Control structure design: What should we control, measure and manipulate? [1]

Sigurd Skogestad

Norwegian University of Science and Technology (NTNU)
Department of Chemical Engineering, 7491 Trondheim, Norway

Abstract

Control structure design deals with the structural decisions of the control system, including what to control and how to pair the variables to form control loops. Although these are very important issues, these decisions are in most cases made in an ad-hoc fashion, based on experience and engineering insight, without considering the details of each problem. In the paper, a systematic procedure for control structure design for complete chemical plants (plantwide control) is presented. It starts with carefully defining the operational and economic objectives, and the degrees of freedom available to fulfill them. Other issues, discussed in the paper, include inventory and production rate control, decentralized versus multivariable control, loss in performance by bottom-up design, and a definition of a the "complexity number" for the control system.

1. Introduction

A chemical plant may have thousands of measurements and control loops. In practice, the control system is usually divided into several layers, separated by time scale, including (see Figure 1)

- scheduling (weeks)
- site-wide optimization (day)
- local optimization (hour)
- supervisory (predictive, advanced) control (minutes)
- regulatory control (seconds)

We here consider the lower three layers.
The local optimization layer typically recomputes new setpoints only once an hour or so, whereas the feedback layers operate continuously. The layers are linked by the controlled variables, whereby the setpoints are computed by the upper layer and implemented by the lower layer. An important issue is the selection of these variables.
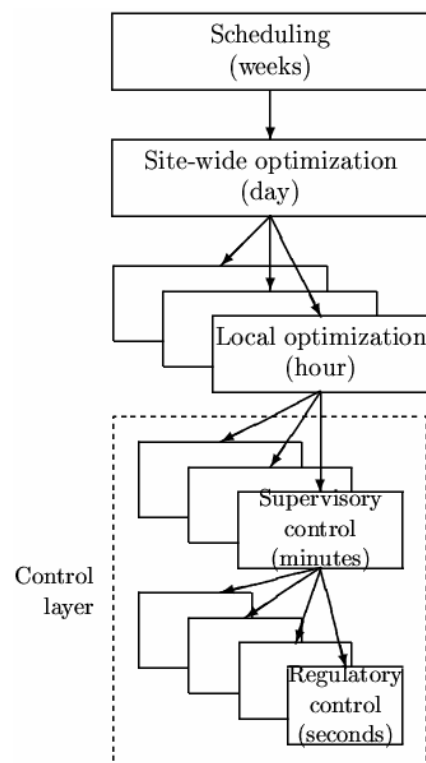

Figure 1: *Typical control hierarchy in a chemical plant.*

Control structure design deals with the structural decisions that must be made *before* we start the controller design, and involves the following *tasks* (Foss, 1973); (Skogestad and Postlethwaite, 1996):

1. *Selection of manipulated variables m* (``inputs")
2. *Selection of controlled variables* (``outputs"; variables with setpoints)
3. *Selection of (extra) measurements* (for control purposes including stabilization)
4. *Selection of control **configuration*** (the structure of the overall controller that interconnects the controlled, manipulated and measured variables)
5. *Selection of controller type* (control law specification, e.g., PID, decoupler, LQG, etc.).

Control structure design for complete chemical plants is also known as *plantwide control*. In practice, the problem is usually solved without the use of existing theoretical tools. In fact, the industrial approach to plantwide control is still very much along the lines described by Page Buckley in 1964 in his chapter on *Overall process control..* The realization that the field of control structure design is underdeveloped is not new. Alan Foss (1973) made the observation that in many areas application was *ahead* of theory, and 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? There is more than a suspicion that the work of a genius is needed here, for without it the control configuration problem will likely remain in a primitive, hazily stated and wholly unmanageable form. The gap is present indeed, but contrary to the views of many, it is the theoretician who must close it."*

A recent review of the literature on plantwide control can be found in Larsson and Skogestad (2000). In addition to Page Buckley and Alan Foss, important contributors in this area include George Stephanopoulos and Manfred Morari (1980-) (synthesis of control structures), William "Bill" Luyben (1975-) ("snowball effect"), Ruel Shinnar (1981-   ) ("dominant variables"), Jim Douglas and Alex Zheng (1985-   ) (hierarchical approach) and Jim Downs (1991-   ) (Tennessee-Eastman challenge process).

This paper is organized as follows. First, we present an expanded version of the plantwide control design procedure of Larsson and Skogestad (2000). A systematic approach to plantwide control starts by formulating the operational objectives. This is done by defining a cost function J that should be minimized with respect to the N_{opt} optimization degrees of freedom, subject to a given set of constraints. In reminder of the paper we go through the procedure step by step with special emphasis on:

- Degree of freedom analysis
- Selection of controlled variables
- Inventory control
- Loss in performance by bottom-up design

Finally, we discuss recycle systems and the so-called snowball effect.

2. Procedure for control structure design for chemical plants

The proposed design procedure is summarized in Table 1. In the table we also give the purpose and typical model requirements for each layer, along with a short discussion on when to use decentralized (single-loop) control or multivariable control (e.g. MPC) in the supervisory control layer. The procedure is divided in two main parts:

I. Top-down analysis, including definition of operational objectives and consideration of degrees of freedom available to meet these (tasks 1 and 2)
II. Bottom-up design of the control system, starting with the stabilizing control layer (tasks 3, 4 and 5 above)

The procedure is generally iterative and may require several loops through the steps, before converging at a proposed control structure.

Table 1: A plantwide control structure design procedure

STEP	Comments, analysis tools and model requirements
I. TOP-DOWN ANALYSIS:	
1. DEFINITION OF OPERATIONAL OBJECTIVES Identify operational constraints, and preferably identify a scalar cost function J to be minimized.	
2. MANIPULATED VARIABLES AND DEGREES OF FREEDOM Identify dynamic and steady-state degrees of freedom (DOF)	May need extra equipment if analysis shows there are too few DOFs.
3. PRIMARY CONTROLLED VARIABLES: Which (primary) variables c should we control? • Control active constraints • Remaining DOFs: Control variables for which constant setpoints give small (economic) loss when disturbances occur.	*Steady-state economic analysis:* • Define cost and constraints • Optimization w.r.t. steady-state DOFs for various disturbances (gives active constraints) • Evaluation of loss with constant setpoints
4. PRODUCTION RATE: Where should the production rate be set? (Very important choice as it determines the structure of remaining inventory control system.)	Optimal location follows from steady-state optimization (step 3), but may move depending on operating conditions.
II. BOTTOM-UP DESIGN: (With given controlled and manipulated variables)	*Controllability analysis*: Compute zeros, poles, pole vectors, gains, disturbance gains, relative gain array, minimum singular values, etc.
5. REGULATORY CONTROL LAYER. *5.1 Stabilization* *5.2 Local disturbance rejection* *Purpose:* "Stabilize" the plant using low-complexity controllers (single-loop PID controllers) such that 1) the plant does not drift too far away from its nominal operating point and 2) the supervisor layer (or the operators) can handle the effect of disturbances on the primary outputs (y_1=c). *Main structural issue*: What more (y_2) should we control? • Select secondary controlled variables (measurements) y_2 • Pair these with manipulated variables m, avoiding m's that saturate (reach constraints)	*5.1 Pole vector analysis* (Havre and Skogestad, 1997) for selecting measured variables and manipulated inputs for stabilizing control. *5.2 Partially controlled plant analysis*. Control secondary measurements (y_2) so that the sensitivity of states (x) to disturbances is small at intermediate frequencies. *Model:* Linear multivariable dynamic model. Steady state usually not important.

6. SUPERVISORY CONTROL LAYER. *Purpose:* Keep (primary) controlled outputs $y_1=c$ at optimal setpoints c_s, using as degrees of freedom (inputs) the setpoints y_{2s} for the regulatory layer and any unused manipulated variables. *Main structural issue*: Decentralized or multivariable control? *6a. Decentralized (single-loop) control* Possibly with addition of feed-forward and ratio control. • May use simple PI or PID controllers. • Structural issue: choose input-output pairing *6b. Multivariable control* Usually with explicit handling of constraints (MPC) • Structural issue: Size of each multivariable application	*6a. Decentralized*: Preferred for noninteracting process and cases where active constraints remain constant. *Pairing analysis:* Pair on RGA close to identity matrix at crossover frequency, provided not negative at steady state. Use CLDG for more detailed analysis *6b. Multivariable:* 1. Use for interacting processes and for easy handling of feedforward control 2. Use MPC with constraints handling for moving smoothly between changing active constraints (avoids logic needed in decentralized scheme 5a) *Model*: see 5
7. OPTIMIZATION LAYER *Purpose:* Identify active constraints and compute optimal setpoints c_s for controlled variables. *Main structural issue*: Do we need real-time optimization (RTO)?	*Model:* Nonlinear steady-state model, plus costs and constraints.
8. VALIDATION	Nonlinear dynamic simulation of critical parts

Model requirements

For the analysis of the *control layers* (step 5 and 6) we need a linear multivariable dynamic model. Since we are controlling variables at setpoints using feedback, the steady-state part of the model is not important (except for controller design with pure feedforward control). For the analysis of the *optimization layer* (steps 3 and 7) a nonlinear steady-state model is required. Dynamics are usually not needed, except for batch processes and cases with frequent grade changes. For modeling, we need to distinguish further between the cases of

1. Control structure design (this paper): "Generic" model sufficient
2. Controller design (tuning of controllers): Specific model needed

Since a good control structure is generally insensitive to parameter changes, it follows that a "generic" model is generally sufficient for our purpose. This is a model where the structural part is correct, but where all the parameters may not match the true plant in question. A first-principle theoretical model, based on material and energy balances, that covers the whole plant is usually recommended for this. For the control system design in case 2 (which is not the concern of this paper) we need a ``specific'' model, for example, based on model identification. Here it is usually sufficient with a local model for the application in question with emphasis on the time scale corresponding to the desired closed-loop response time (of each loop), or, if on-line tuning is used, we may not need any model at all.

Why not a single big multivariable controller?

Most of the steps in Table 1 could be avoided by designing a single optimizing controller that stabilizes the process and at the same time perfectly coordinates all the manipulated variables based on dynamic on-line optimization. There are fundamental reasons why such a solution is not the best, even with tomorrows computing power. One fundamental reason is the cost of modeling and tuning this controller, which must be balanced

against the fact that the hierarchical structuring proposed in this paper, without much need for models, is used effectively to control most chemical plants.

3. Definition of operational objectives and constraints (step 1)

The operational objectives must be clearly defined before attempting to design a control system. Although this seems obvious, this step is frequently overlooked. Preferably, the operational objectives should be combined into a scalar cost function J to be minimized. In many cases J may be simply selected as the operational cost, but there are many other possibilities. Other objectives, including safety constraints, should normally be formulated as constraints.

4. Selection of manipulated variables and degree of freedom analysis (step 2)

Degree of freedom analysis. We start with the number of *dynamic or control degrees of freedom*, N_m (m here denotes manipulated), which is equal to the number of manipulated variables. N_m is usually easily obtained by process insight as the number of independent variables that can be manipulated by external means from step 1 (typically, the number of adjustable valves plus other adjustable electrical and mechanical variables). Note that the original manipulated variables are always extensive variables.

Next, we must identify the N_{opt} *optimization degrees of freedom*, that is, the degrees of freedom that affect the operational cost J. In most cases the cost
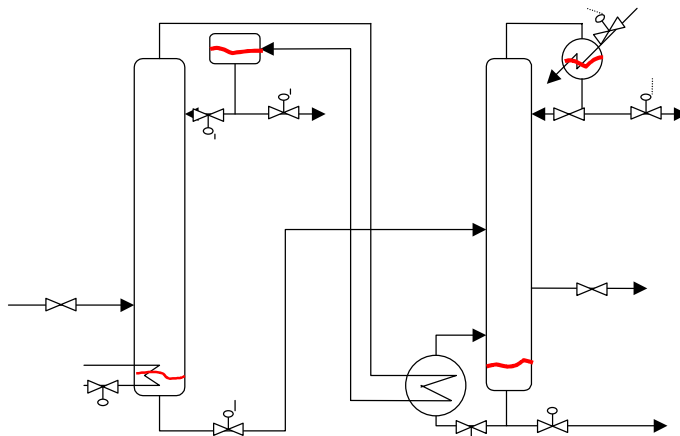
depends on the steady state only, and N_{opt} equals the number of *steady-state degrees of freedom* $N_{ss.}$ To obtain the number of steady-state degrees of freedom we need to *subtract* from N_m:

- N_{0m} = the number of manipulated (input) variables with no steady-state effect (or more generally, with no effect on the cost). Typically, these are "extra" manipulated variables used to improve the dynamic response, e.g. an extra bypass on a heat exchanger.
- N_{0y} = the number of (output) variables that need to be controlled, but which have no steady-state effect (or more generally, no effect on the cost). Typically, these are liquid levels in holdup tanks.

and we have

$$N_{ss} = N_m - (N_{0m} + N_{0y})$$

Example 1. The integrated distillation process in Figure 2 has $N_m = 11$ manipulated variables (including the feedrate), and $N_{0y} = 4$ liquid levels with no steady-state effect, so there are $N_{ss} = 11 - 4 = 7$ degrees of freedom at steady state.

$N_m = 11$ (incl. feed), $N_{0y} = 4$ (levels), $N_{ss} = 11 - 4 = 7$

Figure 2. *Degrees of freedom for integrated distillation process* (Example 1).

The optimization is generally subject to constraints, and at the optimum many of these are usually "active". The number of ``free'' (unconstrained) degrees of freedom that are left to optimize the operation is then $N_{opt} - N_{active}$. This is an important number, since it is generally for the unconstrained degrees of freedom that the selection of controlled variables (task 1 and step 3) is a critical issue.

Need for extra equipment (design change). In most cases the manipulated variables are given by the design, and a degree of freedom analysis should be used to check that there are enough DOFs to meet the operational objectives, both at steady state (step 2) and dynamically (step 5). If the DOF analysis and/or the subsequent design shows that there are not enough degrees of freedom (either for the entire process or locally for dynamic purposes), then degrees of freedom may be added by adding equipment. This may, for example, involve adding a bypass on a heat exchanger, or adding an extra heat exchanger or a surge tank. Note that it is not only the number of variables that is important, but also their range. If a manipulated variable saturates, then it is effectively lost as a degree of freedom

5. What should we control? (steps 3 and 5)

A question that puzzled me for many years was: Why do we control all these variables in a chemical plant, like internal temperatures, pressures or compositions, when there are no *a priori* specifications on many of them? Intuitively, we need to control the *"dominant" variables* for the process. The answer to this question is that we first need to control the variables directly related to ensuring *optimal economic operation* (these are the primary controlled variables $y_1 = c$ in step 3):

- Control active constraints (Maarleveld and Rijnsdorp, 1971; Skogestad, 2000)
- Select unconstrained controlled variables so that with constant setpoints the process is kept close to its optimum in spite of disturbances and implementation errors. (Skogestad, 2000) These are the less intuitive ones, for which the idea of self-optimizing control (see below) is very useful.

In addition, we need to control variables in order to achieve *satisfactory regulatory control* (these are the secondary controlled variables y_2 in step 5):

- With the regulatory control system in place, the plant should not drift too much away from its desired steady-state operation point. This will reduce the effect of nonlinearity, and enable the above supervisory control layer (or the operators) to control the plant at a slower time scale. Preferably, this "basic" control layer should be able to work for a wide range of primary control objectives.

In particular, we should

- Control unstable/integrating liquid levels. This consumes steady-state degrees of freedom since liquid levels have no steady-state effect (but this has already been taken into account in the degree of freedom analysis).
- Stabilize other unstable modes, for example, for an exothermic reactor (these are also usually quite obvious). This involves controlling extra local measurements, but does not consume any degrees of freedom, since the setpoints for the controlled variables replace the manipulated inputs (valve positions) as degrees of freedom.
- Control variables which would otherwise "drift away" due to large disturbance sensitivity (these are sometimes less obvious). This involves controlling extra local measurements, e.g. a tray temperature in a distillation column, and also does not consume any degrees of freedom.

Self-optimizing control (step 3)

The basic idea of self-optimizing control was formulated about twenty years ago by Morari *et al.* (1980) who write that "*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.*" To quantify this more precisely, we define the (economic) *loss* L as the difference between the actual value of the cost function and the truly optimal value, i.e. $L = J(u; d) - J_{opt}(d)$ where $u = f(c,d)$.

Self-optimizing control (Skogestad, 2000) is achieved if a constant setpoint policy results in an acceptable loss L (without the need to reoptimize when disturbances occur).

The main issue here is *not* to find the optimal setpoints, but rather to find the right variables to keep constant. The idea of self-optimizing control is illustrated in Figure 3. We see that a loss results when we keep a constant setpoint rather than reoptimizing when a disturbance occurs.

An additional concern with the constant setpoint strategy is that there is always a difference between the setpoint c_s and the actual value c due to implementation errors caused by measurement errors and imperfect control. To minimize the effect of the implementation errors, the cost surface as a function of c should be as flat as possible, see Figure 4.
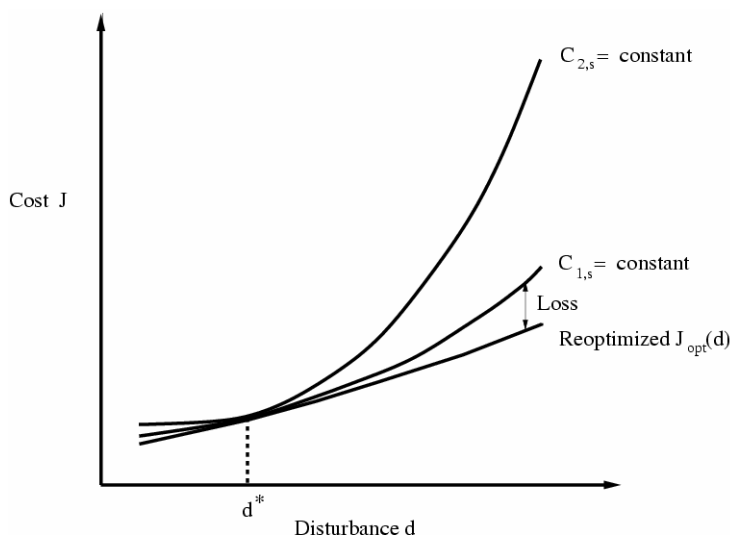

Figure 3. *Loss $L = J - J_{opt}$ (d) imposed by constant setpoint policy: There is a loss if we keep a constant setpoint rather than reoptimizing when a disturbance occurs. For the case in the figure it is better (with a smaller loss) to keep the setpoint c_{1s} constant than to keep c_{2s} constant.*
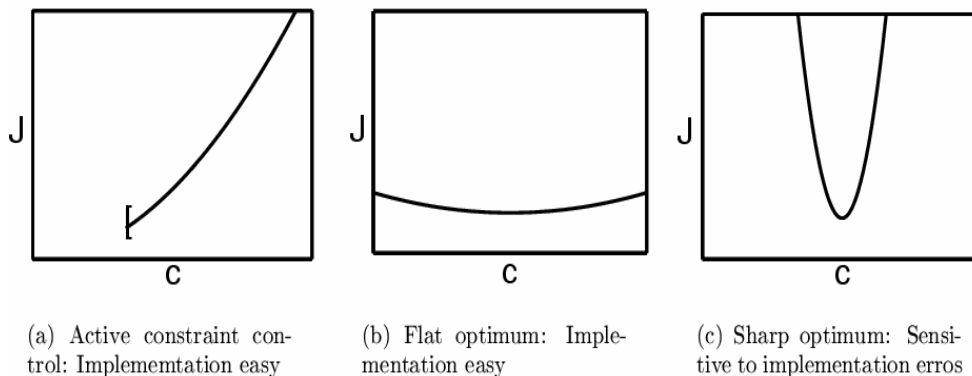

(a) Active constraint control: Implememtation easy

(b) Flat optimum: Implementation easy

(c) Sharp optimum: Sensitive to implementation erros

Figure 4. Implementing the controlled variable: Effect of implementation error on cost

Example sharp optimum. An example of a sharp optimum (Figure 4c) is for high-purity distillation where the controlled variable c is the temperature in the end of the column. In this case even a small change in temperature at the end of the column will imply a large relative change in composition, and thus a large change in cost J for the process.

To select controlled variables for self-optimizing control, one may use the stepwise procedure of Skogestad (2000):

> **Step 3.1** Definition of optimal operation (cost and constraints)
> **Step 3.2** Determine degrees of freedom for optimization
> **Step 3.3** Identification of important disturbances
> **Step 3.4** Optimization (nominally and with disturbances)
> **Step 3.5** Identification of candidate controlled variables
> **Step 3.6** Evaluation of loss for alternative combinations of controlled variables (loss imposed by keeping constant setpoints when there are disturbances or implementation errors)
> **Step 3.7** Evaluation and selection (including controllability analysis)

Note that except for Step 3.7, this procedure normally requires steady-state information only. The procedure has been applied to several applications, including distillation column control (Skogestad, 2000), the Tennessee-Eastman process (Larsson et al., 2001) and the reactor-recycle process (Larsson et al., 2003).

To identify good *candidate* controlled variables, c, one should look for variables that satisfy all of the following *requirements* (Skogestad, 2000):

1. The *optimal value* of c should be *insensitive* to disturbances
2. c should be easy to measure and control (so that the implementation error is acceptable)
3. The *value* of c should be *sensitive* to changes in the manipulated variables (the steady-state degrees of freedom). Equivalently, the optimum (J as a function of c) should be flat.
4. For cases with more than one unconstrained degrees of freedom, the selected controlled variables should be independent.

At least "locally" (for small disturbances), these requirements may be combined into a single rule (which generalizes requirement 3): *Look for variables that maximize the minimum singular value of the appropriately scaled steady-state gain matrix* G from u to c (Skogestad and Postlethwaite, 1996) (Skogestad, 2000). Here u denotes the steady-state degrees of freedom.

If a linearized model is available, then the minimum singular value rule may very useful for *eliminating* poor candidate variables, but it is a local analysis, and for a final selection one should use the above procedure with evaluation of the loss for larger disturbances.

It is stressed that the issue of selecting appropriate controlled variables c for the unconstrained degrees of freedom is equally important when we use multivariable constrained control (MPC) in the supervisory control layer. The setpoints for the selected controlled variables as well as the active constraints, which may vary with time, are then computed by the steady-state optimization layer and supplied to MPC for implementation.

6. Production rate and inventory control (step 4)

In chemical plant mass moves through the process, starting up as feeds and ending up as products. The production rate is commonly assumed to be set at the inlet to the plant, with outflows used for level control. One important reason for this is probably that most of the control structure decisions are done at the design stage (before the plant is built) where we usually fix the feedrate. However, during operation the feedrate is usually a degree of freedom, and very often the economic conditions are such that it is optimal to maximize production. We then have the following rule:

> *Identify the main bottleneck in the plant by optimizing the operation with the feedrate as a degree of freedom (steady state, see step 3). Set the production rate at this location.*

The justification for this rule is that the economic benefits of increasing the production are usually very large (when the market conditions are such), so that it is important to maximize flow at the bottleneck. On the other hand, if market conditions are such that we are operating with a given feed rate or given product rate, then the economic loss imposed by using a outer cascade loop to adjust the production rate at the bottleneck is usually zero, as deviations from the desired feed or production rate can be averaged out over time, provided we have storage tanks for feeds or products

7. Regulatory layer (step 5)

In this paper we use the terms "lower layer", "inner loops" and "secondary loops" as synonyms for the regulatory control layer. The "primary" control system is the same as the supervisory control system.

The regulatory control layer should usually be of *"low complexity"*. Usually it consists of single-input-single-output (SISO) PI control loops. The main objective is to *"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. The controlled variables for stabilization are measured output variables y_2, and their setpoints y_{2s} may be used as degrees of freedom by the layers above.

More generally, the objective of the regulatory control layer is to locally control secondary measurements (y_2), so that the effect of disturbances on the primary outputs ($y_1=c$) can be handled by the layer above (or the operators). In the regulatory control layer we generally avoid using manipulated variables that may saturate, because otherwise control is lost and reconfiguration of loops is required.

The main structural issue in the regulatory layer is to determine *which extra (secondary) variables y_2 to control* in order to stabilize the process and achieve local disturbance rejection (see Figure 5). A good secondary controlled variable (measurement) usually has the following properties:
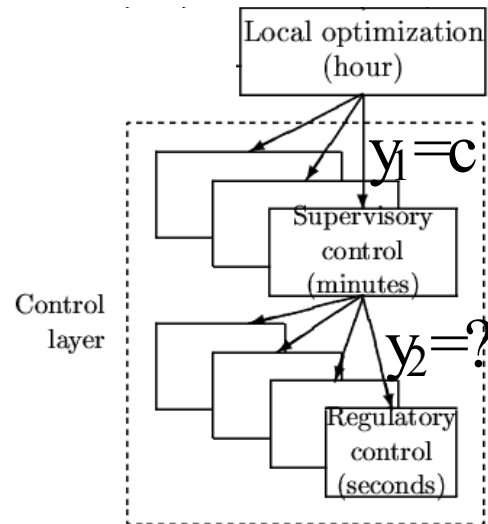

Figure 5. Selection of secondary controlled variables y_2

- The variable is easy to measure
- The variable is easy to control using one of the available manipulated variables (the manipulated variable should have a "direct", fast and strong effect on it)
- For stabilization: The unstable mode should be detected "quickly" by the measurement (compute. for example, the *pole vectors* for a more detailed analysis)
- For local disturbance rejection: The variable is located "close" downstream of an important disturbance (use, for example, a *partial control analysis* for a more detailed analysis).

We have here distinguished between stabilization and local disturbance rejection, but in practice the may be combined into the requirement of avoiding that the states x (or more generally, the weighted states Wx) drift too far away from their desired (nominal) value. The advantages of considering a measure of all the states in the system is that the regulatory control system is then not tied to closely to a particular primary control objective which may change with time. Also, keeping all the states bounded is important to avoid that nonlinear effects give a problem. The sensitivity of the system state to disturbances with the regulatory control loops closed may be analyzed using partial control, as discussed later.

The "unstable'' modes are very often related to inventory in each unit. This includes both the overall inventory (total mass) as well as the inventory of individual components.

- For liquid phase systems, overall inventory in each unit is stabilized by controlling liquid level.
- For gas phase systems, overall inventory (pressure) is controlled in selected units, but in many units it is left uncontrolled (floating), for example, to minimize pressure drop.
- For both gas and liquid phase systems, the inventory of individual components may need to be stabilized. Usually, this involves controlling a composition, or a derived property such as temperature. For example, in a distillation column, a temperature controller is often used to stabilize its otherwise drifting composition profile. Note that we do not need to control the inventory of all components, as there may be only one unstable mode associated with the "drift" of many components. Also, control of a single measurement may stabilize several unstable modes.

As already discussed, the design of the regulatory layer (or more precisely, the assignment of control loops in the regulatory control layer) usually starts by determining where to set the production rate (step 4), and then assigning the stabilizing liquid level loops. For the other unstable modes a *pole vector analysis* (Havre and Skogestad, 1997) may be useful. It requires a linearized model, and to minimize the required input usage, the rule is to select for stabilizing control measured variables and manipulated inputs corresponding to large elements in the pole vectors.

Except for cases where we do final control in the regulatory control layer, no degrees of freedom are lost as the setpoints y_{2s} for the locally controlled variables remain degrees of freedom for the layer above. This assumes that also the setpoints for the liquid levels remain as degrees of freedom.

Partial control

To analyze this in more detail the concept of *partial control* is very useful, and the objective is to minimize the magnitude of the partial control

gain, P_{d1} (see below), which gives the effect of the disturbances on the weighted states (x) with the secondary (regulatory) loops closed. The results in this section are based on Skogestad and Postlethwaite (1996) and Havre and Skogestad (1996), but they considered the effect on the primary outputs ($y_1=c$), whereas we here have a more general view, where $y_1=W$ x denotes the weighted states.. In the case where we are concerned with keeping the plant close to its steady-state, the weight matrix W is a diagonal matrix consisting of the inverse of the allowed variation in its state along the diagonal. However, in general W can be a full non-square matrix.

Let the overall process model be $y = G u + G_d d$. We partition the manipulated inputs u and the measured outputs y into two sets,
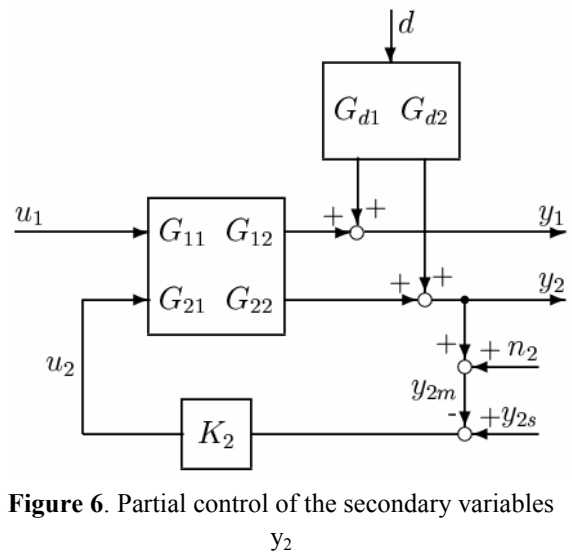
$$y = [y_1 \ y_2], \quad u=[u_1 \ u_2]$$

Figure 6. Partial control of the secondary variables y_2

where $y_1 = Wx$ denotes the weighted states, y_2 the (secondary) measurements, u the inputs (m), u_2 the inputs used in the secondary layer, and u_1 denotes the unused inputs (this is not very important). The plant model G is partitioned correspondingly,

$$y_1 = G_{11} u_1 + G_{12} u_2 + G_{d1} d$$
$$y_2 = G_{21} u_1 + G_{22} u_2 + G_{d2} d$$

By closing the lower-layer (inner) loops involving u_2 and y_2,

$$u_2 = K_2 (y_{2s} - y_2 - n_2),$$

we obtain, as seen from the (supervisory) layer above, *a partially controlled system* with y_{2s} (the setpoints for the locally controlled variables y_2) and u_1 (the unused input) as degrees of freedom. The transfer function for the partially controlled system is

$$y_1 = P_1 u_1 + P_{r1} (y_{2s} - n_2) + P_{d1} d$$

where $P_1 = G_{11} - G_{12} K_2 (I + G_{22} K_2)^{-1} G_{21}$,
$P_{d1} = G_{d1} - G_{12} K_2 (I + G_{22} K_2)^{-1} G_{d2}$,
$P_{r1} = G_{12} K_2 (I + G_{22} K_2)^{-1}$.

The lower layer is assumed to be much faster than the upper layer, so for a preliminary analysis when selecting secondary controlled variables y_2 we may assume that y_2 is perfectly controlled (let $K_2 \to \infty$, or equivalently set $y_2 = y_{2s} - n_2$) and we get

$P_1 = G_{11} - G_{12} G_{22}^{-1} G_{21}$,
$P_{d1} = G_{d1} - G_{12} G_{22}^{-1} G_{d2}$,
$P_{r1} = G_{12} G_{22}^{-1}$

We should then (note that $y_1 = Wx$ here denotes the weighted states):

- Select to control secondary variables y_2 such that the norm of $P_{d1} = G_{d1} - G_{12} G_{22}^{-1} G_{d2}$ is sufficiently small (especially at intermediate and high frequencies beyond the bandwidth of the primary control system, but also the steady state should be considered). If this is satisfied then we say that the plant is "stabilized'' (in the extended sense defined earlier).

It is recommended to plot the magnitude of the elements in P_{d1} as a function of frequency, and if the variables y_1 $(= W x)$ and d are appropriately scaled then "small" means less than 1 (Skogestad and Postlethwaite, 1996).

If we have found a lower-layer control structures which "stabilizes" the plant, then we should as the next step check that it is consistent with the selected set of controlled variables. This may be done by recalculating the partial gains, but now with $y_1 = c$. We then have:
- If we want to use the reference r_2 as a degree of freedom to control the primary outputs $y_1 = c$, then $P_{r1} = G_{12} G_{22}^{-1}$ should

be (sufficiently) large (also at steady-state).
- If r_2 is not a degree of freedom in the primary control layer, then it may be viewed as a disturbance and $P_{r1} = G_{12} G_{22}^{-1}$ should be small so that it can be handled by the supervisory control system.
- If we want to use the unused inputs u_1 as a degree of freedom to control the primary outputs y_1, then $P_1 = G_{11} - G_{12} G_{22}^{-1} G_{21}$ should be (sufficiently) large (also at steady-state).

The use of partial gains with $K_2 \to \infty$ is useful for finding which variables to control in regulatory control layer. However, for a more detailed analysis it may be useful to design a controller K_2. In the simplest case K_2 may be assumed static. Also, it is often desirable to find a simple (low-complexity) way of implementing the controller K_2. For this it is useful to define some measure of complexity as discussed next.

Complexity

It is generally desirable that the complexity of the control system, and in particular of the regulatory control layer should be as small as possible (e.g., Nett, 1989). To quantify this it may be useful to introduce a *structural complexity number* Π_s. This may be defined in many ways, and one possibility is the following:

$$\Pi_s = \text{\#measurements} + \text{\#manipulators} + \text{\#blocks} + \text{\#control-parameters}$$

8. Supervisory control (step 6)

The purpose of the supervisor control layer is to keep the (primary) controlled outputs c at their optimal setpoints c_s, using as degrees of freedom the setpoints y_{2s} in the regulatory layer and any unused manipulated inputs. Which variables to control and their setpoints are determined by the optimization layer above. Note that the variables to control may change if the active constraints change.

For the supervisory control layer, the first structural issue is deciding on whether to use decentralized or multivariable control. Note that there is usually some decentralization, that is, there is often a

combination of several multivariable and single-loop controllers.

Decentralized single-loop control is the simplest. It is preferred for noninteracting process and cases where active constraints remain constant. Advantages with decentralized control:

+ Tuning may be done on-line
+ No or minimal model requirements
+ Easy to fix and change

Disadvantages:

- Need to determine pairing
- Performance loss compared to multivariable control
- Complicated logic required for reconfiguration when active constraints move

The decision on how to pair inputs (y_{2s}, u_1) and outputs (c), and this is often done based on process insight. In more difficult cases a RGA-analysis may be useful, and the rule is pair such that the resulting transfer matrix is close to identity matrix at the crossover expected frequency, provided the element is not negative at steady state. For a more detailed analysis one should also consider disturbances, and compute the closed-loop disturbance gain (CLDG) (Skogestad and Postlethwaite, 1996). One disadvantage with decentralized control is that it may require reconfiguration of loops (with complicated logic) if the active constraints change

Multivariable control is preferred for interacting processes and for processes with changes in active constraint. For the cases where the constraints may change, one needs a multivariable controller with explicit constraint handling (e.g., MPC). This avoids the need for logic, and gives a smooth transition between active constraints. Advantages with multivariable constrained control (MPC):

+ Coordinated control for interactive processes
+ Easy handling of feedforward control
+ Easy handling of changing constraints with no logic required and smooth transition between active constraints

Disadvantages:
- Requires multivariable dynamic model
- Tuning may be difficult

- Generally more sensitive to uncertainty and changes in plant operation
- May be less transparent
- May have a reliability problem: "Everything goes down at the same time"

The optimization in step 2 with various disturbances may be used to set up a table of possible combinations of active constraints, and multivariable constrained control (MPC) should be used if a structure with single-loop controllers will require excessive reconfiguration of loops.

9. Optimization (step 7)

The purpose of the optimization is to identify the active constraints and recompute optimal setpoints c_s for controlled variables.

In addition to deciding on which unconstrained variables to control (see step 3), the main structural issue is to decide if it is necessary to use real-time optimization (RTO), or if manual optimization is sufficient. With RTO new setpoints are typically computed about every hour or so, after the steady-state model has been adjusted to match the current conditions. Real-time optimization (RTO) is costly in the sense that it requires a detailed steady-state computer model to be maintained and continuously updated. If the active constraints do not change, and we are able to find good self-optimizing controlled variables, then RTO gives little benefit and should not be used. There are also situations where the active constraints do change, but where the operators may be able to identify and implement the required changes.

10. Validation (step 8)

After having determined a plantwide control structure, it may be necessary to validate the structure, for example, using nonlinear dynamic simulation of critical parts.

11. Discussion. Bottom-up design (steps 5-7): Any loss in control performance?

We have here assumed that the control system is designed bottom-up starting with the lower regulatory control layer, involving the inputs u_2 ⁻

(denoted m earlier) and the outputs y_2. Does this hierarchical decomposition into control layers impose any loss on the overall achievable control performance in terms of the primary outputs y_1 (denoted c earlier)?

The answer is "no" provided we have full access to the lower (secondary; regulatory) layer from the upper (primary; supervisory) control layer:

Theorem (Larsson and Skogestad, 1998). *The closing of a lower-layer (partial) control system, involving the manipulated input u_2 and the measured and controlled variable y_2, introduces no new control limitations (e.g., in terms of RHP-zeros) provided*
   1. *The setpoints y_{2s} (for y_2) are available as degrees of freedom at the next layer.*
   2. *The measurements y_2 are available at the next layer*
   3. *The controller interconnecting y_2 and u_2 is minimum phase and stable (but may have integrators).*

The proof is trivial because under these conditions we can just invert away the controller K_2 used in the lower layer. Although the theorem is trivial, it has some important practical significance in terms of multivariable control (MPC). It tells us that the presence of the lower-layer control system imposes no limitations on the overall control performance, provided we at the next layer use a multivariable controller with full access to the measurements (y_2) and setpoints (r_2) used in the lower layer.

However, in many practical cases, we want to use a simpler control system, and we may impose limitations by (A) improper pairing, or (B) use of inner cascade loops that actually amplify disturbances.

A. Improper pairing. Assume that we do not have access to r_2 when controlling y_1. For example, this is the situation if we do "final control" in the lower layer, i.e. y_2 is actually a "primary" output. (Alternatively, this is the situation if we use decentralization within the supervisory control layer, and design the single-loop controllers sequentially).

In this case pairing on a negative steady-state RGA-element will impose a fundamental limitation in terms of the control of y_1. More precisely, *if*

   (i) *the pairing between u_2 and y_2 corresponds a negative steady-state RGA-element (in the RGA of G),*
   (ii) *G_{22} has no RHP-zero, and*
   (iii) *we use integral control in K_2,*

then closing the inner loop involving u_2 and y_2 will introduce a RHP-zero in the resulting transfer function P_{11} from u_1 to y_1 (Shinskey 1979, Bristol 1966, Grosdidier and Morari, 1985, Jacobsen, 1999). *The effect of the RHP-zero is less severe (moved to higher frequencies) as we tighten the control in the inner loop* (Jacobsen, 1999, Larsson, thesis 2000).

Note: There are also other reasons for avoiding pairing on negative steady-state RGA-elements, including ensuring failure tolerance and allowing for independent tuning (DIC).

B. Improper cascade control. Assume that the outputs y_2 are "secondary" outputs (extra measurements) which we choose to control in order to stabilize the plant or improve local disturbance rejection. This is a standard cascade control system. In this case the set u_1 is empty, and the layer above uses the setpoints r_2 in order to control the "primary" outputs y_1.

In this case, the main purpose of the lower-layer control system is to improve the control of y_1, but if improperly designed, it may make the situation worse. For example, if we had a case where originally the disturbance had no effect on the output ($G_{d1} = 0$), then the closing of a lower-layer loop may introduce sensitivity to the disturbance (with $P_{d1} = G_{d1} - G_{12} K_2 (I + G_{22} K_2)^{-1} G_{d2}$ nonzero).

12. Case studies

The design procedure described in this paper has been applied to numerous case studies, several of which are found in the thesis by Larsson (2000).

- Larsson et al. (2003) and Govatsmark and Skogestad (2002): Selection of controlled variables for reactor, separator and recycle

process (Steps 1, 2 and 4 plus some on Steps 5 and 6)

- Larsson and Skogestad (1999) and Engelien et al. (2003): Optimization and selection of controlled variables for heat-integrated distillation columns (Steps 1, 3 and 4)

- Larsson et al. (2001): Selection of controlled variables for the Tennessee-Eastman process with focus on how to eliminate poor candidate variables (Steps 1 and 3). (The control system design in this paper was included to show that the proposed controlled variables are workable, but otherwise do not follow the steps in Table 1. For example, there is no thorough analysis on where to locate the throughput manipulator (step 4)).

Skogestad (2000): Optimization (moving active constraints as a function of feedrate) and selection of controlled variables for a propane-propylene distillation column (Steps 1, 3 and 4).

13. Conclusion

The proposed plantwide control design procedure in Table 1 has two main parts:

I. Top-down analysis to identify degrees of freedom and primary controlled variables (look for self-optimizing variables).

II. Bottom-up analysis to determine secondary controlled variables and structure of control system (pairing).

There are many outstanding research issues related to filling in more detailed procedures in Table 1 on what to do in each step of the procedure. In particular this applies to the bottom-up part of the procedure. For example, more work is needed in order to understand how to decompose and coordinate the layers of the control system.

References

Bristol, E.H., 1966, "On a new measure of interaction for multivariable control systems", *IEEE Trans. on automatic control*, 133-134.

Buckley, P.S., 1964. *Techniques of process control.* Wiley.

Engelien, H.K., Larsson. T. and S. Skogestad, 2003, "Implementation of optimal operation for heat-integrated distillation columns"', *Trans. IChemE*, **81** (Part A), 277-281 (2003).

Foss, A.S., 1973, "Critique of chemical process control theory", *AIChE Journal,* **19**, 209-214.

Govatsmark. M.S. and S. Skogestad, 2002, "Selection of controlled variables and robust setpoints", Proc. of IFAC World Congress, Barcelona, 21-26 July 2002, Paper T-Mo-M-11-4.

Grosdidier, P. and M. Morari, 1985, "Closed-loop properties from steady-state gain information", *Ind.Eng.Chem.Res*., **24**, 221-235.

Havre, K. and S. Skogestad, 1996, ``Input/output selection and partial control," IFAC World Congress, San Francisco, July 1996, **M**, 181-186.

Havre, K. and S. Skogestad, 1998, ``Selection of variables for regulatory control using pole vectors", *Proc. IFAC symposium DYCOPS-5,* Corfu, Greece, 614-619.

Jacobsen, E.W., 1999, "On the dynamics of integrated plants - non-minimum phase behavior", *J. Proc. Control*, **9**, 439-451.

Larsson, T. and S. Skogestad, 1998, ``Limitations imposed by lower layer control configurations ", *AIChE Annual Meeting*, Miami Beach, 16-20 Nov. 1998

Larsson, T. and S. Skogestad, 1999, ``Control of industrial heat integrated distillation columns", *AIChE Annual Meeting*, Dallas, 1-5 Nov. 1999.

Larsson, T., 2000. *Studies on plantwide control,* Ph.D. Thesis, Norwegian University of Science and Technology, Trondheim.

Larsson, T. and S. Skogestad, 2000, "Plantwide control: A review and a new design procedure", *Modeling, Identification and Control*, **21**, 209-240. *(Also available from the home page of Sigurd Skogestad).*

Larsson, T., K. Hestetun, E. Hovland and S. Skogestad, 2001, "Self-optimizing control of a large-scale plant: The Tennessee Eastman process'', *Ind.Eng.Chem.Res*., **40,** 4889-4901.

Larsson, T., M.S. Govatsmark, S. Skogestad and C.C. Yu, 2003, "Control structure selection for reactor, separator and recycle processes", *Ind. Eng. Chem. Res.*, **42**, 1225-1234.

Luyben, W.L. 1993. "Dynamics and control of recycle systems. 2. Comparison of alternative process designs." *Ind. Eng. Chem. Res.* **32**, 476-486.

Luyben, W.L. 1994. "Snowball effect in reactor/separator processes with recycle". *Ind. Eng. Chem. Res.* **33**, 299-305.

Maarleveld, A. and J.E. Rijnsdrop, 1970, "Constraint control of distillation columns". *Automatica* **6**, 51-58.

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* **26**, 220–232.

Nett, C.N., 1989, "A quantitative approach to the selection and partitioning of measurements and manipulations for the control of complex systems", *Presentation at the Caltech Control Workshop*, Pasadena, USA, January 1989.

Shinskey, F.G., 1979, *Process control systems*, Mc-Graw Hill

Skogestad, S. and I. Postlethwaite, 1996, *Multivariable feedback control*, Wiley.

Skogestad, S. (2000). "Plantwide control: The search for the self-optimizing control structure". *J. Proc. Control* **10**, 487-507.

Wu, K.L. and C.-C. Yu (1996). "Reactor/separator process with recycle. 1.Candidate control structure for operability". *Computers. Chem. Engng.,* **20**, 1291-1316.