K(j) \\ w_k + \rho s_k - \sum_{k' \neq k} \ln(p_{k'}) \cdot y_{k'k} = 0 & \forall k \in K(j) \\ w_k - \sum_n a_n\,(\lambda_{kn} + \Lambda_{kn}) = 0 & \forall k \in K(j) \\ \sum_n \lambda_{kn} + x_k = 1 & \forall k \in K(j) \\ \sum_n \Lambda_{kn} - x_k = 0 & \forall k \in K(j) \\ \sum_{q \in (QT(k) \cap QC(r))} \hat{x}_{kq} + N_{kr}\,x_k = N_{kr} & \forall k \in K(j),\ \forall r \in R \\ y_{kk'} = 1,\ y_{k'k} = 0 & \forall (k,k') \in A \\ s_k - s_{k'} + U y_{kk'} \leq U - d_k & \forall k \in K(j),\ \forall k' \in K(j) \\ s_k - T_j \leq -d_k & \forall k \in K(j) \\ y_{kk'} + y_{k'k} \leq 1 & \forall k, k' \in K(j) \mid k < k' \\ y_{kk'} + y_{k'k''} + y_{k''k} \leq 2 & \forall k, k', k'' \in K(j) \mid k < k' < k'' \\ y_{k'k} + y_{kk''} + y_{k''k'} \leq 2 & \forall k, k', k'' \in K(j) \mid k < k' < k'' \\ \multicolumn{2}{l}{\hat{x}_{kq} + \hat{x}_{k'q} - y_{kk'} - y_{k'k} \leq 1\ \forall q \in Q, \forall k \in (K(q) \cap K(j)), \forall k' \in (K(q) \cap K(j)) \mid k < k'} \end{array}\right) \vee \left(\begin{array}{ll} \multicolumn{2}{c}{\neg z_j} \\ T_j = 0 & \\ C_k = 0 & \forall k \in K(j) \\ \lambda_{kn} = 0 & \forall k \in K(j), \forall n \\ \Lambda_{kn} = 0 & \forall k \in K(j), \forall n \\ s_k = 0 & \forall k \in K(j) \\ w_k = 0 & \forall k \in K(j) \\ y_{kk'} = 0 & \forall k, k' \in K(j) \\ x_k = 0 & \forall k \in K(j) \\ x_{kq} = 0 & \forall k \in K(j), \forall q \in Q \end{array}\right) \qquad (B1)$$

Using matrix notation, disjunction (B1) can be written as follows:

$$\begin{bmatrix} z_j \\ A\boldsymbol{x} = c \\ \mathrm{B}\mathbf{x} \leq \mathrm{d} \end{bmatrix} \vee \begin{bmatrix} \neg z_j \\ \boldsymbol{x} = \boldsymbol{0} \\ \ \end{bmatrix} \qquad (B2)$$

where $\boldsymbol{x} = [T_j\ C_k\ s_k\ w_k\ \lambda_{kn}\ \Lambda_{kn}\ y_{kk'}\ x_k\ \hat{x}_{kq}]^{\mathrm{T}}$.

Using the convex hull formulation[18, 19], disjunction (B2) can be expressed by constraints (B3) to (B7):

$$\boldsymbol{x} = \boldsymbol{x}^1 + \boldsymbol{x}^2 \qquad (B3)$$

$$A\boldsymbol{x}^1 = cz_j \qquad (B4)$$

$$B\boldsymbol{x}^1 \leq dz_j \qquad (B5)$$

$$\boldsymbol{x}^1 \leq \boldsymbol{U}z_j \qquad (B6)$$

$$\boldsymbol{x}^2 = \boldsymbol{0}(1-z_j) \qquad (B7)$$

Equation (B7) implies that $\boldsymbol{x}^2$ is always zero, which, in turn, implies that $\boldsymbol{x} = \boldsymbol{x}^1$. Thus, we get:

$$A\boldsymbol{x} = cz_j \qquad (B4')$$

$$B\boldsymbol{x} \leq dz_j \qquad (B5')$$

$$\boldsymbol{x} \leq \boldsymbol{U}z_j \qquad (B6')$$

If we now re-write constraints (B4') and (B5') in expanded form $\forall j \in JP$, we get:

$$C_k - c_k \left(\sum_n e^{an} \lambda_{kn}\right) + \hat{c}_k \left(\sum_n e^{an} \Lambda_{kn}\right) = 0 \quad \forall k \in K \qquad (B8)$$

$$w_k + \rho s_k - \sum_{k' \neq k} \ln(p_{k'}) \cdot y_{k'k} = 0 \quad \forall k \in K \qquad (B9)$$

$$w_k - \sum_n a_n (\lambda_{kn} + \Lambda_{kn}) = 0 \quad \forall k \in K \qquad (B10)$$

$$\sum_n \lambda_{kn} + x_k = z_j \quad \forall j \in JP, \forall k \in K(j) \qquad (B11)$$

$$\sum_n \Lambda_{kn} - x_k = 0 \quad \forall k \in K \qquad (B12)$$

$$\sum_{q \in (QT(k) \cap QC(r))} \hat{x}_{kq} + N_{kr} x_k = N_{kr} z_j \quad \forall j \in JP, \forall k \in K(j), \forall r \in R \qquad (B13)$$

$$y_{kk'} = z_j,\ y_{k'k} = 0 \quad \forall j \in JP, \forall k, k' \in K(j), \forall (k,k') \in A \qquad (B14)$$

$$s_k - s_{k'} + U y_{kk'} \leq (U - d_k) z_j \quad \forall j \in JP, \forall k, k' \in K(j) | k \neq k' \qquad (B15)$$

$$s_k - T_j \leq - d_k z_j \quad \forall j \in JP, \forall k \in K(j) \qquad (B16)$$

$$y_{kk'} + y_{k'k} \leq z_j \quad \forall j \in JP, \forall k, k' \in K(j) | k < k' \qquad (B17)$$

$$y_{kk'} + y_{k'k''} + y_{k''k} \leq 2z_j \quad \forall j \in JP, \forall k, k', k'' \in K(j)\ | k < k' < k'' \qquad (B18)$$

$$y_{k'k} + y_{kk''} + y_{k''k'} \leq 2z_j \quad \forall j \in JP, \forall k, k', k'' \in K(j) | k < k' < k'' \qquad (B19)$$

$$\hat{x}_{kq} + \hat{x}_{k'q} - y_{kk'} - y_{k'k} \leq z_j \quad \forall j \in JP, \forall q \in Q, \forall k \in (K(q) \cap K(j)), \forall k' \in (K(q) \cap K(j)) | k < k' \qquad (B20)$$

If potential product $j \in JP$ is not selected, then $z_j=0$ and, thus, constraints (B17) forces $y_{kk'}=0\ \ \forall k \in K(j)$. Setting $y_{kk'}=0$ into (B20), forces $\hat{x}_{kq} = 0\ \ \forall q \in Q,\ \ \forall k \in K(j)$. Moreover, setting $z_j=0$ and $\hat{x}_{kq} = 0\ \ \forall q \in Q,\ \ \forall k \in K(j)$ into (B13), forces $x_k = 0\ \ \ \forall k \in K(j)$. If, in addition, we plug $z_j = 0$ and $x_k = 0$ into (B11) and (B12), we get $\lambda_{kn} = 0$ and $\Lambda_{kn} = 0\ \ \forall k \in K(j),\ \ \forall n$. For $\lambda_{kn} = 0$ and $\Lambda_{kn} = 0\ \ \forall k \in K(j),\ \ \forall n$ (B10) gives $w_k = 0\ \ \forall k \in K(j)$. Finally, if we plug $w_k = 0$ and $y_{kk'} = 0$ into (B9) we get $s_k = 0\ \ \forall k \in K(j)$. Thus, from (B6') only $T_j = 0$ is not implied when $z_j = 0$, and therefore, instead of adding (B6') we can only add the following inequality:

$$T_j \leq U z_j \quad \forall j \in JP \qquad (B21)$$

Constraints (B8) to (B21) correspond to constraints (20)-(34) of the proposed model.

APPENDIX C: Derivation of mixed-integer equations from disjunction in (63)

First, we re-formulate the inner disjunction of (63) *using* the convex hull formulation:

$$\vee_t \begin{bmatrix} zI_{jt} \\ HT_{t-1} \le T_j \le HT_t \end{bmatrix} \qquad (C1)$$

Disjunction (C1) is equivalent to *constrains* (C2) to (C4):

$$T_j = \textstyle\sum_t T_j^t \qquad \forall j\in JP \qquad (C2)$$

$$\textstyle\sum_t zI_{jt} = 1 \qquad \forall j\in JP \qquad (C3)$$

$$HT_{t-1}\, zI_{jt} \le T_j^t \le HT_t\, zI_{jt} \qquad \forall j\in JP,t \qquad (C4)$$

Thus, disjunction (63) can now be written as follows:

$$\begin{bmatrix} z_j \\ T_j = \sum_t T_j^t \\ \sum_t zI_{jt} = 1 \\ HT_{t-1}\, zI_{jt} \le T_j^t \le HT_t\, zI_{jt} \end{bmatrix} \vee \begin{bmatrix} \neg z_j \\ T_j = 0 \\ \\ \end{bmatrix} \qquad (C5)$$

Using the convex hull formulation[18, 19], disjunction (B5) is expressed in mixed integer form as follows:

$$T_j = T_j^1 + T_j^2 \qquad \forall j\in JP \qquad (C6)$$

$$T_j^1 = \textstyle\sum_t T_j^t \qquad \forall j\in JP \qquad (C7)$$

$$\textstyle\sum_t zI_{jt} = z_j \qquad \forall j\in JP \qquad (C8)$$

$$HT_{t-1}\, zI_{jt} \le T_j^t \le HT_t\, zI_{jt} \qquad \forall j\in JP,\ \forall t\in T \qquad (C9)$$

$$T_j^1 \le U z_j \qquad \forall j\in JP \qquad (C10)$$

$$T_j^2 = 0(1-z_j) \qquad \forall j\in JP \qquad (C11)$$

Since, from (C11), T_j^2 is always zero, (C6) implies that $T_j = T_j^1,\ \forall j\in JP,$ and we can simplify constraints (C6) to (C11) as follows:

$$T_j = \textstyle\sum_t T_j^t \qquad \forall j\in JP \qquad (C7')$$

$$\textstyle\sum_t zI_{jt} = z_j \qquad \forall j\in JP \qquad (C8')$$

$$HT_{t-1}\, zI_{jt} \le T_j^t \le HT_t\, zI_{jt} \qquad \forall j\in JP,\ \forall t\in T \qquad (C9')$$

$$T_j \le U z_j \qquad \forall j\in JP \qquad (C10')$$

Furthermore, if $z_j = 0$, (C9') implies that $T_j^t = 0\ \forall t\in T$. If $T_j^t = 0\ \forall t\in T$, then (C7') implies that $T_j = 0$. Constraint (C10'), therefore, is redundant **(*)**. Thus, the convex hull formulation of disjunction (65) consists of constraints (C7') to (C9') which correspond to constraints (64)-(68) of the model.

* This means that constraint (B21) of the previous derivation is also redundant.

Literature Cited

(1) Schmidt, C.W.; Grossmann, I.E. Optimization Models for the Scheduling of Testing Tasks in New Product Development. *Ind. Eng. Chem. Res*. **1996**, 35, 3498-3510.
(2) Jain, V; Grossmann, I. E. Resource-constrained Scheduling of Tests in New Product Development. *Ind. Eng. Chem. Res.* **1999**, 38, 3013-3026.
(3) Honkomp, S.J.; Reklaitis, G.V.; Pekny, G.F. Robust Planning and Scheduling of Process Development Projects Under Stochastic Conditions. AIChE Annual Meeting, Los Angeles, CA, 1997.
(4) Subramanian, D.; Pekny, J.F.; Reklaitis, G.V. A Simulation-Optimization Framework for Addressing Combinatorial and Stochastic Aspects of an R&D Pipeline Management Problem. *Computers Chem*. *Engng*. **2000**, 24, 1005-1011.
(5) Blau, G.; Metha, B.; Bose, S.; Pekny, J.; Sinclair, G.; Keunker, K.; Bunch, P. Risk Management in the Development of New Products in Highly Regulated Industries. *Computers Chem*. *Engng*. **2000**, 24, 659-664.
(6) Bassett, M. Assigning Projects to Optimize the Utilization of Employees' Time and Expertise. *Computers Chem*. *Engng*. **2000**, 24, 1013-1021.
(7) Elmaghraby, S.E.; Herrolen, W.S. The Scheduling of Activities to Maximize the Net Present Value of Projects. *Eur. J. Opr. Res.* **1990**, 49, 35.
(8) Papageorgiou, L.G.; Rotstein, G.E.; Shah, N. Strategic Supply Chain Optimization for the Pharmaceutical Industries. *Ind. Eng. Chem. Res.* **2001**, 40, 275-286.
(9) Shah, N.; Samsalti, N.J.; Sharif, M.; Borland, J.N.; Papageorgiou, L.G. Modeling and Optimization for Pharmaceutical and Fine Chemical Process Development. AIChE Symp. Ser., (323), 96, 31-35.
(10) Sahinidis, N.V.; Grossmann, I. E. Multiperiod Investment Model for Processing Networks with Dedicated and Flexible Plants. *Ind. Eng. Chem. Res.* **1991**, 30, 1165-1171.
(11) Norton, L. C.; Grossmann, I. E. Strategic Planning Model for Complete Process Flexibility. *Ind. Eng. Chem. Res.* **1994**, 33, 69-76.
(12) Henning, G.P.; Camussi, N.B.; Cerda, J. Design and Planning of Multipurpose Plants Involving Nonlinear Processing Networks. *Computers Chem*. *Engng*. **1994**, 18, 129-152.
(13) Reklaitis G.V. Progress and Issues in Computer-aided Batch Process Design. FOCPD Proc., 241-275. Elsevier: NY, 1990.
(16) Straub, D. A.; Grossmann, I. E. Integrated Stochastic Metric of Flexibility for Systems with Discrete State and Continuous Parameter Uncertainties. *Computers Chem. Engng.* **1990**, 14, 967-985.
(15) Birge, J.R.; Louveaux, F. *Introduction to Stochastic Programming;* Springer: New York, 1997.
(16) Balas, E. Disjunctive Programming and Hierarchy of Relaxations for Discrete Optimization Problems. *SIAM J. Alg. Disc. Math.* **1985**, 35, 2611-2623.
(17) Turkay, M.; Grossmann, I.E. Disjunctive Programming Techniques for the Optimization of Process Systems with Discontinuous Investment Costs-Multiple Size Regions. *Ind. Eng. Chem. Res.* **1996**, 35, 2611-2623.
(18) Raman, R.; Grossmann, I.E. Relation between MILP Modeling and Logical Inference for Chemical Process Synthesis. *Computers Chem. Eng*ng. **1991**, 15, 73-84.
(19) Fisher, M.L. The Lagrangean Relaxation Method for Solving Integer Programming Problems. *Management Science* **1981**, 27, 1-18.
(20) Guignard, M.; Kim, S. Lagrangean Decomposition: A Model Yielding Stronger Lagrangean Bounds. *Mathematical Programming* **1987**, 39, 215-228.
(21) Gupta, A.; Maranas, C.D. A Hierarchical Lagrangean Relaxation Procedure for Solving Midterm Planning Problems. *Ind. Eng. Chem. Res.* **1999**, 38, 1937-1947.
(22) Asenjo, J.A.; Montana, J.M.; Vecchietti, A.R.; Iribarren, O.A.; Pinto, M.P. Strategies for the Simultaneous Optimization of the Structure and the Process Variables of a Protein Production Plant. *Computers Chem. Engng.* **2000**, 24, 2277-2290.