On Defining the Partial Control Problem: Concepts and Examples

Mayuresh V. Kothare
Dept. of Chemical Engineering, Lehigh University, Bethlehem, PA 18015

Reuel Shinnar and Irven Rinard
Dept. of Chemical Engineering, City College of New York, New York, NY 10031

Manfred Morari
Automatic Control Laboratory, Swiss Federal Institute of Technology (ETH), CH-8092 Zürich, Switzerland

Many complex chemical processes having a large number of process variables and poorly understood models can be controlled reasonably well by controlling only a small subset of process variables using an equally small number of manipulated variables. This is the central premise of this article and is referred to as partial control. Knowingly or unknowingly, this idea has been and continues to be applied to successfully control numerous complex industrial processes. Despite its widespread use, partial control has never been explicitly formulated. The partial control problem is defined. A number of terms is introduced such as process variable dominance, modelable responses, practical degrees of freedom, and sufficiency of partial control. The new framework allows incorporation of both engineering-based decisions and more rigorous theoretical tools to achieve the goals of partial control. A number of practical examples illustrate the applicability of these concepts.

Introduction

The chemical and petrochemical industry is characterized by processes which can be classified as *highly complex*. This complexity arises from several sources:

- *Process model uncertainty:* The development of detailed and accurate models for chemical processes is rarely possible or prohibitively expensive. For example, the kinetics and mechanisms of many chemical reactions are often far too complex to allow an exact mathematical description. As a consequence, the model of the chemical process, which is based on this imperfect understanding of the kinetics, is inherently imperfect and has a large uncertainty associated with it.
- *Inherent process nonlinearity:* The steady state and dynamic behavior of chemical processes is generally highly nonlinear, implying the potential existence of complex behavior such as multiple steady states, instabilities, and limit cycles. Linear models are rarely suitable for describing these processes except in an unrealistically narrow operating regime.
- *Process constraints:* The manipulated variables are limited in magnitude due to equipment limitations such as pump and compressor throughput limits and actuator limits. On the other hand, the process states and outputs are constrained to lie between prespecified limits of their reference values; these limits are arising due to stringent product specifications dictated by market demands, safety limits, or environmental regulations on effluent concentrations.
- *Process integration:* The need for minimizing material and energy costs has led to the use of multiple material recycle loops and complex energy management networks. As a result, chemical processes have evolved into highly *integrated*, and, at times, unmanageable, monolithic systems.
- *Large-scale system aspects:* The total number of controlled and manipulated process variables is generally large and typically, the former is significantly larger than the latter. It is important to have a methodology to decide which of the large number of process and state variables are critical from the point of view of ensuring stability, rejecting disturbances, and meeting specifications.

Correspondence concerning this article should be addressed to M. V. Kothare.

From a control perspective, the last fifteen years have seen an enormous research effort in areas which focused on the first three aspects of this complexity. Some of these areas include linear *robust process control* (Morari and Zafiriou, 1989), model predictive control (MPC) (Camacho and Bordons, 1995; Morari et al., 1997; Soeterboek, 1991), and the general area of constrained control (Kothare, 1997) and nonlinear process control (Bequette, 1991; Henson and Seborg 1997, Isidori, 1995).

The aforementioned research efforts have contributed much to our understanding of controller design for complex chemical processes. However, the critical issue of the choice of the *control system configuration*, that is, the choice of the manipulated and measured/controlled variables and their interconnection, which has a significant impact on addressing the control issues in integrated and large-scale complex processes mentioned above, has received far less attention. Most current research in process control focuses predominantly on the controller design problem, assuming that the control system configuration has been determined *a priori*.

Existing techniques which address the problem of control system configuration and large-scale system aspects rely almost entirely on *ad hoc* techniques drawn from the experience and ingenuity of a few skilled practitioners. Morari (1982) and Morari et al. (1980) outlined numerous decomposition schemes to divide the overall integrated plant control problem into a number of tractable subproblems. Åström and Wittenmark (1997) discuss the *structuring* of the control problem for large-scale complex processes using either a *top-down* or *bottom-up* approach or some combination of both, the so-called *inside-out-outside-in* approach. Rinard and Downs (1992) reviewed several plant-wide control strategies for general large-scale systems. The review included the seminal work of Buckley (1964) involving decoupled material balance control and quality control, the work of Hougen (1979) based on fourteen heuristic control strategy principles, the decomposition approach of Morari et al. (1980), the work of Govind and Powers (1982), the hierarchical approach to conceptual design by Douglas (1988), and the mixed integer linear programming (MILP) approach of Georgiou and Floudas (1989). Recent reviews by Ng and Stephanopoulos (1998) and Skogestad and Larsson (1998) have argued that the concepts of *feedback optimizing control structure* and *self-optimizing control structure* provide a formal framework for designing plant-wide control systems. A recent book by Luyben et al. (1998) develops a number of guidelines and control principles applicable to practical plant-wide control problems. Some of these principles are attributed to the authors of this article. Similarly, the thermodynamically motivated method for partial control developed in Tyreus (1999a) builds on the key ideas of Shinnar and coworkers. Its application to the Tennessee Eastman Challenge process (Tyreus, 1999b) very clearly illustrates the power of the method. A plant-wide hierarchical control strategy has been reported recently in Zheng et al. (1999).

Partial control

The basic idea that a large number of variables in a complex uncertain system can be controlled reasonably well by controlling only a small subset of process variables was first expressed by Shinnar (1981) in the context of control of complex chemical reactors. In many complex chemical processes this is standard practice. Indeed, it is the only alternative available since the number of process variables that one would like to control at their specified set points is significantly larger than the number of manipulated variables at our disposal.

In such cases, due to insufficient degrees of freedom, offset-free set point tracking of all the specification variables is generally infeasible. Instead, an attempt is made to maintain all the process variables within an *acceptable range* of their set points by controlling a small subset of *dominant variables* at their exact set points. This is the problem of interest for this article and will be referred to as the *partial control* problem. In the context of partial control, the proper choice of the control system configuration, in particular, the choice of the small number of dominant state variables whose control ensures overall process stability, disturbance rejection, and satisfaction of performance specifications, assumes even greater significance.

The concept of partial control was illustrated recently through extensive case studies on an industrial-scale fluidized-bed catalytic cracker (FCC) by Arbel et al. (1996, 1997). These case studies showed that for the FCC with about seventeen output variables to be controlled using only seven manipulated variables, good "partial control" of all output variables could be achieved by controlling only two dominant variables using two manipulated variables (we will see more details later). In particular, the interplay between the initial FCC design and its control was illustrated in Arbel et al. (1997) by emphasizing the impact of the number of manipulated variables on the achievable partial control performance. In a different context, Shinnar (1990a,b) examined the applicability of partial control for explaining control mechanisms in economic systems by identifying dominant variables in the economy and by elucidating numerous similarities with complex chemical systems.

The objective of this article is to formally define the concept of partial control and its associated terminology. A major emphasis is on setting up a framework for partial control which is rigorous but also practically motivated to be understandable to the process control engineer. This will allow a more systematic incorporation of engineering-based decisions into the problem solution than the currently employed experience-based approaches. There are, however, inherent limitations to the rigor that can be accomplished. The key problem is one of incomplete information; when only qualitative knowledge about the system behavior is available it is impossible to define quantitative criteria. Indeed, one could even argue that the whole issue of partial control becomes irrelevant when complete information about the system is available. The concepts and definitions of partial control introduced in the article are therefore most relevant for the case when model information is incomplete, which is the realistic case in complex systems.

Motivation

Consider Figure 1 which shows a typical FCC used in present day refineries. Its function is to convert heavy hydrocarbon petroleum fractions into a number of more useful products such as gasoline, middle distillates, and light olefins. In

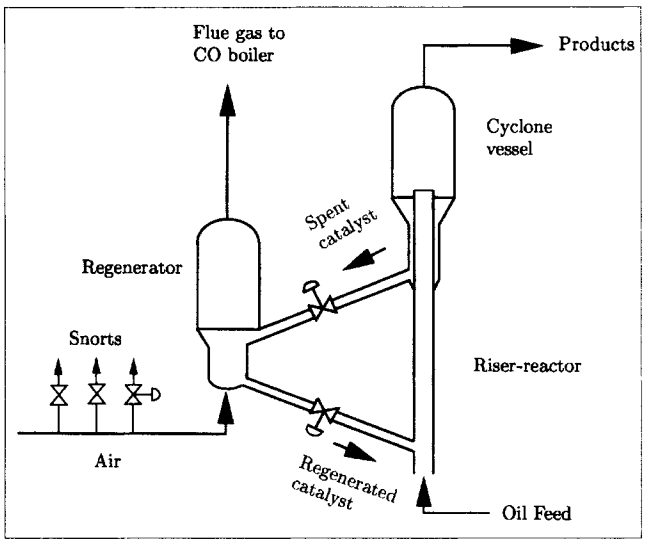

Figure 1. Modern FCC unit.

the riser-reactor, oil feed is brought in contact with hot catalyst from the regenerator. During the cracking reaction, carbonaceous material or *coke* deposits on the catalyst and deactivates it. In the regenerator, the spent catalyst is regenerated by burning off this coke. The heat of combustion raises the catalyst temperature, thereby providing the heat required for the endothermic cracking reaction in the riser-reactor. A detailed discussion of model development, nonlinear dynamics and control aspects for the FCC can be found in a series of articles by Arbel et al. (1995, 1996, 1997).

The overall objective of the unit is to achieve an economically profitable conversion of the feed to the desired products, while maintaining key process variables within operating limits. This translates to specifications on the system variables given in Table 1.

The manipulated variables commonly employed are the air-flow rate F_{air} to the regenerator and the catalyst recirculation rate F_{cat}. Sometimes, a feed preheater may also be available. In addition, there are also available "slow" manipulated variables such as catalyst replacement rate and catalyst properties, and catalyst additives and feed quality which have a much slower effect on the plant operation. The measurements available are the temperatures of the regenerator and riser-top, and the exiting flue gas. Again, "slower" measurements such as composition analyzer records for the product stream may be used.

We immediately note that the number of process variables that need to be controlled to achieve the process objectives is much larger than the number of available manipulated variables. Further complicating the scenario is the fact that a detailed and accurate model describing the FCC is rarely available, although good working models which capture the key aspects of the process have been reported (see Arbel et al. (1995) and references therein). Controller design for the FCC is clearly a nontrivial task.

Table 1. Specification Variables y_p for the FCC

Conversion of oil feed	Riser top temp., T_{ris}
Gasoline yield, octane no.	Regenerator temp., T_{rgn}
Light and heavy fuel oil yields	Temp. rise in the regenerator, ΔT
Yield of olefins	Flue-gas temp., T_{sg}
Excess oxygen	Emission Levels of CO, SO_2, NO_x, etc.

As we will show later in detail, despite this evident complexity, stabilization and reasonable *partial control* of the entire process can still be achieved by controlling only three *dominant* variables—T_{rgn}, T_{ris} and C_{rgn} (see Arbel et al., (1996); Morari, (1982); Shinnar, (1981) for other choices of the dominant variables) although for some purposes (Arbel et al., 1996; Morari, 1982; Shinnar, 1981), it suffices to control just one variables (T_{rgn}) or two variables (T_{rgn}, T_{ris}). This choice of the control system configuration was arrived at after extensive simulation studies and from an understanding of the fact that process temperatures are strongly related to almost all the specification variables. The key requirement of achieving stability by maintaining the heat balance between the riser and regenerator in an overall adiabatic process provided further justification for this choice. The set points for the dominant variables and the appropriate steady-state values of the "slow" manipulated variables are determined by an external supervisory controller. Most industrial practitioners apply similar ideas, based on their extensive experience and expertise, to come up with satisfactory *partial control* strategies for many large-scale complex nonlinear processes.

The FCC case studies strongly suggest that there is merit in further exploring the applicability of partial control to general large-scale complex nonlinear processes beyond the current experience-based approach.

Concept of Partial Control

Let the process to be controlled be represented by the following nonlinear, nonautonomous, first-order ordinary differential equation

$$\dot{x}(t) = f(x(t), u(t), d(t)). \tag{1}$$

Here, $x \in \Re^{n_x}$ is the system, $u \in \Re^{n_u}$ is the manipulated variable, $d \in \Re^{n_d}$ is the external state disturbance, and f is in general a nonlinear function of its arguments. We will assume that the state vector is comprised of physically meaningful process variables.

Quite often, the state vector x has a very large dimension and many of the states may not be measurable. Moreover, rarely is it necessary to directly control all the elements of x. Instead, we are more interested in an output vector $y \in \Re^{n_y}$ consisting of the process variables to be controlled and also all available measurements from the plant

$$y(t) = g(x(t), u(t), n(t)), \tag{2}$$

where $n \in \Re^{n_n}$ is the output measurement noise and $g(.,.,.)$ is in general allowed to be a nonlinear function of its arguments. The measured subset of y is denoted by $y_m \in \Re^{n_{y_m}}$ and is given by the nonlinear relation

$$y_m(t) = g_m(x(t), u(t), n(t)), \tag{3}$$

where $g_m(.,.,.)$ is the appropriate component of $g(.,.,.)$ in Eq. 2 relating to y_m. y may also contain some elements of the

state vector x which are measurable or are of interest for system performance.

Control system design for a chemical process typically involves the following steps:

- Identification of process objectives.
- Development of a model for the process.
- Choice of the control system configuration.
- Choice of the control algorithm.
- Evaluation if the proposed design can meet the process specifications.

A unique feature of chemical processes is the fact that their operation is dominated by process objectives and specifications, and often the objectives change depending on the market demand. Understanding process objectives is therefore crucial for designing a control system which can meet changing specifications.

Process objectives are generally in the form of specifications and constraints, both steady-state and dynamic, that need to be met by certain process variables. The process variables (excluding the manipulated variables u) which define the process specifications and all state variables of interest such as those which characterize process stability are denoted by $y_p \in \Re^{n_{y_p}}$ and are included in the output vector y. We will assume that y_p is given by the nonlinear relationship

$$y_p(t) = g_p(x(t), u(t), n(t)) \tag{4}$$

where $g_p(.,.,.)$ is the appropriate component of $g(.,.,.)$ in Eq. 2 relating to y_p. We will assume, as is generally the case, that all elements of y_p are measurable, possibly with different sampling times, in which case, y_p is a subset of y_m and $g_p(.,.,.)$ is a component of $g_m(.,.,.)$ in Eq. 3.

A typical steady-state process specification is the requirement that y_p be maintained at a set point y_p^s without offset, that is,

$$\lim_{t \to \infty} y_p(t) = y_p^s. \tag{5}$$

Typical dynamic or time-domain specifications are expressed in the form of bounds or inequality constraints on the trajectories of the elements of y_p, that is,

$$y_{p_j\min} \le y_{p_j}(t) \le y_{p_j\max}, \quad \forall t \ge 0,\ j = 1, \ldots, n_{y_p}. \tag{6}$$

Standard manipulated variable constraints are also expressible in this form

$$u_{j\min} \le u_j(t) \le u_{j\max}, \quad \forall t \ge 0, \quad j = 1, \ldots, n_u. \tag{7}$$

Definition 1 (Exact Control). *The system (Eq. 1), (Eq. 2), without any constraints on u, is said to be exactly controllable if the performance variable y_p can be moved to and maintained at an arbitrarily prescribed set point y_p^s without offset, starting from an arbitrary initial point, by an appropriate (possibly nonunique) choice of the steady-state value of the manipulated variable u.*

We may note that this is a *global* definition of exact control and in the presence of input constraints, it can be appropriately modified to a *local* definition by restricting the class of set points and initial points to those that do not cause violation of input constraints.

We note that for exact control, the set of steady-state equations

$$f(x, u, d) = 0 \tag{8}$$

$$g_p(x, u, n) = y_p^s \tag{9}$$

must have a solution for any arbitrary choice of y_p^s, n, and d. Invoking the Implicit Function Theorem, we conclude that this is equivalent to the requirement that

$$\operatorname{rank}\begin{bmatrix} \dfrac{\partial f}{\partial x} & \dfrac{\partial f}{\partial u} \\ \dfrac{\partial g_p}{\partial x} & \dfrac{\partial g_p}{\partial u} \end{bmatrix} = n_x + n_{y_p} \tag{10}$$

at *every* steady-state operating point (x_s, u_s). Therefore, as a necessary condition, this condition requires that $n_u \ge n_{y_p}$, that is, there should be at least as many manipulated variables u as there are specification variables y_p.

For the class of large-scale complex processes of interest to this article, the number of manipulated variables n_u is in general much less than the number of performance variables n_{y_p}. In this case, exact control of all the elements of y_p is not feasible in general. In this situation, we must relax the requirement of exact control on y_p and accept something less stringent, namely, that the elements of y_p be maintained within an *acceptable range* around their set points.

Definition 2 (Partial control). *The system (Eq. 1), (Eq. 2), without any constraints on u, is said to be partially controllable if the performance variable y_p can be moved to and maintained within an acceptable range of an arbitrarily prescribed set point y_p^s, starting from an arbitrary initial point, by an appropriate (possibly nonunique) choice of the steady-state value of the manipulated variable u. This is equivalent to requiring that*

$$y_{p_j\min}^s \le \lim_{t \to \infty} y_{p_j}(t) \le y_{p_j\max}^s, \quad j = 1, 2, \ldots, n_{y_p}, \tag{11}$$

where the interval $[y_{p_j\min}^s, y_{p_j\max}^s]$ contains the set point, if any, of the variable y_{p_j}.

As with Definition 1, the above definition can be modified for the case with input constraints and for the case with a finite operating space of realistic set points. The interval limits $y_{p_j\min}^s, y_{p_j\max}^s$ are user specified, depending on how much loss of "exact control" is tolerable.

Partial control structure

The goal of the control system is to achieve a combination of exact control requirements on some y_{p_j} and partial control requirements on the remaining y_{p_j}, as discussed above, while maintaining stable plant operation. The set points $y_{p_j}^s$ for achieving exact control of some y_{p_j} typically arise from higher level plant-wide optimization criteria. The issue of how these set points are computed and what is the optimality criterion employed is beyond the scope of this article. The interval limits $[y_{p_j\min}^s, y_{p_j\max}^s]$ for achieving partial control of the remaining y_{p_j} arise either from relaxation of exact control requirements or from higher level plant-wide operational criteria such as minimum product purity specifications, maximum allowable effluent concentrations, and so on.

To achieve these goals of partial control, we adopt an axiomatic approach involving, as a first step, an appropriate decomposition of process variables—both measured and manipulated—and, as a second step, a corresponding decomposition of the control system. This structure is justified from both a theoretical and practical standpoint and is indeed commonly employed in a variety of industrial applications.

Process Variable Decomposition. The classical decomposition of y_m and u employed in practice is the following (Morari, 1982)

$$y_m = \begin{bmatrix} y_{m1} \\ y_{m2} \end{bmatrix}, \quad u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \tag{12}$$

where the subscript 1 denotes process variables which are measured (or manipulated) continuously or with very small sampling times compared to the overall time scale of the process, and the subscript 2 denotes variables which are infrequently measured (or manipulated). The ease and cost of measurement at high sampling rates and the underlying time scale of response of a particular variable are both important factors in choosing the frequency of measurement of a variable. Similarly, for a manipulated variable, in addition to the ease and cost of manipulation at high frequencies, it is the underlying speed with which it can affect the process that also determines the above partitioning of u into "fast" and "slow" components. The classical two-level structuring of control for this decomposition has been discussed in Morari (1982).

For the purpose of partial control, the decomposition of the process variables is somewhat different and is dictated by its end use in arriving at a suitable partial control structure. Two subsets of the measured output y_m are identified:

- y_p: These are as defined in Eq. 4 and relate to the performance specifications. The elements of y_p can belong to either y_{m1} or y_{m2}, that is, they can be fast or slow measurements.
- $y_d \in \Re^{n_{y_d}}$: These are defined as *dominant* variables and are given by the following nonlinear relation

$$y_d(t) = g_d[x(t), u(t), n(t)] \tag{13}$$

where $g_d(.,.,.)$ is the appropriate component of $g(.,.,.)$ in Eq. 2 relating to y_d. The dominant variables y_d are such that controlling them closely or exactly (see Definition 1) allows us to control y_p partially (see Definition 2). Elements of y_d may or may not belong to y_p, but they must necessarily belong to y_{m1}, that is, they must be measurable at fast sampling rates. Typically, $n_{y_d} \ll n_{y_p}$.

The decomposition of the manipulated variables u remains the same as that in Eq. 12, but we will adopt the more intuitive notation

$$u = \begin{bmatrix} u_d \\ u_p \end{bmatrix} \tag{14}$$

where $u_d \in \Re^{n_{u_d}}$ are the fast or *dynamic* manipulated variables which directly control y_d and $u_p \in \Re^{n_{u_p}}$ are slow manipulated variables which are adjusted based on the measurement of the performance variables y_p. Typically, $n_{u_d} \geq n_{y_d}$. Even though u_p is adjusted at slower frequencies, this does not necessarily reflect the speed with which it affects the process. Sometimes, manipulated variables having a fast impact on the process are deliberately placed in u_p based on economic and cost considerations.

The external disturbance d can be partitioned similarly into slow and fast components

$$d = \begin{bmatrix} d_d \\ d_p \end{bmatrix} \tag{15}$$

where $d_d \in \Re^{n_{d_d}}$ are fast or dynamic disturbances and $d_p \in \Re^{n_{d_p}}$ are slow disturbances. We will assume that $d_d \to 0$ as $t \to \infty$, in other words, that the transient part of the disturbances is assumed to decay to zero. The persistent steady-state part of the disturbances is captured by d_p.

The distinction between the proposed decomposition and the classical decomposition (Eq. 12) is that it is not based merely on speed of measurement and manipulation of variables, but more on the functional aspects of the two sets of resulting variables. First, as described below, it leads naturally to an appropriate two-level control structure which is the essence of almost any practical process control structure, and, second, the reduction in the number of variables y_d that need to be controlled dynamically to achieve the partial control goals results in considerable simplification of the modeling effort and the overall controller design.

In order to achieve good partial control of the variables y_p, particular care needs to be taken in choosing the dominant variables y_d and the corresponding dynamic manipulated variables u_d. Typically, a combination of physical insight, mathematical criteria, and engineering judgment is required in making this choice, and, many times, this choice is very much dependent on the specific goals and objectives of a particular process. The concept of process variable dominance is one of the central ideas of partial control and will be discussed in greater detail in subsequent sections.

Control System Decomposition. Having defined the decomposition of process variables, the partial control strategy is now structured into a two-level or cascade structure shown in Figure 2. The dominant variables y_d are employed in the computation of the manipulated variables u_d such that y_d is maintained at its set point y_d^s while rejecting fast disturbances d_d. We will refer to this as the fast *primary control loop* with the following (possibly nonlinear) dynamic control law

$$u_d(t) = C_d[y_d(t), y_d^s, d_d(t)]. \tag{16}$$

The primary control law (Eq. 16) is generally required to be square, that is, $n_{u_d} = n_{y_d}$, and $\lim_{t \to \infty} d_d(t) = 0$, thereby allowing exact control of y_d at its set point y_d^s.

The performance variables y_p (both slow and fast) are used to compute the values of the slow manipulated variables u_p and the set points y_d^s of the dominant variables y_d such that

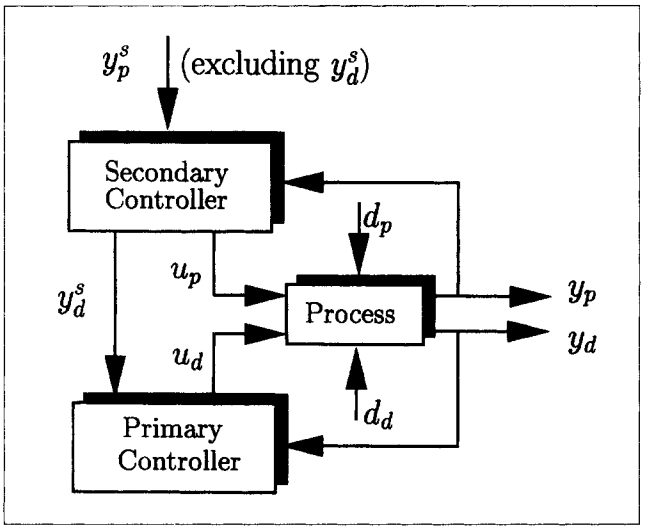

Figure 2. Partial control diagram.

the slow disturbances d_p are rejected and the partial control requirements on y_p are met. This loop is referred to as the slow *secondary control loop*, with the outputs of the secondary controller given by the following, generally nonlinear, function $C_p(.,.)$

$$\begin{bmatrix} u_p \\ y_d^s \end{bmatrix} = C_p(y_p, y_p^s, d_p). \tag{17}$$

The secondary control law (Eq. 17) is not required to be square, since it serves only to achieve partial control of y_p by computing y_d^s and u_p. Thus, in effect, the secondary control law (Eq. 17) defines the steady-state operating point of the overall system, and in particular, the steady-state target y_d^s for the primary control loop. Most control literature focuses on the primary loop. The potential and importance of the secondary loop is often overlooked.

Remark 1. *The terms primary and secondary or slave control are commonly used in classical cascade control structures (Luyben, 1990; Stephanopoulos, 1984). However, the use and interpretation of these terms in the context of the partial control structure of Figure 2 is exactly opposite to that used in cascade control. The primary controller of Figure 2 satisfies the "primary" regulatory objectives of stabilization and fast disturbance rejection, while the secondary or supervisory controller attempts to achieve the "secondary" objectives of prescribing the most desirable or optimum operating point and rejection of the steady-state disturbance d_p.*

Remark 2. *The decomposition discussed here need not be restricted to two layers, but can be extended to multiple layers involving higher level site-wide optimizations (Skogestad and Larsson, 1998). For example, the highest level might involve a site-wide optimization, followed by a layer of plant-wide optimization, with successively lower layers of supervisory/predictive controllers and finally, regulatory PI/PID controllers. The broad distinction into a stabilizing primary control layer and an optimizing secondary control layer with appropriately decoupled time scales should still hold. Each of these layers can in turn have sublayers organized along the same principles, leading to multiple layers.*

Primary control loop

The primary control loop is required to satisfy three principal goals:

- *Stabilization:* For open-loop unstable processes, it should stabilize the nonlinear process (Eq. 1) around the desired steady state prescribed by the secondary control loop.
- *Regulation:* It should reject fast disturbances d_d and regulate y_d at its set point y_d^s.
- *Containment:* It should keep the dominant variables in a range where the process behaves linearly and where oscillations, multiple steady states, and other nonlinear phenomena do not occur.

The requirement of stabilization implies that, necessarily, all the unstable modes of the process (Eq. 1) should be observable from the output y_d and controllable with the manipulated inputs u_d. Interpreting stability in a broader sense, the primary control must ensure that the process stays in a permissible range of operation avoiding explosive limits or conditions where the nature of the process changes.

For open-loop stable processes, the above requirements of controllability and observability should hold with respect to the "slow" modes of the process (Eq. 1) to ensure that an acceptable time scale of response can be achieved with feedback control. At the primary control level, a linear feedback controller with disturbance feedforward and constraint protection usually suffices. The nonlinear control tasks are delegated to the secondary controller. Generally, the choice of the dominant variables y_d and the corresponding manipulated variables u_d to be employed is more critical than the choice of the control algorithm to be used.

While linear control is normally sufficient for the primary control loops, its function is not limited to stabilization and disturbance rejection. The primary control loop facilitates the achievement of the nonlinear secondary control-loop objectives. The set points of the primary control loop are the manipulated variables of the secondary control loop. This interrelation between the two control loops may result in a different choice of the control system configuration than would result from classical linear multivariable control theory. This will become clear in the following sections.

Secondary control loop

The secondary control loop operates at a time-scale which is intermediate between that of the plant-wide or higher level optimizer and that of the primary control loop. It sets the overall control goals of the process and is required to satisfy the following objectives, each of which has individual motivation:

- *Process Monitoring:* By measuring the variables y_p, it should monitor the process and control system performance.
- *Steady-State Control:* Using the measurement y_p, it should update the values of the *slow* manipulated variables u_p so as to improve process performance and minimize the effect of slow disturbances d_p on y_p.

• *Partial Control:* By prescribing the set points y_d^s of an appropriately chosen subset of dominant variables y_d, it should be able to have a predictable effect on the specification variables y_p, thereby ensuring that the partial control goals of the process are met.

The requirement of partial control implies that the steady-state specifications or the limits on y_d should not be "tight" to begin with, in order to give flexibility in varying the set point y_d^s over a wide enough range. Moreover, the choice of the dominant variables y_d should be such that changes in the set points y_d^s should have a meaningful and quantifiable effect on the achieved steady-state values of y_p. In physical terms, this implies that the steady-state behavior of y_d should be strongly coupled with the behaviors of all elements of y_p, thereby allowing all elements of y_p to be influenced by making changes in y_d^s.

Since the secondary control loop operates at a much slower time scale, the system can be considered to be at steady or pseudo-steady state. A more rigorous and generally nonlinear steady-state model is used here

$$\begin{aligned} \dot{x} &= \tilde{f}(x, u_d, u_p, d_p) = 0 \\ y_p &= \tilde{g}_p(x, u_d, u_p, d_p) \\ y_d &= \tilde{g}_d(x, u_d, u_p, d_p) \end{aligned} \tag{18}$$

where $\sim$ refers to the model, which is in general different from the real plant (Eqs. 1, 2, and 13). Here, d_p is the asymptotic value of the disturbance affecting the system in steady state. If d_p is not measurable, an estimate of d_p can be used in the above model. This being a steady-state model, the system is not affected by d_d since $d_d(t) = 0$ at $t = \infty$.

A suitable optimization problem can now be formulated for computing u_p and y_d^s as outputs of this loop

$$(u_p, y_d^s) = \arg \min_{u_p, y_d} \Phi(y_p, y_p^s) \tag{19}$$

subject to $\tilde{f}(x, u_d, u_p, d_p) = 0$,

$$y_p = \tilde{g}_p(x, u_d, u_p, d_p)$$

$$y_d = \tilde{g}_d(x, u_d, u_p, d_p)$$

$$u_{j\min} \le u_j \le u_{j\max},\ j = 1, \ldots, n_u$$

and the partial control constraints

$$y^s_{p_j\min} \le y_{p_j} \le y^s_{p_j\max},\ j = 1, \ldots, n_{y_p}. \tag{20}$$

Here, y_p^s is a set of target values for the performance variables y_p (not necessarily all of them and excluding y_d^s) specified either by the operator or by a higher level plant-wide optimizer, based on optimizing an economic performance objective. The objective function Φ can take numerous forms, depending on the significance of each of the performance variables y_{p_j} and how critical it is for each of them to achieve their target values.

In order to eliminate the need for solving the optimization problem and the implied nonlinear modeling problem it has been proposed by various authors to select y_d such that keeping y_d^s constant is "almost" optimal for wide ranges of disturbances and set points (Skogestad and Larsson, 1998).

Remark 3. *In many industrial applications, the values of u_p and y_d^s are determined by the plant operator based on intuition and process understanding, without recourse to any sophisticated nonlinear optimizations of the type discussed above. However, the current industrial trend is towards incorporation of computer-intensive model-based nonlinear real-time optimization (RTO) routines (Georgiou et al., 1997; Hendon, 1997; Mudt et al., 1995) in process operations and our framework for partial control provides appropriate flexibility for capturing these developments.*

Remark 4. *y_d^s is an input to the primary control loop. If y_d belongs to y_p, it should not have tight constraints and it should preferably not have a target value specification in the objective function $\Phi(y_p, y_p^s)$. In other words, y_d^s should be available as a degree of freedom so that it can be moved over a significant range and not be restricted to satisfying a tight bound or an offset-free specification through Φ.*

In most control applications, the process specifications which are stated in terms of the performance variables y_p do not have equal priority. For example, keeping certain variables y_{p_j} within safe operating limits is generally higher in priority than maintaining other variables y_{p_j} within tight limits from a profitability point of view. The relative importance of the constraints (Eq. 20) on the elements of y_p can be set by appropriately prioritizing these constraints. In such a formulation, an attempt is made to satisfy the constraints on as many y_{p_j} as is possible, with the highest priority constraints being satisfied first. When not all constraints can be satisfied, an attempt is made to minimize the extent to which the highest priority variable violates its constraints. As shown in Tyler and Morari (1999), this formulation involving prioritized constraints can be expressed as a set of mixed integer linear constraints in y_{p_j}. Alternatively, the Modular Multivariable Control (MMC) approach of Meadowcroft et al. (1992) can be applied.

Admissibility criteria for partial control

In the preceding sections, we discussed the basic structure of partial control and the functional aspects of the primary and secondary controllers. We will focus next on the issues involved in designing such a control system. We begin with a discussion of the ability of the chosen partial control structure to meet the process specifications and constraints which is captured by the notion of sufficiency.

Sufficiency. An important consideration in the design of the proposed control system is to match the capability of the controller to the target specifications and goals. The goals of a process are determined by economic considerations which in turn are determined by a changing market demand. On the other hand, process capabilities are determined by the initial plant design and by the nature and number of manipulated variables that are available for control.

Definition 3 (Sufficiency). *The ability of the partial control system to maintain stable operation and meet the changing process goals, expressed as appropriately relaxed or partial control*

specifications (Eq. 11), in the face of model uncertainty, manipulated variable constraints and external disturbances, is termed as sufficiency of the partial control structure.

Except for the ability to maintain stable operation, which is a minimum requirement of a partial control structure, sufficiency is closely linked with the process specifications. For example, the first FCC built in the late 1930s had only one specification of a minimum desirable conversion. A partial control structure which was sufficient to meet this minimal specification is very unlikely to be sufficient for a modern FCC which has many more stringent specifications on the product purities, effluent concentrations, and emission levels.

In the event that the chosen partial control structure is *insufficient* in maintaining stable operation and meeting the process specifications, one has several options: (a) change the partial control structure by changing the choices of y_d and u_d; (b) add more manipulated variables through a design change; (c) relax the specifications; or (d) change the original process design. This interplay between design and partial control should be done iteratively. Recent efforts have focused on minimizing the iterations (Shinnar et al., 2000).

Arriving at a partial control strategy which is sufficient involves three central concepts which provide guidelines for controller design and also give insight into the working of the approach. These are: the concept of *process variable dominance* which leads to the choices of the primary loop variables y_d, u_d; the concept of *process response modelability* which addresses the issue of the required model complexity and the associated model uncertainty in the two loops; and the concept of *effective* or *practical degrees of freedom* available for controlling the process which suggests the *number* of manipulated variables to be employed in the two control loops.

Process Variable Dominance. Partial control hinges on the observation that many process systems have a few independent variables that tend to dominate the overall *dynamic* behavior of the system. Moreover, by varying the steady-state values of these small numbers of dominant variables, one can also significantly affect the *steady-state* behavior and performance of the overall system. There is no *a priori* reason or law that a large complex system with many state and output variables is dominated by a few process variables and therefore can be controlled by a few manipulated variables. A simple counter example is a system of independent and noninteracting parallel subsystems each of which must necessarily be separately controlled for the overall system to be controllable, thereby precluding partial control.

An obvious requirement for the dominance of a few process variables is that there are large interactions between all the variables or subsets of them. Furthermore, a few manipulated inputs impact the dominant process variables which are the main carriers of these interactions. A typical example is the temperature in an exothermic reactor: it simultaneously dominates the dynamic behavior and stability properties of the reactor and also the steady-state properties such as reactor productivity, selectivity, and yield. Similarly, a single sensitive tray temperature in a conventional distillation column (Tyreus, 1999a) simultaneously affects overall dynamic control as well as steady-state product purities and operating costs. These observations can be summarized in the following definition of dominance.

Definition 4 (Dominance). *A subset of process variables is said to be dominant for a given process if this subset is preponderant in achieving or in aiding to achieve the specified process objectives.*

Thus, process variables are dominant if their control ensures that all performance variables y_p are held in their desired range. Dominance naturally depends on what outputs are to be controlled (y_p), but it does not depend on the specifications. Eventually we need to judge if the control scheme meets the specifications, or in other words, if it is *sufficient*.

An important question is how one can identify potential dominant variables and their impact on the specification variables in a realistic design setting. If a detailed process model is obtainable then the quantitative procedure described below can be applied. This may be the case when a new process is developed by a minor modification of a well-established and well-studied process.

More generally, what would be given is not a model but just the available technology, such as the catalyst, the reaction scheme, and so on. On this basis, the task of the design engineer would be to come up with a working processing scheme. There may be some limited reaction data, but it would not be sufficient to identify the dominant variables. Based on a preliminary design, the engineer would request more lab data to successively refine the design and eventually determine a set of dominant variables. This process is what we refer to as concurrent engineering and design (Shinnar et al., 2000) in the present context.

Possible dominant variables may be suggested directly by the laboratory experiments which are needed to establish design data. If the chemist seeks to determine the characteristics of a new catalyst, it will be necessary to do so under the different conditions which can occur in real operation. Based on these experiments, the chemist/chemical engineer can identify the ranges of the critical operating parameters which affect the catalyst performance, such as temperature, pressure, space velocity, and so on. These critical parameters are the prime candidates for dominant variables, because, almost by definition, controlling these dominant variables in real operation will control the product specifications. This is much harder to evaluate in a pilot plant or industrial unit. Therefore, laboratory data give the basis for control despite the uncertainties in scale-up.

It is more difficult to model the *exact* impact of the dominant variables on y_p. The importance of auto-catalysis may be overlooked. Backmixing and the effects of scale-up are usually hard to model. These uncertainties can sometimes be reduced by building larger pilot plants coupled with more research. Alternatively, one can look at designs that minimize the deviation from the laboratory operation. Finally, uncertainties in performance can be compensated by control; with control, the impact of the dominant variables can become well defined. As a rule, data decrease uncertainty and, with decreased uncertainty, the design can become simpler. Thus, there is a trade-off between the cost of obtaining more data and the cost/complexity of the design.

The set of dominant variables is a "property" of the processing scheme and the underlying technology. The designer cannot "create" new dominant variables. For example, let temperature be the dominant variable for a certain reaction carried out in a CSTR. Putting two CSTRs in series does not

increase the number of dominant variables, but it allows the designer to set these dominant variables to different values and thus refine their impact on the specification variables.

One of the engineering challenges is to design the process such that these dominant variables can indeed be controlled. Contrary to the laboratory, however, the control of all dominant variables may not be possible in real operation for a variety of technological reasons. Also, depending on the expected disturbances, the control of all the dominant variables may not be needed. Some dominant variables may have no importance for the specific process goals in the sense that they may impact outputs that have little relevance. This can change when specifications change. Furthermore, even those dominant variables that need to be controlled often have to be controlled in the slow secondary control loop. For example, in a tubular reactor, temperature is dominant for stability. Since one cannot control the temperature dynamically in thousands of reactor tubes with varying flow rates, one would design a tubular reactor with a large heat-transfer area and a small temperature drop such that the effective temperature is that of the cooling medium, which is dynamically controlled to be practically constant (see Shinnar et al., 1992). The reactor temperature is still dominant for the process, but is only infrequently adjusted in the secondary control loop.

If control of all dominant variables is not *sufficient*, it will be necessary for the design engineer to suggest design alternatives which in turn may require additional lab data leading to concurrent engineering and design (Shinnar et al., 2000)

Dynamic Dominant Variables. Since the goal of the dynamic primary controller is to stabilize the unstable modes of the process and to improve the closed-loop speed of response, we have the following definition of dominant dynamic variables.

Definition 5 (Dynamic Dominance). *The dominant dynamic process variables are those variables which under closed-loop control determine the overall system stability properties and dynamic response characteristics. The corresponding manipulated variables are the dominant dynamic manipulated variables.*

We note here that stability should not be interpreted in a standard linear sense, but implies stable operation in the expected range of operating conditions and disturbances. Therefore, dynamic dominance can be interpreted as dominance for stability and nonlinearity. For a nonlinear system in which a variable has a highly nonlinear effect on the overall system behavior and stability, maintaining this variable in a very narrow range by feedback control can effectively linearize the process and enable stabilization of the overall system. Such a process variable or set of process variables can be considered as a dominant dynamic variable. In fact, many nonlinear processes can be controlled using linear model-based methods due to this property.

Although the definition is not restricted to the linear case, it can be illustrated through the following simple linear system example

$$\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\end{bmatrix}=\begin{bmatrix}1 & 0\\ 1 & -1\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix}+\begin{bmatrix}3 & 4\\ 1 & 1\end{bmatrix}\begin{bmatrix}u_1\\ u_2\end{bmatrix} \tag{21}$$

Although the system is stabilizable using either u_1 or u_2 and detectable using measurements x_1 or x_2, the dominant variable which characterizes the instability in the system is the state x_1 corresponding to the unstable mode. Either u_1 and u_2 are dominant manipulated variables.

Dynamic process variable dominance is an intrinsic property of a process, independent of the manipulated variables that may be employed in the control system. Consequently, a straightforward method to identify dominant variables is by considering the corresponding *autonomous* system description. For example, in the linear system (Eq. 21) above, rewritten in the autonomous form

$$\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\end{bmatrix}=\begin{bmatrix}1 & 0\\ 1 & -1\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix} \tag{22}$$

setting the initial condition $x_1(0)=0$ eliminates the equations for $\dot{x}_1$ and makes the reduced x_2 subsystem, $\dot{x}_2=-x_2$, stable. However, setting $x_2(0)=0$ does not eliminate the equation for $\dot{x}_2$ because the unstable x_1 subsystem, $\dot{x}_1=x_1$, drives the x_2 subsystem to instability. Thus, x_1 is the dominant dynamic state variable.

A significant number of nonlinear processes can have multiple steady states, some of which are locally stable. Usually, there is only one useful steady state and in the vast majority of cases it is a stable one. Moreover, there is a sizable operating region inside which the desired steady state is asymptotically stable. If the process conditions move out of this region, the process will either jump to the null state, which is expensive to restart, or to a stable "upper" steady state which could have severe penalties. Therefore, it is important to keep the process within this useful region.

Most instabilities share one feature. They are the result of internal feedback processes that cause the instability. The variable that is at the heart of this internal feedback process is the dynamic dominant variable for stability. An example is reactor temperature in an exothermic reaction system with high activation energies and heat release that is run adiabatically. This instability is driven by temperature and would not occur in isothermal reactors. Besides temperature, there are other mechanisms that can drive an instability. The nucleation rate in a crystallizer is an example (see Sherwin et al., 1967; Shinnar et al., 2000).

The problems for the operation and design are twofold.

1. Instability is hard to recognize in the laboratory. Processes that are unstable in the laboratory or in a pilot plant are seldom considered for further development. Continuous crystallizers are almost always stable in the pilot plant, but frequently exhibit limit cycles when scaled up to commercial size. All we can do is use our knowledge of the chemistry and physics of the processes to identify potential sources of instability and the dominant variables associated with them. We can then test these potentially dominant variables in the laboratory.

2. We very seldom have sufficient model information accurate enough to clearly define the stable region. We must use our "learning models" (Shinnar, 1981) to provide reasonable guidelines. What makes this feasible is that the number of potentially dominant variables for stability is very limited, although there are many input variables that affect the position of the stable region.

This analysis is very helpful as it gives us a tool to provide for stable operation in the face of large model uncertainty. The key is to provide for independent control of the dominant variables over a sufficiently wide range. As a simple example, if the dominant variable were temperature, an isothermal reactor would solve the problem. If this is too expensive, independent control of the temperature over a reasonable range will compensate for any uncertainty. As a minimum, we must ensure direct and sufficiently fast control of this temperature.

A nontrivial practical example would be an unstable distillation column (Dorn et al., 1998). Because unstable columns had neither been simulated nor operated prior to the publication of this article, there was no experience or even intuition on how to stabilize such a system. There is no easily identifiable internal feedback process which causes the instability. From steady-state simulation results, column profile control was conjectured to be sufficient for stabilization. On this basis, tray temperature was identified as an inferential variable and later confirmed experimentally as the dominant dynamic variable.

Remark 5. *Generally, there are several alternative choices of the dominant variables y_d and u_d. In such cases, additional information such as the requirements imposed by the secondary loop should provide guidelines for evaluating these alternatives. Physical insight, the ease and cost of measuring and manipulating certain process variables, and the available range and rate of change of the candidate manipulated variables should also help in sifting through these alternative choices.*

Steady-State Dominant Variables. Most chemical processes are open-loop stable and obtaining a fast dynamic response is rarely the overriding operational concern. In such situations, the choice of the dominant variables y_d is dictated largely by the steady-state requirements of the secondary control loop.

Let us recall that the overall goal of the process is to meet specifications on the performance variables y_p. As discussed before, since the number of available manipulated variables $n_u = n_{u_p} + n_{u_d}$ is less than n_{y_p}, we resort to the partial control strategy. The primary control loop achieves all the *primary* or regulatory objectives such as those related to rejection of fast disturbances, stabilization, and other goals which are of concern from the point of view of process operability. This is necessarily a square control problem involving an equal and small number of measured variables y_d and manipulated variables u_d.

Once these primary objectives are satisfied, the remaining degrees of freedom are used to adapt the operating conditions so as to stay at the most desirable or optimum operating point and eliminate the effect of slow disturbances d_p. This *secondary* objective is carried out in the secondary control loop, using u_p and y_d^s as the optimization variables. Of these two, generally, u_p has a much slower impact and, in many cases, is expensive to manipulate even at the slower frequency of operation of the secondary control loop. A good example of this is the catalyst activity in the FCC (see the Motivation section) which can be influenced by slow acting catalyst additives or by expensive catalyst replacement.

If a disturbance d_p enters the plant, the secondary control loop should change the set points y_d^s so that if the process achieves these set points (via the primary controller), the effect of the disturbance d_p on *all* the performance variables y_p is eliminated or minimized. This suggests the following definition of steady-state dominant variables.

Definition 6 (Steady-State Dominance). *Process variables which under feedback control and in the presence of disturbances determine the steady-state values of (preferably all) the performance variables y_p in a well-defined and quantifiable manner are called steady-state dominant variables.*

As in the dynamic case, steady-state process variable dominance is an intrinsic property of a process, independent of whether appropriate manipulated variables exist in the primary control loop to achieve the desired steady states of these dominant variables. For example, as discussed before, reactor temperature is a dominant steady-state variable which impacts the steady-state reactor productivity, selectivity and yield, independent of the manipulated variable used to control it. Indeed, the identification of a particular dominant variable may suggest to the design engineer what manipulated variable is needed. If such a manipulated variable cannot be identified, an alternate dominant variable may have to be found.

The sets of steady-state dominant variables and dynamic dominant variables may overlap, as is the case for the exothermic reactor where temperature determines both the stability and the specifications. This is desirable because it reduces the minimal complexity of the required control system, but there is no reason why this should be true in general. Regardless of this overlap, all the dominant variables need to be controlled.

As discussed above, it may be possible to determine a candidate set of steady-state dominant variables in collaboration with the laboratory establishing the experimental data for the design ("concurrent design"). Once a mathematical model is available, we can identify candidate dominant variables from the model. We need to evaluate the sensitivity

$$\frac{\partial y_{p_j}}{\partial y_{d_j}} \tag{23}$$

for a chosen set of $y_d \in \Re^{n_{y_d}}$, for all the performance variables y_{p_j}. We can do this differentially as implied by the formula or by running a set of cases for various perturbations. For this purpose, we consider the set of differential equations and eliminate the equation for dy_{d_j}/dt. In this manner y_{d_j} becomes an independent parameter in the rest of the equations rather than a dependent variable. We can solve the system of equations in steady state (with the values for the manipulated variables and disturbances fixed at their steady-state values) and calculate y_{p_j} for various choices of y_{d_j}. These results will allow us to identify a set of dominant variables which are independent of any manipulated variables which may or may not be present.

The challenge for the control engineer is to come up with manipulated variables which are suitable for controlling these dominant variables. In some situations, the available manipulated variables may do the job. More likely, additional manipulated variables may have to be created through appropriate design changes and additions. Also, the identified dominant variables may not be measurable and alternative inferential dominant variables may have to be found. Finally, it may simply not be possible to achieve satisfactory control of the dom-

inant variables and the whole design may have to be modified —for example, a different catalyst may have to be chosen or the reactor design may have to be changed.

Rarely will these assessments be clear cut. It will generally not be possible to identify a set of manipulated variables where each one affects one and only one dominant variable. (An example of a fortunate case would be the control of FCC regenerator temperature, a dominant variable, by adding a catalyst cooler.) More commonly, each manipulated variable will affect all dominant variables and also other process variables in some manner. Clearly, we need a quantitative method to evaluate the effectiveness of the different options. We describe here a differential approach based on the system model equations (Eq. 18), but an equivalent finite-perturbation approach may be more appropriate for highly nonlinear systems.

A first-order perturbation of y_d in Eq. 18 gives

$$\delta y_d = \left(\frac{\partial \tilde{g}_d}{\partial u_d}\right)\delta u_d + \left(\frac{\partial \tilde{g}_d}{\partial u_p}\right)\delta u_p + \left(\frac{\partial \tilde{g}_d}{\partial d_p}\right)\delta d_p. \quad (24)$$

Here, the partial derivatives are assumed to be under the constraint $\tilde{f}(x, u_d, u_p, d_p) = 0$ so that the dependence on x is eliminated. For (u_d, y_d) to be a meaningful choice of measured and manipulated variables in the primary control loop, we must have $\partial \tilde{g}_d/\partial u_d$ nonsingular. Solving Eq. 24 for δu_d gives

$$\delta u_d = \left(\frac{\partial \tilde{g}_d}{\partial u_d}\right)^{-1}\delta y_d - \left(\frac{\partial \tilde{g}_d}{\partial u_d}\right)^{-1}\left(\frac{\partial \tilde{g}_d}{\partial u_p}\right)\delta u_p - \left(\frac{\partial \tilde{g}_d}{\partial u_d}\right)^{-1}\left(\frac{\partial \tilde{g}_d}{\partial d_p}\right)\delta d_p. \quad (25)$$

Similarly, a first-order perturbation of y_{p_i} along with the expression of δu_d above gives

$$\begin{aligned}\delta y_d &= \left(\frac{\partial \tilde{g}_p}{\partial u_d}\right)\delta u_d + \left(\frac{\partial \tilde{g}_p}{\partial u_p}\right)\delta u_p + \left(\frac{\partial \tilde{g}_p}{\partial d_p}\right)\delta d_p \\ &= \left(\frac{\partial \tilde{g}_p}{\partial u_d}\right)\left(\frac{\partial \tilde{g}_d}{\partial u_d}\right)^{-1}\delta y_d + \left[\left(\frac{\partial \tilde{g}_p}{\partial u_d}\right) - \left(\frac{\partial \tilde{g}_p}{\partial u_d}\right)\left(\frac{\partial \tilde{g}_d}{\partial u_d}\right)^{-1}\left(\frac{\partial \tilde{g}_d}{\partial u_p}\right)\right]\delta u_p \\ &\quad + \left[\left(\frac{\partial \tilde{g}_p}{\partial d_p}\right) - \left(\frac{\partial \tilde{g}_p}{\partial u_d}\right)\left(\frac{\partial \tilde{g}_d}{\partial u_d}\right)^{-1}\left(\frac{\partial \tilde{g}_d}{\partial d_p}\right)\right]\delta d_p. \quad (26)\end{aligned}$$

The selection criterion can now be deduced from the coefficient of δy_d on the righthand side of the above equation. y_d is a dominant variable for steady-state partial control if the matrix

$$\left(\frac{\partial \tilde{g}_p}{\partial u_d}\right)\left(\frac{\partial \tilde{g}_d}{\partial u_d}\right)^{-1} \quad (27)$$

has as many singular values of order unity or larger as there are dominant variables. (We assume here and in the following, of course, that all the variables y_p, y_d, u_d, u_p are scaled so that all "significant" changes are of order unity.) If this condition is satisfied, the pair (u_d, y_d) is a feasible candidate for determining the overall decomposition of the partial control system. Additionally, if

$$\left\| \left(\frac{\partial \tilde{g}_{p_i}}{\partial d_p}\right) - \left(\frac{\partial \tilde{g}_{p_i}}{\partial u_d}\right)\left(\frac{\partial \tilde{g}_d}{\partial u_d}\right)^{-1}\left(\frac{\partial \tilde{g}_d}{\partial d_p}\right) \right\| \quad (28)$$

is small for all $j = 1, \ldots, n_{y_p}$, then the choice of (u_d, y_d) also ensures good rejection of the steady-state disturbance d_p. Here, the matrix norm $\|\cdot\|$ roughly reflects the fact that changes in d_p should not affect y_p significantly.

Here, we have considered only the magnitude of $(\partial \tilde{g}_p/\partial u_d)(\partial \tilde{g}_d/\partial u_d)^{-1}$ to evaluate dominance. Additional information such as the sign of the terms in the vector $(\partial \tilde{g}_{p_i}/\partial u_d)(\partial \tilde{g}_d/\partial u_d)^{-1}$ at a given operating point can be used to determine the direction in which the values of y_{p_i} will move with changes in y_d. This is particularly important when some variable y_{p_i} is close to or outside its constraints and changes in y_d need to be made to move y_{p_i} inside its feasible partial control limits.

One should note that the method described above for identifying dominant variables is strongly dependent on the quality and accuracy of the underlying nonlinear model and needs to be backed by our knowledge of the process, as we seldom have an accurate model at the time of design. The rigorous method should therefore be looked upon as playing the same role as a "learning model" (Shinnar, 1978) for complex systems with poor models.

We can summarize the discussion on process variable dominance as follows:

1. Dominant process variables exist in every complex process. It is the extent to which they are dominant in a given process that varies and it is this variation which makes some processes more suitable for partial control than others.

2. Dominant dynamic process variables are the unstable and "slow" modes of the process. An understanding of the physical mechanism for instability, slow dynamics and dominant time scales provides additional insight for identifying them.

3. Preliminary screening of dominant steady-state variables can be carried out using standard sensitivity analysis and first-order perturbations through case-studies on the nonlinear steady-state model. Conclusions regarding the magnitude and direction in which a given process variable y_{p_i} changes locally with changes in the dominant output variable y_d can be made with this analysis. For highly nonlinear processes, equivalent case studies employing finite perturbation are possible.

4. The choice of dominant variables y_d is also dictated by model uncertainty: when the variables y_d are kept constant at a prescribed set point y_d^s, they should minimize the impact of dynamic model uncertainty (see the next subsection) on the overall system, and, when moved to a new set point, they should have a predictable steady-state effect on the performance variables y_p (see Steady State Modelability in the next subsection). This consideration of model uncertainty in the

choice of y_d is often neglected in the optimization-based approaches for choosing y_d that are reported in the literature.

Modelability of Process Responses. As with dominance, the term "modelability" by itself intuitively captures the notion of model availability for describing process behavior. Our discussion and analysis in the preceding sections was based on the assumption that we have at our disposal a dynamic model in the primary control loop and a general nonlinear steady-state model (Eq. 18) in the secondary control loop. However, for complex chemical processes, obtaining such models is a nontrivial and generally expensive task. However, the need for a minimum amount of model information is unavoidable in making decisions about structuring the control task.

Definition 7 (Modelability). *The relation between two process variables y_1 and y_2 is said to be modelable if, for a given change (increase or decrease) in y_1 at a given operating point, we can predict at least two aspects of the corresponding change in y_2:*

- *Trend — increase or decrease*
- *Order of magnitude — large, small or negligible.*

The above definition encapsulates the notion of *minimum model information.* The concept is general enough to be applicable to either the dynamic or steady-state relationship between any two variables—measured, manipulated, or disturbance. The "more" modelable the behavior between two process variables, the more accurately one can predict the response of one to changes in the other beyond this minimum requirement, and, consequently, the more complex is the model describing this relationship. The extent of modelability and the desired accuracy of model predictions is dictated by the end use of these predictions. There are two distinct aspects to modelability in the context of partial control:

- the *dynamic* relationship $[u_d(t), d_d(t)] \rightarrow y_d(t)$ in the primary control loop; and
- the *static* relationship $(y_d^s, u_p, d_p) \rightarrow y_p$ in the secondary control loop.

Dynamic Modelability. An understanding of the dynamic relationship $(u_d(t), d_d(t)) \rightarrow y_d(t)$ is required for designing the dynamic primary control law (Eq. 16). The complexity of the model describing this relationship is dictated by the performance requirements for the primary loop.

There will always be model uncertainty, that is, mismatch between the model predictions and the actual process response, caused by unmodeled dynamics, nonlinearities, uncertain parameters and time-varying plant characteristics. For linear time-invariant controllers, a well-established theory exists (Morari and Zafiriou, 1989) for carrying out such a robustness analysis and also for incorporating uncertainty information in arriving at *robust controller designs.* For constraint-handling controllers such as MPC and anti-windup (Kothare et al., 1994), until recently, such an analysis was considered difficult, but there has been significant progress in the past few years (see Kothare et al. (1996), Kothare and Morari (1999), and references therein). For nonlinear dynamic controllers, a general theory for robustness analysis is still in its infancy.

Steady-State Modelability. An understanding of the steady-state relationship $(y_d^s, u_p, d_p) \rightarrow y_p$ is required for designing the secondary control law (Eq. 17). For manual operation where a plant operator chooses the values of u_p and y_d^s by monitoring the performance variable y_p, a model of minimum complexity, in the sense of Definition 7, would suffice to make good decisions. For more formal optimizations of the form (Eq. 19), the model required is significantly more complex, involving fundamental conservation equations, thermodynamic principles, and reaction kinetics. Two critical factors dictate the required complexity of the steady-state model:

- *Feasibility:* The ability to predict steady states which are feasible for the *real* process over a wide range of operating conditions, feedstocks and plant parameter variations.
- *Partial controllability:* The ability to predict, with sufficient accuracy, the effect of changes in the steady-state values of y_d on the performance variables y_p.

For illustration, suppose we can identify a vector $\theta \in \Re^{n_\theta}$ of process model parameters such as heat-transfer coefficients, kinetic parameters, activity factors, feedstock characteristics, and so on which are not known exactly but lie in pre-specified bounds, $\theta_{j\min} \le \theta_j \le \theta_{j\max}$, $j = 1, 2, \ldots, n_\theta$. Let us now rewrite the steady-state model (Eq. 18) and the constraints with explicit dependence on the parameter vector θ

$$\begin{aligned} \text{Model} \quad & \tilde{f}(x, u_d, u_p, d_p, \theta) = 0 \\ & y_p = \tilde{g}_p(x, u_d, u_p, d_p, \theta) \\ & y_d = \tilde{g}_d(x, u_d, u_p, \theta) \\ \text{Constraints} \quad & u_{j\min} \le u_j \le u_{j\max},\ j = 1, \ldots, n_u \\ & y^s_{p_j\min} \le y_{p_j} \le y^s_{p_j\max},\ j = 1, \ldots, n_{y_p}. \end{aligned} \tag{29}$$

Assume further that this model is structurally the same as the "real" process, differing only in the value of θ. Then, the feasibility problem is the requirement that there exists a (hopefully large) set of possible (u_p, y_d^s) such that for every allowable θ the above model equations have a feasible solution that also satisfies feasibility for any other θ in its expected range. Since the model and the real process are assumed to be structurally the same, with different values of θ, this implies the existence of a feasible (u_p, u_d) for the real process. This feasibility discussion need not be restricted to parametric uncertainty, but can be extended to general uncertainty models.

This "robust feasibility" requirement is important in ensuring that the set points y_d^s which are downloaded to the primary controller are indeed feasible for the real process so that the primary controller can realistically achieve them without plant upsets. As shown in Arbel et al. (1996), infeasible set points for $y_d = [T_{\text{ris}}, T_{\text{rgn}}]$ in the FCC can cause the FCC to crash to the cold steady state, or lead to manipulated variable constraints, or cause the FCC to move to different and undesirable steady-state values of the dominant variables.

The partial controllability requirement is self-explanatory. Recall earlier that the selection criterion for dominance of y_d is that n_{y_d} singular values of the sensitivity matrix $\partial y_p / \partial y_d$ be of order one or larger. For the case with parametric uncertainty, if the steady-state models $\tilde{g}_p(x, u_p, u_d, d_p, \theta)$, $\tilde{g}_d(x, u_p, u_d, \theta)$, $\tilde{f}(x, u_p, u_d, d_p, \theta)$ are excessively sensitive to variations in θ, this criterion will give incorrect recommendations for the selection of y_d. A simple modification would be that the n_{y_d} singular values of the sensitivity matrix *when*

evaluated over all possible values of θ should be of order one or larger. This modification need not be restricted to parametric uncertainty, but can be extended to any general uncertainty models in a similar fashion.

Beyond the issue of selection of dominant variables, the accurate prediction of the impact of a change in the set point y_d^s on achieved steady-state values of y_p is very crucial for the synthesis of the secondary control law (Eq. 19). The issue of how steady-state uncertainty information, either structural or parametric, can be incorporated in the synthesis of the secondary control law is beyond the scope of this article (see Biegler et al. (1997) for a more refined treatment of this topic). However, this is a very important practical problem in present-day implementations of on-line steady-state real-time optimization (RTO) routines (Georgiou et al., 1997; Hendon, 1997; Mudt et al., 1995) and our modelability discussion serves to motivate such studies.

Practical Degrees of Freedom. The number of available manipulated variables $n_u = n_{u_d} + n_{u_p}$ is generally referred to as the *degrees of freedom* for control (Arbel et al., 1997). For the class of complex systems under consideration, this number is much less than the number of performance variables y_p. In partial control, we choose to directly and dynamically control only a small subset of process variables y_d using an equally small subset of manipulated variables u_d. The remaining manipulated variables u_p along with the set-point targets of y_d are then used to achieve as many partial control specifications on y_p as is possible.

In this scenario, merely knowing the available degrees of freedom for control is not enough to evaluate the ability of the control system to achieve the partial control goals, for the following reasons:

1. *Slow/fast degrees of freedom:* Different manipulated variables impact the process at differing speeds. However, despite their fast impact on the process, some manipulated variables are inherently expensive and difficult to manipulate at fast rates than others. Therefore, classification of u into slow variables u_p and fast variables u_d is not based only on the speed with which they impact the process, but also on the ease and cost of their manipulation at a desired frequency.

2. *Ineffective degrees of freedom:* Some potential manipulated variables may have no real impact on y_p. Moreover, from an operational point of view, some manipulated variables may be required to be essentially always at their constraint limit, thereby rendering them ineffective for control. And finally, two or more manipulated variables may have near identical effect on y_p. All these situations lead eventually to loss of some degrees of freedom for control.

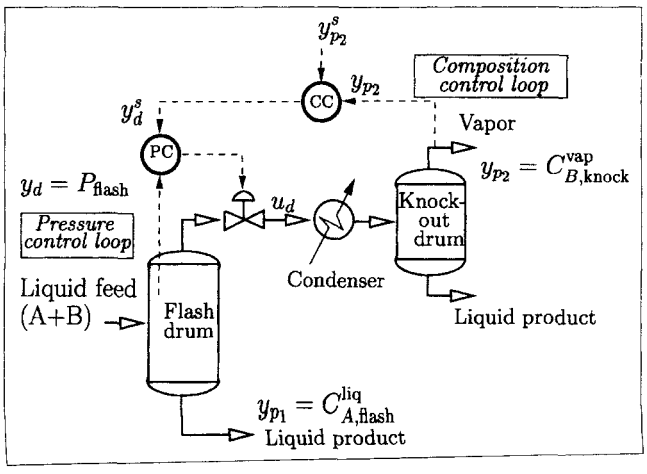

Figure 3. Partial control of a flash process.

Clearly, the concept of *degrees of freedom* needs to be modified to account for these situations. In partial control, the set points y_d^s are used to achieve the steady-state specifications on y_p. Ignoring u_p, this motivates the following definition:

Definition 8. *The effective or practical degrees of freedom for controlling the performance variables y_p in a process is given by the number of dominant variables y_d which can independently influence y_p at steady state and which are at the same time amenable to direct dynamic control.*

Illustrative Example

In this section, we will demonstrate the use of the various terms introduced in this article with the help of a simple flash process. The example is deliberately chosen to be simple so that the reader can become familiar with the numerous terms and concepts involved.

Consider the flash process shown in Figure 3. Liquid feed containing components A and B is introduced into the flash drum. The goal is to separate the volatile impurity A (such as hydrogen or nitrogen) from the nonvolatile condensible liquid product B (such as a light hydrocarbon). The low pressure in the flash drum causes the liquid stream to vaporize. Ideally, we would like to separate all the volatile impurity A as vapor and all the condensible product B as liquid. However, depending on the relative volatilities of A and B, there will be some B in the vapor stream and some A in the liquid stream leaving the flash drum.

The product B in the vapor stream is recovered by passing the vapor through a condenser and a knock-out drum. Once again, ideally, the vapor stream from the knock-out drum should ideally contain only the noncondensible A and the liquid stream should contain only the condensible B. However, both components will be present in both streams exiting the knock-out drum. The performance specifications can be expressed as follows:

In the face of changing inlet concentrations of A and B:

(1) Regulate the impurity concentration $y_{p_1} = C_{A,\text{flash}}^{\text{liq}}$ of A in the flash drum liquid stream at an acceptable low level, that is, $y_{p_1} \approx y_{p_1}^s$, where $y_{p_1}^s$ is a set point;

(2) Minimize unrecoverable B by regulating the concentration $y_{p_2} = C_{B,\text{knock}}^{\text{vap}}$ of B in the knockout drum vapor stream at an acceptable low level, that is, $y_{p_2} \approx y_{p_2}^s$, where $y_{p_2}^s$ is a set point.

The system has only one manipulated variable, the flow rate $u = F_{\text{flash}}^{\text{vap}}$ of the vapor leaving the flash drum. Thus, $n_u = 1 < n_{y_p} = 2$, thereby precluding exact control of both y_{p_1}, y_{p_2}. A solution based on partial control can be found by noting that the single *dominant* variable which strongly affects both y_{p_1} and y_{p_2} is the flash drum pressure P_{flash}, that is, $y_d = P_{\text{flash}}$. Assuming that there is some slack in the pressure rundown between the upstream and downstream vessels, P_{flash} can be varied over some range while still ensuring proper fluid flow. If P_{flash} is at its lowest allowable limit, we can flash a maximum amount of liquid to vapor, including the

product B, thereby meeting the requirement on the impurity concentration $C^{\text{liq}}_{A,\text{flash}}$, at the same time violating the requirement on the product concentration $C^{\text{vap}}_{B,\text{knock}}$ in the knock-out vapor stream. The opposite effect occurs when P_{flash} is operated at its highest allowable limit. The final design pressure is therefore arrived at by a compromise between these two conflicting requirements.

To account for process disturbances in the form of changing inlet concentrations of A and B, we can measure the concentration $C^{\text{vap}}_{B,\text{knock}}$ in the knock-out drum vapor stream or the concentration $C^{\text{liq}}_{A,\text{flash}}$ in the flash drum liquid stream, or both, and adjust the pressure set point to the pressure control loop, as shown in Figure 3. It may be possible, although sometimes expensive and even unnecessary, to add a heater to control the temperature of the feed entering the flash drum and potentially get complete control of the two desired variables $C^{\text{vap}}_{A,\text{knock}}$, $C^{\text{liq}}_{B,\text{flash}}$. As before, the set points for the inlet temperature and drum pressure would be derived by measuring y_{p_1} and y_{p_2}. In more complex systems, adding manipulated variables in this manner is often too expensive to be justifiable.

In summary, the main ingredients of partial control for the flash process are the following:

(1) $n_{y_p} = 2 > n_u = 1$, thereby precluding exact control of y_{p1} and y_{p2}.

2. Good *partial control* of y_{p_1}, y_{p_2}, that is, $y_{p1} \le C^{\text{liq}}_{A,\max}$, $y_{p_2} \le C^{\text{vap}}_{B,\max}$, can still be achieved by dynamically controlling P_{flash} which strongly influences both y_{p_1}, y_{p_2}. P_{flash} is the *dominant* variable for partial control, that is, $y_d = P_{\text{flash}}$. Here, $C^{\text{liq}}_{A,\max}$ and $C^{\text{vap}}_{B,\max}$ are acceptable and relaxed upper bounds on y_{p_1} and y_{p_2} (see Definition 2).

(3) y_d is controlled by $u_d = F^{\text{vap}}_{\text{flash}}$ in the fast dynamic *primary pressure control loop.*

(4) The set point y^s_d for P_{flash} is derived from the measurements y_{p_1}, y_{p_2} in the slow *secondary concentration control loop.*

(5) Our ability to use the measurements y_{p_1}, y_{p_2} to specify the set point y^s_d is based on our *modelability* of the steady-state relation $y^s_d \to (y_{p_1}, y_{p_2})$, that is, our knowledge of how the operating drum pressure affects y_{p_1} and y_{p_2}. Similarly, the *dynamic modelability* of the relation $u_d \to y_d$ allows us to define and operate the primary pressure control loop.

(6) The number of *independent* dominant variables y_d which can influence y_{p_1}, y_{p_2} (*practical degrees of freedom*) is one. An additional degree of freedom (y_{d_2}), the feed temperature, can be added if a feed preheater (u_{d_2}) is provided.

Partial Control of Complex Chemical Processes

Having familiarized ourselves with the various terms introduced in the article, we will next discuss applicability of partial control to two industrial-scale chemical processes of significant complexity. The emphasis will be on the concepts and how they apply to the problem of structuring the controller design.

Continuous crystallization with nucleation trap

Crystallization is a process where it is nearly impossible to obtain a reasonably reliable process model. As we will see, one of the dominant variables in crystallization is the rate of nucleation which determines particle-size distribution and various other crystal properties. The inherent difficulties in modeling the rate of nucleation have been well-documented (Rawlings et al., 1993) for most crystallizers of commercial interest. As a result, the corresponding control problem in turn becomes difficult.

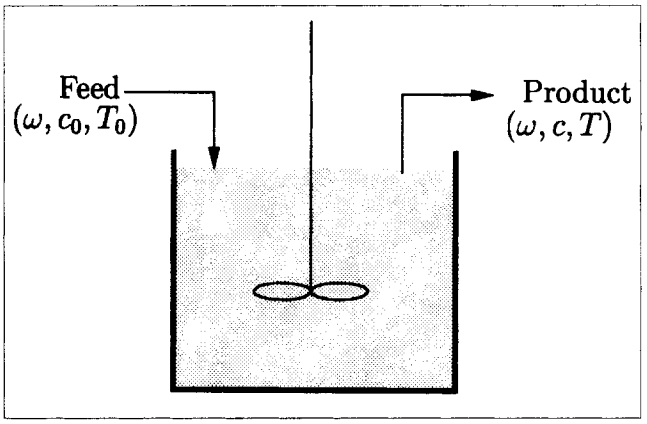

Figure 4. Continuous stirred-tank crystallizer.

Figure 4 shows a typical cooling crystallization process being carried out in a continuous well-mixed isothermal stirred tank. A saturated feedstream containing the solute to be crystallized is continuously fed to the tank of volume V at flow rate ω and solute concentration c_0. Cooling of the feed with cooling coils or by vaporization of the solvent in vacuum results in supersaturation of the solution, thereby providing the driving force for nucleation of the solute into crystals and the subsequent crystal growth. The mixed product slurry or crystal magma is continuously withdrawn from the tank at flow rate ω and solute concentration c.

The overall objective is to ensure stable production of the crystalline product with uniform properties by controlling the following performance variables y_p: particle-size distribution, average crystal size, uniformity of size distribution, crystal purity, the crystal form and structure and its mechanical and optical properties. The dimension of y_p is thus quite large. The crystal-size distribution has a strong impact on all other variables. The manipulated variables available to control y_p are feed rate ω, feed concentration c_0, feed temperature T_0 and crystallizer temperature T. The basic mechanism of crystal formation is believed to involve two general steps (Lei et al., 1971; Liss and Shinnar, 1976; Sherwin et al., 1967):

(a) nucleation, that is, formation of nuclei or *small* crystals from clear solution;

(b) crystal growth, that is, increase in size of the nuclei formed.

One of the earliest quantitative models for crystallizers was formulated by Sherwin et al. (1967), which can be written as an overall material balance

$$\frac{d}{dt}[\epsilon c + (1-\epsilon)\rho] = \frac{1}{\theta}c_0 - \frac{1}{\theta}[\epsilon c + (1-\epsilon)\rho]. \quad (30)$$

Here, $\theta = V/\omega$ is the residence time; ϵ is the fractional volume occupied by the solution; ρ is the crystal density. We have assumed that the feed does not contain any seed crys-

tals. A balance of particles of size r is given by

$$\frac{\partial f(r,t)}{\partial t} + \frac{\partial}{\partial r}\{G[c,c_s,f(r,t),\omega]f(r,t)\} = \epsilon B[c,c_s,f(r,t),\omega]\delta(r) - \frac{1}{\theta}f(r,t). \quad (31)$$

Here, c_s is the saturation of the solute at the given temperature; $f(r,t)$ is the particle-size distribution at time t such that, $f(r,t)dr$ is the number density of crystals per unit volume having radii in the range r, $r+dr$; $B[c,c_s,f(r,t),\omega]$ is the nucleation rate, that is, the number of new crystals formed per unit time per unit liquid volume; $G[c,c_s,f(r,t),\omega]$ is the growth rate dr/dt of an existing crystal. It is assumed that new crystals appear as a Dirac delta function $\delta(r)$, or that all new crystals are formed at a nominal size $r=0$. Numerous variations of this basic model have been studied in Liss and Shinnar (1976), Lei et al. (1971), and elsewhere. A recent extensive review of other modeling approaches for crystallization can be found in Rawlings et al. (1993).

Both $B(.)$ and $G(.)$ are complex functions of supersaturation $(c-c_s)$, size distribution $f(r,t)$, as well as many other often unknown factors such as agitation and mixing. One cannot determine $B(.)$ and $G(.)$ from simple experiments. In their simplest forms, $B(.)$ and $G(.)$ are functions of supersaturation $c-c_s$ only, although this is rarely found to be true in practice (Liss and Shinnar, 1976). The nucleation rate $B(.)$ is very sensitive to the mixing in the entrance zone of the feed. For most reactors, a stirred tank is a sensible modeling assumption as long as the mixing time is reasonably short compared to residence time. The sensitivity of $B(.)$ to the entrance effects makes it much more sensitive to scale-up and therefore much harder to predict.

Model-based multivariable control of y_p with the available manipulated variables is, therefore, difficult, if not impossible, due to the inherently large model uncertainty, or poor modelability of the process. Also, decentralized control is difficult because of the strong interactions between the controlled and manipulated variables. In addition, as shown by Shinnar and co-workers (Sherwin et al., 1967; Lei et al., 1971; Liss and Shinnar, 1976), industrial crystallizers can exhibit unstable behavior in the form of cyclic variation of the average crystal particle size, leading to loss of production rate and lowering of average product crystal size. Stabilization, optimal operation, and control of crystallizer present formidable challenges from a theoretical and practical point of view.

Mechanism of Instability and Partial Control. An understanding of the mechanism of instability in conjunction with the partial control concepts introduced in this article, provides guidelines to stabilize a large class of crystallizers and achieve partial control of y_p.

At equilibrium, there is a steady production of nuclei, their growth and subsequent removal as crystalline product. However, under certain operating conditions (Lei et al., 1971; Liss and Shinnar, 1976; Sherwin et al., 1967), the nucleation rate is extremely sensitive to positive deviations from the equilibrium solute concentration, leading to a significantly large production of nuclei in the short term. As the nucleated crystals are small in size, there is a significant time lag before they grow enough to provide a large surface area for growth. At this point, the rapid growth results in consumption of the solute and reduction of the supersaturation, thereby leading to a drop in the nucleation rate $B(.)$ which depends on the supersaturation. This internal feedback between supersaturation and $B(.)$ causes less than equilibrium formation of nuclei. Continuous withdrawal of the crystalline product reduces the available area for growth and the continuous addition of fresh feed raises the solute concentration. This gives rise to a large number of nuclei and repetition of the cycle. This instability manifests in the form of cyclic variations in the mean crystal size, leading to production of a nonuniform, off-spec crystalline product and lowering of the mean crystal size.

One can easily infer that the nucleation rate $B(.)$ is the single process variable which has a direct impact on stability and achievable product particle-size distribution and can be considered as a *dominant* variable y_d for partial control. Unfortunately, it is not possible to directly measure $B(.)$ nor is it possible to identify the functional form of $B(.)$. However, since variations in $B(.)$ lead to varying amounts of nuclei, an alternative dominant variable which can be used as an inferential measurement for $B(.)$ is the number of crystals below a certain critical size r_c. These are referred to as "fines" and are representative of the nucleation rate $B(.)$ in the limit as $r_c \to 0$. Thus, in the notation of partial control, y_d = fines concentration.

By directly controlling y_d without interaction with other process variables, one can stabilize the crystallizer and directly control the achievable particle-size distribution $f(r,t)$ which determines the crystal product properties in the performance vector y_p. One such method for independent control of y_d is by a nuclei or "fines" trap (Lei et al., 1971), as shown in Figure 5. Only fine particles less than a certain size r_c reach the top of the baffle where a stream is drawn off to a heater at a flow rate ω_0, while large particles settle at the bottom. By measuring the concentration and size of the fines (by light adsorption using a particle counter) in the *primary* fines control loop, one can adjust the flow ω_0 through the trap, thereby regulating the effective nucleation rate. In the notation of partial control $u_d = \omega_0$. The fines are then dissolved by heating and clear solution is recycled to the crystallizer. The set point y_d^s for the fines concentration y_d is derived from slow measurements of the product property vector y_p in the steady-state *secondary* control loop (not shown in the figure).

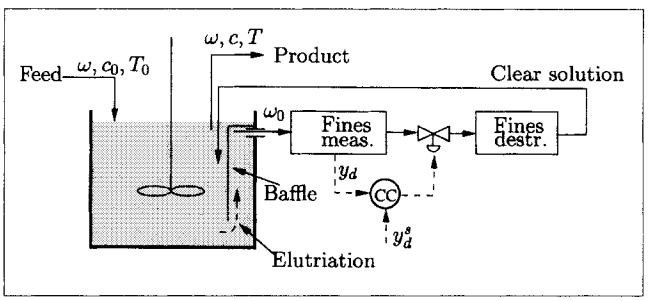

Figure 5. Partial control of a crystallizer with a fines trap.

Thus, by choosing a dominant variable—the fines concentration or equivalently, the nucleation rate—as the controlled variable and introducing into the process the fines flow rate ω_0 as the dominant manipulated variable, we are able to decouple the interactions, stabilize this complex system and directly control the product particle-size distribution $f(r,t)$ despite a very high model uncertainty. The remaining manipulated variables, that is, feed rate ω, feed temperature T_0, feed concentration c_0 and crystallizer temperature T can be used to achieve good partial control of the performance vector y_p, that is, in the notation of partial control $u_p = [\omega\ T_0\ c_0\ T]^T$. This partial crystallizer control scheme is patented (U.S. Patent No. 3649782-A, Shinnar (1972)) and used commercially.

Fluidized-bed catalytic cracking

Consider once again the fluidized-bed catalytic cracker (FCC) in Figure 1 which we discussed earlier. The two basic elements of the FCC are the reactor and the regenerator joined by the recirculating flow of catalyst. Oil feed enters the riser-reactor along with the hot, regenerated catalyst which cracks the oil into lower molecular weight products. During the catalytic cracking operation, the catalyst gets coated with carbonaceous material or *coke*, thereby causing the catalyst to deactivate. The circulating catalyst from the reactor is reactivated in the regenerator by burning off the coke with air. The combustion products leave the regenerator as stack gases and regenerated catalyst is recycled back to the reactor.

Critical to the operation of the FCC is the thermal balance between the reactor and regenerator: the heat released in the exothermic combustion reaction of coke and air in the regenerator raises the catalyst temperature, thereby providing the heat required for the endothermic catalytic cracking reaction in the reactor. In addition to stable operation, we need to control a number of variables which determine the performance and economic value of the process. These have been listed in Table 1 and have been discussed earlier.

To identify dominant variables in the process, we begin with the reactor which determines product distribution and is therefore the critical unit from an economic perspective.

FCC reactor. The reactor is an adiabatic riser-reactor with a high throughput and, therefore, short contact time. It is usually modeled as a plug-flow reactor (Arbel et al., 1995). The reaction rates in the riser, and, consequently, the product distribution in the exiting stream from the riser, are totally controlled by the inlet conditions to the riser. Therefore, the potential dominant process variables at the riser inlet are: inlet temperature (T_{mix}) which is the temperature of the mixture of the regenerated catalyst stream and the fresh feed stream; pressure; feed composition; catalyst activity and catalyst composition. Candidate dominant manipulated variables are: feed rate and catalyst flow rate (F_{cat}) both of which determine the residence time.

Due to reasons of overall refinery management, feed rate and feed composition are not allowed to be varied fast but can only be varied at very slow rates (usually, few times a day or slower). These two variables are therefore not suitable for the dynamic primary control of the rise-reactor and need to be included in the slower secondary optimizing control loop. Similarly, inherent catalytic activity and catalyst composition can only be controlled by addition of fresh catalyst or removal of nonregenerable catalyst, both of which can be accomplished only in the slow secondary control loop. Since the catalyst is reactivated by burning off coke in the regenerator, a meaningful variable which can be used to maintain a desired level of catalyst is C_{rgn}, the coke content on the regenerated catalyst.

This leaves us with T_{mix}, C_{rgn}, and pressure as the candidate dominant process variables for the reactor, with F_{cat} as the candidate dominant manipulated variable. Referring to Eqs. A-1, A-2, A-3, A-4, A-5, A-6 in the FCC model reported by Arbel et al. (1995), we note that there is no impact of pressure on the reaction rates, coke deposition, and the endothermic heat requirements for the catalytic reactions taking place in the riser-reactor. Moreover, fast control of pressure is needed to maintain the pressure balance between the regenerator and reactor which implies that riser pressure is not available as a dominant dynamic process variable for the riser. This leaves us with T_{mix}, C_{rgn} as the dominant process variables for the FCC riser-reactor, with F_{cat} as the dominant manipulated variable. T_{max} has the strongest impact on product composition in the exit stream from the riser-reactor and is the most dominant variable. Generally, the riser top temperature T_{ris} is used as an inferential measurement for T_{mix}.

It is interesting to note that the impact of the three candidate dominant variables (including the manipulated variable F_{cat}) can be directly evaluated in a laboratory-scale pilot-plant reactor where each variable can be varied independently of the other two. This facilitates better plant-scale modelability, based on laboratory-scale pilot-plant data, thereby enabling scale-up and concurrent design (see Shinnar et al. (2000) for details).

FCC Regenerator. The purpose of the regenerator is to reactivate the catalyst by burning off the coke formed on the catalyst during the cracking reactions in the reactor, and to supply the heat for the endothermic cracking by raising the temperature of the circulating catalyst before it enters the reactor. The three main exothermic combustion reactions taking place in the regenerator with their reaction rate expressions are given below (Arbel et al., 1995)

$$\mathrm{C}+\frac{1}{2}\mathrm{O_2}\rightarrow\mathrm{CO},\quad r_1=(1-\epsilon)\rho_c k_1\frac{C_{\text{rgn}}}{MWc}P_{\mathrm{O_2}}$$

$$\mathrm{C}+\mathrm{O_2}\rightarrow\mathrm{CO_2},\quad r_2=(1-\epsilon)\rho_c k_2\frac{C_{\text{rgn}}}{MW_c}P_{\mathrm{O_2}}$$

$$\mathrm{CO}+\frac{1}{2}\mathrm{O_2}\rightarrow\mathrm{CO_2},\quad r_3=\left[x_{pt}(1-\epsilon)\rho_c k_{3c}+\epsilon k_{3h}\right]P_{\mathrm{O_2}}P_{\mathrm{CO}}.$$

Here, ϵ is the regenerator bed void fraction, ρ_c is catalyst density, MW_c is the coke molecular weight, $P_{\mathrm{O_2}}$ and P_{CO} are the partial pressures of oxygen and carbon monoxide, k_1, k_2, k_{3h}, k_{3c} are rate constants with standard Arrhenius type temperature dependence, and x_{pt} is relative catalytic combustion rate. The combustion of CO to $\mathrm{CO_2}$ consists of a homogeneous free radical combustion reaction and a heterogeneous catalytic combustion, each of which combine to give the overall rate expression for CO combustion.

We immediately see that regenerator temperature T_{rgn} is a dominant process variable for the three reactions through the

Table 2. Dominant Variables for the FCC

	Dominant Variables	
	Process	Manipulated
Riser-reactor	$T_{mix}(T_{ris})$, C_{rgn}	F_{cat}
Regenerator	T_{rgn}, C_{rgn}, P_{O_2}	F_{cat}, F_{air}

temperature dependent rate constants k_1, k_2, k_{3h}, k_{3c}. Similarly, partial pressure P_{O_2} of oxygen is also dominant for all the three reactions in partial combustion mode when oxygen supply is insufficient for the combustion of coke and carbon monoxide. In complete combustion mode with excess oxygen, P_{O_2} is not dominant and C_{rgn} becomes dominant in controlling the CO_2/CO ratio and average carbon combustion rate. Thus, candidate dominant process variables are T_{rgn}, P_{O_2}, C_{rgn}. The dominant manipulated variables which have a direct and strong impact on these dominant process variables are the air feed rate F_{air} and the catalyst circulation rate F_{cat}. F_{air} directly affects P_{O_2} and indirectly affects T_{rgn} through the heat released in the exothermic reactions. F_{cat} controls the space velocity of the catalyst in the regenerator, thereby controlling C_{rng} through the coke balance and T_{rgn} through the heat balance.

Because coke on catalyst C_{rgn} is expensive to measure directly an inferential measurement is needed. As argued above the CO_2/CO ratio is a good indicator of C_{rgn}. The temperature rise (ΔT) across the cyclones at the exit of the regenerator reflects the CO_2/CO ratio and, therefore, C_{rgn}, but only if T_{rgn} is tightly controlled. Otherwise, the relationship is not unique (Arbel et al., 1995, 1996).

In summary for the FCC, the choices of the dominant process variables and the corresponding dominant manipulated variables for the riser-reactor and the regenerator are given in Table 2. As noted before, T_{ris} serves as an inferential measurement for T_{mix}. We note that C_{rgn} is a dominant process variable in both the reactor and the regenerator, and the catalyst circulation rate F_{cat} is a dominant manipulated variable in both units. The CO_2/CO ratio in the regenerator flue-gas stream is also an important dominant variable since it determines the relative extents of the CO and CO_2 formation reactions, and, hence, the heat release in the system. However, in many FCC units, the unconverted CO needs to be combusted in a CO boiler. As a result, the CO/CO_2 ratio becomes much less important as compared to C_{rgn}, and, therefore, it has not been listed above.

Sources of uncertainty. The main uncertainty in the riser-reactor is the complex mixing of the regenerated catalyst and the fresh feed taking place at the bottom of the riser. Typically, hot catalyst (1,200–1,300°F) is mixed with cold feed (400–700°F). The total residence time in the riser-reactor is about four seconds and it is virtually impossible to completely mix the large amount of solid catalyst and oil feed in a few seconds at the riser inlet. While the mixing time can be minimized by proper design, it is neither modelable nor predictable. This introduces a large uncertainty in the amount of coke formed on the catalyst, which, in turn, introduces uncertainty in the heat balance when this coke is combusted in the regenerator. Additionally, there is considerable uncertainty in the kinetics of the catalytic cracking reactions, and, consequently, in the product distribution in the riser exit stream. However, contrary to the mixing uncertainty, the kinetic uncertainty can be reduced by laboratory-scale experiments.

The major uncertainty in the regenerator arises from the inability to predict the CO_2/CO ratio of the flue-gas stream since this ratio determines the relative extents of the CO and CO_2 formation reactions, and, hence, the heat release and the amount of air needed to combust the coke on the catalyst. However, the largest uncertainty in the process is due to the integration of the two units through the recirculating catalyst. As discussed above, one cannot predict exactly the amount of coke formed on the catalyst in the riser-reactor, and, therefore, one cannot predict the coke input to the regenerator. This introduces uncertainty in the overall heat balance between the two units.

Current FCC catalysts have a self-adjusting property: the catalyst activity is a strong inverse function of C_{rgn}. As a result, both cracking and coke production in the riser-reactor decrease with increasing C_{rgn}. Since the coke balance is related to the heat balance between the two units, this self-adjusting property makes the overall process stable over a wide operating range. This does not eliminate the need for good control since the performance variables y_p (see Table 1) are still unpredictable and controlling them in the presence of uncertainty requires good control.

FCC partial control. From the point of view of control, we need to identify dominant variables for stability. There are no known cyclic instabilities (Arbel et al., 1995) for the type of endothermic reactions that occur in the riser-reactor. However, it is known (see Arbel et al. (1995)) that catalytic CO combustion in the regenerator can be cyclic. Since this instability is driven by the positive internal feedback of the impact of temperature on reaction rates in an exothermic reaction system, we conclude that regenerator temperature T_{rgn} is the dominant variable for stability.

T_{rgn} can be directly and independently controlled by adding a catalyst cooler (Arbel et al., 1997) to the regenerator. Cooling is achieved by circulation of a part of the hot regenerated catalyst through an additional catalyst bed equipped with a heat exchanger. The rate of circulation of the catalyst, F_{cool} in this loop can be controlled by a slide valve. Note that F_{cool} was not listed as an available manipulated variable, but was introduced here as a design change to specifically stabilize T_{rgn} independently. We may note the distinct similarity between the T_{rgn}-F_{cool} loop and the fines/nuclei control loop in the crystallizer in the previous section.

Having independent control of T_{rgn} allows us to control C_{rgn} (or its inferential measurement, CO_2/CO ratio) using F_{air} as the manipulated variable. Once T_{rgn} and C_{rgn} are controlled independently, all interactions between the riser-reactor and the regenerator are effectively eliminated. The dominant variable for the riser-reactor, which is T_{mix} (or its inferential measurement T_{ris}), can be controlled by F_{cat}. In the notation of partial control, we have

$$y_d = \begin{bmatrix} T_{rgn} \\ T_{ris} \\ C_{rgn} \end{bmatrix}, \quad u_d = \begin{bmatrix} F_{cool} \\ F_{cat} \\ F_{air} \end{bmatrix} \tag{32}$$

with pairings ($T_{rgn} - F_{cool}$, $T_{ris} - F_{cat}$, $C_{rgn} - F_{air}$). This partial control scheme has been shown to not only stabilize the pro-

cess in the face of considerable model uncertainty (Arbel et al., 1996, 1997) but also reject disturbances due to changes in feedstock and catalyst activity. y_p is now controllable in the slow secondary control loop by changing the set points y_d^s. The first FCC, built in the late 1930s without any pilot-plant studies and with very few laboratory experiments, had the same basic control structure. For the FCC of that time, the only specification for satisfactory performance was a minimum conversion. Despite this single specification, the need for stabilization in the presence of uncertainty led to the need for the $y_d - u_d$ control structure discussed above.

Conclusions

The concept of partial control was recognized nearly two decades ago by Shimar (1981). An explicit use of the basic idea did not occur until a few years ago when the concept was applied by Shinnar and co-workers (Arbel et al., 1995, 1996, 1997) in a series of articles to develop control strategies for the FCC, a highly complex and nonlinear petrochemical process. Extensive case studies showed how partial control can lead to a meaningful structuring of the control strategy for the FCC.

Motivated by these recent developments, this article is an attempt to formalize the notion of partial control in a rigorous framework. The basic premise adopted is that many complex processes having a large number of process variables, poorly understood models, and a limited number of manipulated variables can be controlled reasonably well by dynamically controlling only a small subset of process variables using an equally small number of manipulated variables. Central to the partial control framework are the concepts of *dominance, modelability, practical degrees of freedom* and *sufficiency*, each of which is defined. The proposed framework is general enough to incorporate various two-level hierarchies used in industrial practice such as a higher level secondary nonlinear optimizer (RTO) and a lower level primary multivariable predictive controller (MPC) or a higher level supervisory predictive controller and a lower layer of primary regulatory PI/PID controllers. In general, the formulation can be extended recursively to multi-level hierarchies.

The framework allows incorporation of engineering-based decisions in addition to formal techniques in addressing the control problem for complex systems. The presented formulation of the partial control problem and the illustrative examples should provide useful tools for developing control strategies for complex processes and at the same time provide stimulation for further research in the area of control of large-scale complex processes.

Acknowledgment

Financial support for this project from the U.S. Dept. of Energy, Division of Basic Energy Sciences, is gratefully acknowledged. The first author would like to acknowledge financial support from the Swiss Federal Institute of Technology (ETH), Zürich, during his stay at the City College of New York in the spring of 1997.

Literature Cited

Arbel, A., Z. Huang, I. H. Rinard, R. Shinnar, and A. V. Sapre, "Dynamics and Control of Fluidized Catalytic Crackers. 1. Modeling of the Current Generation of FCC's," *Ind. Eng. Chem. Res.*, **34**, 1228 (1995).

Arbel, A., I. Rinard, R. Shinnar, and A. Sapre, "Dynamics and Control of Fluidized Catalytic Crackers. 2. Multiple Steady States and Instabilities," *Ind. Eng. Chem. Res.*, **34**, 3014 (1995).

Arbel, A., I. H. Richard, and R. Shinnar, "Dynamics and Control of Fluidized Catalytic Crackers: 3. Designing the Control System: Choice of Manipulated and Measured Variables for Partial Control," *Ind. Eng. Chem. Res.*, **35**(7), 2215 (1996).

Arbel, A., I. H. Rinard, and R. Shinnar, "Dynamics and Control of Fluidized Catalytic Crackers: 4. The Impact of Design on Partial Control," *Ind. Eng. Chem. Res.*, **36**, 747 (1997).

Åström, K. J., and B. Wittenmark, *Computer Controlled Systems. Theory and Design (3rd ed.)*, Prentice Hall, Upper Saddle River, NJ (1997).

Bequette, B. W., "Nonlinear Control of Chemical Processes: A Review," *Ind. Eng. Chem. Res.*, **30**, 1391 (1991).

Biegler, L. T., I. E. Grossmann, and A. W. Westerberg, *Systematic Methods of Chemical Process Design*, Prentice Hall, Upper Saddle River, NJ (1997).

Buckley, P. A., *Techniques of Process Control*, Wiley, New York (1964).

Camacho, E. F., and C. Bordons, *Model Predictive Control in the Process Industry*, Springer, New York (1995).

Dorn, C., T. E. Güttinger, G. Wells, M. Morari, A. Kienle, E. Klein, and E.-D. Gilles, "Stabilization of an Unstable Distillation Column," *Ind. & Eng. Chemistry Res.*, **37**, 506 (1998).

Douglas, J. M., *Conceptual Design of Chemical Processes*, McGraw-Hill, New York (1988).

Georgiou, A., and C. A. Floudas, "Structural Analysis and Synthesis of Feasible Control Systems," *Chem. Eng. Res. Des.*, **67**, 600 (1989).

Georgiou, A., P. Taylor, R. Galloway, and L. Casey, "Plant Wide Closed Loop Real Time Optimization and Advanced Control of Ethylene Plant—CLTRO Improves Plant Profitability and Operability," *Proc. of NPRA Comput. Conf.*, New Orleans, LA (Nov. 16–19, 1997).

Govind, R., and G. J. Powers, "Control System Synthesis Strategies," *AIChE J.*, **28**, 60 (1982).

Hendon, S. R., "Equation-based Modeling and Optimization: The Cornersone for Integrated Site-Wide Optimization," *Proc. of Process Optimization Conf.*, Gulf Publishing and Hydrocarbon Processing, Houston, TX (April 8–10, 1997).

Henson, M. A. and. D. E. Seborg, eds., *Nonlinear Process Control*, Prentice Hall, Upper Saddle River, NJ (1995).

Hougen, J. O., *Measurements and Control Applications*, 2nd ed., Instrument Soc. of Amer. (1979).

Isidori, A., *Nonlinear Control Systems*, Springer-Verlag, Berlin, Germany (1995).

Kothare, M. V., "Control of Systems Subject to Constraints," PhD thesis, California Institute of Technology, Pasadena (March 1997).

Kothare, M. V., V. Balakrishnan, and M. Morari, "Robust Constrained Model Predictive Control Using Linear Matrix Inequalities," *Automatica*, **32**, 1361 (October 1996).

Kothare, M. V., P. J. Campo, M. Morari, and C. N. Nett, "A Unified Framework for the Study of Anti-Windup Designs," *Automatica* **30**, 1869 (1994).

Kothare, M. V., and M. Morari, "Multiplier Theory for Stability Analysis of Anti-Windup Control Systems," *Automatica*, **35**, 917 (May 1999).

Lei, S. J., R. Shinnar, and S. Katz, "The Stability and Dynamic Behavior of a Continuous Crystallizer with a Fines Trap," *AIChE J.*, **17**, 1459 (Nov. 1971).

Liss, B., and R. Shinnar, "The Dynamic Behavior of Crystallizers in which Nucleation and Growth Depend on Properties of the Crystal Magma," *AIChE Symp. Ser.*, **72**(153), 28 (1976).

Luyben, W. L., *Process Modeling, Simulation, and Control for Chemical Engineers*, 2nd ed., McGraw-Hill, New York (1990).

Luyben, W. L., B. D. Tyreus, and M. L. Luyben, *Plantwide Process Control*, McGraw Hill, New York (1998).

Meadowcroft, T. A., G. Stephanopoulos, and C. Brosilow, "The Modular Multivariate Controller: Part I: Steady-State Properties," *AIChE J.*, **38**, 1254 (1992).

Morari, M., "Integrated Plant Control: A Solution at Hand or a Research Topic for the Next Decade?," *Chem. Process Control—CPC II*, 467 (1982).

Morari, M., Y. Arkun, and G. Stephanopolous, "Studies in the Synthesis of Control Structures for Chemical Processes: I. Formulation of the Problem. Process Decomposition and the Classification

of the Control Tasks. Analysis of the Optimizing Control Structures," *AIChE J.*, **26**, 220 (March 1980).
Morari, M., C. E. Garcia, J. H. Lee, and D. M. Prett, *Model Predictive Control*, Prentice-Hall, Englewood Cliffs, NJ (in preparation) (2000).
Morari, M., and E. Zafiriou, *Robust Process Control*, Prentice-Hall, Englewood Cliffs, NJ (1989).
Mudt, D. R., T. W. Hoffman, and S. R. Hendon, "The Closed-Loop Optimization of a Semi-Regenerative Catalytic Reforming Process," AIChE Meeting, Houston, TX (March 1995).
Ng, C., and G. Stephanopoulos, "Plant-Wide Control Structures and Strategies," *Preprints of IFAC Symp. on Dynamics and Control of Process Systems (DYCOPS-5)*, C. Georgakis, ed., Corfu, Greece, pp. 1–16 (June 8–10, 1998).
Rawlings, J. B., S. M. Miller, and W. R. Witkowski, "Model Identification and Control of Solution Crystallization Processes—A Review," *Ind. & Eng. Chemistry Res.*, **32**, 1275 (July 1993).
Rinard, I. H., and J. J. Downs, "Plant Wide Control. A Review and Critique," AIChE Meeting (Spring 1992).
Sherwin, M., S. Katz, and R. Shinnar, "Dynamic Behavior of the Well-Mixed Isothermal Crystallizer," *AIChE J.*, **13**, 1141 (Nov. 1967).
Shinnar, R., "Chemical Reactor Modelling—The Desirable and the Achieveable," *Chem. Reaction Eng. Rev.*, D. Luss and J. V. W. Weekman, eds., ACS Symp. Ser., No. 72, p. 1, Houston, TX (1978).
Shinnar, R., "Chemical Reactor Modelling for Purposes of Controller Design," *Chem. Eng. Com.*, **9**, 73 (1981).
Shinnar, R., "The Partially Controlled Economy: 1. A Control Engineer Views Today's Economic Choices," *ChemTech*, **20**, 280 (1990a).
Shinnar, R., "The Partially Controlled Economy: 2. What is the Range of Today's Economic Choices?," *Chem. Tech.*, **20**, 354 (1990b).
Shinnar, R., B. Dainson, and I. H. Rinard, "Partial Control 5: A Systematic Approach to the Concurrent Design and Scale Up of Complex Processes. The Role of Control System Design in Compensating for Significant Model Uncertainties," *Ind. Eng. Chem. Res.*, **39**, 103 (2000).
Shinnar, R., F. J. Doyle, III, H. M. Budman, and M. Morari, "Design Considerations for Tubular Reactors with Highly Exothermic Reactions," *AIChE J.*, **38**, 1729 (1992).
Skogestad, S., and T. Larsson, "A Review of Plant-Wide Control," Internal Report, Dept. of Chem. Eng., Norwegian Univ. of Science and Technology, Trondheim, Norway (Aug. 1998).
Soeterboek, R., *Predictive Control—A Unified Approach*, Prentice Hall, Englewood Cliffs, NJ (1991).
Stephanopoulos, G., *Chemical Process Control. An Introduction to Theory and Practice*, Prentice Hall, Englewood Cliffs, NJ (1984).
Tyler, M. L., and M. Morari, "Propositional Logic in Control and Monitoring Problems," *Automatica*, **35**, 565 (April 1999).
Tyreus, B. D., "Dominant Variables for Partial Control: 1. A Thermodynamic Method for Their Identification," *Ind. Eng. Chem. Res.*, **38**, 1432 (1999a).
Tyreus, B. D., "Dominant Variables for Partial Control: 2. Application to the Tennessee Eastman Challenge Process," *Ind. Eng. Chem. Res.*, **38**, 1444 (1999b).
Zheng, A., R. V. Mahajanam, and J. M. Douglas, "Hierarchical Procedure for Plantwide Control System Synthesis," *AIChE J.*, **45**, 1255 (June, 1999).

Manuscript received Jan. 25, 2000, and revision received June 2, 2000.