A review of plantwide control

Sigurd Skogestad* and Truls Larsson
Department of Chemical Engineering
Norwegian University of Science and Technology
N–7034 Trondheim Norway

August 4, 1998

Abstract

Most (if not all) available control theories assume that a control structure is given at the outset. They therefore fail to answer some basic questions that a control engineer regularly meets in practice (Foss 1973): "Which variables should be controlled, which variables should be measured, which inputs should be manipulated, and which links should be made between them?" These are the question that plantwide control tries to answer.

There are two main approaches to the problem, a mathematically oriented approach (control structure design) and a process oriented approach. Both approaches are reviewed in the paper. Emphasis is put on the selection of controlled variables, and it is shown that the idea of "self-optimizing control" provides a link between steady-state optimization and control.

We also provide some definitions of terms used within the area of plantwide control.

This paper is an unpublished internal report which was written as part of the European Union CAPE.NET project (working group 5). We welcome any comments you may have.

*E-mail: skoge@chembio.ntnu.no; phone: +47-7359-4154; fax: +47-7359-4080

Contents

1 Introduction

A chemical plant may have thousands of measurements and control loops. By the term *plantwide control* it is *not* meant the tuning and behavior of each of these loops, but rather the *control philosophy* of the overall plant with emphasis on the *structural decisions*. The structural decision include the selection/placement of manipulators and measurements as well the *decomposition* of the overall problem into smaller subproblems (the control configuration).

Thus, a very important (if not the most important) problem in plantwide control is the issue of determining the *control structure*:

- Which "boxes" (controllers; usually consisting of a data handling and/or decision making part) should we have and what information should be send between them?

Note that that we are here not interested in what should be inside the boxes (which is the controller design or tuning problem). More precisely, **control structure design** is defined as the *structural decisions* involved in control system design, including the following tasks ((Foss 1973); (Morari 1982); (Skogestad and Postlethwaite 1996))

1. *Selection of controlled outputs* c (variables with setpoints)
2. *Selection of manipulated inputs* m
3. *Selection of measurements* v (for control purposes including stabilization)
4. *Selection of control* **configuration** (a structure interconnecting measurements/setpoints and manipulated variables, i.e. the structure of the controller K which interconnects the variables c_s and v (controller inputs) with the variables m)
5. *Selection of controller type* (control law specification, e.g., PID, decoupler, LQG, etc.).

In most cases the control structure is solved by a mixture between a top-down consideration of control objectives and which degrees of freedom are available to meet these (tasks 1 and 2), combined with a bottom-up design of the control system, starting with the stabilization of the process (tasks 3,4 and 5). In most cases the problem is solved without the use of any theoretical tools. In fact, the industrial approach to plantwide control is still very much along the lines described by Page Buckley in his book from 1964.

Of course, the control field has made many advances over these years, for example, in methods for and applications of on-line optimization and predictive control. Advances has also been made in control theory and in the formulation of tools for analyzing the controllability of a plant. These latter tools can be most helpful in screening alternative control structures. However, a systematic method for generating promising alternative structures has been lacking. This is related to the fact the plantwide control problem itself has not been well understood or even defined.

The control structure design problem is difficult to define mathematically, both because of the size of the problem, and the large cost involved in making a precise problem definition, which would include, for example, a detailed dynamic and steady-state model. An alternative to this is to develop heuristic rules based on experience and process understanding. This is what will be referred to as the *process oriented approach*.

The realization that the field of control structure design is underdeveloped is not new. In the 1970's several "critique" articles where written on the gap between theory and practice in the area of process control. The most famous is the one of (Foss 1973) who made the observation that in many areas application was *ahead* of theory, and he stated that

> The central issue to be resolved by the new theories is the determination of the control system structure. Which variables should be measured, which inputs should be manipulated and which links should be made between the two sets. ... The gap is present indeed, but contrary to the views of many, it is the theoretician who must close it.

A similar observation that applications seems to be ahead of formal theory was made by Findeisen *et al.* (1980) in their book on hierarchical systems (p. 10).

The issue is well illustrated by the following personal anecdote of Jack Ponton (Ponton and Liang 1993):

> Some years ago, when a fairly junior academic, but nonetheless with several years of teaching experience, he took an 'industrial sabattical' as a process engineer on a large project. The work in general had nothing to do with control, computers or any of the author's other main research interests. However, at one point he was asked what subjects he had taught, and mentioned that these included process control. "Ah!" said his questioner, "I have a control problem for you". The author was then presented with a process flowsheet an asked to put the control loops on it.
>
> The author was nonplussed and embarrassed. Despite having taught differential equations, Laplace transforms, Bode diagrams, root locus plots, and the other appurtenances of a traditional control course, he was at loss even as to start this task. So must have been generations not just of the author's students, but graduates of most chemical engineering degree courses. And this is *the* control task which process engineers in industry are most frequently called upon to perform.

Many authors point out that the need for a plant-wide perspective on control is mainly due to changes in the way plants are designed – with more heat integration and recycle and less inventory. Indeed, these factors lead to more interactions and therefore the need for a perspective beyond individual units. However, we would like to point out that even without any integration there is still a need for a plant-wide perspective as a chemical plant consists of a string of units connected in series, and one unit will act as a disturbance to the next, for example, all units must have the same through-put at steady-state.

Optimization and control

Maybe the most important reason for the slow progress in plantwide control theory, and in particular when it comes to the selection of which variables to control, is that most people do not realize that there is an issue. But ask the question:

> Why are we controlling hundreds of temperatures, pressures and compositions in a chemical plant, when there is no specification on most of these variables? Is it just because we can measure them or is there some deeper reason?

The starting point for any formalized procedure is a definition of the problem. So why do we do control? Well, first there is the issue of stabilization and then of keeping the operation within given constraints. However, even after some of the original degrees of freedom have been used to stabilize levels with no steady-state effect and satisfy product specifications, there are generally many degrees of freedom left. What should these be used for?

Loosely speaking, they should be used to "optimize the operation". There may be many issues involved, and to trade them off against each other in a systematic manner we usually quantify a scalar performance index J which should be minimized. In many cases this performance index is an economic measure, e.g. the operation cost. Since the economics of plant operation usually are determined mainly by steady-state issues, the analysis of how to use the remaining degrees of freedom can often be based on steady-state considerations, and their optimal values may be found using steady-state optimization. The resulting optimization problem may be very large, with hundreds of thousands of equations, and hundreds of degrees of freedom. However, with todays computers and optimization methods this problem is easily solvable, and it is indeed solved routinely in some plants, such as ethylene plants.

However, it is often much less clear how the optimal solution should actually be implemented in practice. Three alternative solutions are shown in Figure 1:

(a) Open-loop optimization.

(b) Closed-loop implementation with a separate control layer.

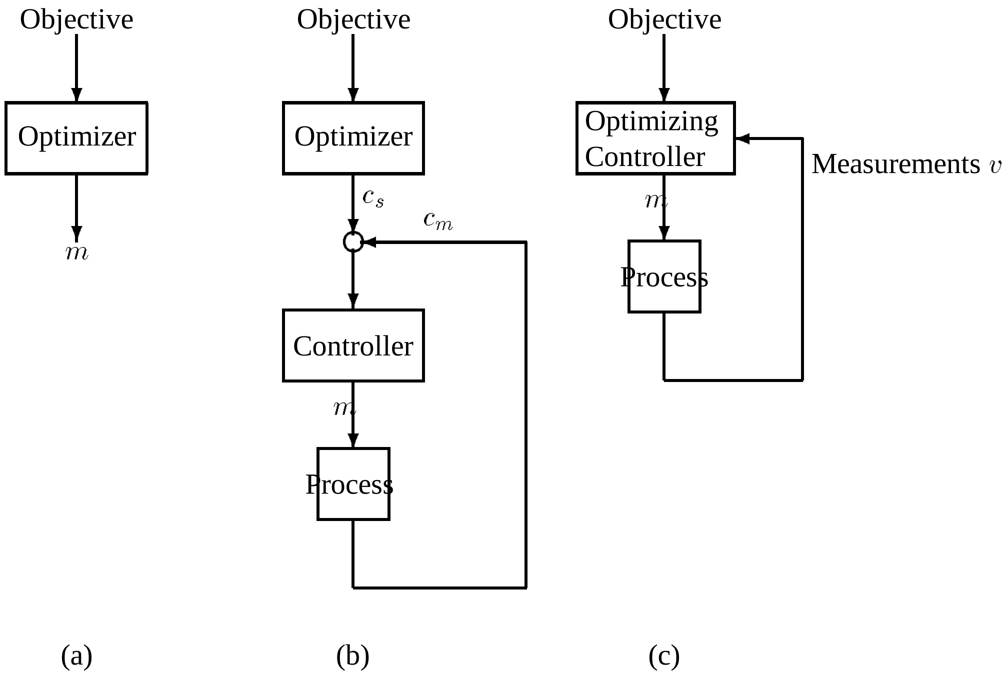

Figure 1: Alternative structures for optimization and control.

(c) Integrated optimization and control.

It should be stressed that in all the cases "optimization" may be performed manually be operators or engineers.

The open-loop implementation (a) can generally not be used because of sensitivity to uncertainty. In practice, the hierarchical feedback implementation (b) is preferred. It consists of

- *optimization layer* — computes setpoints c_s for the controlled variables c
- *control layer* — implements this in practice, with the aim of achieving $c \approx c_s$.

The optimization layer typically recomputes new setpoints c_s only every hour or so, whereas the feedback layer operates continuously. However, the data and model used by the optimizer are uncertain and there are disturbances entering the plant between each re-optimization, and the objective of the feedback layer is therefore to keep the plant close to its optimal operating point in spite of this uncertainty. One important issue, which will be discussed in detail, is to select the variables c which are to be controlled (task 1 in control structure design).

Of course, we could imagine using solution (c) above in which we have a centralized optimizing controller which stabilizes the process while at the same time perfectly coordinates all the manipulated inputs based on dynamic on-line optimization. There are fundamental reasons why such a solution is not the best, even with todays and tomorrows computing power. One fundamental reason is the cost of modeling, and the fact that feedback control, without much need for models, is very effective when performed locally. In fact, by cascading feedback loops, it is possible to control large plants with thousands of variables without the actual need to develop any models. However, the traditional single-loop control systems can sometimes be rather complicated, especially if the cascades are heavily nested or if the presence of constraints during operation make it necessary to use logic switches. Thus, model based control should be used when the modeling effort gives enough pay-back in terms of simplicity and/or improved performance, and this will usually be at the higher layers in the control hierarchy.

The resulting control system is usually divided into more than an optimization and a control layer. Typically, layers include include scheduling (weeks), site-wide optimization (day), local optimization (hour), supervisory/predictive control (minutes) and regulatory control (seconds); see Figure 2.

It is important to note that when we close a loop somewhere in the hierarchy (e.g. a loop for level control), then we do not really use any degrees of freedom (since the setpoint for the level is still a degree

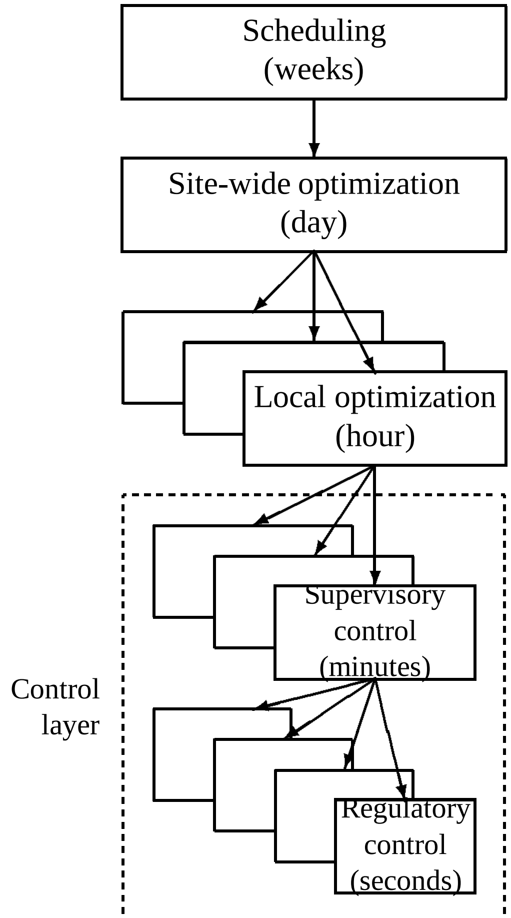

Figure 2: Typical control hierarchy in a chemical plant.

of freedom). Thus, there are generally many degrees of freedom left for optimization. However, note that some variables (e.g. levels in buffer tanks) may have no steady-state effect, so the setpoints for these variables can only be used dynamically.

Finally, let us remark that widespread use of model predictive control does not eliminate the control structure design problem. True, with model predictive control (or with alternative sequential modular approaches such as that proposed by Meadowcroft *et al.* (1992)) we can solve larger problems with many inputs and outputs, but unless we plan on implementing a single large nonlinear model predictive controller for the entire plant, we will still have to worry about how the various model predictive controllers interact.

Related work

Parts of this review are based on Chapter 10 in the book of Skogestad and Postlethwaite (1996). In addition, we have made use of some unpublished work by Skogestad and coworkers on *self-optimizing control*. The latter work is planned to be published as a series of papers with the following tentative titles:

Part 1. The basic issues in self-optimizing control (selection of controlled outputs to make implementation of the optimal solution insensitive to uncertainty).

Part 2. Taylor series analysis.

Part 3. Theoretical basis for using the minimum singular value for output selection.

Part 4. Partial and indirect control with application to selection of temperature measurements in distillation.

Part 5. Constraints and feasibility.

Except for the book of Skogestad and Postlethwaite (1996), preliminary versions of the above work are available in the Ph.D. theses of Morud (1995), Glemmestad (1997) and Havre (1998), as well as in a number of conference publications. These references are available on the Internet[1].

Outline

We will first discuss in more detail some of the terms used above and provide some definitions. We then present a review of some of the work on plantwide control. In section 4 we discuss the mathematically oriented approach (control structure design). Then, in section 5 we look at the process oriented approach. In section 6 we consider a fairly simple plant consisting of reactor, separator and recycle. Finally, in section 7 we consider the most studied plantwide control problem, namely the Tennessee Eastman problem introduced by Downs and Vogel (1993), and we discuss how various authors have attempted to solve the problem.

2 Terms and definitions

We here make some comments on the terms introduced above, and also attempt to provide some more precise definitions.

Let us first consider the terms *plant* and *process* which are almost synonymous terms. In the control community as a whole, the term plant is somewhat more general than process: A *process* usually refers to the "process itself" (without any control system) whereas a *plant* may be *any system to be controlled* (including a a partially controlled process). This is how we will use these terms in this paper (however, note that in the chemical engineering community the term plant has a somewhat different meaning, namely as the whole factory which consists of many process units; the term *plantwide* control is derived from this meaning of the word plant.)

Let us then discuss the closely related terms *layer* and *level* which are used in hierarchical control. Following the literature (e.g. Findeisen et al. (1980) the correct term for us is *layer*. This term is used when the control system is split into layers that act at different time scales. Each layer has some feedback or information from the process and follows setpoints given from layers above. A lower layer may not know the criterion of optimality by which the setpoint has been set. A *multi-layer* system cannot be strictly optimal because the actions of the higher layers are discrete and thus unable to follow the strictly optimal continuous time pattern. On the other hand, in a *multilevel* system there is no time scale separation and the levels are coordinated such that there are no performance loss. Multilevel decomposition may be used in the optimization algorithm but otherwise is of no interest here.

The design of a technical system may for our purposes be divided into two main steps:

I Process design

II Control system design

The latter activity may be divided into many steps of which the main ones are:

IIa Control structure design (structural decisions)

IIb Controller design (parametric decisions)

IIc Implementation

The term *control structure design* is now commonly used in the control community as a whole, and it refers to the structural decisions involved when designing the control system as defined by the five tasks given in the introduction:

[1]http://www.chembio.ntnu.no/users/skoge/

1. *Selection of controlled outputs* (c with setpoints c_s).
2. *Selection of manipulated inputs* (m).
3. *Selection of measurements* (v)
4. *Selection of control* **configuration**
5. *Selection of controller type*

The result from the *control structure design* is the *control structure* (alternatively denoted the *control strategy* or *control philosophy* of the plant).

The term *plantwide control* is used only in the process control community. One could regard *plantwide control* as the *process control version of control structure design*, but this is probably a bit too limiting. In fact, Rinard and Downs (1992) refer to the control structure design problem as defined above as the "strict definition of plantwide control", and they point out that there are other important issues such as the interaction with the operators, issues related to startup, changeover and shut-down, fault detection and performance monitoring and design of safety and interlock systems. This is more in line with the discussion in (Stephanopoulos, 1982).

Maybe a better distinction is the following: *Plantwide control* refers to the structural and strategic decisions involved in the control system design of a complete chemical plant (factory). The systematic (mathematical) approach to solving this problem is called *control structure design*.

The control configuration is defined as the restrictions imposed by the overall controller K by decomposing it into a set of local controllers (sub-controllers), units, elements, blocks) with predetermined links and possibly with a predetermined design sequence where sub-controllers are designed locally.

Operation involves the behavior of the system once it has been build, and this includes a lot more than control. More precisely, the control system is designed to aid the operation of the plant. *Operability* is the ability of the plant (together with its control system) to achieve acceptable operation (both statically and dynamically). Operability includes switchability and controllability as well as many other issues.

Flexibility refers to ability to obtain feasible steady-state operation at a given set of operating points. This is a steady-state issue, and we will assume it to be satisfied at the operating points we consider. It is not considered any further in this paper.

Switchability refers to the ability to go from on operating point to another in an acceptable manner usually with emphasis on feasibility. It is related to other terms such optimal operation and controllability for large changes, and is not considered explicitly in this paper.

We will assume that the "quality (goodness) of operation" can be quantified in terms of a scalar performance index (objective function) J, which should be minimized. For example, J can be the operating costs.

Optimal operation usually refers to the *nominally* optimal way of operating a plant as it would result by applying steady-state and/or dynamic optimization to a model of the plant (with no uncertainty), attempting to minimize the cost J by adjusting the degrees of freedom.

In practice, we cannot obtain optimal operation due to uncertainty. The difference between the actual value of the objective function J and its nominally optimal value is the *loss*.

The two main sources of *uncertainty* are (1) *signal uncertainty* (includes disturbances (d) and measurement noise (n)) and (2) *model uncertainty*.

Robust means insensitive to uncertainty. *Robust optimal operation* is the actual optimal way of operating a plant (with uncertainty considerations included).

Control is the adjustment of available degrees of freedom (manipulated inputs u) to assist in achieving acceptable operation (including stability) in spite of the presence of uncertainty.

Integrated optimization and control (or optimizing control) refers to a system where optimization and its control implementation are integrated. In theory, it should be possible to obtain robust optimal operation with such a system. In practice, one often uses an *hierarchical decomposition* with separate layers for optimization and control. In making this split we assume that for the control system the goal of "acceptable operation" has been translated into "keeping the controlled variables (c) within specified bounds from their setpoints (c_s)". The optimization layer sends setpoint values (c_s) for selected controlled outputs (c) to the control layer. The setpoints are updated only periodically. (The tasks, or parts of the tasks, of either of these layers may be performed by humans.) The control layer may be further divided, e.g. into supervisory control and regulatory control. In general, in a hierarchical system, the lower layers work on a shorter time scale.

In addition to keeping the controlled variables at their setpoints, the control system must *"stabilize"* the plant. We have here put stabilize in quotes because we use the word in an extended meaning, and include both modes which are mathematically unstable as well as slow modes ("drift") that need to be "stabilized" from an operator point of view. Usually, stabilization is done within a separate (lower) layer of the control system, often called the regulatory control layer. The controlled outputs for stabilization are measured output variables, and their setpoints may be used as degrees of freedom for the layers above.

For each layer in a control system we use the terms *controlled output* (y with setpoint y_s) and *manipulated input* (u). Correspondingly, the term "plant" refers to the system to be controlled (with manipulated inputs u and controlled outputs y). The layers are often structured hierarchically, such that the manipulated input for a higher layer (u_1) is the setpoint for a lower layer (y_{2s}), i.e. $y_{2s} = u_1$. (These controlled outputs need in general not be measured variables, and they may include some of the manipulated inputs (u).)

From this we see that the terms plant, controlled output (y) and manipulated input (u) takes on different meaning depending on where we are in the hierarchy. To avoid confusion, we reserve special symbols for the variables at the top and bottom of the hierarchy. Thus, as already mentioned, the term *process* is often used to denote the uncontrolled plant as seen from the bottom of the hierarchy. Here the the manipulated inputs are the physical manipulators (e.g. valve positions), and are denoted m. Correspondingly, at the top of the hierarchy, we use the symbol c to denote the controlled variables for which the setpoint values (c_s) are determined by the optimization layer.

Input-Output controllability of a plant is the ability to achieve acceptable control performance, that is, to keep the controlled outputs (y) within specified bounds from their setpoints (r), in spite of signal uncertainty (disturbances d, noise n) and model uncertainty, using available inputs (u) and available measurements. In other words, the plant is controllable if there exists a controller which satisfies the control objectives.

This definition of controllability may be applied to the control system as a whole, or to parts of it (in the case the control layer is structured). The term controllability generally assumes that we use the best possible multivariable controller, but we may impose restrictions on the class of allowed controllers (e.g. consider "controllability with decentralized PI control").

A plant is *self-regulating* if we with constant inputs ($u = 0$) can keep the controlled outputs within acceptable bounds. (Note that this definition may be applied to any layer in the control system, so the plant may be a partially controlled process). "True" self-regulation is defined as the case where no control is ever needed at the lowest layer. It relies on the process itself to dampen the disturbances, e.g. by having large buffer tanks. We rarely have "true" self-regulation because it may be very costly.

A process is optimizable if it is possible to keep the loss within an acceptable bound in spite of uncertainty. In other words, a process is optimizable if there exists an integrated optimization and control system with an acceptable loss.

A process together with its control structure is *self-optimizing* if we with constant setpoints for the optimized variables (c_s) can keep the loss within an acceptable bound within a specified time period (that is, the sensitivity of the economic objective to uncertainty is less than the accepted limit). We may also view self-optimization as as special case of self-regulation when viewed from the optimization layer (with $u = c_s$). "True" self-optimization is defined as the case where no re-optimization is ever needed (so c_s can be kept constant always), but this objective is usually not satisfied. On the other hand, we must require that the process is self-optimizable within the time period between each re-optimization, or else we cannot use separate control and optimization layers.

A process is self-optimizable if there exists a set of controlled outputs (c) such that if we with keep constant setpoints for the optimized variables (c_s), then we can keep the loss within an acceptable bound within a specified time period. A steady-state analysis is usually sufficient to analyze if we have self-optimality. This is based on the assumption that the closed-loop time constant of the control system is smaller than the time period between each re-optimization (so that it settles to a new steady-state) and that the value of the objective function (J) is mostly determined by the steady-state behavior (i.e. there is no “costly” dynamic behavior e.g. imposed by poor control).

Remark 1. Most of the terms given above are in standard use and the definitions are mostly follows those of Skogestad and Postlethwaite (1996). Some of the terms, like self-optimization, have previously not been formally defined.

Remark 2. Luyben (1988) introduced the term “eigenstructure” to describe the inherently best control structure (with the best self-regulating and self-optimizing property). However, he did not really define the term, and also the name is unfortunate since “eigenstructure” has a another unrelated mathematical meaning in terms of eigenvalues.

3 General reviews and books on plantwide control

We here presents a brief review of some of the previous reviews and books on plantwide control.

Morari (1982) presents a well-written review on plantwide control, where he discusses why modern control techniques were not (at that time) in widespread use in the process industry. The four main reasons were believed to be

1. Large scale system aspects.
2. Sensitivity (robustness).
3. Fundamental limitations to control quality.
4. Education.

He then proceeds to look at how two ways of decompose the problem:

1. Multi-layer (vertical), where the difference between the layers are in the frequency of adjustment of the input.
2. Horizontal decomposition, where the system is divided into noninteracting parts.

Stephanopoulos (1982) states that the synthesis of a control system for a chemical plant is still to a large extent an art. He asks: “Which variables should be measured in order to monitor completely the operation of a plant? Which input should be manipulated for effective control? How should measurements be paired with the manipulations to form the control structure, and finally, what the control laws are?” He notes that the problem of plantwide control is “multi-objective” and “There is a need for a systematic and organized approach which will identify all necessary control objectives”. The article is comprehensive, and discusses many of the problems in the synthesis of control systems for chemical plants.

Rinard and Downs (1992) review much of the relevant work in the area of plantwide control, and they also refer to important papers that we have not referenced. They conclude the review by stating that “the problem probably never will be solved in the sense that a set of algorithms will lead to the complete design of a plantwide control system”. They suggests that more work should be done on the following items: (1) A way of answering whether or not the control system will meet all the objectives, (2) Sensor selection and location (where they indicate that theory on partial control may be useful), (3) Processes with recycle. The also welcome computer-aided tools, better education and good new test problems.

The book by Balchen and Mummé (1988) attempts to combine process and control knowledge, and to use this to design control systems for some common unit operations and also consider plantwide control. The book provides many practical examples, but there is little in terms of analysis tools or a systematic framework.

The book “Integrated process control and automation” by Rijnsdorp (1991), contains several subjects that are relevant here. Part II in the book is on optimal operation. He distinguishes between two situations, sellers marked (maximize production) and buyers marked (produce a given amount at lowest possible cost). He also have a procedure for design of a optimizing control system.

Loe (1994) presents a systematic way of looking at plants with the focus is on functions. The author covers “qualitative” dynamics and control of important unit operations.

van de Wal and de Jager (1995) lists several criteria for evaluation of control structure design methods: generality, applicable to nonlinear control systems, controller-independent, direct, quantitative, efficient, effective, simple and theoretically well developed. After reviewing they concludes that such a method does not exist.

The book by Skogestad and Postlethwaite (1996) has two chapters on controllability analysis and one chapter on control structure design. Many of the ideas presented in this paper are based on this work.

The coming monograph by Ng and Stephanopoulos (1998*a*) deals almost exclusively with plantwide control.

There also exists a large body of system-theoretic literature within the field of large scale systems, but most of it has little relevance to plantwide control. One important exception is the book by Findeisen et al. (1980) on “Control and coordination in hierarchical systems” which probably deserves to be studied more carefully by the process control community.

4 Control Structure Design

In this section we look at the mathematically oriented approach to plantwide control.

Structural methods

There are some methods that use structural information about the plant as a basis for control structure design. Central concepts are structural state controllability, observability and accessibility. Based on this, sets of inputs and measurements are classified as viable or non-viable. Although the structural methods are interesting, they are not quantitative and usually provide little information other than confirming insights about the structure of the process that most engineers already have. For a recent review of these methods we refer to the coming monograph of Ng and Stephanopoulos (1998*a*).

In the reminder of this section we discuss the five tasks of the control structure design problem, listed in the introduction. Emphasis is put on selection of controlled outputs (task 1).

4.1 Selection of controlled outputs (c)

By selection of controlled outputs we here refer to the controlled variables c for which the setpoints c_s are determined by the optimization layer. There will also be other (internally) controlled outputs which result from the decomposition of the controller into blocks or layers (including controlled measurements used for stabilization), but these are related to the control configuration selection, which is discussed as part of task 4.

The issue of selection of controlled outputs (task 1), is probably the least studied of the tasks in the control structure design problem. In fact, it seems from our experience that most people do not consider it (selection of controlled outputs) as being an issue at all. The most important reason for this is probably that it is a structural decision for which there has not been much theory. Therefore the decision has mostly been based on engineering insight and experience, and the validity of the decision has seldom been questioned by the control theoretician.

In the introduction we asked the question: *Why are we controlling hundreds of temperatures, pressures and compositions in a chemical plant, when there is no specification on most of these variables?* After some thought, one realizes that the main reason for controlling all these variables is that one needs to specify the available degrees of freedom in order to keep the plant close to its optimal operating point. But there is a follow-up question: *Why do we select a particular set c of controlled variables? (e.g.,* *why specify (control) the top composition in a distillation column, which does not produce final products, rather than just specifying its reflux?)* The answer to this second question is less obvious, because at first it seems like it does not really matter which variables we specify (as long as all degrees of freedom are consumed, because the remaining variables are then uniquely determined). However, this is true only when there is no uncertainty caused by disturbances and noise (signal uncertainty) or model uncertainty. When there is uncertainty then it does make a difference how the solution is implemented, that is, which variables we select to control at their setpoints.

Thus, when selecting controlled outputs (task 1) one should aim at finding a set of variables which achieves *self-optimizing control*. After having made this realization, we can formalize the approach. Before we proceed, let us make it clear that we may as a special case include some of the manipulated inputs (m's) in the set of controlled outputs (c's). Thus, rather than controlled "outputs" it may be better to use the more general term *controlled variables*. The two methods given below for selecting controlled outputs were first presented in Chapter 10.3 of Skogestad and Postlethwaite (1996), but they are here derived in a slightly different way, and are complemented by a simple example.

Method 1: Evaluating the performance index (cost) J

Assume that the optimal operation problem is defined in terms of a performance index (cost) J, which is a scalar function to be minimized with respect to the available degrees of freedom. J may be a purely economic objective, but is more generally a weighted sum of the various control objectives. For the optimization itself it does not really matter which variables we use as degrees of freedom as long as they form an independent set. Let the "base set" for the degrees of freedom be denoted u (these may consist, for example, of a subset the physical manipulators m). In addition, the cost will depend on the disturbances d (which here is assumed to include uncertainty in the model and uncertainty in the optimizer). We can then write $J(u, d)$. The nominal value of the disturbances is denoted d_0, and we can solve the nominal operating problem and obtain $u_{opt}(d_0)$ for which

$$min_u J(u, d_0) = J_{opt}(d_0) = J(u_{opt}(d_0), d_0)$$

From this we can obtain a table with the corresponding optimal value of any other dependent variable, including $c_{opt}(d_0)$.

The issue is now to decide how to best implement the optimal policy in the presence of uncertainty by selecting the right set of controlled variables c with constants setpoints $c_s = c_{opt}(d_0)$. As already mentioned, if there where no uncertainty it would make no difference which variable c that was chosen.

We assume that the controlled variables can be controlled within accuracy e (where e is at least as large as the measurement noise for the variable c). Then the set of variables c we are looking for are the ones which minimize some mean value of the performance index

$$J(\underbrace{c_s + e}_{c}, d)$$

for the expected set of disturbances $d \in \mathcal{D}$, and expected set of control error $e \in \mathcal{E}$.

In the simplest case we select the setpoints as $c_s = c_{opt}(d_0)$, but the value of c_s may also be the subject to an optimization.

If we are already performing a steady-state optimization of our plant, then the objective function J is already defined, and except for the issue of combinatorial complexity, it is straightforward to find the optimal set of controlled outputs c that minimize the mean value of the performance index J thus having the best "self-optimizing" property.

Instead of evaluating the mean value of the performance index, it may be better to evaluate the always positive loss function. The loss function expresses the difference between the actual operating costs (while keeping $c = c_{opt}(d_0) + e$) and the optimal operating cost (while keeping $c = c_{opt}(d)$),

$$L(c, d) = J(c, d) - J_{opt}(d)$$

where

$$J_{opt}(d) = \min_u J(u,d)$$

The loss has the advantage of providing a better "absolute scale" on which to judge whether a given set of controlled variables c is "good enough", and thus is self-optimizing.

Constraint problems

The approach outlined above may be extended to include problems with constraints,

$$\begin{array}{lll} \min\limits_{u} & J(u,d) & \\ \text{subject to} & g_1(u,d) & = 0 \\ & g_2(u,d) & \leq 0 \end{array} \tag{1}$$

Problems with equality constraints are relatively straightforward, especially if we can identify a single variable (manipulated or measured) directly related to the constraint; this should then be included as a controlled variables c ("active constraint control"). The main effect is then that each constraint removes a degree of freedom for the optimization. The same argument holds for inequality constraints where the optimal policy is always to keep them active (i.e. satisfy them as equalities for any disturbance). The more difficult problems are when we have inequality constraints which are active only under certain conditions (disturbances). For such cases one must be careful to avoid infeasibility during implementation. The on-line optimization is usually for simplicity based on the nominal disturbance (d_0), and in this case two approaches to avoid infeasibility are (1) to use back-offs for the controlled variables during implementation, or (2) to add safety margins to the constraints during the optimization (Narraway *et al.* (1991); Glemmestad (1997)). Alternatively, one may solve the "robust optimization problem", where on also optimizes c_s for all the possible disturbances. A different approach is to track the active constraint. In particular, model predictive control is very well suited and much used for tracking active constraints.

A discussion of degrees of freedom for optimization is also given in section 5.1.

Method 2: Maximizing the minimum singular value

Let the matrix G represent the effect of a small change in the "base set" of independent variables (u; often the manipulated inputs) on the selected set of controlled variables (c), i.e.

$$\Delta c = G \cdot \Delta u$$

Then, a common criterion (rule) in control structure design is to select the set of outputs which maximizes the minimum singular value of the gain matrix, $\underline{\sigma}(G)$ (Yu and Luyben (1986) refer to this as the "Morari Resiliency Index") Previously, this rule has had little theoretical justification, and it has not been clear how to scale the variables. We will now derive the rule by considering a local approximation of the loss function.

It is desirable to select the controlled variables such that the loss is minimized. For a given disturbance d, a Taylor series expansion of the loss around the optimal value $u_{opt}(d)$ gives

$$\Delta L = J(u,d) - J(u_{opt},d) = \frac{1}{2}(u-u_{\text{opt}})^T \left(\frac{\partial^2 J}{\partial u^2}\right)_{\text{opt}} (u-u_{\text{opt}}) \tag{2}$$

(where we have assumed that the problem is unconstrained, so that the first-order term $\partial J/\partial u$ is zero.) Thus, the loss depends on the quantity $u - u_{opt}$ which we obviously want as small as possible. Now, for small deviations from the optimal operating point we have that the candidate output variables are related to the independent variables by $c - c_{opt} = G(u - u_{opt})$, or

$$u - u_{opt} = G^{-1}(c - c_{opt}) \tag{3}$$

Since we want $u - u_{opt}$ as small as possible, it therefore follows that we should select the set of controlled outputs c such that the product of G^{-1} and $c - c_{opt}$ is as small as possible. Thus, the correct statement of the rule is:

> *Assume we have scaled each output c such that the expected $c - c_{opt}$ is of magnitude 1 (including the effect of both disturbances and control error), then select the output variables c which minimize the norm of G^{-1}, which in terms of the two-norm is the same as maximizing the minimum singular value of G, $\underline{\sigma}(G)$.*

Interestingly, we note that this rule does not depend on the actual expression for the objective function J, but it does enter indirectly through the variation of c_{opt} with d, which enters into the scaling. Also note that in the multivariable case we should scale the inputs u such that the Hessian $\left(\frac{\partial^2 J}{\partial u^2}\right)$ is close to unitary; see Skogestad and Postlethwaite (1996) for details. Also note that use of the rule may be computationally much simpler than evaluating the mean value of J or the loss function.

Example

To give a simple "toy example", let $J = (u - d)^2$ where nominally $d_0 = 0$. For this problem we always have $J_{opt}(d) = 0$ corresponding to $u_{opt}(d) = d$. Let is now consider three alternative choices for the controlled output (e.g. we can assume they are three alternative measurements)

$$c_1 = 0.1(u - d); \quad c_2 = 20u; \quad c_3 = 10u - 5d$$

For the nominal case with $d_0 = 0$ we have in all three cases that $c_{opt}(d_0) = 0$ so we select in all three cases $c_s = 0$. However we note that the he optimal value for the three alternative controlled outputs as a function of the disturbance are (recall that $u_{opt}(d) = d$) $c_{1opt}(d) = 0$, $c_{2opt}(d) = 20d$ and $c_{3opt} = 5d$.

Method 1. The losses can in for this example be evaluated analytically, and we find for the three choices

$$L_1 = (10e_1)^2; \quad L_2 = (0.05e_2 - d)^2; \quad L_3 = (0.1e_3 - 0.5d)^2$$

(For example, in case 3, we have $u = (c_3 + 5d)/10$ and with $c_3 = c_{3s} + e_3 = e_3$ we get $J = (u - d)^2 = (0.1e_3 + 0.5d - d)^2$). If we further assume that the variables have been scaled such that $|d| \leq 1$ and $|e_i| \leq 1$ then the worst-case values of the losses are $L_1 = 100$, $L_2 = 1.05^2 = 1.1025$ and $L_3 = 0.6^2 = 0.36$, and we find that *output c_3 is the best overall choice for self-optimizing control*. However, with no control error c_1 is the best, and with no disturbances c_2 is the best.

Method 2. For the three choices of controlled outputs we have $G_1 = 0.1$, $G_2 = 20$ and $G_3 = 10$, and $\underline{\sigma}(G_1) = 0.1$, $\underline{\sigma}(G_2) = 20$ and $\underline{\sigma}(G_3) = 10$. This would indicate that c_2 is the best choice, but this is only correct with no disturbances. The reason for the error is that we have not scaled the output variables properly; in particular, we have not take into account the effect of the disturbances on the magnitude of $c - c_{opt}(d)$.

Let us now scale the variables properly. We have $u_{opt} = d$ so we have $c_{1,opt} = 0$, $c_{2,opt} = 20d$ and $c_{3,opt} = 5d$. For c_1 we then have that $|c_1 - c_{1,opt}| = 1 + 0$ (the control error is 1 plus the variation in $c_{1,opt}(d)$ due to disturbances is 0), and we find that

$$|G_1^{-1}(c_1 - c_{1,opt})| = \frac{1}{0.1} \cdot (1 + 0) = 10$$

Similarly,

$$|G_2^{-1}(c_2 - c_{2,opt})| = \frac{1}{20} \cdot (1 + 20) = 1.05$$

$$|G_3^{-1}(c_3 - c_{3,opt})| = \frac{1}{10} \cdot (1 + 5) = 0.6$$

and we find as expected that c_3 is the best choice. Thus, the two methods agree.

In general, method 1 is more accurate that method 2. The main limitation with method 2, is that for the multivariable case the particular value of $c - c_{opt}(d)$ corresponding the direction of the minimum singular value of G may not occur in practice, that is, there is no disturbance in this direction. Method 2 may therefore eliminate some viable control structures.

Other work on selection of controlled variables

As mentioned, the ideas above are based on Skogestad and Postlethwaite (1996) and unpubished works by Skogestad and coworkers. There seams to be very little other work in this area. With the possible exception of the work by Morari *et al.* (1980), it does not seem that anyone previously has addressed the problem of selection controlled variables to minimizing the sensitivity to uncertainty.

Morari *et al.* (1980) write that *in attempting to synthesize a feedback optimizing control structure, our main objective is to translate the economic objectives into process control objectives. In other words, we want to find a function c of the process variables which when held constant, leads automatically to the optimal adjustments of the manipulated variables, and with it, the optimal operating conditions. ... This means that by keeping the function $c(u, d)$ at the setpoint c_s, through the use of the manipulated variables u, for various disturbances d, it follows uniquely that the process is operating at the optimal steady-state* $J = J_{\mathrm{opt}}$. This is a precise description of the best self-optimizing control structure, except that they do not consider the effect of implementation error $e = c - c_s$. Unfortunately, it seems that very few people, including the authors themselves, have picked up on the idea.

As a minor remark we mention that Morari *et al.* (1980) claim that "ideally one tries to select c in such a way such that some or all the elements in c are independent of the disturbances d." This statement is generally *not* true, as illustrated above by our simple toy example. A more reasonable requirement, which holds when there is no contror error, is that the *optimal* values of some or all the elements in c are independent of the disturbances d (which is probably what Morari *et al.* (1980) had in mind).

Another related work which is that of Shinnar (1981). He defines the output variables Y_p as the set of process variables that define the product and process specifications, and the variables Y_d as the set of dynamically measured process variables. The goal is to maintain Y_p within prescribed limits. He writes that to achieve this goal "we choose in most cases a small set Y_{cd} [a subset of Y_d] and try to keep these at a fixed set of values by manipulating [the dynamic input variables] U_d". He writes that the overall control algorithm can normally be decomposed into a dynamic control system (which adjust U_d) and a steady-state control which determines the set points of Y_{cd} as well as the values of U_s [the latter are the manipulations which only can be changed slowly]. This a special case of the ideas outlined above if we interpret J as consisting of some weighted sum of the variables in Y_p, and we interpret the controlled variables c as the set of Y_{cd} and U_s. In the paper a case study of a fluidized catalytic cracker (FCC) is presented where the controlled variables are selected mainly based on process insight ("our main concern is to control the heat balance and the set of Y_{cd} is chosen accordingly"). In a more recent paper, by the same author (Arbel *et al.* 1996) some additional heuristics are presented for selecting controlled outputs; one is that it is necessary to obtain information about how the specifications Y_p are related to the setpoints Y_{cd} ("modelability"), and another is to select "dominant" variables as controlled variables Y_{cd}. Arbel *et al.* (1996) refer to the selection of controlled outputs c (Y_{cd} in their notation) as the *partial control* problem, because the system as seen from the optimizer with c_s (Y_{cd}^s in their notation) as independent variables and J as the objective (Y_p in their notation) is a partially controlled system. Actually, the concept of partial control can be used at various layers in the control hierarchy, and it is discussed in more detail below when we look at the control configuration.

The minimum singular value has previously been used as a tool for selecting control structures in some case studies, but with little theoretical justifaction. It has been previously shown (Morari (1983)) that the minimum singular value is related to input saturation, but this is not relevant for selecting controlled outputs since the various choices for controlled variables c do not differ in this respect (the optimal solution has a unique value of the manipulated inputs m). Yu and Luyben (1986) propose to use the minimum singular value to select between input sets. They claim that the minimum singular value "is a measure of the plant inherent ability to handle disturbances, model plant mismatches, change in operating conditions, etc.", but the claim seems to be based on experience or intuition since no further justification is given. A related idea is presented in Chang and Yu (1990) who propose to use the column sum for non-square plants for selecting controlled outputs. The idea is that the set of outputs with the largest row sum will lead to small steady-state sum of square error. Cao *et al.* (1998) also uses the minimum singular value to select between inputs. The derivation above provides, for the first time a theoretical justification for the use of the mimimum singular value for selecting controlled variables.

The idea that the selection of the controlled variables is somehow related to the steady-state performance index is not new, although it does not seem to have been stated clearly. Maarleveld and Rijnsdrop (1970) state that the steady-state optimum usually is constrained, and that we therefore we should control that variable. Arkun and Stephanopoulos (1980) reach the same conclusion and provides a good discussion on the advantages of active constraint control. In the book by Rijnsdorp (1991), he gives on page 99 a stepwise design procedure. One step is to "transfer the result into on-line algorithms for adjusting the degrees of freedom for optimization". But he states that "we have not yet come up with an automatic procedure for generating optimizing control systems for process units. Process insight remains of vital importance, and we cannot see any escape from that."

Luyben and coworkers (e.g. Luyben (1975), Yi and Luyben (1995), Luyben (1988)) have studied unconstrained problems, and some of the examples presented point in the direction of the selection methods presented in this paper. In particular, this applies to the distillation case study in Luyben (1975).

Fisher *et al.* (1988) discuss plant economics in relation to control. They provide some interesting heuristic ideas. In particular, hidden in their HDA example in part 3 (p. 614) one finds an interesting discussion on the selection of controlled variables, which is quite closely related to the approach presented as Method 1 above.

Finally, Narraway and Perkins ((Narraway *et al.* 1991), (Narraway and Perkins 1993) and (Narraway and Perkins 1994)) stress the need to base the selection of the control structure on economics. However, they consider the entire problem, including the selection of the control configuration and controller tunings, as one single optimization problem. They do not explicitly discuss the issue of selecting controlled variables, and it is only included as a integer parameter in the optimization.

Controllability issues

Of course, steady-state issues related to the cost J are not the only ones to be considered when selecting controlled outputs. It may happen that the "optimal" controlled outputs from a steady-state point of view, may result in a difficult control problem, so that dynamic control performance is poor. This may analyzed using an input-output controllability analysis. For example, in distillation column control it is well-known (Skogestad 1997) that controlling both product compositions may be difficult due to strong two-way interactions. In such cases, one may decide to control only one composition ("one-point control") and use, for example, constant reflux L/F (although this may not be optimal from a steady-state point of view). Alternatively, one may choose to over-purify the products to make the control problem easier (reducing the sensitivity to disturbances).

4.2 Selection of manipulated inputs (m)

By manipulated inputs we refer to the physical degrees of freedom, typically the valve positions or electric power inputs. Actually, selection of these variables is usually not much of an issue at the stage of control structure design, since these variables usually follow as direct consequence of the the design of the process itself.

However, there may be some possibility of adding valves or moving them. For example, if we install a bypass pipeline and a valve, then we may use the bypass flow as an extra degree of freedom for control purposes.

Finally, let us make it clear that the possibility of not actively using some manipulated inputs (or only moving them rarely), is a decision that is included above in “selection of controlled outputs”.

4.3 Selection of measurements (v)

Controllability considerations, including dynamic behavior, are important when selecting which variables to measure. There are often many possible measurements we can make, and the number, location and accuracy of the measurement is a tradeoff between cost of measurements and benefits of improved control. A controllability analysis is usually very important. In most cases the selection of measurements must be considered simultaneously with the selection of the control configuration. For example, this applies to the issue of stabilization and the use of secondary measurements.

4.4 Selection of control configuration

The issue of control configuration selection, including decentralized control, is discussed in Hovd and Skogestad (1993) and in sections 10.6, 10.7 and 10.8 of Skogestad and Postlethwaite (1996), and we will here discuss mainly issues which are not covered there.

The control configuration is the structure of the controller K that interconnects the variables v, c_s and m. The controller can be structured (decomposed) into blocks both in an vertical (hierarchical) and horizontal (decentralized control) manner.

Why is the controller decomposed? (1) The first reason is that it requires less computation. This reason may be relevant in some decision making systems where there is limited capacity for transmitting and handling information (like in most systems where humans are involved), but it does not hold in todays chemical plant where information is centralized and computing power is abundant. Two other reasons often given are (2) failure tolerance and (3) the ability of local units to act quickly to reject disturbances (e.g. Findeisen et al., 1980). These reasons may be more relevant, but, as pointed out by Skogestad and Hovd (1995) there are probably even more fundamental reasons. The most important one is probably (4) to reduce the cost involved in defining the control problem and setting up the detailed dynamic model which is required in a centralized system with no predetermined links. Also, (5) decomposed control systems are much less sensitive to model uncertainty (since they often use no explicit model). In other words, by imposing a certain control configuration, we are implicitly providing information about the behavior of the process, which we with a centralized controller would need to supply explicitly through the model.

4.4.1 Stabilizing control

Instability requires the active use of manipulated inputs (m) using feedback control. There exist relatively few systematic tools to assist in selecting a control structure for stabilizing control. Usually, single-loop controllers are used for stabilization, and issues are which variable to measure and which input to use. One problem in stabilization is that measurement noise may cause large variations in the input such that it saturates. Havre and Skogestad (1996, 1998) have shown that the *pole vectors* may be used to select measurements and manipulated inputs such that this problem is minimized.

4.4.2 Secondary measurements

Extra (secondary) measurements are often added to improve the control. These variables may be used as follows:

1. Centralized controller: All measurements are used and a the controller calculates the input. This controller has implicitly an estimator hidden inside it.
2. Inferential control: Based on the measurements, an estimate of the primary output (e.g. a controlled output c) is constructed. The estimate is send to a separate controller.

3. Cascade control: The secondary measurements are controlled and their setpoints are used as degrees of freedom at some higher layer in the hierarchy.

The subject of estimation and measurements selection for estimation is beyond the scope of this review article; we refer to Ljung (1987) for a control view and to Martens (1989) for a chemometrics approach to this issue. However, we would like point out that the control system should be designed for best possible control of the primary variables (c), and not the best possible estimate.

For cascade control Havre (1998) has shown how to select secondary measurements such that the need for updating the setpoints is small. The issues here are similar to that of selecting controlled variables (c) discussed above. One approach is to minimize some norm of the transfer function from the disturbance and control error in the secondary variable to the control error in the primary variable. A simpler, but less accurate, alternative is to maximize the minimum singular value in the transfer function from secondary measurements to the input used to control the secondary measurements. A similar problem is considered by Lee and Morari ((Lee and Morari 1991), (Lee *et al.* 1995) and (Lee *et al.* 1997)), but they use a more rigorous approach where model uncertainty is explicitly considered and the structured singular value is used as a tool.

4.4.3 Partial control

Most control configurations are structured in a hierarchical manner with fast inner loops, and slower outer loops that adjust the setpoints for the inner loops. Control system design generally starts by designing the inner (fast) loops, and then outer loops are closed in a sequential manner. Thus, the design of an "outer loop" is done on a *partially controlled system.* We here provide some simple but yet very useful relationships for partially controlled systems. We divide the outputs into two classes:

- y_1 – (temporarily) uncontrolled output
- y_2 – (locally) measured and controlled output (in the inner loop)

We have inserted the word *temporarily* above, since y_1 is normally a controlled output at some higher layer in the hierarchy. We also subdivide the available manipulated inputs in a similar manner:

- u_2 – inputs used for controlling y_2 (in the inner loop)
- u_1 – remaining inputs (which *may* be used for controlling y_1)

Skogestad and Postlethwaite (1996) distinguish between the following four cases of partial control:

		Meas./Control of y_1 ?	Control objective for y_2 ?
I	Indirect control	No	No
II	Sequential cascade control	Yes	No
III	"True" partial control	No	Yes
IV	Sequential decentralized control	Yes	Yes

In all cases there is a control objective associated with y_1 and a measurement of y_2. The first two cases are probably the most important as they are related to vertical (hierarchical) structuring. The latter two cases (where y_2 has its own control objective so that the setpoints y_{2s} are not adjustable) gives a horizontal structuring.

With these definitions the linear model for the plant can be written

$$y_1 = G_{11}(s)u_1 + G_{12}(s)u_2 + G_{d1}(s)d \tag{4}$$
$$y_2 = G_{21}(s)u_1 + G_{22}(s)u_2 + G_{d2}(s)d \tag{5}$$

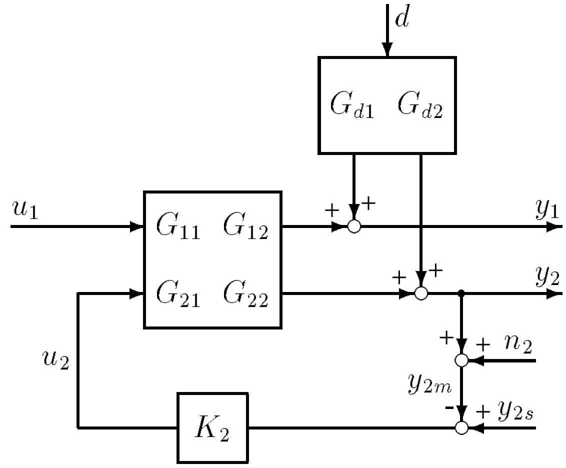

Figure 3: Block diagram of a partially controlled plant

A block diagram of the partially controlled system resulting from closing the loop involving u_2 and y_2 with the local controller $K_2$is shown in Figure 3.

To derive transfer functions for the partially controlled system we simply solve (5) with respect to u_2 (assuming that $G_{22}(s)$ is square an invertible at a given value of s)[2]

$$u_2 = G_{22}^{-1}(s)\,(y_2 - G_{21}(s)u_1 - G_{d2}(s)d) \tag{6}$$

Substituting (6) into (4) then yields

$$\boxed{y_1 = P_u(s)u_1 + P_d(s)d + P_y(s)y_2} \tag{7}$$

where

$$P_u(s) \stackrel{\text{def}}{=} G_{11}(s) - G_{12}G_{22}^{-1}G_{21}(s) \tag{8}$$
$$P_d(s) \stackrel{\text{def}}{=} G_{d1}(s) - G_{12}G_{22}^{-1}G_{d2}(s) \tag{9}$$
$$P_y(s) \stackrel{\text{def}}{=} G_{12}G_{22}^{-1}(s) \tag{10}$$

Here P_d is the *partial disturbance gain*, P_y is the gain from y_2 to y_1, and P_u is the partial input gain from the unused inputs u_1. If we look more carefully at (7) then we see that the matrix P_d gives the effect of disturbances on the primary outputs y_1, when the manipulated inputs u_2 are adjusted to keep y_2 constant, which is consistent of the original definition of the partial disturbance gain given by Skogestad and Wolff (1992). Note that no approximation about perfect control has been made when deriving (7). Equation (7) applies for any fixed value of s (on a frequency-by-frequency basis).

Sometimes it is useful to write

$$y_2 = y_{2s} + e_2$$

where e_2 is the overall control error. There are two independent contributions to the control error,

$$e_2 = e_{2m} + n_2$$

where $e_{2m} = y_2 - y_{2m}$ is the offset from the measurement caused by imperfect control, and n_2 is the measurement noise. The case of "perfect control" corresponds to achieving $y_{2m} = y_{2s}$ ($e_{2m} = 0$), so that the control error is equal to the measurement noise, $e_2 = n_2$.

Note that y_{2s} may be viewed as an independent variable ("input") for the "outer loop" in the hierarchy in cases I and II, whereas it may be viewed as a disturbance in cases III and IV. The control error e_2 may be viewed as a disturbance in all cases.

[2]The assumption that G_{22}^{-1} exists for all values of s can be relaxed by replacing the inverse with the pseudo-inverse.

The above equations are simple yet very useful. Relationships containing parts of these expressions have been derived by many authors, e.g. see the work of Manousiouthakis *et al.* (1986) on block relative gains and the work of Häggblom and Waller (1988) on distillation control configurations.

Note that this kind of analysis can be performed at each layers in the control system. At the top layer we may assume that the cost J is a function of the variables y_1, and we can then interpret y_2 as the set of controlled outputs c. If c is never adjusted then this is a special case of indirect control, and if c is adjusted at regular intervals (as is usually done) then this may be viewed as a special case of sequential cascade control.

5 The Process Oriented Approach

We here reviews procedures for plantwide control which are based on using process insight, that is, methods that are unique to process control.

The first comprehensive discussion on plantwide control was given by Page Buckley in his book "Techniques of process control" in a chapter on *Overall process control* (Buckley 1964). The chapter introduces the main issues, and presents what is still in many ways the industrial approach to plantwide control. In fact, when reading this chapter, 35 years later one is struck with the feeling that there has been relatively little development in this area. Some of the terms which are introduced and discussed in the chapter are material balance control (in direction of flow, and in direction opposite of flow), production rate control, buffer tanks as low-pass filters, indirect control, and predictive optimization. He also discusses recycle and the need to purge impurities, and he points out that you cannot at a given point in a plant control inventory (level, pressure) and flow independently since they are related through the material balance. In summary, he presents a number of useful engineering insights, but there is really no overall procedure. As pointed out by Ogunnaike (1995) the basic principles applied by the industry does not deviate far from Buckley (1964).

Wolff and Skogestad (1994) review previous work on plantwide control with emphasis on the process-oriented decomposition approaches. They suggest that plantwide control system design should start with a "top-down" selection of controlled and manipulated variables, and proceed with a "bottom-up" design of the control system. At the end of the paper ten heuristic guidelines for plantwide control are listed.

There exists other more or less heuristics rules for process control; e.g. see Hougen and Brockmeier (1969) and Seborg *et al.* (1995).

5.1 Degrees of freedom for control and optimization

A starting point for plantwide control is to establish the number of degrees of freedom for operation; both dynamically (for control, N_c) and at steady-state (for optimization, N_{ss}). Fortunately, it is in most cases relatively straightforward to establish these numbers from process insight, e.g. see (Ponton and Liang 1993) and (Luyben 1996). The basis is that the number of independent variables for control (N_c) equals the number of variables that can be manipulated by external means (N_m), i.e., $N_c = N_m$. In process control N_m equals the number of number of adjustable valves plus the number of other adjustable electrical and mechanical variables (electric power, etc.)

The number of degrees at freedom at steady-state (N_{ss}) is generally less than this. We have

$$N_{ss} = N_c - N_0 \tag{11}$$

where $N_0 = N_{m0} + N_{y0}$ is the number of variables with no steady-state effect (on the cost function). Here

N_{m0} is the number of manipulated inputs (u's), or combinations thereof, with no steady-state effect.

N_{y0} is the number of manipulated inputs that are used to control variables with no steady-state effect.

The latter usually equals the number of liquid levels with no steady-state effect, including most buffer tank levels. However, note that some liquid levels *do* have a steady-state effect, such as the level in a non-equilibrium liquid phase reactor, and levels associated with adjustable heat transfer areas. Also, we should *not* include in N_{y0} any liquid holdups that are left uncontrolled, such as internal stage holdups in distillation columns.

We find that N_{y0} is nonzero for most chemical processes, whereas we often have $N_{m0} = 0$. A simple example where N_{m0} is non-zero is a heat exchanger with bypass on both sides, (i.e. $N_c = N_m = 2$). However, at steady-state $N_{ss} = 1$ since there is really only one operational degree of freedom, namely the heat transfer rate Q (which at steady-state may be achieved by many combinations of the two bypasses), so we have $N_{m0} = 1$.

The optimization is generally subject to several constraints. First, there are generally upper and lower limits on all manipulated variables (e.g. fully open or closed valved). In addition, there are constraints on many dependent variables; due to safety (e.g. maximum pressure or temperature), equipment limitations (maximum throughput), or product specifications. Most of these constraints are in terms of inequalities. In some cases all constraints can not be met simultaneously, and the problem is infeasible. During operation, one will then need to relax one or more of the constraints. For cases with a feasible solution, one usually finds that the optimal solution has many "active" constraints (being satisfied as equalities.[3] The number of unconstrained variables "left for optimization" is then equal to

$$N_{ss,opt} = N_{ss} - N_{active}$$

where N_{active} is the number of active constraints. Note that the term "left for optimization" may be misleading, since the the decision to keep some constraints active, really follows as part of the optimization; thus all N_{ss} variables are really used for optimization. In some simple cases with $N_{ss,opt} = 0$ we can identify from physical insight which constraints are active, and no on-line optimization is needed. However, as illustrated by the distillation example below, it may not be clear if the constraints will be active at the optimal operating point. Also, even for cases with $N_{ss,opt} = 0$ it may be difficult to identify which constraints are active. Indeed, this is exactly the problem to be solved in linear programming (where the cost and the constraints are linear; and the optimal solution always is constrained, i.e. with $N_{ss,opt} = 0$).

Example: Degrees of freedom in distillation. Consider a conventional two-product distillation column with a given feed (this is the main disturbance). The column has five manipulable valves (flows); these are the reflux, distillate, bottom, boilup (heating) and cooling flows. Thus, there are five dynamic (control) degrees of freedom,

$$N_c = N_m = 5$$

To find the number of steady-state degrees of freedom, we subtract the variables with no steady-state effect. There are two such variables that need to be controlled for stabilization; namely the condenser and reboiler drum levels. This leaves three degrees of freedom for optimization, $N_{ss} = N_c - N_0 = 3$. The three degrees of freedom may be chosen as the pressure, the distillate composition and the bottom composition. The cost function to be optimized should be related to the behavior of the overall plant and will involve the value of the products from the column, the cost of energy, etc..

The pressure is often given (as an equality constraint). If pressure is free we often find that the optimal choice is to have maximum cooling corresponding to minimum pressure ("floating pressure control" as suggested by Shinskey (1984). The reason is that in most cases the relative volatility is improved when pressure is lowered.

There are often inequality constraints on the two product compositions (of the kind "maximum 1% impurities allowed"), and we then often find that the optimum is to keep them active at their specifications. There are then no degrees of freedom left for optimization, so $N_{ss,opt} = 0$, and since we know which constraints are active there is no need for on-line optimization.

[3] Note that here there are no model equations which must be satisfied as equality constraints. This is because we write the cost function as a function of the "true" independent variables, $J(u, d)$, that is, without any "internal state variables" which would otherwise need to be related to u and d through model equations.

However, if the values of the products are different and the cost of over-purification is low (typically, if the column has many stages), then it may happen that it may be optimal to over-purify the least valuable product, in order to get more of the valuable product, and we have $N_{ss,opt} = 1$. In another case, where both products become more valuable when they are purer then it may be optimal to over-purify both ends and $N_{ss,opt} = 2$. Both these cases are discussed by Gordon (1986).

Remark on design degrees of freedom. Above we have discussed operational degrees of freedom. The design degree of freedom (which is not really a concern of this paper) includes all the N_{ss} operational degrees of freedom plus all parameters related to the size of the equipment, such as the number of stages in column sections, area of heat exchangers, etc.

5.2 Production rate

Identifying the major disturbances is very important in any control problem, and for process control the production rate (throughput) is often the main disturbance. In addition, the location of where the production rate is actually set ("throughput manipulator"), usually determines the control structure for the inventory control of the various units. For a plant running at maximum capacity, the location where the production rate is set is usually somewhere inside the plant, (e.g. caused by maximum capacity of a heat exchanger or a compressor). Then, downstream of this location the plant has to process whatever comes in (given feed rate), and upstream of this location the plant has to produce the desired quantity (given product rate). To avoid any "long loops", it is preferably to use the input flow for inventory control upstream the location where the production rate is set, and to use the output flow for inventory control downstream this location.

From this it follows that it is critical to know where in the plant the production rate is set. In practice, the location may vary depending on operating conditions. This may require reconfiguring of many control loops, but often supervisory control systems, such as model predictive control, provide a simpler and better solution.

5.3 Decomposition of the problem

The task of designing a control system for complete plants is a large and difficult task. Therefore most methods will try to decompose the problem into manageable parts. Four common ways of decomposing the problem are

1. Decomposition based on process units
2. Decomposition based on process structure
3. Decomposition based on control objectives (material balance, energy balance, quality, etc.)
4. Decomposition based on time scale

The first is a horizontal (decentralized) decomposition whereas the three latter latter three provide hierarchical decompositions. Most practical approaches contain elements from several categories.

Many of the methods described below perform the optimization at the end of the procedure after checking if there degrees of freedom left. However, as discussed above, it should be possible to identify the steady-state degrees of freedom initially, and make a preliminary choice on controlled outputs (c's) before getting into the detailed design.

It is also interesting to see how the methods differ in terms of how important inventory (level) control is considered. Some regard inventory control as the most important (as is probably correct when viewed purely from a control point of view) whereas Ponton (1994) states that "inventory should normally be regarded as the least important of all variables to be regulated" (which is correct when viewed from a design point of view). We feel that there is a need to integrate the viewpoints of the control and design people.

5.3.1 The unit based approach

The unit-based approach, suggested by Umeda *et al.* (1978), proposes to

1. Decompose the plant into individual unit of operations
2. Generate the best control structure for each unit
3. Combine all these structures to form a complete one for the entire plant.
4. Eliminate conflicts among the individual control structures through mutual adjustments.

This approach has always been widely used in industry, and has its main advantage that many effective control schemes have been established over the years for individual units (e.g. Shinskey (1988)). However, with an increasing use of material recycle, heat integration and the desire to reduce buffer volumes between units, this approach may result in too many conflicts and become impractical.

As a result, one has to shift to plant-wide methods, where a hierarchical decomposition is used. The first such approach was Buckley's (1964) division of the control system into material balance control and product quality control, and three plantwide approaches partly based on his ideas are described in the following.

5.3.2 Hierarchical decomposition based on process structure

The hierarchy given in Douglas (1988) for process design starts at a crude representation and gets more detailed:

Level 1 Batch vs continuous

Level 2 Input-output structure

Level 3 Recycle structure

Level 4 General structure of separation system

Level 5 Energy interaction

Fisher *et al.* (1988) propose to use this hierarchy when performing controllability analysis, and Ponton and Liang (1993) point out that this hierarchy, (e.g. level 2 to level 5) could also be used for control system design. This framework enables parallel development for the process and the control system. Within each of the levels above any design method might be applied.

Douglas (1988) present a different hierarchy for control system design. In this hierarchy the view point is not one the flowsheet but on steady-state, normal dynamic response and abnormal dynamic operation.

Ng and Stephanopoulos (1998*b*) propose to use a similar hierarchy for control structure design. The difference between Douglas (1988) and Ng and Stephanopoulos (1998*b*)'s hierarchy is that level 1 is replaced by a preliminary analysis and level 4 and on is replaced by more and more detailed structures. At each step the objectives identified at an earlier step is translated to this level and new objectives are identified. The focus is on construction of mass and energy balance control. The method is applied to the Tennessee Eastman case.

All these methods have in common that at each level a key point is to check if there are enough manipulative variables to meet the constraints and to optimize operation. The methods are easy to follow and gives a good process understanding, and the concept of a hierarchical view is possible to combine with almost any design method.

5.3.3 Hierarchical decomposition based on control objectives

The hierarchy based on control objectives is sometimes called the tiered procedure. This bottom-up procedure focuses on the tasks that the controller has to perform. Normally one starts by stabilizing the plant, which mainly involves placing inventory (mass and energy) controllers.

Price *et al.* (1993) build on the ideas that was introduced by Buckley (1964) and they introduce a tiered framework. The framework is divided into four different tasks:

I Inventory and production rate control.

II Product specification control

III Equipment & operating constraints

IV Economic performance enhancement.

Their paper does not discuss points III or IV. They perform a large number (318) of simulations with different control structures, controllers (P or PI), and tunings on a simple process consisting of a reactor, separator and recycle of unreacted reactant. The configurations are ranked based on integrated absolute error of the product composition for steps in the disturbance. From these simulation they propose some guidelines for selecting the through-put manipulator and inventory controls. (1) Prefer internal flows as through-put manipulator. (2) the through-put manipulator and inventory controls should be self-consistent (self-consistency is fulfilled when a change in the through-put propagates through the process by "itself" and does not depend on composition controllers). They apply their ideas on the Tennessee Eastman problem (Price *et al.* 1994).

Ricker (1996) comments upon the work of Price *et al.* (1994). Ricker points out that plants are often run at full capacity, corresponding to constraints in one or several variables. If a manipulated variable that is used for level control saturates, one looses a degree of freedom for maximum production. This should be considered when choosing a through-put manipulator.

Luyben *et al.* (1997) point out three limitations of the approach of Buckley. First, he did not explicitly discuss energy management. Second, he did not look at recycle. Third, he placed emphasis on inventory control before quality control. Their plantwide control design procedure is listed below:

1. Establish control objectives.
2. Determine the control degrees of freedom by counting the number of independent valves.
3. Establish energy inventory control, for removing the heats of reactions and to prevent propagation of thermal disturbances.
4. Set production rate. The production rate can only be increased by increasing the reaction rate in the reactor. One recommendation is to use the input to the separation section.
5. Product quality and safety control. Here they recommend the usual "control close"-rule.
6. Inventory control. Fix a flow in all liquid recycle loops. They state that all liquid levels and gas pressures should be controlled.
7. Check component balances. (After this point it might bee necessary to go back to item 4.)
8. Unit operations control.
9. Optimize economics or improve dynamic controllability.

Step 3 comes before determining the throughput manipulator, since the reactor is typically the heart of the process and the methods for heat removal are intrinsically part of the reactor design. In order to avoid recycling of disturbances they suggest to set a flowrate in all recycles loops; this is discussed more in section 6. They suggest in step 6 to control all inventories, but this may not be necessary in all cases; e.g. it may be optimal to let the pressure float (Shinskey 1988). They apply their procedure on several test problems; the vinyl acetate monomer process, the Tennessee Eastman process, and the HDA process.

5.3.4 Hierarchical decomposition based on time scales

Buckley (1964) proposed to design the quality control system as high-pass filters for disturbances and to design the mass balance control system will as low pass filters. If the resonance frequency of the quality control system is designed to be an order of magnitude higher than the break frequency of the mass balance system then the two loops will be non-interacting.

McAvoy and Ye (1994) divide their method into four stages:

1. Design of inner cascade loops.
2. Design of basic decentralized loops, except those associated with quality and production rate.
3. Production rate and quality controls.
4. Higher layer controls.

The decomposition in stages 1-3 is based on the speed of the loops. In stage 1 the idea is to locally reduce the effect of disturbances. In stage 2 there generally are a large number of alternatives configurations. These may be screened using simple controllability tools, such as the RGA. One problem of selecting outputs based on a controllability analysis is that one may end up with the outputs that are easy to control, rather than the ones that are important to control. The method is applied to the Tennessee Eastman test problem.

6 The reactor, separator and recycle plant

A common feature of most plants is the presence of recycle. A simple example is distillation, with recycle ("reflux") of liquid from the top of the column and of vapor from the bottom of the column.

In this section, we consider the reactor and separator process with recycle of unreacted feed from a reactor. This kind of problem has lately been studied by many authors, (Papadourakis *et al.* 1987), (Wolff *et al.* 1992), (Price *et al.* 1993), (Luyben 1994), (Luyben and Floudas 1994), (Mizsey and Kalmar 1996), (Wu and Yu 1996), (Hansen 1998), (Ng and Stephanopoulos 1998*a*) and many more. It may be difficult to follow all the details in the case studies presented, so instead we aim in this section to gain some basic insight into the problem.

In the simplest case, let the reactor be a CSTR where component A is converted to a product and the amount converted is given by

$$P = k(T) z_A M \; [mol A/s]$$

The amount of unreacted A is separated from the product and recycled back to the reactor. To increase the conversion P one then has three options:

1. Increase the temperature which increases the reaction constant k
2. Increase the amount of recycle, which indirectly increases the fraction of A in the reactor, z_A.
3. Increase the reactor holdup M.

In a liquid phase system the reactor holdup is determined by the reactor level, and in a gas phase system by the reactor pressure. Here we will assume that the temperature is constant, so there are two options left.

Since at steady-state with given product specifications the conversion of A in the reactor is given by the feed rate, it follows that only one of the two remaining options can be controlled independently (or more generally, one variables that influences these options), and we must let the second variable "float" and adjusts itself.

Two common control strategies are then

(A) Keep the reactor holdup constant (and let the recycle flow float)

(B) Keep the recycle flow constant (and let the reactor holdup float).

In case (A) one may encounter the so-called "snowball effect" where the recycle goes to infinity. This occurs because at infinite recycle flow we have $z_A = 1$ which gives the highest possible production. In effect, the snowball effect occurs because the reactor is too small to handle the given feed rate, so it is really a steady-state design problem.

Luyben (1992, 1994), has studied liquid phase systems and has concluded that a variant of control strategy (B) (where the reactor level is allowed to vary) with one flow fixed in the recycle loop should be used to avoid the ”snowball effect”.

Wu and Yu (1996) also studies the snowball effect for the reactor/separator and recycle plant. They propose as a remedy the snowball effect to distribute the “work” evenly between the different units. To achieve this they suggest to (C) keep the reactor composition constant. Also in this case the reactor volume varies depending on the disturbance.

However, from an economic point of view one should usually for liquid phase systems keep the reactor level at its maximum value. This maximizes the conversion per pass and results in the smallest possible recycle, which generally will reduce the operational cost. Thus, the so-called Luyben rule, to fix one flow in the recycle loop, has an economic penalty which it seems that most researchers so far have neglected.

On the other hand, for gas phase systems, there is usually an economic penalty from compression costs involved in increasing the reactor holdup (i.e. the reactor pressure), and strategy (B) where we let the holdup (pressure) float may in fact be optimal. Indeed, such schemes are used in industry, e.g. in ammonia plants. For example, for processes with gas recycle and purge, Fisher *et al.* (1988) recommends to keep the gas recycle constant at its maximum value.

Wolff *et al.* (1992) have studied a similar plant. They include a inert and look on the effects of recycle on the controllability of the process. Their conclusion is that the purge stream flow should be used to control the composition of inerts. They did not consider the reactor holdup as a possible controlled variable.

All the above works have in common that the authors are searching for the right controlled variables to keep constant (recycle flow, reactor volume, composition, etc.). However, a common basis for comparing the alternatives seems to be lacking. In terms of future work, we propose that one first needs to define clearly the objective function (cost) J for the operation of the reactor system. Only when this is given, may one decide in a rigorous manner on the best selection of controlled outputs.

7 Tennessee Eastman Problem

7.1 Introduction to the test problem

The problem of Downs and Vogel (1993) was first proposed the problem at an AIChE meeting in 1990 and has since been studied by many authors. The process has four feed streams, one product streams, and one purge stream to remove inerts. The reactions are

$$\begin{aligned} \text{A(g) + C(g) + D(g)} &\rightarrow \text{G(liq)}, && \text{Product 1,} \\ \text{A(g) + C(g) + E(g)} &\rightarrow \text{H(liq)}, && \text{Product 2,} \\ \text{A(g) + E(g)} &\rightarrow \text{F(liq)}, && \text{Byproduct,} \\ \text{3D(g)} &\rightarrow \text{2F(liq)}, && \text{Byproduct,} \end{aligned}$$

All reactions are irreversible, exothermic and temperature dependent via the Arrhenius expression. The process has five major units; a reactor, a product condenser, a vapour-liquid separator, a recycle compressor and a product stripper; see Figure 4. There are 41 measurements and 12 manipulated variables. For a more detailed description see (Downs and Vogel 1993) [4]. We will here mainly look at the approaches used to solve the problem, not at the solutions themselves.

7.2 McAvoy and Ye solution

In stage 1, they close inner cascade loops involving eight flows and two temperature. This reduces the effect of the disturbances associated with these loops. At stage 3 they use a simple mass balance of the

[4] See also http://weber.u.washington.edu/ control/LARRY/TE/download.html

Figure 4: Tennessee Eastman process flowsheet

plant. This gives some constraints for stage 2, for example, that either the C-feed or the product flow must be left for the third stage.

In stage 2 decentralized loops are closed. They start with the level loops since they are the most important loops. There are three level loops; reactor, separator and stripper, and they they consider four possible level configurations. Three of the configurations were ruled out based on controllability analysis. The alternative where the E-feed is used for reactor level control is analyzed in greater detail. They look at three 6x6, eighteen 5x5, and fifteen 4x4 systems, where the outputs seem to be rather randomly chosen. After an analysis involving RGA, Niederlinski index and linear valve saturation, only four alternatives are left. These are further screened on their steady-state behavior for a range of disturbances.

7.3 Lyman, Georgakis and Price's solution

Georgakis and coworkers have studied the problem in several papers (Lyman and Georgakis 1995), (Price *et al.* 1994). They start by identifying the primary path, which is from the raw materials, through the reactor, condenser, the stripper, and to the product flow. They do not consider the C-feed since it is in excess in the recycle. In (Price *et al.* 1994) they lists all candidates for through-put manipulations along the primary path: The feed streams, flow of coolant to reactor condenser, the separator drum bottoms flows and final product flow. Of the feeds only D is considered. As noted by the authors one possible through-put manipulator is missing, the C-feed since it was assumed not to be on the primary path. Next, they list the inventories that need to be controlled; pressure, reactor level, separator level and stripper level. Inventory controls are chosen so to construct a self-consistent path (which does not depend on quality controllers). At this point they have four different structures. After this reactor temperature controller and quality controllers are added.

Their procedure is simple and clear to follow. The result is a control system that is fairly simple to understand.

7.4 Ricker's solution

Ricker (1996) starts by listing the variables that must be controlled: Production rate, Mole % G in product, reactor pressure, reactor liquid level, separator level and stripper level. The production rate is chosen as the input that most likely is going to saturate; namely a combination of D and E. The reaming control system follows by applying quality control and inventory control. After that overrides are installed.

7.5 Luyben's solution

Luyben *et al.* (1997) look on two cases, control of through-put with the product flow and control with the A-feed. Here we only look at the case where they set the product flow as through-put manipulator. At step 3 they look at energy inventory control, which in this case is to control the reactor temperature with the reactor cooling water. In step 5 they assign the stripper steam stream to control stripper temperature, and therefore also the product compositions. Since the pressure of the reactor must be kept within bounds, they use the largest gas feed (the feed of C) to control the reactor pressure. Step is 7 the check of component balances, which gives a composition controller for inerts using the purge flow and a composition controller for A using the A-feed. After doing some simulations they add a controller for control of the condenser, using the reactor temperature.

The resulting control system is simple, but there could have been a better justification on what outputs to control.

7.6 Ng and Stephanopulos's solution

Ng and Stephanopoulos (1997, 1998) starts by stabilizing the reactor. Then they proceed to look at the input/output level of the plant, where the central point is to fulfill material and energy balances. At this level it should have been possible to say something about how the feeds should be adjusted in order to achieve the right mix of G and H, but they do not. Rather they look at which feed or exit flows that should be used maintain material balance control.

At the final level they translates the control objectives to measureaments. Here material balance control is translated into inventory controllers, like product flow to control stripper level and bottom flow to control separator level. The next objective is then reactor pressure where purge is assigned. Finally feed E is assigned to control of product ratio, and E is assigned to through-put control. The A and C feed is used for controlling composition of A and C.

The method is somewhat difficult to follow and they seem to repeat many of the arguments in each phase.

7.7 Other work

The above review is by no means complete, and there are many authors who have worked on this problem, some are (Wu and Yu 1997), (Banerjee and Arkun 1995) and (Scali and Cortonesi 1995). In addition there are several others that has looked at other aspects of the Tennessee Eastman plant.

7.8 Other test problems

There are several other test problem, in addition to the the Tennessee Eastman problem, that are suitable for studying issues related to plantwide control. These include the HDA-plant (Douglas 1988), the vinyl acetate monomer process (Luyben and Tyreus 1997), the reactor/separator and recycle plant (Wu and Yu 1996), (Price *et al.* 1993) and the Luyben& Luyben plant (Luyben and Luyben 1995).

8 Conclusion

In this paper we have given a review on plantwide control with emphasis on the following tasks that make up the control structure design problem:

1. *Selection of controlled outputs* (c with setpoints c_s).

2. *Selection of manipulated inputs* (m).
3. *Selection of measurements* (v)
4. *Selection of control configuration*
5. *Selection of controller type*

The main emphasis has been on the the selection of controlled outputs, where we have seen that the use of a steady-state economic criteria is very useful. It appears that the solution to this task provides the much needed link between steady-state optimization and process control, and that the idea of "self-optimizing control" to reduce the effect of uncertainty provides a very useful framework for making the right decision. We thus propose that the design of the control system should start with the optimization (or at least identifying what the control objectives really are) and thus providing candidate sets for the controlled outputs. The control problem is then defined, and one may proceed to analyze (e.g. using an input-output controllability analysis, whether the control objectives can be met).

The actual design of the control system is done at the the end, after the control problem has been defined, including the classification of all variables (into inputs, disturbances, controlled variables, etc.). Control system design usually starts with stabilizing control where it is usually important to avoid input saturation. The control system is then build up in a hierarchical manner such that each controller is of limited size (usually with as few inputs and outputs as possible). Emphasis should be on avoiding "long" loops, that is, one should pair inputs and outputs with are "close" to another.

Most of the proposed process oriented procedures have elements from this way of thinking, although some procedures focus mostly on control and operation and seem to skip lightly over the phase where the overall control problem is defined, including.

Several case studies have been proposed, which is in itself very good. However, some of the work on these case studies seem to provide little general insight, and their value may therefore be questioned. A more systematic approach and a common ground of comparison is suggested for future work.

In summary, the field of plantwide control is still at a relatively early stage of its development. However, the progress over the last few years, both in terms of case studies and theoretical work, bears promise for the futher. There is still a need for a clearer definition of the issues, and it is hoped that this paper may be useful in this respect. In the longer term, where automatic generation and analysis of alternative structures may become more routine, the main problem will probably be to be able to generate models in an efficient way, and to provide efficient means for their analysis (e.g. using input-output controllability analysis).

References

Arbel, A., I.H. Rinard and R. Shinnar (1996). 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.* pp. 2215–2233.

Arkun, Y. and G. Stephanopoulos (1980). Studies in the synthesis of control structures for chemical processes: Part iv. design of steady-state optimizing control structures for chemical process units. *AIChE Journal* **26**(6), 975–991.

Balchen, J.G. and K.I. Mummé (1988). *Process Control, Structures and applications*. Van Nostrand Reinhold.

Banerjee, A. and Y. Arkun (1995). Control configuration design applied to the tennessee eastman plantwide control problem. *Computers. chem. Engng.* **19**(4), 453–480.

Buckley, P.S. (1964). *Techniques of Process control*. Chap. 13. John Wiley & Sons.

Cao, Y., D. Rossiter and D.H. Owens (1998). Globally optimal control structure selection using the branch and bound method. In: *Dycopds-5*.

Chang, J.W. and C.C. Yu (1990). The relative gain for non-square multivariable systems. *Chem. Eng. Sci.* pp. 1309–1323.

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

Downs, J.J. and E.F. Vogel (1993). A plant-wide industrial process control problem. *Computers chem. Engng.* pp. 245–255.

Findeisen, W, F.N. Bailey, M. Brdys, K. Malinowski, P. Tatjewski and A. Wozniak (1980). *Control and coordination in Hierarchical Systems*. John Wiley & sons.

Fisher, W.R., M.F. Doherty and J.M. Douglas (1988). The interface between design and control. 1, 2 and 3.; 1: Process controllability, 2: Process operability 3: Selecting a set of controlled variables.. *Ind. Eng. Chem. Res.* **27**(4), 597–615.

Foss, C.S. (1973). Critique of chemical process control theory. *AIChE Journal* **19**(2), 209–214.

Glemmestad, B. (1997). Optimal operation of integrated processes, studies on heat recovery systems. PhD thesis. Norwegian University of Science and Technology.

Gordon, M.L. (1986). Simple optimization for dual composition control. *Hydrocarbon Process* pp. 59–62.

Häggblom, K.E. and K.V. Waller (1988). Transformations and consistency relations of distillation control structures. *AIChe J.* pp. 1634–1648.

Hansen, J.E. (1998). Plant wide dynamic simulation and control of chemical processes. PhD thesis. Danmarks Tekniske Universitet.

Havre, K. (1998). Studies on controllability analysis and control structure design. PhD thesis. NTNU Trondheim. Aviable from http://www.chembio.ntnu.no/users/skoge/.

Havre, K. and S. Skogestad (1996). Selection of variables for regulatory control using pole directions. *1996 AIChE Annual Meeting, Chicago, Illinois; Paper 45f.*

Havre, K. and S. Skogestad (1998). Selection of variables for regulatory control using pole vectors. *DYCOPS-5, 5th IFAC Symposium on Dynamics and Control of Process Systems, Corfu, Greece, June 8-10, 1998* pp. 614–619. See C98-1.

Hougen, J. O. and N. F. Brockmeier (1969). Developing Process Control Strategies - I: Eleven Basic Principles. *Instrumentation Technology* pp. 45–49.

Hovd, M. and S. Skogestad (1993). Procedure for Regulatory Control Structure Selection with Application to the FCC Process. *AIChE Journal* pp. 1938–1953.

Lee, J.H. and M. Morari (1991). Robust measurements selection. *Automatica* pp. 519–527.

Lee, J.H., P. Kesavan and M. Morari (1997). Control structure selection and robust control system design for a high-purity distillation column. *IEEE Transactions on control systems technology* pp. 402–416.

Lee, J.H., R.D. Braatz, M. Morari and A. Packard (1995). Screening tools for robust control structure selection. *Automatica* **31**(2), 229–235.

Ljung, L. (1987). *System Identification - Theory for the User*. Prentice-Hall.

Loe, I. (1994). System-strukturer i prosessindustrien. Lecture notes, for a course given at Telemark College. In Norwegian.

Luyben, M.L. and B.D. Tyreus (1997). An industrial design/control study for the vinyl acetate monomer process. *Submitted to Computers chem. Engng.*

Luyben, M.L. and C.A. Floudas (1994). Analyzing the interaction of design and control. 2. reactor separator recycle system. *Computers & Chemical Engineering* pp. 971–994.

Luyben, M.L. and W.L. Luyben (1995). Design and control of a complex process involving two reaction steps, three distillation columns, and two recycle streams. *Ind. Eng. Chem. Res.* **34**(11), 3885–3898.

Luyben, M.L., B.D. Tyreus and W.L. Luyben (1997). Plantwide control design procedure. *AICHE journal* pp. 3161–3174.

Luyben, W.L. (1975). Steady-state energy conservation aspects of distillation column control system design. *Ind. Eng. Chem. Fundam.* pp. 321–325.

Luyben, W.L. (1988). The concept of eigenstructure in process control. *Ind. Eng. Chem. Res.* pp. 206–208.

Luyben, W.L. (1992). Design and control of recycle processes in ternary systems with consecutive reactions. In: *Interactions between process design and process control*. IFAC Workshop. Pergamon Press. pp. 65–74.

Luyben, W.L. (1994). Snowball effect in reactor/separator processes with recycle. *Ind. Eng. Chem. Res.* pp. 299–305.

Luyben, W.L. (1996). Design and control degrees of freedom. *Ind. Eng. Chem. Res.* pp. 2204–2214.

Lyman, P.R. and C. Georgakis (1995). Plant-wide control of the Tennessee Eastman Problem. *Computers chem. Engng.* pp. 321–331.

Maarleveld, A. and J.E. Rijnsdrop (1970). Constraint control of distillation columns. *Automatica* pp. 51–58.

Manousiouthakis, V., R. Savage and Y. Arkun (1986). Synthesis of decentralized process control structures. *AIChE Journal* pp. 991–1003.

Martens, H. (1989). *Multivariate calibration*. Wiley.

McAvoy, T.J. and N. Ye (1994). Base control fro the tennessee eastman problem. *Computers chem. Engng.* pp. 383–413.

Meadowcroft, T.A., G. Stephanopoulos and C. Brosilow (1992). The modular multivariable controller: I: Steady-state properties. *AIChE Journal* pp. 1254–1278.

Mizsey, P. and I. Kalmar (1996). Effects of recycle on control of chemical processes. *ESCAPE-6, 26-29 May 1996, Rhodes, Greece; Supplement to Computers & Chemical Engineering* pp. S883–S888.

Morari, M. (1982). Integrated plant control: A solution at hand or a research topic for the next decade?. In: *CPC-II*. pp. 467–495.

Morari, M. (1983). Design of Resilient Processing Plants -III. *Chemical Engineering Science* pp. 1881–1891.

Morari, M., G. Stephanopoulos and Y. Arkun (1980). Studies in the synthesis of control structures for chemical processes. Part I: Formulation of the problem. Process decomposition and the classification of the control task. Analysis of the optimizing control structures.. *AIChE Journal* pp. 220–232.

Morud, J. (1995). Studies on the dynamics and operation of integrated plants. PhD thesis. University of Trondheim.

Narraway, L. and J. Perkins (1994). Selection of process control structures based in economics. *Computers chem. Engng* pp. S511–S515.

Narraway, L.T. and J.D. Perkins (1993). Selection of process control structure based on linear dynamic economics. *Ind. Eng. Chem. Res.* pp. 2681–2692.

Narraway, L.T., J.D. Perkins and G.W. Barton (1991). Interaction between process design and process control: economic analysis of process dynamics. *J. Proc. Cont.* pp. 243–250.

Ng, C. and G. Stephanopoulos (1998*a*). Plant-wide control structures and strategies. To be published in Process System Engineering Series of Academic press.

Ng, C. and G. Stephanopoulos (1998*b*). Plant-Wide control structures and strategies. In: *Preprints Dycops-5*. IFAC. pp. 1–16.

Ogunnaike, B.A. (1995). A contemporary industrial perspective on process control theory and practice. *Dycord+ '95, 4th IFAC Symposium on Dynamics and Control of Chemical Reactors, Distillation Columns, and Batch Processes, Preprints, 7-9 June 1995* pp. 1–8.

Papadourakis, A., M.F. Doherty and J.M. Douglas (1987). Relative gain array for units in plants with recycle. *Ind. Eng. Chem. Res.* pp. 1259–1262.

Ponton, J.W. (1994). Degrees of freedom analysis in process control. *Chemical Engineering Science*.

Ponton, J.W. and D.M. Liang (1993). Hierarchical approach to the design of process control systems. *Chemical engineering Research and Design* pp. 181–188.

Price, R.M., P.R. Lyman and C. Georgakis (1993). Selection of throughput manipulators for plant-wide control structures. *ECC '93* pp. 1060–1066.

Price, R.M., P.R. Lyman and C. Georgakis (1994). Throughput manipulation in plantwide control structures. *Ind. Eng. Chem. Res.* pp. 1197–1207.

Ricker, N.L. (1996). Decentralized control of the Tennessee Eastman Challenge Process. *J. Proc. Cont.* pp. 205–221.

Rijnsdorp, J.E. (1991). *Integrated Process Control and automation*. Elsevier.

Rinard, I.H. and J.J. Downs (1992). Plant wide control: A review and critique. *AIChE Spring Meeting 1992, New Orleans, paper 67f*.

Scali, C. and C. Cortonesi (1995). Control of the tennessee eastman benchmark: Performance versus integrity tradeoff. *Proc. of 3rd European Control Conference, Rome, Italy, 1995* pp. 3913–3918.

Seborg, D.E., T.F. Edgar and D.A. Mellichamp (1995). *Process dynamics and control*. Wiley.

Shinnar, R. (1981). Chemical reactor modelling for purposes of controller design. *Cheng. Eng. Commun.* pp. 73–99.

Shinskey, F.G. (1984). *Distillation Control*. 2 ed.. McGraw-Hill Book Company.

Shinskey, F.G. (1988). *Process Control Systems*. McGraw-Hill.

Skogestad, S. (1997). Dynamics and control of distillation columns - a tutorial. In: *Preprints Distillation and absorption '97*. IChemE. pp. 1–25.

Skogestad, S. and E.A. Wolff (1992). Controllability measures for disturbance rejection. *IFAC Workshop, London, Sept. 7-8 1992* pp. 23–30. Later printed in *Modeling, Indentification and Control, 1996, pp 167-182* .

Skogestad, S. and I. Postlethwaite (1996). *Multivariable Feedback Control*. John Wiley & Sons.

Skogestad, S. and M. Hovd (1995). Letter to the editor on the decentralized versus multivariable control. *J. Proc. Cont.* pp. 499–400.

Stephanopoulos, G. (1982). Synthesis of control systems for chemical plants - a challenge for creativity. *Computers & Chemical Engineering* pp. 331–365.

Umeda, T., T. Kuriyama and A. Ichikawa (1978). A logical structure for process control system synthesis. *Proc. IFAC Congress (Helsinki) 1978*.

van de Wal, M. and B. de Jager (1995). Control structure design: A survey. In: *Proceedings of the American control conference*. pp. 225–229.

Wolff, E. and S. Skogestad (1994). Operability of integrated plants. *PSE '94, Korea 30 May-3 June 1994* pp. 63–69.

Wolff, E.A., S. Skogestad and M. Hovd (1992). Controllability of integrated plants. AIChE Spring National Meeting Paper 67a.

Wu, K.L. and C.-C. Yu (1996). Reactor/separator process with recycle-1. candidate control structure for operability. *Computers. chem. Engng.* pp. 1291–1316.

Wu, K.L. and C.-C. Yu (1997). Operability for processes with recycles: Interaction between design and operation with application to the tennessee eastman challenge process. *Ind. Eng. Chem. Res.* **36**(6), 2239–2251.

Yi, C.K. and W.L. Luyben (1995). Evaluation of plant-wide control structures by steady-state disturbance sensitivity analysis. *Ind. Eng. Chem. Res.* pp. 2393–2405.

Yu, C.C and W.L. Luyben (1986). Design of multiloop siso controllers in multivariable processes. *Ind. Eng. Chem. Process Des. Dev.* pp. 498–503.