Optimal Production Planning under Time-sensitive Electricity Prices for Continuous Power-intensive Processes

Sumit Mitra,[*] Ignacio E. Grossmann[*][†] Jose M. Pinto[‡] Nikhil Arora[§]

June 16, 2011

Abstract

Power-intensive processes can lower operating expenses when adjusting production planning according to time-dependent electricity pricing schemes. In this paper, we describe a deterministic MILP model that allows optimal production planning for continuous power-intensive processes. We emphasize the systematic modeling of operational transitions, that result from switching the operating modes of the plant equipment, with logic constraints. We prove properties on the tightness of several logic constraints. For the time horizon of one week and hourly changing electricity prices, we solve an industrial case study on air separation plants, where transitional modes help us capture ramping behavior. We also solve problem instances on cement plants where we show that the appropriate choice of operating modes allows us to obtain practical schedules, while limiting the number of changeovers. Despite the large size of the MILPs, the required solution times are small due to the explicit modeling of transitions.

1 Introduction

The profitability of industrial power-intensive processes is affected by the availability and pricing of electricity supply. Nowadays, two major trends increase the complexity of managing power-intensive processes. First, deregulation in the 1990s introduced hourly as well as seasonal variations. Second, the environmental pressure to reduce CO_2 emissions and diminishing natural resources lead to an increasing share of renewable energies, which

---

[*]Center for Advanced Process Decision-making, Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, PA 15213

[†]Corresponding author. Email address: grossmann@cmu.edu

[‡]Praxair Inc., Danbury, CT 06810

[§]Praxair, Inc., Tonawanda, NY 14150

intensifies the aforementioned problem. These trends have added a considerable amount of uncertainty and variability in the daily operating expenses of power-intensive industries, which in turn affect their competitiveness.

One important component of the current and future power system is the concept of Demand Side Management (DSM), consisting of Energy Efficiency (EE) and Demand Response (DR). A report released by The World Bank (Charles River Associates, 2005) defines DSM as the "systematic utility and government activities designed to change the amount and/or timing of the customer's use of electricity for the collective benefit of the society, the utility and its customers." While EE aims for permanently reducing demand for energy, DR focusses on the operational level (Voytas et al., 2007). The official classification of DR by the North American Electric Reliability Corporation (NERC) distinguishes between dispatchable and non-dispatchable programs (see Figure 1).

Dispatchable DR programs include any kind of demand response that is according to instructions from a grid operator's control center. They are divided into capacity services, such as load control and interruptible demand, and ancillary services, such as spinning and nonspinning reserves as well as regulation. The control actions, which balance the electricity supply and demand, differ on the time scale and usually range from a few seconds to an hour. Hence, participation in one of these dispatchable DR programs requires the process to be highly flexible, while process feasibility and safety have to be maintained. Nowadays, chemical companies, which operate flexible processes like chlor-alkali synthesis, market already a few percent of their total load as operative capacity reserve (e.g. in Germany (Paulus and Borggrefe (2011)). The potential of ancillary services for aluminum production was recently evaluated in a case study by ALCOA (Todd et al., 2009). However, both processes, chlor-alkali synthesis and aluminum production, are examples of capital-intensive processes that are operated at a high level of capacity utilization. Thus, these processes usually only shift production on a minute level around a predefined setpoint.

In contrast to dispatchable DR programs, non-dispatchable DR programs do not involve instructions from a control center. Instead, the electricity consumption of industrial customers is influenced by the market price of electricity. Typical examples of time-sensitive electricity prices are time-of-use (TOU) rates and real-time prices (RTP). While TOU rates are usually specified in terms of on-peak, mid-peak and off-peak hours, real-time prices vary every hour and are quoted either on a day-ahead or hourly basis. Other pricing models exist but strongly depend on the characteristics of the regional market (NERC study (Voytas et al., 2007).

Non-dispatchable DR programs allow industrial customers to perform production planning based on predefined hourly prices. At first glance it may seem that production planning due to price fluctuations is only attractive for processes that are operated significantly below the process ca-
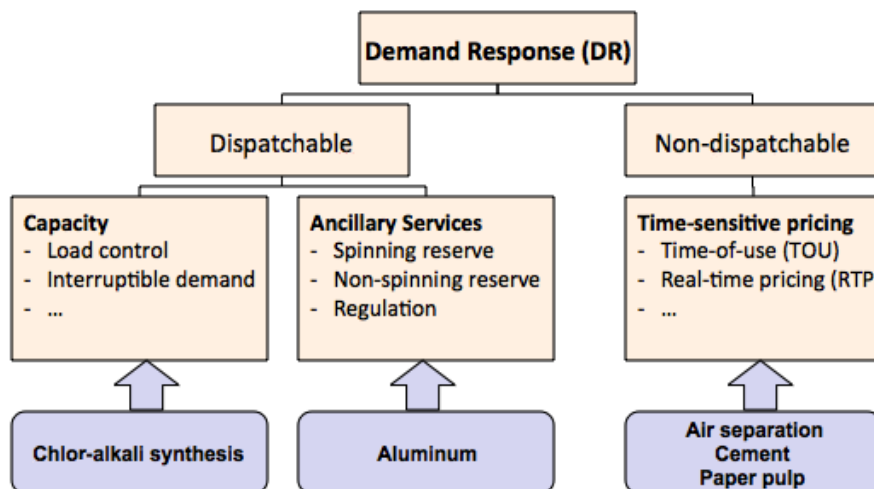
Figure 1: **Illustration of DR applications for chemical processes according to the classification of DR programs by NERC (Voytas et al., 2007); reduced diagram.**

pacity, and therefore have operational flexibility. However, major demand drops due to economic changes, such as the 2008 recession, can lead to over-capacities, which in turn make a systematic production planning more attractive. Promising examples can be found in the industrial gases sector (cryogenic air separation plants) and in the cement industry.

The purpose of this paper is to describe a general model that helps decision-makers for power-intensive production processes to optimize their production schedules with respect to operating costs that are due to fluctuations in electricity prices, which are in turn caused by non-dispatchable DR programs.

2 Literature Review

The economic potential for DSM of industrial processes in developed countries has been recognized by numerous institutions and authors (World Bank report (Charles River Associates, 2005), NERC study (Voytas et al., 2007), Klobasa (2006), Gutschi and Stigler (2008), Paulus and Borggrefe (2011)). In these publications, we can find a list of various chemical processes where the consumption of electricity is due to different unit operations: grinding (cement), compression (air separation), electrolysis (chlor-alkali, aluminum) and drying (paper pulp production). Figure 1 illustrates possible DR applications for these processes. In chemical engineering, two different lines of research address optimizing operations of these processes according to time-sensitive pricing.

The first line of research proposes a control approach. For the economic optimization of air separation plants, Zhu et al. (2010) developed a model based on heat and mass balances, including nonlinear thermodynamics. Huang (2010) proposed the application of economically-oriented nonlinear model predictive control. While these approaches have the advantage of providing an accurate process representation, they cannot readily incorporate discrete decisions such as equipment shutdowns and startups and corresponding operational constraints such as enforcing a minimum up- or downtime. Furthermore, the resulting NLP problems tend to be large, and therefore problems of only 24-48 hours time horizon can be efficiently solved. The problems become even harder to solve when parallel production lines or whole process networks are included.

The second line of research comes from the area of planning and scheduling. An initial approach by Daryanian et al. (1989) shows the economic benefit of demand response for air separation plants, but does not consider discrete operating decisions such as equipment shutdowns and startups. Ierapetritou et al. (2002) as well as Karwan and Keblis (2007) extend the methodology to incorporate discrete operating decisions. However, their models do not capture the transient behavior between different operating modes. Moreover, some of their logic constraints show room for improvement with respect to the tightness of the linear relaxation.

For continuous paper pulp production, Pulkkinen and Ritala (2008) develop a scheduling model to determine the number of operating units and production levels for each time period. They use stochastic constraints to handle uncertainty in model parameters, introducing nonlinearities. However, their optimization procedure is not rigorous in the sense that heuristic methods such as simulated annealing and genetic algorithms are employed. Furthermore, the model does not include a detailed description of plant dynamics in terms of transitions and logic constraints.

For cement plants, Castro et al. (2009, 2011) extend the MILP formulation of the Resource Task Network (RTN) to incorporate hourly changing electricity prices. In their paper from 2009, Castro and co-workers propose a discrete time formulation (DT) that cannot be solved to optimality within a reasonable amount of time for some case studies, although the final gap is small. Additionally, the obtained schedules may contain a significant number of changeovers. Castro and co-workers also show a continuous time formulation (CT), which is unfortunately computationally intractable. In their paper from 2011, Castro and co-workers develop an aggregate formulation (AG) and a rolling horizon formulation (RH). While AG and RH can be solved considerably faster and yield schedules with fewer transitions than DT, the obtained schedules may be suboptimal under power restrictions. Furthermore, AG seems to only adequately work in a setting with time-of-use (TOU) prices, since it aggregates periods with the same price level.

In this paper, we first introduce the generic problem statement for a continuous power-intensive process in section 3. In section 4, we generalize the approaches proposed by Ierapetritou et al. (2002) and Karwan and Keblis (2007). We discuss the concept of transitional modes to allow for a more detailed modeling of transitional behavior that is due to plant dynamics. Furthermore, we enhance the computational efficiency of their formulations by improving the formulation of logic constraints. In section 5, we apply the proposed methodology to a variety of power-intensive processes. We solve an industrial case study on air separation plants where transitional modes help us capture ramping behavior, as well as problem instances on cement plants that are reported in Castro et al. (2009, 2011). We show that the appropriate choice of operating modes allows us to solve the cement problem instances with a discrete time model to optimality within a reasonable amount of time, while limiting the number of changeovers.

The focus of this paper is to obtain an efficient deterministic formulation that can handle transient plant behavior. Therefore, we do not consider uncertainty in electricity prices explicitly. However, it should be mentioned that our formulation can be used as a basis to also incorporate uncertainty in electricity prices, e.g. in a rolling horizon fashion (e.g. Karwan and Keblis (2007)) or with a stochastic programming approach (e.g. Ierapetritou et al. (2002)). Furthermore, note that seasonal electricity tariffs, e.g. time-of-use (TOU) contracts, contain no uncertainty for the timespan of the operational planning (usually a few days or a week). Hence, the deterministic optimization model allows to obtain the optimal production plan for seasonal electricity tariffs.

3 Generic Problem Statement

Given is a set of products $g \in G$ that can be produced in different continuously operated plants or production lines $p \in P$. While some products can be stored on-site ($g \in Storable$), other products must be directly delivered to customers ($g \in Nonstorable$). The plants have to satisfy demand for the products that can be specified on a weekly, daily or hourly basis. The costs of production vary for every hour $h \in H$ and are related to electricity costs, which also undergo seasonal changes. We assume that a seasonal electricity price forecast for a typical week is available on an hourly basis. The problem is to determine production levels, modes of operation (including start-up and shut-down of equipment), inventory levels and product sales on an hourly basis, so that the given demand is met. The objective is to minimize the operating costs that are mostly due to energy expenses.

4 Model Formulation

The traditional way of modeling a process involves heat and mass balances for each individual equipment, which requires the detailed description of the system's performance (e.g. thermodynamics, kinetics). The disadvantage of this approach is that the model can become prohibitively hard to solve due to its nonlinearities and its size for longer time horizons.

An alternative approach is to model the plant in a reduced space, e.g. the product space. To determine the feasible region of the plant in the product space, either a sequence of steady-state simulations or the insights of flexibility theory by Swaney and Grossmann (1985), Grossmann and Floudas (1987) and Ierapetritou (2001) can be used. While it is still useful to have a detailed plant model based on heat and mass balances, we shift the computational burden of evaluating the feasible region to offline computations to obtain a surrogate model. A similar idea is employed by Sung and Maravelias (2007) for batch planning and scheduling.

In absence of a detailed plant model, the surrogate model can also be built from historic plant data. The evaluation of the production levels for the feasible region must be performed according to the chosen time discretization Δt, which is usually on an hourly basis due to the hourly price changes of non-dispatchable DR programs.

The model can be classified into three sets of constraints. The first set deals with the previously mentioned surrogate description for the feasible region in the product space. The second set contains constraints for transitions between different operating points and modes. The third and last set describes the mass balances that capture inventory relationships and demand.

4.1 Production Modes and Feasible Region

To describe the feasible region of a power-intensive plant, the unit operations that consume electricity have to be first identified. The equipment operation involves discrete as well as continuous operating decisions. A discrete operating decision refers to the state of an equipment, e.g. "off", "production mode" or "ramp-up transition". A continuous operating decision refers to production levels, i.e. flowrates of material.

To distinguish discrete operating decisions of a plant, we formally introduce the concept of a mode:

A mode is a set of operating points, for which the same discrete operating decisions are active (i.e. selection of running equipment). The operating points vary in terms of the continuous variables in the product space. The feasible region of each mode is approximated by a convex region consisting of known operating points. In addition, the energy consumption is approximated by a correlation for the entire feasible region of a mode. In any
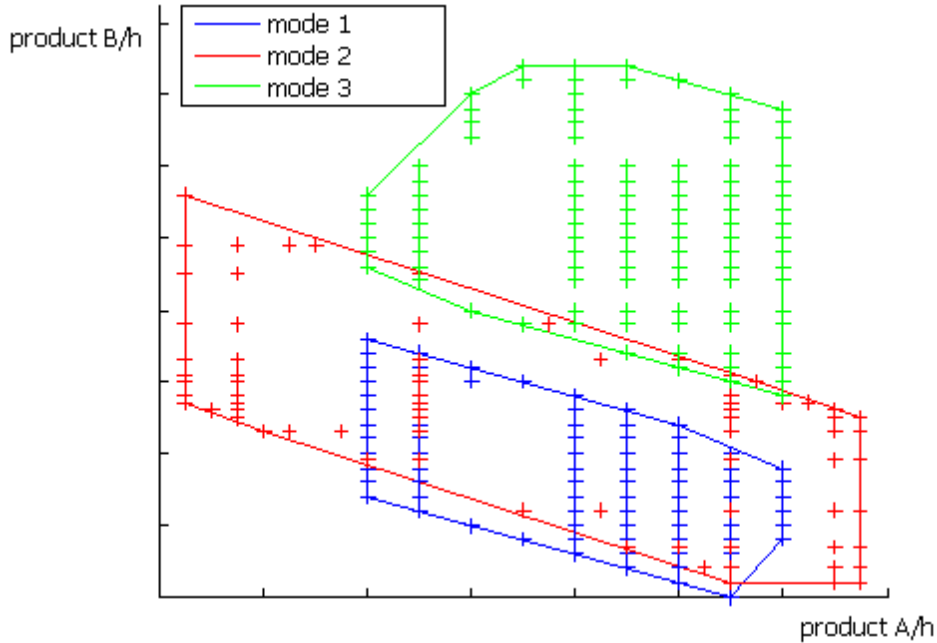
Figure 2: **Feasible region with distinct operating modes.**

given time period only one mode can be active, in other words the modes are disjoint. However, modes may overlap in the product space while usually exhibiting different correlations. In Figure 2, a plant can be seen that produces two products (A and B) having three distinct production modes.

We assume that offline computations or plant measurements provide the data that describes the disjunctive set of modes for the feasible region. We can formulate constraint (1) to describe the disjunction between the modes $m \in M$ for each plant p and every hour h (see nomenclature section). Each term of the disjunction is defined in terms of the binary variable $y_{p,m}^h$, which when true ($y_{p,m}^h = 1$) defines the term that applies in the disjunction. The data for all modes is represented as a collection of operating points (slates). $x_{p,m,i,g}$ are the extreme points i of mode m of plant p in terms of the products g. These extreme points have to be determined a-priori by using an appropriate tool such as MATLAB (The MathWorks Inc., 2010) or PORTA (Christof and Lobel, 1997). The convex combination of the extreme points, with weight factors $\lambda_{p,m,i}^h$, determines the production $Pr_{p,g}^h$ at hour h for each plant p and product g.

$$\bigvee_{m \in M} \begin{pmatrix} \sum_{i \in I} \lambda^h_{p,m,i} x_{p,m,i,g} = Pr^h_{p,g} \quad \forall g \\ \\ \sum_{i \in I} \lambda^h_{p,m,i} = 1 \\ 0 \leq \lambda^h_{p,m,i} \leq 1 \\ y^h_{p,m} = 1 \end{pmatrix} \quad \forall p \in P, h \in H \qquad (1)$$

The disjunction (1) is reformulated using the convex hull (Balas (1985)) to allow for solving the program with an MILP solver. The convex hull reformulation can be written with constraints (2) - (7). Note that $M_{2,p,g,m}$ is the maximum production of product g in mode m.

$$\sum_{i \in I} \lambda^h_{p,m,i} x_{p,m,i,g} = \bar{P}r^h_{p,m,g} \quad \forall p \in P, m \in M, g \in G, h \in H \qquad (2)$$

$$\sum_{i \in I} \lambda^h_{p,m,i} = y^h_{p,m} \quad \forall p \in P, m \in M, h \in H \qquad (3)$$

$$0 \leq \lambda^h_{p,m,i} \leq 1 \quad \forall p \in P, m \in M, i \in I, h \in H \qquad (4)$$

$$Pr^h_{p,g} = \sum_{m \in M} \bar{P}r^h_{p,m,g} \quad \forall p \in P, g \in G, h \in H \qquad (5)$$

$$\bar{P}r^h_{p,m,g} \leq M_{2,p,g,m} y^h_{p,m} \quad \forall p \in P, m \in M, g \in G, h \in H \qquad (6)$$

$$\sum_{m \in M} y^h_{p,m} = 1 \quad \forall p \in P, h \in H \qquad (7)$$

4.2   Transitions

Due to the hourly changing electricity prices, it might be economically desirable to run the plant at different time-dependent operating points. If the operating point is changed, a transition occurs. Transitions deserve special attention because they are closely related to the system dynamics. There are two major types of transitions:

1. Transitions between operating points of different modes - these transitions are described first and are the major focus of this section

2. Transitions between operating points that are both part of the same mode - these transitions are described at the end of this section (rate of change constraints)

4.2.1   Transitions between operating points of different modes

A typical case of a transition between operating points of different modes is the start-up of a plant. During the first few hours the plant might produce
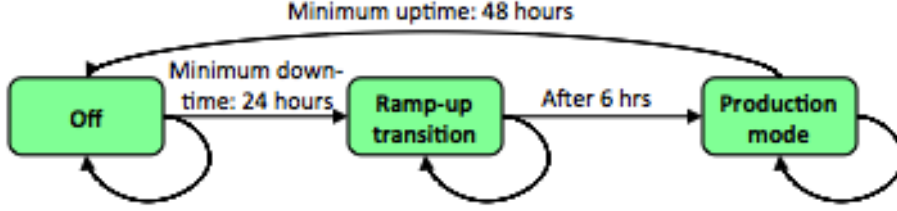
Figure 3: **State graph of a plant with one production and one transitional ramp-up mode. The edges indicate the direction of the allowed transitions and operational constraints of the system.**

a specified amount of off-spec product until a steady-state is reached. Furthermore, the plant cannot be turned off immediately after. The dynamics of the plant in terms of switching behavior can be visualized with a state graph as shown in Figure 3. Each node represents the state of the system. The edges represent the direction of the allowed transitions and show the operational constraints.

We can model the ramp-up characteristics using three modes: "off", "ramp-up transition" and "production mode". In the given example, the plant can switch from "off" to "ramp-up transition" immediately, but has to stay in the "ramp-up transition" mode for 6 hours. Then, a transition occurs to the mode "production mode", in which the plant has to stay at least 24 hours. The plant can switch from "production mode" to mode "off" immediately, but if the plant is shut down, the plant has to stay in the mode "off" at least 48 hours. All transitional constraints faced by the plant operator, who decides on a sequence of states as a control input, have to be expressed in terms of logic constraints.

4.2.1.1 Switch variables constraints To model the transitional constraints efficiently, we introduce the binary transitional variable $z^h_{p,m,m'}$, which is true if and only if a transition from state m to state m' at plant p occurs from time step $h-1$ to h. Hence, the logic relationship to express is the following:

$$(y^{h-1}_{p,m} \wedge y^h_{p,m'}) \Leftrightarrow z^h_{p,m,m'} \quad \forall p \in P, m \in M, m' \in M, h \in H \qquad (8)$$

Using propositional logic (Raman and Grossmann, 1991), the following constraints can be derived for the "$\Leftarrow$" direction:

$$y^{h-1}_{p,m} \geq z^h_{p,m,m'} \quad \forall p \in P, m \in M, m' \in M, h \in H \qquad (9)$$

$$y^h_{p,m'} \geq z^h_{p,m,m'} \quad \forall p \in P, m \in M, m' \in M, h \in H \qquad (10)$$

9

In a similar fashion, the constraint for the "$\Rightarrow$" direction can be derived:

$$y_{p,m}^{h-1} + y_{p,m'}^{h} - 1 \leq z_{p,m,m'}^{h} \quad \forall p \in P, m \in M, m' \in M, h \in H \qquad (11)$$

Note that the constraints (9) - (11) are also used in Ierapetritou et al. (2002) and Karwan and Keblis (2007) to model an equivalent relationship for air separation plants. However, as shown by Sahinidis and Grossmann (1991) as well as by Erdirik-Dogan and Grossmann (2008), the logic relationship (8) can be more effectively modeled by the following set of constraints:

$$\sum_{m' \in M} z_{p,m',m}^{h} = y_{p,m}^{h} \quad \forall p \in P, h \in H, m \in M \qquad (12)$$

$$\sum_{m' \in M} z_{p,m,m'}^{h} = y_{p,m}^{h-1} \quad \forall p \in P, h \in H, m \in M \qquad (13)$$

The reformulation (12)-(13) requires the same number of variables but fewer constraints $((2 * |M|) * |P| * |H|$ vs $3|M|^2 * |P| * |H|)$. Furthermore, its relaxation is tighter. Note that the logic is expressed differently, for more details see the aforementioned references, where the modes are equivalent to tasks in batch scheduling.

4.2.1.2 Minimum Stay Constraint Some modes have a minimum time that the plant has to stay in that particular mode, depending on the previous mode that was active before the plant switched. An example can be found in Figure 3, where the plant has to stay for six hours in mode "ramp-up transition" after it switched from mode "off". Formally, this behavior can be expressed by the following logic relationship:

$$z_{p,m,m'}^{h} \Rightarrow y_{p,m'}^{h+\theta} \quad \forall (p, m, m') \in AL, \forall h \in H, \theta \in MinStay(m, m')$$

Let AL be the set of allowed transitions of plant p from mode m to mode m'. Some of the allowed transitions require a minimum stay constraint. For these transitions, let $MinStay(m, m')$ be the associated placeholder, which describes a set that contains the elements $\left\{0, 1, ..., K_{m,m'}^{min} - 1\right\}$ to model the duration of the minimum stay constraint. $K_{m,m'}^{min}$ is the minimum amount of time the plant has to stay in mode m' after a transition from mode m. Examples for substitutes of $MinStay(m, m')$ include the set TT (to model the duration of a transition), the set UT (minimum uptime for a production mode) and the set DT (minimum downtime after a shut-down). Note that $MinStay(m, m')$ is the empty set for transitions that do not require a minimum stay constraint. The previous logic statement can be converted into the following constraint:

$$y_{p,m'}^{h+\theta} \geq z_{p,m,m'}^h \quad \forall (p,m,m') \in AL, \forall h \in H, \theta \in MinStay(m,m') \qquad (14)$$

Note that Ierapetritou et al. (2002) (constraints (8) and (9) in their paper) and Karwan and Keblis (2007) (constraint (11) in their paper) used a different constraint for the same relationship. With our nomenclature their constraint can be written as:

$$\sum_{\theta=0}^{K_{m,m'}^{min}-1} \sum_{m'' \in M, m'' \neq m'} y_{p,m''}^{h+\theta} \leq K_{m,m'}^{min}(1 - z_{p,m,m'}^h) \quad \forall (p,m,m') \in AL, \forall h \in H$$

$$(15)$$

We show in appendix I that despite fewer number of constraints, formulation (15) is weaker in the LP relaxation compared to that of constraint (14).

4.2.1.3 Maximum Stay Constraint   If the duration that a mode can be operated is restricted, one has to express a maximum stay constraint. An example can be found for an auxiliary equipment that cannot be operated longer than a specified amount of time. Once a transition occurs from mode m to m', the plant cannot stay in mode m' after $K_{m,m'}^{max}$ number of hours. However, the maximum stay constraint only applies if no transition $z_{p,m,m'}^h$ from mode m to m' is active between $h + K_{m,m'}^{min}$ and $h + K_{m,m'}^{max}$. The corresponding logic relationship is:

$$(z_{p,m,m'}^h \wedge \neg z_{p,m'',m'}^{h+K_{m,m'}^{min}} \wedge ... \wedge \neg z_{p,m'',m'}^{h+K_{m,m'}^{max}}) \Rightarrow \neg y_{p,m'}^{h+K_{m,m'}^{max}+1}$$
$$\forall (p,m,m') \in AL, m'' \in M, m'' \neq m', \forall h \in H$$

This logic statement can be reformulated to the following constraint:

$$1 - y_{p,m'}^{h+K_{m,m'}^{max}+1} \geq z_{p,m,m'}^h - \sum_{\theta=K_{m,m'}^{min}}^{K_{m,m'}^{max}} \sum_{m'',m'' \neq m'} z_{p,m'',m'}^{h+\theta} \quad \forall (p,m,m') \in AL, \forall h \in H$$

$$(16)$$

4.2.1.4 Transitional Mode Constraints   As seen in the state graph in Figure 3, the transition from "off" to "ramp-up transition" is coupled with the transition to "production mode" after 6 hours. More formally, let $(p,m,m',m'') \in Trans$ be the set of transitions from mode m to mode m'', which require the plant to stay in the transitional mode m' for a certain specified time $K_{m,m'}^{min}$, as defined before. Aside from a minimum stay constraint
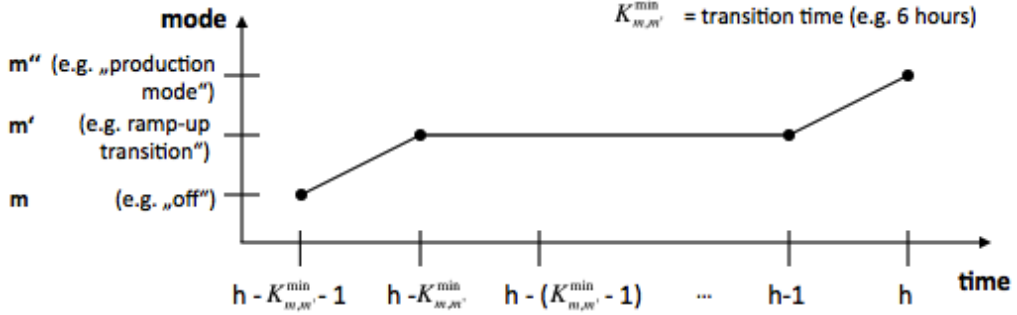
Figure 4: **Illustration of transitions involving transitional mode m' and minimum stay constraint for mode m'. Following the terminology of Figure 3, the plant has to stay $K_{m,m'}^{min}$ (six) hours in mode m' (ramp-up transition) and switches thereafter in mode m'' (production mode).**

for the transitional mode m', the transition from m to m' and the transition from m' to m'' have to be coupled, which is conceptually visualized in Figure 4. In terms of propositional logic, we can express it as follows:

$$z_{p,m,m'}^{h-K_{m,m'}^{min}} \Leftrightarrow z_{p,m',m''}^{h} \quad \forall (p,m,m',m'') \in Trans, \forall h \in H$$

This can be written as:

$$z_{p,m,m'}^{h-K_{m,m'}^{min}} - z_{p,m',m''}^{h} = 0 \quad \forall (p,m,m',m'') \in Trans, \forall h \in H \qquad (17)$$

4.2.1.5 Forbidden Transitions Since the logic relationship (8) implies $\neg(y_{p,m}^{h-1} \wedge y_{p,m'}^{h}) \Leftrightarrow \neg z_{p,m,m'}^{h} \quad \forall p,m,m',h$, it is sufficient to fix $z_{p,m,m'}^{h}$ to zero for transitions that are not allowed (denoted by the set DAL):

$$z_{p,m,m'}^{h} = 0 \quad \forall (p,m,m') \in DAL, \forall h \in H \qquad (18)$$

The concept of the $z_{p,m,m'}^{h}$ variables is very useful to describe systems with slow dynamics, i.e. where the impact of transitions on later time periods has to be modeled. However, if the systems dynamics are fast and there are no minimum stay restrictions, then the $z_{p,m,m'}^{h}$ variables can be avoided. Thus, forbidden transitions have to be modeled in a slightly different way.

Let us denote the set of all combinations for possible transitions of plant p from mode m to mode m' by $(p,m,m') \in PT$. We can partition PT into two disjoint subsets: The subset of allowed transitions AL and the subset of disallowed transitions DAL. Note that $AL \cap DAL = \emptyset$.

To model forbidden transitions, one of the following logic relationships, which are equivalent since AL and DAL are disjoint sets and can be translated by negating, has to be satisfied.

$$y_{p,m'}^h \Rightarrow \bigvee_{m:(p,m,m')\in AL} y_{p,m}^{h-1} \quad \forall p \in P, m' \in M, \forall h \in H \tag{19}$$

$$y_{p,m'}^h \Rightarrow \bigwedge_{m:(p,m,m')\in DAL} \neg y_{p,m}^{h-1} \quad \forall p \in P, m' \in M, \forall h \in H \tag{20}$$

These logic statements (19) and (20) can be expressed by the corresponding constraints (21) and (22):

$$\sum_{m:(p,m,m')\in AL} y_{p,m}^{h-1} \geq y_{p,m'}^h \quad \forall p \in P, m' \in M, \forall h \in H \tag{21}$$

$$y_{p,m}^{h-1} + y_{p,m'}^h \leq 1 \quad \forall (p,m,m') \in DAL, \forall h \in H \tag{22}$$

Constraint (21) has the advantage that fewer constraints in total are needed ($|M| * |P| * |H|$ instead of $|M| * |M'| * |P| * |H|$ ($M' =$ disallowed m')) and it is furthermore tighter than constraint (22). The proof can be found in appendix II.

Note that Ierapetritou et al. (2002) do not consider the case of forbidden transitions. Karwan and Keblis (2007) model forbidden transitions using constraint (22).

4.2.2 Transitions between operating points that are both part of the same mode

4.2.2.1 Rate of change constraints From a control perspective, we have assumed so far that for a given discrete operating decision, each point of the feasible region is reachable within the given time discretization. However, for transitions between operating points that are both part of the same mode, the rate of change might be restricted by a parameter $r_{p,m,g}$. To model these restrictions in terms of the production variables $\bar{P}r_{p,m,g}^h$, the following constraints on the production levels from hour h to hour $h+1$ can be formulated:

$$\bar{P}r_{p,m,g}^{h+1} - \bar{P}r_{p,m,g}^h \leq r_{p,m,g} \quad \forall p \in P, m \in M, g \in G, h \in H \tag{23}$$

Note that $r_{p,m,g}$ is specified in $[mass/\Delta t]$, where Δt is the chosen time discretization (e.g. one hour).

4.3 Mass balances

A mass balance has to be enforced to describe the relationship between production levels, inventory levels and satisfied demand. The demand can be specified on an hourly, daily or weekly basis. The products G can be partitioned into two sets: the storable products *Storable* and the non-storable products *Nonstorable*.

The amount of each product g that is stored at each plant p for each hour h is denoted by $INV_{p,g}^h$ and the hourly outputs for each product are described by $S_{p,g}^h$. The mass balance for the storable products is:

$$INV_{p,g}^h + Pr_{p,g}^h = INV_{p,g}^{h+1} + S_{p,g}^h \quad \forall p \in P, g \in Storable, h \in H \qquad (24)$$

The inventory levels $INV_{p,g}^h$ are constrained by the outage levels ($INV_{p,g}^L$, lower bound) and by the storage capacities ($INV_{p,g}^U$, upper bound):

$$INV_{p,g}^L \le INV_{p,g}^h \le INV_{p,g}^U \quad \forall p \in P, g \in Storable, h \in H \qquad (25)$$

If a product g cannot be stored, the inventory level is simply $INV_{p,g}^h = 0, \quad \forall h$ and the mass balance becomes:

$$Pr_{p,g}^h = S_{p,g}^h \quad \forall p \in P, g \in Nonstorable, h \in H \qquad (26)$$

We have to satisfy demand, that is usually specified on an hourly basis:

$$S_{p,g}^h \ge d_{p,g}^{h,hourly} \quad \forall p \in P, g \in G, h \in H \qquad (27)$$

However, it is also possible to specify demand for the storable products on a daily or weekly basis ($d_{p,g}^{weekly}$, $d_{p,g}^{daily}$), e.g. if the exact sales points for the products are unknown:

$$\sum_{h \in H_{week}} S_{p,g}^h \ge d_{p,g}^{weekly} \quad \forall p \in P, g \in Storable \qquad (28)$$

$$\sum_{h \in H_{day}} S_{p,g}^h \ge d_{p,g}^{daily} \quad \forall day, p \in P, g \in Storable \qquad (29)$$

Note that stockouts are not allowed, but the extension can easily be made by introducing slack variables in the demand constraints with corresponding penalty terms in the objective function. Furthermore, the notation used above assumes that every plant p has its own demand. However, it is also possible to aggregate demand of the customers for multiple plants. Due to space limitation the corresponding equations are omitted.

4.4　Objective Function

The objective function minimizes the total cost that consists of production cost, inventory cost and transition cost for every hour h. Hence, we need a correlation between the power consumption and the production levels $\bar{P}r^h_{p,m,g}$. We denote by $\Phi_{p,m,g}$ the coefficients that have to be determined by a multivariate regression. The inventory cost coefficients are denoted by $\delta_{p,g}$ and the cost coefficients related to transition cost are described by $\zeta_{p,m,m'}$. Hence, the objective function can be written as follows:

$$
OBJ = \sum_{p,h} e^h_p \left(\sum_{m,g} \Phi_{p,m,g} \bar{P}r^h_{p,m,g} \right) + \sum_{p,g} \delta_{p,g} \sum_h INV^h_{p,g} + \sum_{p,m,m'} \zeta_{p,m,m'} \sum_h z^h_{p,m,m'}
$$

(30)

4.5　Power Availability Restrictions

If the available power is restricted, e.g. due to a contracted power curve, the previously defined correlation between the power consumption and the production levels $\bar{P}r^h_{p,m,g}$ can be used to model the following constraint, where PW^h_{max} is the maximum power consumption at hour h:

$$
\sum_{p,m,g} \Phi_{p,m,g} \bar{P}r^h_{p,m,g} \leq PW^h_{max} \quad \forall h \in H
$$

(31)

5　Case Studies

5.1　Air separation plants

The first case study is on air separation plants, which produce liquid oxygen (LO2), nitrogen (LN2) and argon (LAr) as well as gaseous oxygen (GO2) and nitrogen (GN2). All liquid products can be stored on-site in storage tanks. In contrast to that, the gaseous products must be provided to industrial on-site customers according to previously negotiated contracts on an hourly basis and cannot be stored if the air separation plant is in proximity to the consumer. Therefore, plants that are connected to a gas pipeline have fewer degrees of freedom in their daily operation.

In our case study, we optimize two different air separation plants that supply the liquid merchant market without on-site customers. Thus, the power consumption for cryogenic air separation is mostly due to the compressors that are required to provide cooling for liquefaction by exploiting the Joule-Thompson effect. Hence, each operating mode can be related to a liquefier in the air separation plant. If the plant ramps up production, a transitional mode is active in which a specified amount of off-spec product is generated and vented. If the plant has a second liquefier installed and

switches its operation from one to two liquefiers, the plant also undergoes a transitional mode for a few hours.

5.1.1 Model Formulation

The feasible region for each mode is modeled with constraints (2) - (7). The switch variables constraints (12) and (13) are used to link mode variables $y_{p,m}^h$ and transitional variables $z_{p,m,m'}^h$. Minimum stay constraint (14) is used to enforce a minimum uptime once the plant is switched on, a minimum downtime once the plant is switched off and transition times for transitional modes. Constraints (17) and (18) are used to express logic relationships between different modes. The mass balance for the storable liquid products (LO2, LN2, LAr) is expressed by constraints (24) and (25), the mass balance for the non-storable gaseous products (GO2, GN2) is described by constraint (26). The demand for all products is specified on an hourly basis using constraint (27).

5.1.2 Objective function

For air separation plants, Karwan and Keblis (2007) as well as Ierapetritou et al. (2002) suggest linear objective functions. The structure of our linear correlation is the same as the one reported by Ierapetritou et al. (2002), namely between the equivalent liquid rate for liquid production, the gaseous oxygen production, the gaseous nitrogen production and the power consumption. Startup costs are described with respect to the energy consumption of the corresponding transitional modes.

The equivalent liquid rate, $\bar{P}r_{p,LIQ}^h$, is defined as follows, with conversion parameters $\hat{\alpha}_{LN2}, \hat{\alpha}_{LO2}, \hat{\alpha}_{LAR}$:

$$\bar{P}r_{p,m,LIQ}^h = \hat{\alpha}_{LN2}\bar{P}r_{p,m,LN2}^h + \hat{\alpha}_{LO2}\bar{P}r_{p,m,LO2}^h + \hat{\alpha}_{LAR}\bar{P}r_{p,m,LAR}^h \quad (32)$$

A multivariate regression, based on a given set of operating points, was applied to obtain the model parameters for each operating mode of each plant. Note that the statistical significance of this cost model can be measured by several metrics. Student's t-test for a 95% confidence interval reported t-values greater than 2 for all parameters of every plant configuration. Also, the R^2 value, the standard deviation in the residual and the overall F-test were analyzed. With fitting parameters $\alpha_{p,m}, \beta_{p,m}, \gamma_{p,m}$ and the hourly electricity prices at each plant p (e_p^h), the expenses for electricity can be formulated based on (30) as:

$$OBJ = (\sum_{p,m,h} e_p^h(\alpha_{p,m}\bar{P}r_{p,m,LIQ}^h + \beta_{p,m}\bar{P}r_{p,m,GO2}^h + \gamma_{p,m}\bar{P}r_{p,m,GN2}^h)) \quad (33)$$

Note that storage technology allows to keep the liquid products without additional cooling and negligible losses. Hence, we neglect inventory cost. The transitional cost term of (30) is removed because transitional cost are implicitly modeled by a higher energy consumption in transitional modes, which reflects the efficiency losses during mode transitions. Therefore, equation (33) is the objective function of our optimization model, which we intend to minimize.

5.1.3 Results

Based on real industrial data, we modeled two different air separation plants: a plant with one liquefier (P1) and a plant with two liquefiers (P2). We tested the computational performance of our model for five different cases (A-E) that differ in the demand that is specified for the liquid products on a six hour basis (which is similar to assume that a certain number of tanker trucks arrive within that time window) for a time horizon of one week (168 hours). The electricity price forecast of the real-time prices (see Figure 5) was given on an hourly basis. Due to confidentiality issues we cannot disclose the actual LO2 and LN2 demands and production levels. Therefore, scales are omitted, but the total demand is reported as a function of the plants' production capacity in Table 1.

For the same demand data, we solved the problem using time intervals of one hour for two different setups: for a non-cyclic schedule and a cyclic schedule. An optimal non-cyclic schedule might suggest to shut down the plant at the end of the week. This is an undesirable situation if the decision maker can assume that the market conditions will be similar in the following week, because a shut-down hinders successful operations in the following week, e.g. due to minimum downtime restrictions. In contrast to the non-cyclic schedule, the cyclic schedule also satisfies all mass balances and transitional constraints in a cyclic manner throughout the week.

Table 1: **Savings of the optimal noncyclic schedules for all test cases as function of the demand/capacity ratio**

case	demand/ capacity	savings
P1A	82%	4.58%
P1B	74%	12.02%
P2E	95%	3.76%
P2D	85%	4.90%
P2C	72%	7.44%
P2B	51%	13.78%

To assess the economic impact of our optimization model, we compare our solutions for the noncyclic schedules with a heuristic: we assume con-
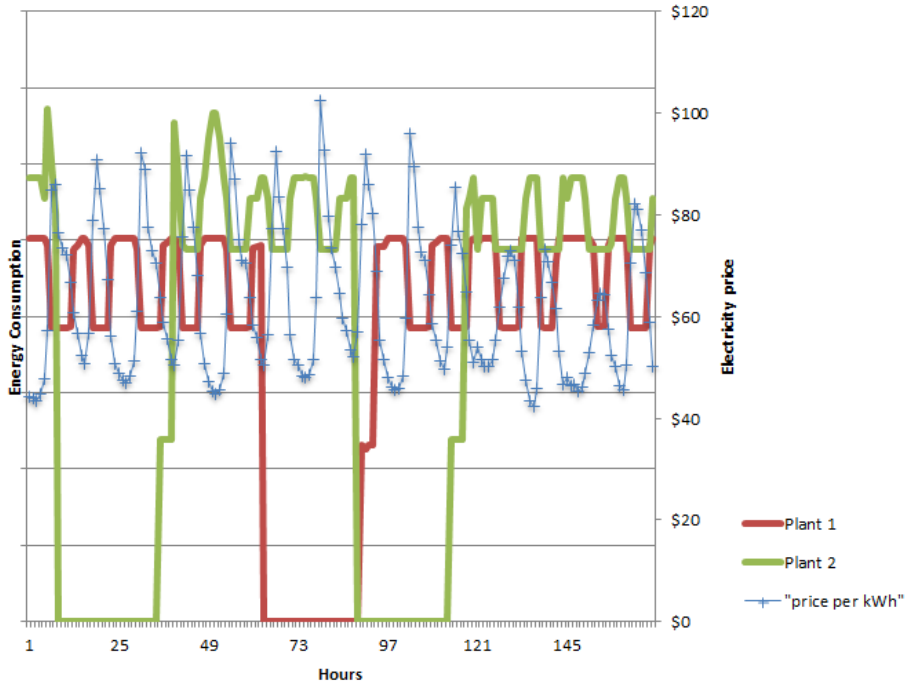
Figure 5: **Electricity prices and energy consumption profiles for optimal cyclic schedules of plants P1 and P2 (test case B).**

stant operation throughout the week and find a set point that locally minimizes the energy consumption while satisfying the demand constraints. In Table 1 we can see that for both plants the potential savings increase with decreasing demand that is reported in percent of the total production capacity. For test cases A and B the realized savings are mostly due to plant shutdowns, which is illustrated in Figure 5 for the electricity consumption profiles of plants P1 and P2 for test case B. We can also see that the electricity consumption is reduced during hours of high electricity prices - due to higher operational flexibility plant P2 can produce 5.2% cheaper than plant P1. The corresponding inventory profiles for plant P1 are reported in Figure 6 and the inventory profiles for plant P2 are shown in Figure 7. The lower bound corresponds to the outage level and the upper bound is the storage tank capacity. We can clearly observe how the demand is served from the storage tank during plant shutdowns.

For test cases C, D and E no shutdowns are reported since the storage tank capacity is not large enough to store the products required for the minimum downtime of the plant. Nevertheless, savings are observed that are due to the presence of plant P2 second liquefier that is turned on during hours of low electricity prices and turned off during hours of peak prices.

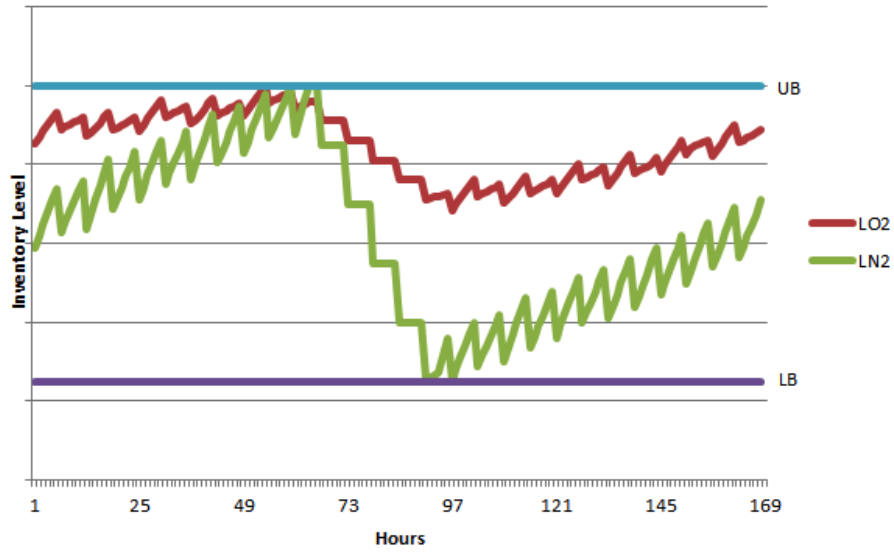In Table 2 the problem sizes are reported. The problem sizes differ signif-

18

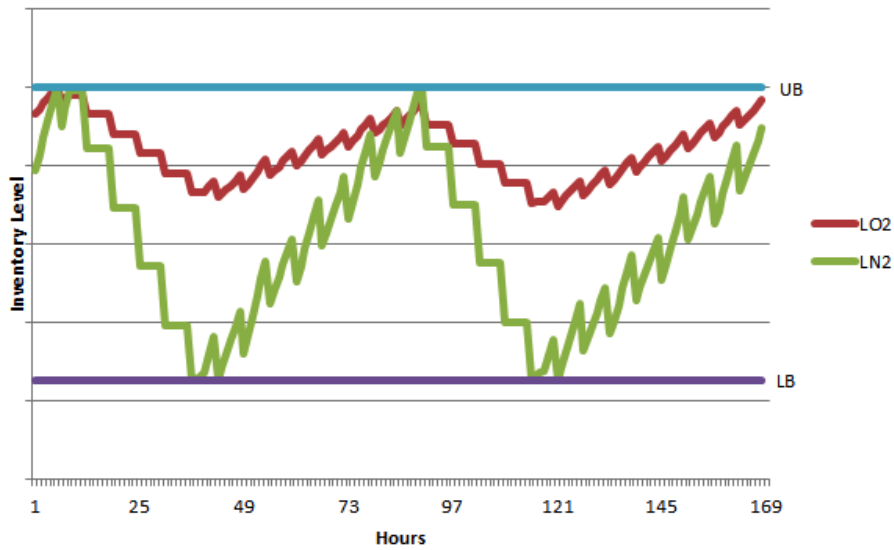Figure 6: **Optimal LO2 and LN2 inventory profiles for plant P1, test case B, cyclic schedule**

Figure 7: **Optimal LO2 and LN2 inventory profiles for plant P2, test case B, cyclic schedule**

Table 2: **Problem sizes for two air separation plants**

plant	type of schedule	# constraints	# variables	# binary
P1	noncyclic	22508	16129	1512
P1	cyclic	22513	16129	1512
P2	noncyclic	44173	29401	3528
P2	cyclic	44185	29401	3528

icantly for the two plants since plant P1 has only three operating modes, in contrast to plant P2, which has six operating modes (including transitional modes). However, the problem sizes for the cyclic and noncyclic schedules differ only slightly, because the implementation for the noncyclic schedule includes a few redundant constraints that could be filtered with conditional statements. However, the pre-solve operations of the MILP solver eliminate those constraints reliably.

Table 3: **Computational results for test cases A-E, plants P1 and P2. cyc: cyclic schedule, ncyc: noncyclic schedule, RMIP gap% shows the tightness of the linear relaxation.**

case	cyc/ncyc	RMIP	MIP		CPU times(s)		RMIP gap%	final gap%	
			CPLEX	*XPRESS*	*CPLEX*	*XPRESS*		*CPLEX*	*XPRESS*
P1A	ncyc	88416	**90129**	**90129**	2	3	1.9006%	0.0000%	0.0000%
P1A	cyc	89197	**90448**	**90448**	3	26	1.3831%	0.0000%	0.0000%
P1B	ncyc	72806	**74932**	**74932**	4	5	2.8372%	0.0000%	0.0000%
P1B	cyclic	73305	**76011**	**76011**	15	27	3.5600%	0.0588%	0.0000%
P2B	ncyc	67659	**70017**	**70017**	135	47	3.3678%	0.0995%	0.0118%
P2B	cyc	67173	**72027**	**72027**	844	462	6.7391%	0.0977%	0.0936%
P2C	ncyc	98645	**100796**	100837	7	10	2.1340%	0.0955%	0.0740%
P2C	cyc	99726	101280	**101242**	6	8	1.4974%	0.0684%	0.0085%
P2D	ncyc	113521	**114715**	**114715**	4	4	1.0408%	0.0076%	0.0000%
P2D	cyc	115173	116125	**116122**	5	6	0.8172%	0.0079%	0.0016%
P2E	ncyc	128019	**128054**	**128054**	4	4	0.0273%	0.0000%	0.0000%
P2E	cyc	130186	**130186**	**130186**	4	4	0.0000%	0.0000%	0.0000%

The computational results can be found in Table 3. We solved all test cases with the commercial solvers CPLEX 12.2.0.1 and XPRESS 21.01 (both with default settings, no parallel computing features were used) in GAMS 23.6.2 on a Intel i7 (2.93GHz) machine with 4GB RAM, using a termination criterion of 0.1% optimality gap. Except for test case P2B, all test cases can be solved within less than 30 seconds. CPLEX and XPRESS perform similarly, except for test cases A and B with advantages on both solvers. We can observe that with increasing demand the problems become easier to solve since operational flexibility decreases. This can be seen in terms of CPU time and also in terms of the tightness of the LP relaxation (RMIP). For cases P1A and P2E with the highest demands the initial gap is in fact 0 %. Furthermore, the cyclic schedule is harder to obtain when compared to the non-cyclic schedule.

5.2 Cement plant

Castro et al. (2009 and 2011) study the case of cement plants, where the final processing stage is a power-intensive grinding operation. Based on the resource task network (RTN) formulation, Castro and co-workers present solution methods that most importantly differ in their time representation (see literature review). One of the time representations they use is a discrete formulation with a one-hour time grid. They outline three major issues with this approach. First, due to the time discretization the obtained schedules might be slightly suboptimal because changeovers cannot take place between two time points. Second, due to solution degeneracy the obtained schedules might have a large number of undesirable changeovers. Third and lastly, the presence of shared storage adds even more symmetry to the problem. Altogether these issues lead to the observation that the discrete time formulation of Castro et al. cannot be solved to optimality within a reasonable time. However, they report only a small final gap in all their test cases.

In the following we will describe how our framework can be used to model the cement plant with a discrete time formulation that can be solved to optimality within a reasonable amount of time, while limiting the number of changeovers.

5.2.1 Model Formulation

We make the same basic assumptions that are reported in Castro et al. (2009) and use the same model parameters. We model the parallel lines as individual manufacturing facilities denoted by index p. In all cases, we use constraints (2) - (6) to define the feasible region of the operating modes. However, the main challenge is the appropriate definition of the operating modes. This "procedure" can be seen as choosing how and which of the capacity constraints of plant p are projected in the product space for each mode.

5.2.1.1 Single Product Modes (model SPM) The most intuitive way to define the operating modes is to have one mode for each product. The concept is illustrated for two products in the left part of Figure 8, where each node represents a single product mode ("Mode P1" and "Mode P2"). The feasible region of each mode m is the interval $[0, \rho_{p,g}^{max}]$, where $\rho_{p,g}^{max}$ is the maximum processing rate of production line p of product g. This approach is very similar to the DT formulation of Castro et al. since we only allow for transitions at the predefined event points (every hour). Obviously, the number of modes in model SPM scales linearly with the number of products $|G|$.
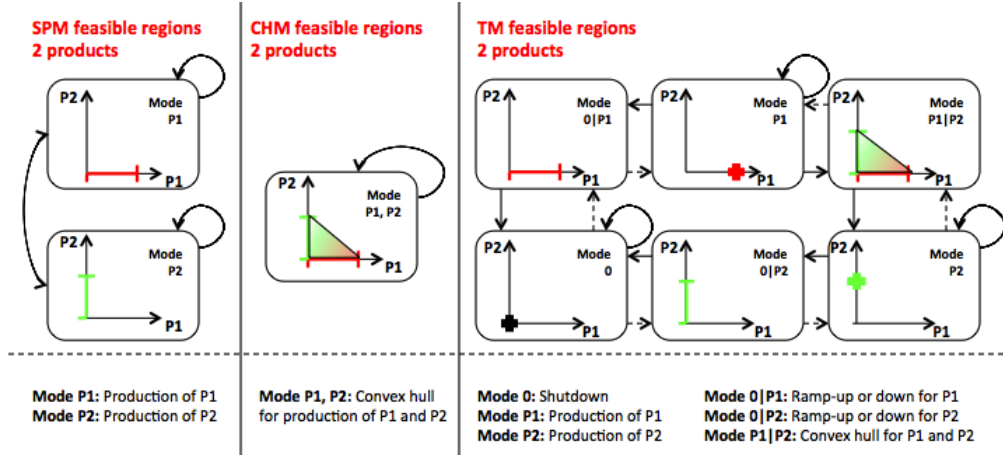
Figure 8: **Different choices of defining modes and corresponding feasible regions for a cement plant with two products. SPM: model with single product modes, CHM: model with convex hull (one mode), TM: model with transitional modes, single product modes and zero production mode.**

5.2.1.2 Convex Hull Mode (model CHM) The feasible region for each production line p can also be described with only one mode, as illustrated in the middle part of Figure 8 by "Mode P1, P2". "Mode P1, P2" is the convex hull of all maximum production levels for each product and the origin. In other words, all capacity constraints of plant p are projected in the product space for one single mode. The feasible region of model CHM for two products is also shown in the middle part of Figure 8. Note that this definition of the feasible region allows for product transitions at any point in time and there are no restrictions on the batch processing time. Hence, we can obtain the lowest possible production costs. The drawback of this feasible region is that we cannot limit the number of transitions that are occurring since the transitions are not linked to any binary variables.

5.2.1.3 Transitional Modes (model TM and extensions) If we want to combine SPM with the ability to allow for transitions at any point in time, we need to introduce ideas from model CHM, e.g. transitional modes that account for changeovers between two products. In the right part of Figure 8, model TM is illustrated with the state graph for two products. As one can see in the upper right corner of , model CHM is used for each pair of products to derive the corresponding feasible region of the transitional mode ("Mode P1|P2"). More formally, the feasible region of the transitional mode m of production line p that accounts for the transition from product g to g' is the convex combination of the origin, $\rho_{p,g}^{max}$ and $\rho_{p,g'}^{max}$. We modify the feasible region of the single product modes ("Mode P1" and "Mode P2") as
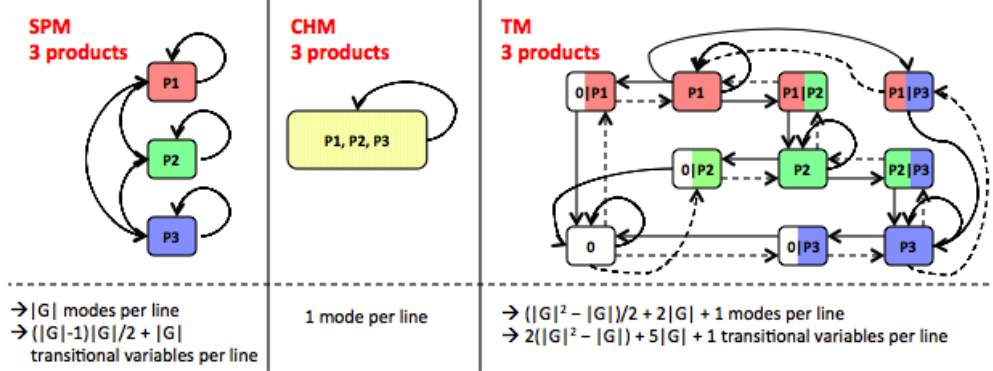
22

Figure 9: **Scalability of different choices for operating modes for cement plants as a function of the number of products $|G|$, here illustrated for three products. In the lower part of the figure, the number of modes and transitional variables per production line is reported. SPM: model with single product modes, CHM: model with convex hull (one mode), TM: model with transitional modes, single product modes and zero production mode.**

well. In order to avoid batch interruptions that may result from a definition of the feasible region as in SPM, we define the feasible region for each product g as a single point, namely $[\rho_{p,g}^{max}]$. Therefore, we have to introduce a "Mode 0" corresponding to a shutdown of the plant, which only contains $[0]$ in the feasible region. We also define transitional modes ("Mode 0|P1" and "Mode 0|P2") to allow for production ramp-ups at any point in time. The feasible region of these transitional modes is the interval $[0, \rho_{p,g}^{max}]$. We use the switch variables constraints (12) and (13) to link mode variables $y_{p,m}^h$ and transitional variables $z_{p,m,m'}^h$. Forbidden transitions are implemented with constraint (18). To ensure the coupling of transitions between transitional and single product modes as well as shutdown, transitional and production modes, constraint (17) is applied. In Figure 8, for each transitional mode the two transitions with dashed lines as well as the two transitions with solid lines are coupled as previously described in section 4.2.4.

With the setup of modes in model TM, we can limit the number of transitions in different ways:

1. **Model TM+MS**, where we utilize the minimum stay constraint (14) to enforce a minimum uptime once the plant is switched on and a minimum downtime once the plant is switched off (e.g. at least two subsequent time periods).

2. **Model TM+RT**, which restricts the total number of transitions that occur.

23

3. **Model TM+MSRT**, which is the combination of both aforementioned methods.

Alternatively, changeover costs can be introduced, which we do not further explore in this paper.

In Figure 9, we illustrate how the number of modes increases depending on the number of products $|G|$ with the case of three products. We observe that despite model TM growing $(|G|^2 - |G|)/2 + 2|G| + 1$ in the number of modes as a function of the number of products $|G|$, the computational results show favorable CPU times. Note that the restricted power availability can be described in terms of equation (32).

5.2.2 Mass balances for shared storage constraints

To model the shared storage and the associated mass balances we need additional constraints. $X_{st,g}^h$ is a binary variable that indicates which product g is stored in storage st at hour h. $Store_{p,st,g}^h$ keeps track how much of product g is transferred from production line p to storage tank st in hour h. Therefore, independent of the operating modes, the material transfers are tracked on an hourly basis. The shared storage constraints can be written as:

$$Pr_{p,g}^h = \sum_{st \in ST} Store_{p,st,g}^h \quad \forall p \in P, g \in G, h \in H \tag{34}$$

$$INV_{st,g}^h + S_{st,g}^h \leq INV_{st}^U X_{st,g}^h \quad \forall st \in ST, g \in G, h \in H \tag{35}$$

$$\sum_g X_{st,g}^h = 1 \quad \forall st \in ST, h \in H \tag{36}$$

$$INV_{st,g}^{h-1} + \sum_p Store_{p,st,g}^h = INV_{st,g}^h + S_{st,g}^h \quad \forall st \in ST, g \in G, h \in H \tag{37}$$

In addition, we use constraint (27) to enforce the hourly specified demand for each product. The objective function is based on (30). The terms for inventory cost and transition cost are neglected, analogous to the papers by Castro and co-workers (2009, 2011).

5.2.3 Results

We compare the different model choices for the operating modes with respect to the known issues of the discrete time scheduling model of Castro et al. (2009): potentially suboptimal solutions, including a small final gap since optimality cannot be proven within a reasonable amount of time, as well as large number of changeovers. All test cases were solved with the commercial

solver GUROBI 4.0.0 (with default settings, no parallel computing features were used) in GAMS 23.6.2 on a Intel i7 (2.93GHz) machine with 4GB RAM, using a termination criterion of 10^{-6} optimality gap (as in Castro et al. (2009)). GUROBI was found to be more efficient than CPLEX and XPRESS on these problems. Our computational results for test cases Ex5-Ex10 (the "harder" problem instances) are reported in Table 4.

5.2.3.1 Single Product Modes (model SPM)   The first observation is that model SPM yields exactly the same suboptimal objective values as the discrete time formulation of Castro et al. (2009), except for Ex8. Moreover, the number of binary variables, which indicate the status of the equipment and the storage tanks, is the same. Despite the newer hard- and software, we cannot solve the test cases to optimality within 2 hours neither (except Ex7).

5.2.3.2 Convex Hull Mode (model CHM)   On the contrary, model CHM yields the same optimal objective values as the ones reported in Castro et al. (2011) for all test cases. Since each production line has only one operating mode, the problem size is significantly smaller and CHM can be solved in less than 20s for all test cases. Unfortunately, the number of transitions cannot be controlled with model CHM, which leads to schedules that are hard to implement.

5.2.3.3 Transitional Modes (model TM and extensions)   The computational results show that the ability to obtain the optimal solution due to the introduction of transitional modes in conjunction with the tightness of the linear relaxation are key to the success of model TM.

Test cases Ex5-7 and Ex9 can be solved to optimality in less than 30s, yielding the same objective value as CHM. For test case Ex8, TM yields a slightly suboptimal solution within 420s. This is due to the fact that TM implicitly assumes a batch processing time of at least one hour, but Ex8 (restricted power availability) requires a few batches with batch processing times less than 1 hour. If we use the convex hull mode instead of the previously described transitional mode for production ramp-ups, we can allow for batch processing time less than one hour and we obtain the optimal objective value within 36s (model TMb).

For test case Ex10, the optimal solution of TM requires 4079s. We can speed up the solution process if we strengthen the relaxation of TM by adding the objective value that we computed with model CHM as a lower bound. The corresponding **model TMS** can be solved within 118s (including 17s from CHM).

We use **model TM+MS** to enforce a minimum uptime once the plant is switched on and a minimum downtime once the plant is switched off, both

Table 4: Computational results for cement production. SPM: single product modes, CHM: convex hull mode, TM: model with transitional modes, single product modes and zero production mode, TMS: TM strengthened with bound from CHM. Extensions: +MS: minimum stay constraints on up- and downtime, +RT: restricted number of transitions, +MSRT: both aforementioned restrictions. *=Castro et al. (2011), Intel Core2 Duo T9300 (2.5 Ghz), 4GB RAM, Windows Vista Enterprise, CPLEX 11.1 with default options; **=maximum computational time; [a]=17s required to solve CHM already included.

Case	Model	constr.	vars.	bin.	RMIP	MIP	CPUs	Gap%
Ex5	*Castro DT*	*6739*	*10953*	*2016*	*26738*	*26780*	*7200**	*0.04*
	SPM	12433	12097	2016	26738	26780	7200**	0.02
	CHM	6385	8401	1008	26738	**26758**	1	0
	TM	58389	59473	13776	26738	**26758**	5	0
	TM+MS	62421	59473	13776	26738	26791	15	0
	TM+RT	58390	59473	13776	26738	**26758**	5	0
	TM+MSRT	62422	59473	13776	26738	26791	19	0
Ex6	*Castro DT*	*8423*	*14155*	*2520*	*43250*	*43259*	*7200**	*0.02*
	SPM	13609	14617	2520	43250	43259	7200**	0.02
	CHM	7561	10921	1512	43250	**43250**	1	0
	TM	59565	61993	14280	43250	**43250**	4	0
	TM+MS	63597	61993	14280	43250	**43250**	8	0
	TM+RT	59566	61993	14280	43250	**43250**	16	0
	TM+MSRT	63598	61993	14280	43250	**43250**	44	0
Ex7	*Castro DT*	*10780*	*18534*	*3528*	*68282*	***68282***	*19.9**	*0*
	SPM	19489	22681	3528	68282	**68282**	1	0
	CHM	10417	17137	2016	68282	**68282**	1	0
	TM	88423	93745	21168	68282	**68282**	8	0
	TM+MS	94471	93745	21168	68282	**68282**	8	0
	TM+RT	88424	93745	21168	68282	**68282**	12	0
	TM+MSRT	94472	93745	21168	68282	**68282**	11	0
Ex8	*Castro DT*	*12464*	*21736*	*4032*	*101139*	*104622*	*7200**	*0.22*
	SPM	20665	25705	4032	104375	104607	7200**	0.09
	CHM	11593	20161	2520	104375	**104375**	5	0
	TM	89599	96769	21672	104375	104502	420	0
	TMb	89599	99793	21672	104375	**104375**	36	0
	TMb+MS	95647	99793	21672	104570	104891	350	0
	TMb+RT	89600	99793	21672	104375	**104375**	115	0
	TMb+MSRT	95648	99793	21672	104570	104891	852	0
Ex9	*Castro DT*	*13810*	*24092*	*4704*	*87817*	*87868*	*7200**	*0.06*
	SPM	29569	32257	4704	87817	87868	7200**	0.06
	CHM	13441	22681	2688	87817	**87817**	2	0
	TM	149845	184969	32928	87817	**87817**	23	0
	TM+MS	159925	184969	32928	87817	87838	32	0
	TM+RT	149846	184969	32928	87817	**87817**	32	0
	TM+MSRT	159926	184969	32928	87817	87838	101	0
Ex10	*Castro DT*	*16840*	*29650*	*5880*	*86505*	*86582*	*7200**	*0.09*
	SPM	41665	42841	5880	86505	86582	7200**	0.04
	CHM	16465	28225	3360	86505	**86550**	17	0
	TM	232924	331633	47208	86505	**86550**	4079	0
	TMS	232924	331633	47208	86550	**86550**	118[a]	0
	TMS+MS	248044	331633	47208	86550	**86550**	401[a]	0
	TMS+RT	232925	331633	47208	86550	**86550**	734[a]	0
	TMS+MSRT	248045	331633	47208	86550	**86550**	870[a]	0

for at least two subsequent time periods. Despite the additional constraints, we can obtain the same optimal objective value as CHM for Ex6, Ex7 and Ex10. For Ex5, Ex8 and Ex9 the objective is slightly larger. In all cases, the optimization runs terminate in less than 6 minutes.

As previously mentioned, another possibility to limit the number of occurring transitions is restricting the total number of transitions (**model TM+RT**). Thus, we have to find an appropriate value for the total number of allowed transitions. One possibility is counting the number of orders and estimate the number of batches required to fulfill each order. Ramping up production, production of a single batch and shutting down production requires 4 transitions. If we assume that each order needs between 1 and 2 batches, reasonable values for the total number of transitions can be found in the range [4×no.orders, 8×no.orders]. In Table 5, we report the parameter values for the number of allowed transitions that we used for our computational experiments.

Table 5: **Parameter values for the maximum number of allowed transitions for models TM+RT and TM+MSRT that we used in our computational experiments as a function of the number of orders.**

case	Ex5	Ex6	Ex7	Ex8	Ex9	Ex10
# of orders	9	9	17	18	21	22
# of allowed transitions	70	70	120	140	150	150

The computational results for model TM+RT show that we can obtain the optimal solution with the same objective as CHM in less than 2 minutes for all test cases except Ex10, which takes less than 8 minutes to solve.

Combining the minimum up- and downtime constraints with the restriction of the total number of transitions leads to **model TM+MSRT**, which can produce even more practical schedules. For the number of allowed transitions, we use again the values reported in Table 5. The obtained objective values are the same as for model TM+MS, the solution times are longer than the ones reported for model TM+MS but smaller than 15 minutes. In Figure 10 we show a practical schedule obtained with model TM+MSRT for Ex10. One can see that the schedule avoids production during hours of peak prices and has a comparable number of batches to the schedule obtained by the rolling horizon approach in Castro et al. (2011).

6   Conclusions

In this paper, we have presented a model for the optimal operational production planning for continuous power-intensive processes that participate in non-dispatchable demand response programs. We described a determinis-
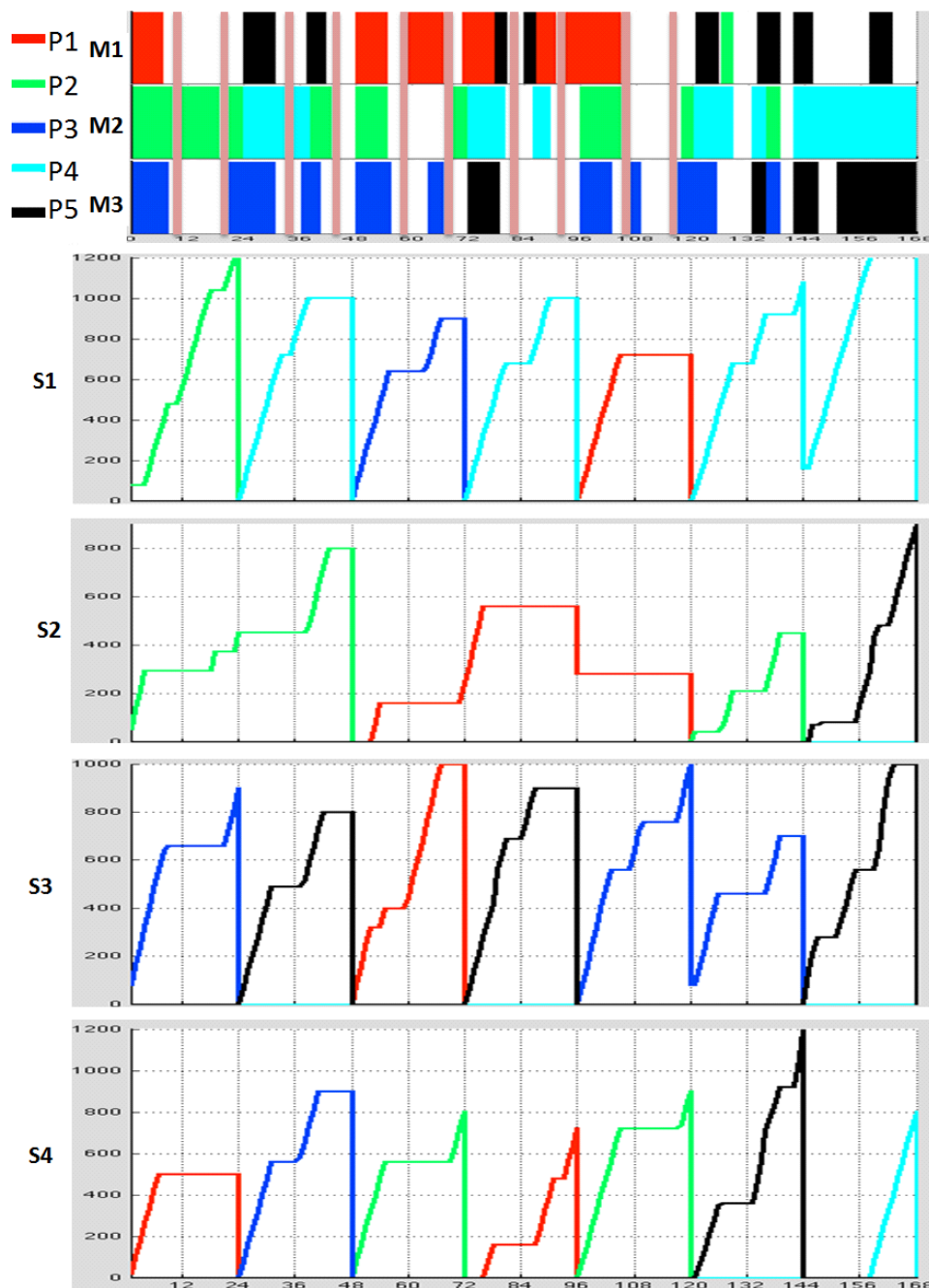
Figure 10: **Cement production: optimal schedule and storage profiles for problem instance Ex10 using model TMS+MSRT. The optimal schedule avoids production during hours of peak prices (brown vertical blocks). M1-M3 are the three parallel production lines, S1-S4 are the four storage tanks, P1-P5 indicate the five different products.**

28

tic MILP model that allows an accurate and efficient modeling of transitions between operating modes using a discrete time representation. Properties on the tightness of several logic constraints were proved. We successfully applied the model to two different real-world air separation plants that supply to the liquid merchant market, as well as cement plants.

In the air separation case study, the operational optimization model produced savings of more than 10% when compared to a simple heuristic. We also learned that operational flexibility, in terms of production and storage capacity, is the key to lower operating expenses. For the same demand profile, a plant with two liquefiers was able to save 5% on costs when compared with a single liquefier plant.

We observed in the case of the cement plant that the introduction of transitional modes resulted in a large-scale model. Nevertheless, the model was superior compared to a smaller model that only had single product modes, and thus lacked the capability of limiting the number of occurring transitions, and could not be solved to optimality. Despite the large size of the MILP model with transitional modes, the required solution times to obtain the optimal solution were small for all test cases. Furthermore, the obtained schedules were practical to implement because we were able to limit the number of occurring transitions.

Acknowledgments

Nomenclature

Sets

- P (index p): The set of plants or production lines

- $M(p)$ (index m), abbreviated as M: The set of modes, depending on plant p

- $I(p, m)$ (index i), abbreviated as I: The set of extreme points that relate to mode m of plant p

- G (index g): The set of products. For air separation plants it is $\{LO2, LN2, LAr, GO2, GN2\}$.

- $Storable \subseteq G$: The subset of products that are storable. For air separation plants it is $\{LO2, LN2, LAr\}$.

- *Nonstorable* $\subseteq G$: The subset of products that are not storable. For air separation plants it is $\{GO2, GN2\}$.

- *ST*: The set of shared storage tanks

- *H* (index *h*): The set of hours of a week in the operational model

- $H_x \subseteq H$: Subset of hours, where *x* can stand for e.g. a certain day

- *MinStay(m,m')* a placeholder for the sets *UT*, *DT* and *TT*

- *UT(p)*, abbreviated as *UT*: The set of hours that a plant *p* has stay online, once it was started

- *DT(p)*, abbreviated as *DT*: The set of hours that a plant *p* has stay offline, once it was shut down

- *TT(p)*, abbreviated as *TT*: The set of hours that a plant stays in an transitional mode

- *Trans*(p, m, m', m''): The set of possible transitions for plant *p* from mode *m* to a production mode m'' with the transitional mode m' in between

- *AL*(p, m, m'): The set of allowed transitions for plant *p* from mode *m* to another mode m'

- *DAL*(p, m, m'): The set of disallowed transitions from mode *m* to mode m' of plant *p*

Variables

Binary variables

- $X_{st,g}^h$: Indicates which product *g* is stored in storage *st* at hour *h* (for shared storage)

- $y_{p,m}^h$: Determines whether plant *p* operates in mode *m* in hour *h*

- $Z_{p,m,m'}^h$: Indicates whether there is a transition from mode *m* to mode m' at plant *p* from hour $h-1$ to *h*

Continuous variables

- $\bar{Pr}_{p,m,g}^h$: Production amount of product *g* in mode *m* at plant *p* in hour *h*

- $Pr_{p,g}^h$: Total production of product *g* at plant *p* in hour *h*

- $\lambda_{p,m,i}^h$: Variable for the convex combination of slates i to describe the feasible region of the plant m of mode m in hour h

- $INV_{p,g}^h$: Inventory level of product g at plant p in hour h

- $S_{p,g}^h$: Sales of product g from plant p in hour h

- $Store_{p,st,g}^h$: Keeps track how much of product g is transferred from production line p to storage tank st in hour h (for shared storage)

- OBJ: Objective function variable

Parameters

- $\hat{\alpha}_{LN2}, \hat{\alpha}_{LO2}, \hat{\alpha}_{LAR}$: Conversion parameters for equivalent liquid rate

- $\alpha_{p,m}, \beta_{p,m}, \gamma_{p,m}$: Cost coefficients for mode m of plant p

- $\delta_{p,g}$: Cost coefficient for inventory of product g at plant p

- $\zeta_{p,m,m'}$: Cost coefficient for transitions from mode m to m' at plant p

- e_p^h: Electricity prices for plant p in hour h

- $x_{p,m,i,g}$: Extreme points of the convex hull of the feasible regions

- $M_{2,p,g,m,o}$: BigM constant for bounds on production for plant p (i.e. max. production of product g of option o in mode m)

- $K_{m,m'}^{min}$: Number of hours the plant has to stay in mode m' after a transition from mode m

- $K_{m,m'}^{max}$: Number of hours the plant can stay at most in mode m' after a transition from mode m

- $r_{p,m,g}$: Maximum rate of change for product g at plant p in mode m

- $d_{p,g}^{daily}, d_{p,g}^{weekly}$: Daily/weekly demand for the products g of plant p. For air separation plants, it is the demand for the liquid products.

- $d_{p,g}^{h,hourly}$: Hourly demand for the products g of plant p in hour h. For air separation plants, it is the demand for the gaseous products.

- $INV_{p,st,g}^U$: Tank capacity of tank st for product g at plant p

- PW_{max}^h: Maximum power consumption at hour h (under restricted power availability)

Appendices for Proofs

I. On the tightness of constraint (14) for the minimum stay relationship

We prove that constraint (14) is tighter in the LP relaxation than constraint (15). If we add over index θ all $K_{m,m'}^{min}$ constraints of constraint (14) we get the following inequality:

$$\sum_{\theta=0}^{K_{m,m'}^{min}-1} y_{p,m'}^{h+\theta} \geq K_{m,m'}^{min} z_{p,m,m'}^h \quad \forall (p,m,m') \in AL, h \in H \qquad (38)$$

We can substitute each $y_{p,m'}^{h+\theta}$ using the equality constraint (7), by $y_{p,m'}^{h+\theta} = 1 - \sum_{m'' \in M, m'' \neq m'} y_{p,m''}^{h+\theta}$ and get:

$$K_{m,m'}^{min} - \sum_{\theta=0}^{K_{m,m'}-1} \sum_{m'' \in M, m'' \neq m'} y_{p,m''}^{h+\theta} \geq K_{m,m'}^{min} z_{p,m,m'}^h \quad \forall (p,m,m') \in AL, \forall h \in H$$

$$(39)$$

Note that there are exactly $K_{m,m'}^{min}$ different $y_{p,m'}^{h+\theta}$ variables involved. By rearranging the equation we get the surrogate constraint:

$$\sum_{\theta=0}^{K_{m,m'}^{min}-1} \sum_{m'' \in M, m'' \neq m'} y_{p,m''}^{h+\theta} \leq K_{m,m'}^{min}(1 - z_{p,m,m'}^h) \quad \forall (p,m,m') \in AL, \forall h \in H$$

$$(40)$$

which is exactly the constraint proposed by Ierapetritou et al. (2002) and Karwan and Keblis (2007). Therefore, the set of constraints (14) implies constraint (15). Furthermore, let us consider the following non-integral point: assume one of the $y_{p,m''}^{h+\theta}$ in constraint (15) is 1. $z_{p,m,m'}^h = (1 - \frac{1}{K_{m,m'}^{min}})$ is feasible for constraint (15), but it is infeasible in constraint (14) if $MinStay(m,m')$ is nonempty. Hence, constraint (14) is tighter. $\square$

II. On the tightness of constraint (21) for forbidden transitions

We prove that constraint (21), which we rewrite here as

$$y_{p,m'}^h - \sum_{m:(p,m,m') \in AL} y_{p,m}^{h-1} \leq 0 \quad \forall p \in P, m' \in M, \forall h \in H \qquad (41)$$

is tighter in the LP relaxation than constraint (22),

$$y_{p,m}^{h-1} + y_{p,m'}^h \leq 1 \quad \forall (p,m,m') \in DAL, \forall h \in H$$

Note that both constraints are equivalent for integer values since AL and DAL are disjoint sets. The key to the proof is equality constraint (7), which results from the disjunction of modes. We can rewrite (7) for a given m' in the following way:

$$\sum_{m \in M} y_{p,m}^h = \sum_{m \in AL} y_{p,m}^h + \sum_{m \in DAL} y_{p,m}^h = 1 \quad \forall p \in P, h \in H \qquad (42)$$

If we add over index m all K constraints (K = number of forbidden transitions for m') of constraint (22) we get the following inequality:

$$\sum_{m \in DAL} y_{p,m}^{h-1} + K y_{p,m'}^h \le K \quad \forall p \in P, m' \in M, \forall h \in H \qquad (43)$$

Using equality (42) we can rewrite (43) as

$$\sum_{m \in DAL} y_{p,m}^{h-1} + K y_{p,m'}^h \le K \Big(\sum_{m \in AL} y_{p,m}^{h-1} + \sum_{m \in DAL} y_{p,m}^{h-1} \Big) \quad \forall p \in P, m' \in M, h \in H \ (44)$$

$$\Leftrightarrow y_{p,m'}^h - \sum_{m:(p,m,m') \in AL} y_{p,m}^{h-1} - \frac{K-1}{K} \sum_{m \in DAL} y_{p,m}^{h-1} \le 0 \quad \forall p \in P, m' \in M, h \in H \ (45)$$

If we compare the resulting inequality (45) with constraint (41), we can see that the additional term $-\frac{K-1}{K} \sum_{m \in DAL} y_{p,m}^{h-1}$ makes (45) weaker compared to (41) because $y_{p,m}^{h-1}$ are binary (nonnegative). $\square$

References

[1] Balas, E. Disjunctive Programming and a Hierarchy of Relaxations for Discrete Optimization Problems. *SIAM Journal on Algebraic and Discrete Methods*, 6:466–486, 1985.

[2] Brooke, A.; Kendrick, D.; Meeraus, A. GAMS: A Users Guide, Release 23.3. *The Scientific Press, South San Francisco*, 2010.

[3] Castro, P.M.; Harjunkoski I.; Grossmann I.E. New Continuous-Time Scheduling Formulation for Continuous Plants under Variable Electricity Cost. *Industrial & Engineering Chemistry Research*, 48:6701–6714, 2009.

[4] Castro, P.M.; Harjunkoski I.; Grossmann I.E. Optimal scheduling of continuous plants with energy constraints". *Computers & Chemical Engineering*, 35:372–387, 2011.

[5] Charles River Associates. Primer on Demand-Side Management, With an emphasis on price-responsive programs, CRA No. D06090. Technical report, The World Bank, 2005.

[6] Christof, T.; Lobel, A. The PORTA manual page v. 1.4.1. Technical report, ZIB, Berlin, 1997.

[7] Daryanian, B.; Boln, R.E.; Tabors, R.D. Optimal demand-side response to electricity spot prices for storage-type customers. *IEEE Transactions on Power Systems*, 4:897–903, 1989.

[8] Erdirik-Dogan, M.; Grossmann, I.E. Simultaneous planning and scheduling of single-stage multi-product continuous plants with parallel lines. *Computers & Chemical Engineering*, 32:2664–2683, 2008.

[9] Grossmann, I.E.; C.A. Floudas. Active Constraint Strategy for Flexibility Analysis in Chemical Processes. *Computers & Chemical Engineering*, 11:675, 1987.

[10] Gutschi, C.; Stigler, H. Potenziale und Hemmnisse fuer Power Demand Side Management in Oesterreich. In *10. Symposium Energieinnovation, Technische Universitaet Graz*, 2008.

[11] Huang, R. *Nonlinear Model Predictive Control and Dynamic Real Time Optimization for Large-scale Processes*. PhD thesis, Carnegie Mellon University, 2010.

[12] Ierapetritou, M.G.; Wu, D.; Vin, J.; Sweeny P.; Chigirinskiy M. Cost Minimization in an Energy-Intensive Plant Using Mathematical Programming Approaches. *Industrial & Engineering Chemistry Research*, 41:5262–5277, 2002.

[13] Karwan, K.; Keblis M. Operations planning with real time pricing of a primary input. *Computers & Operations Research*, 34:848–867, 2007.

[14] Klobasa, M. *Dynamische Simulation eines Lastmanagements und Integration von Windenergie in ein Elektrizitaetsnetz auf Landesebene unter regelungstechnischen und Kostengesichtspunkten.* PhD thesis, ETH Zurich, 2007.

[15] Paulus, M.; Borggrefe, F. The potential of demand-side management in energy-intensive industries for electricity markets in Germany. *Applied Energy*, 88:432–441, 2011.

[16] Pulkkinen, P.; Ritala, R. TMP production scheduling under uncertainty: Methodology and case studies. *Chemical Engineering and Processing*, 47:1492–1503, 2008.

[17] Raman, R.; Grossmann, I.E. Modeling and Computational Techniques for Logic Based Integer Programming. *Computers & Chemical Engineering*, 18:563, 1993.

[18] Sahinidis, N.V.; Grossmann, I.E. MINLP model for cyclic multiproduct scheduling on continuous parallel lines. *Computers & Chemical Engineering*, 15:85–103, 1991.

[19] Sung, C.; Maravelias, C. T. An Attainable Region Approach for Effective Production Planning of Multi-product Processes. *AIChE Journal*, 53:1298–1315, 2007.

[20] Swaney, R.E.; Grossmann, I.E. An Index for Operational Flexibility in Chemical Process Design, Part I: Formulation and Theory. *AlChE Journal*, 31:621, 1985.

[21] The MathWorks Inc. MATLAB 7.10.0.499 (R2010a), 2010.

[22] Todd, D, Caufield, M., Helms, B., Starke M., Kirby B., Kueck, J. Providing Reliability Services through Demand Response: A Preliminary Evaluation of the Demand Response Capabilities of Alcoa Inc. Technical report, U.S. Department of Energy, 2009.

[23] Voytas et al. Data Collection for Demand-Side Management for Quantifying its Influence on Reliability. Technical report, North American Electric Reliability Corporation (NERC), 2007.

[24] Zhu, Y.; Legg S.; Laird, C. A Multiperiod Nonlinear Programming Approach for Operation of Air Separation Plants with Variable Power Pricing. *AIChE Journal*, accepted, 2010.