Techniques in the Control of Interconnected Plants

Morten Hovd*
Fantoft Prosess
Teknostallen
N-7030 Trondheim, Norway

Sigurd Skogestad†
Chemical Engineering
University of Trondheim-NTH
N-7034 Trondheim, Norway

To appear in:
C. T. Leondes (Ed.): Gordon and Breach International Series in Engineering, Technology, and Applied Science Volumes on "Mechatronic Systems Techniques and Applications".
This version printed on September 9, 1996

*Fax: (+47) 73 94 38 61, E-mail: morten.hovd@fantoft.no, Phone: (+47) 73 54 02 11.
†Fax: (+47) 73 59 40 80, E-mail: skoge@kjemi.unit.no, Phone: (+47) 73 59 41 54.

1 INTRODUCTION

For many real-life multivariable control problems, traditional single-loop control theory is clearly incomplete. This occurs when it is difficult or impossible to select and design a set of single-loop controllers which together satisfy the control objectives. Indeed, from our background in chemical process control, it appears that the proportion of such control problems is increasing. This is caused by tighter integration of heat and mass in modern chemical plants, and the decreased volumes (or removal) of buffer tanks between processing steps. Thus, disturbances will spread more rapidly and widely throughout a modern chemical plant than in older plants. This puts greater demands on the control system in modern plants than in older plants.

There exists sophisticated theories for the synthesis of multivariable controllers. Model based predictive control even makes it relatively simple to handle process constraints in a systematic fashion. Nevertheless, modern chemical plants still have a very large number of single-loop controllers. In a typical chemical plant, the control system can be decomposed into a hierarchical structure, as depicted in Fig. 1.

- The fundamental layer of the control system is the **Regulatory control** layer. This consists mainly of single-loop controllers, together with some feedforward and ratio controllers. Truly multivariable controllers are rare at the regulatory control level. The regulatory control level keeps a number of controlled variables close to setpoints determined by operators or higher levels in the control system. This serves to stabilize the plant, and it is usually possible for trained operators to keep the plant in operation with only the help of the regulatory control level.

- Above the regulatory control level there is often a **Supervisory control** level. This level coordinates the action of the individual controllers in the regulatory control layer. This is the layer where truly multivariable control is most often implemented, and some simplified optimization of individual chapters of the plant is often performed.

- The highest level of the control system is the **Plant optimization** level. At this level the operation of the entire plant is optimized to maximize profits subject to constraints such as equipment capacity, safety and environmental regulations.

Note that there is a separation in timescale between the different layers of the control system. The purpose of the regulatory control level is to stabilize and maintain safe operation conditions, and commonly operates on a timescale of minutes or seconds (which is considered fast in the chemical processing industry) or even faster if necessary. The supervisory control system normally operates on a timescale of several minutes or hours. This is sensible since

process variations on the timescale of seconds seldom affect the profitability of the plant - unless they cause plant shutdown, which a well designed regulatory control system should avoid. The production planning level typically operates on the timescale of days.

Although the higher levels in the control hierarchy can be said to be more advanced than the regulatory control level, it would be misleading to conclude that the higher levels are more important - since the higher levels depend on the regulatory control level for implementing the changes to the process that the higher levels have determined to be desirable. Thus, a well designed regulatory control level will make the design of the higher levels in the control system simpler. Conversely, a poor regulatory control system can make it impossible to achieve the process improvements found by the higher levels in the control system.

Having observed that that it is customary to impose *structure* on control systems for large plants, it is pertinent to ask the question whether this is a sensible thing to do. After all, it is clear that imposing structure on the control system restricts the controller design, and in most cases the theoretically optimal controller will be one big centralized controller which perfectly coordinates all control actions. It should come as no surprise to the reader that we believe that a structured control system is actually an advantage. Some reasons are:

- **Economics.** The concept of optimality common in control theory only refers to controllers *in operation*, and does not cover the costs of modeling a plant and updating such models. Single loop feedback controllers get most of the information about the plant from the feedback itself, and often very little explicit process model information is required for implementing such controllers. On the other hand, in order to successfully coordinate control actions, a centralized controller would require a very accurate model of the plant. With the separation in timescale of the layers of the control system, the modeling required for the supervisory control level is greatly simplified, since the fast dynamics of the plant are normally not relevant for the supervisory control system.

- **Robustness.** Even after spending considerable effort in modeling a system, it is not reasonable to expect the resulting model to be perfect unless the system considered is quite trivial. Since the supervisory control level operates within the bandwidth of the regulatory control level, the regulatory control level effectively removes uncertainty a seen from the supervisory control level. The regulatory control level is relatively robust to model uncertainty, for two reasons: a) it relies mostly on feedback and not on a model for information about the plant, and b) the structure of the regulatory control system, consisting mostly of single loop controllers, makes it relatively robust at the expense of (theoretical) performance (see e.g. [59]).

- **Startup and shutdown.** The behavior of a plant during startup or shutdown is

normally very different from the behavior in the normal operating region. Developing models that include startup and shutdown as well as normal operating conditions is usually very challenging, and even if such a model is obtained it may be very difficult to design a controller that performs adequately both in normal operation and during startup and shutdown.

During plant startup, it is normal practice to put most controllers in manual. The controllers of the regulatory control system can then be put in operation one by one as conditions allow. When the regulatory control system for a section of the plant is operating adequately, the supervisory control system can be put in operation. The reverse sequence of events take place during controlled shutdowns. Structuring the control system thus facilitates easier startup and shutdown.

- **Redesign and retuning.** It may be necessary to modify the control system as a result of changing operating conditions, raw materials, or equipment. With a structured control system such modifications are much easier to accomplish, and it will often be possible to make the necessary changes in the regulatory level only, thus leaving the higher levels in the control system unchanged.

- **Operator acceptance and understanding.** It is important that the operators understand the control system. Operators with insufficient understanding of the control system may otherwise cause plant shutdowns. Such shutdowns should rightly be blamed on the control system (or on the people who designed and implemented the control system), since operators are an integral part of the plant operation. Structured control systems, and particularly the regulatory control level, are easier for operators to understand than a centralized control system.

The discussion above has focused on control systems for large plants. For smaller plants, as well as non-critical (in terms of economics) sections of large plants, there may actually be no supervisory control level at all. Fulfilling the objectives of the supervisory control system is then a task for the plant operators. Obviously, a well-designed regulatory control system will be of great help to the operators in such cases.

Anyway, the task of designing a control system includes much more than just designing a set of controllers for controlling a given set of measurements with a given set of manipulated variables. The task of control system design actually starts with determining what manipulated variables and what measurements should be used for control. This part of the control system design has traditionally received little attention in the control literature, but may often be quite involved. This is especially the case when designing new plants, and one

needs to determine both the types and locations of both measurements and actuators - in this case the number of alternatives can be very large.

The topics addressed in this chapter are:

- Limitations on achievable performance in control systems.
- Measures of interaction in control systems.
- Selection og controlled and manipulated variables for control.
- Integrity to control loop failure.
- Choice of pairing of controlled and manipulated variables for fully decentralized (i.e., single loop) control.
- Approaches to the tuning of decentralized controllers

The reader is assumed to be familiar with basic linear algebra and frequency domain analysis - at least for single-input single-output systems. The tools presented in this chapter require a model of the plant to be available; this may appear contradictory to the statements made above about the cost of obtaining models and the issue of robustness. However, some of the tools can give useful information about the control properties of the plant even with very simple models, like for instance steady state models. Also, very simple tools can give indications of problems with robustness. Whether these indications of robustness problems are themselves robust to model uncertainty is not a critical issue - any indication of robustness problems is sufficient cause for searching for an alternative control structure.

The motivation and examples used in this chapter are taken from the area where the authors are most at home, that is chemical process control. However, it is our firm belief that similar problems can be found in other areas where control engineering is applied, and that the techniques described in this chapter can be put to successful use in these areas. There are no limitations in the techniques themselves that restrict their applicability to chemical process control. For instance, readers with a background from aircraft control may see parallels between the regulatory control level and the *stabilization* layer in aircraft control systems, and between supervisory control and *guidance*.

1.1 Notation

The model of the plant is given by

$$y(s) = G(s)u(s) + G_d(s)d(s) \tag{1}$$

where:

y is the vector of measurements,

u is the vector of manipulated variables,

d is the vector of disturbances, and

G and G_d are transfer function matrices of appropriate dimensions. s is the Laplace variable, which will be suppressed when not needed for clarity. g_{ij} is the element in the i'th row and j'th column of G, similarly g_{dij} is the element in row i and column j of G_d. G^{ij} can be obtained from G by deleting row i and column j. The controller is denoted K. For decentralized control, both K and G are square, and it is assumed that the order of inputs and outputs has been rearranged such that the elements of G that connect the paired inputs and outputs appear on the main diagonal of G. K will then be a diagonal matrix, and the i'th element on the diagonal is denoted k_i.

The vector of references (or setpoints) is denoted r and the vector of control offsets, $y-r$, is denoted e. We then have $e = S(d-r)$ and $y = SG_d d + Tr$, where S is the sensitivity function and $T = I - S$ the complementary sensitivity function. These are given by

$$S = (I + GK)^{-1} \tag{2}$$

$$T = GK(I + GK)^{-1} \tag{3}$$

The matrix $\tilde{G}$ consists of the diagonal elements of G, and for decentralized control the matrices of sensitivity functions and complementary sensitivity functions for the individual loops are given by $\tilde{S} = (I+\tilde{G}K)^{-1}$ and $\tilde{T} = \tilde{G}K(I+\tilde{G}K)^{-1}$. Note that the diagonal elements of $\tilde{S}$ differ from the diagonal elements of S.

When we want to distinguish between the *model* of the plant and the *true* plant, the true plant is denoted P.

1.2 Scaling

All interpretations and examples in this paper assumes that appropriate scaling has been performed. In general, the variables should be scaled to be within the interval -1 to 1, that is, their expected magnitudes should be normalized to be less than 1. This is done by dividing the unscaled signals by their *expected maximum allowed change*. Let G' and G'_d denote the unscaled transfer matrices and u', y', d', e' and r' denote the unscaled inputs, outputs, disturbances, errors and references. Define the maximum allowed change in each of the signal, $u'_{i,max}$, $d'_{i,max}$, $e'_{i,max}$, $r'_{i,max}$ and collect the scalers in the diagonal matrices

$$D_u = \mathrm{diag}\{u'_{i,max}\}, \quad D_d = \mathrm{diag}\{d'_{i,max}\},$$
$$D_e = \mathrm{diag}\{e'_{i,max}\}, \quad D_r = \mathrm{diag}\{r'_{i,max}\} \tag{4}$$

then the scaled error can be written

$$e = \underbrace{D_e^{-1} G' D_u}_{G} u + \underbrace{D_e^{-1} G'_d D_d}_{G_d} d - \underbrace{D_e^{-1} D_r}_{R} \tilde{r} \tag{5}$$

where $u' = D_u u$, $d' = D_d d$, $e' = D_e e$, $r' = D_r \tilde{r}$, $y' = D_e y$ and $r' = D_e r$, and we have $|e| < 1$, $|u| < 1$, $|d| < 1$ and $|\tilde{r}| < 1$. The scaling of variables are depicted in Fig. 2. Note the appearance of the matrix $R = D_e^{-1} D_r$ this matrix is equal to the identity matrix if the error and the reference signals are scaled with the same magnitude.

1.3 Use of the term 'Controllability'

The term 'controllability', as used in this chapter, is a plant property which reflects how easy it is to control the plant. The controllability of a plant depends on many different aspects, such as: time delays, transmission zeroes in the right half of the complex plane, actuator constraints, disturbance magnitudes and sensitivity to uncertainty. This use of the term controllability is in accordance with how Ziegler and Nichols [70] used the term.

In contrast, in much of control literature the term 'controllability' has only weak connections to the ease with which a plant can be controlled. A state x is termed 'controllable' if, for any initial state $x(0) = x_0$, any time $t_1 > 0$, and final state x_1, there exists an input $u(t)$ such that $x(t_1) = x_1$. We will use 'state controllability' to distinguish this interpretation of the term from the way we interpret it.

It can be perfectly feasible to control the outputs from a plant with some states which are not state controllable. It may also be difficult in practice to obtain acceptable control of a plant even if all states are state controllable. The clearest connection between controllability and state controllability is that any *unstable* state must be both state controllable and state observable, i.e., it must be possible to close feedback paths around all unstable states and thereby stabilize them. In addition, for decentralized control, an unstable state must not correspond to a *decentralized fixed mode* [63] for the chosen pairing of inputs and output. That is, the decentralized feedback structure must allow for feedback around an unstable state; state controllability and observability is not always sufficient to ensure this.

2 LIMITATIONS TO FEEDBACK PERFORMANCE

2.1 Universal Limitations - Constraints on S and T

For the closed loop system to follow setpoints and reject disturbances, we would like the sensitivity function S to be small. On the other hand, sensitivity to measurement noise is reduced if the complementary sensitivity function T is small. Clearly, these objectives can not both be met at the same frequency, since

$$S(s) + T(s) = I \tag{6}$$

Thus, at a given frequency one is forced to give priority to either setpoint following and disturbance rejection or to the sensitivity to measurement noise. Usually, setpoint following and disturbance rejection is more important at low frequencies, whereas sensitivity to measurement noise take priority at high frequencies. There will of course also be other reasons why the system bandwidth is finite, and T will roll off at high frequencies.

Another constraint on S is the well known *Bode Sensitivity Integral*. Assume that the open loop SISO transfer function $L = GK$ is stable and has at least two more poles than zeros. Then the following result from Bode [6] holds:

$$\int_0^\infty \ln |S(j\omega)| d\omega = 0 \tag{7}$$

This can be generalized to MIMO systems [21], giving

$$\int_0^\infty \ln |\det S(j\omega)| d\omega = 0 \tag{8}$$

Equations (7) and (8) may at first glance not appear to impose any strong restriction on the achievable control quality. Although improved control at one frequency (smaller S) must necessarily imply poorer control at other frequencies, it appears that S (or $\det S$) can be made only infinitesimally larger than unity over an infinite frequency range at high frequencies. Freudenberg and Looze [22] point out that this is naive; in practice bandwidth constraints and roll-off requirements force most of the positive area in Eqs. (7) and (8) to be around the bandwidth frequency. As a result, one must also expect a larger *peak* in the sensitivity function when S is made smaller in a certain frequency range.

The presence of RHP poles or zeros in the plant $G(s)$ cause modifications to Eqs. (7) and (8) above. These modifications are presented below in the relevant sections.

2.2 Plant Characteristics Limiting Feedback Performance

In this section we discuss inherent limitations to feedback performance, that is, fundamental limitations that are independent of the type of controller used. In the cases where the implications for decentralized and centralized control differ, such differences are pointed out.

A useful concept for understanding the limitations of feedback control is the *Internal Model Control* structure [29]. In Figure 3 both the conventional and the Internal Model Control (IMC) feedback structures are shown. With IMC, we use a model of the plant in parallel with the real plant, and the feedback signal is just the difference between the real plant and the model. The feedback signal thus arises only from errors in the model, unmeasured disturbances or measurement noise. This is intuitively appealing, if we have a perfect model and exact knowledge about all disturbances, we can control the plant outputs with feedforward control, and feedback is not necessary.

Note that the IMC structure and the conventional feedback structure are equivalent, the IMC controller Q and the conventional feedback controller K are related by

$$Q = K(I + GK)^{-1} \tag{9}$$

$$K = Q(I - GQ)^{-1} \tag{10}$$

From Fig. 3 we see that the nominal (i.e., assuming the plant model to be perfect, $G = P$) transfer function from reference r to plant output y is $y = Hr = PQr$. We also see that nominal stability is guaranteed if both the plant and the IMC controller are stable.

Perfect control can therefore be obtained by choosing $Q = G^{-1}$. Any phenomenon which makes it impossible to construct the plant inverse is therefore a cause for imperfect control. Such causes for imperfect control are:

1. The plant contains time delays, and it's inverse therefore contains predictive elements (G^{-1} is not causal). Perfect predictions about the future is obviously impossible.

2. The plant contains RHP transmission zeros which will become poles in the inverse. The transfer function from reference (or disturbance) to controller output therefore becomes unstable, and the controller output will increase until some physical constraint in the manipulated variable is met.

3. Constraints in the manipulated variables. If the *magnitudes* of the manipulated variables are constrained, this will limit the magnitudes of reference signals or disturbances for which perfect control can be achieved, since $u = Q(r - d)$. On the other hand, if the *rates of change* of the manipulated variables are constrained, this will limit the frequency range for which perfect control can be achieved. Exactly how the frequency range of perfect control is limited will depend on the magnitudes and dynamics of the reference signals and disturbances. Both the magnitudes and the rates of will be constrained for most real systems, but it varies from case to case whether such constraints seriously limit the control quality obtained. Constraints in the manipulated variables

will also limit the region in the state space for which stability is achievable when the plant is open loop unstable[1].

4. A related cause for imperfect control is pole excess in the plant. The number of plant poles exceed the number of plant zeros for real plants. The inverse would therefore contain more zeros than poles, resulting in an improper controller. Infinite controller power would be required for implementation. A proper IMC controller can be obtained by adding a low pass filter (of sufficiently high order) to the controller.

5. If the model is inaccurate, perfect control is clearly not obtained by choosing the IMC controller to be the inverse of the model. Model uncertainty therefore limits achievable performance. Particularly disastrous effects on performance (and even stability) can result if the plant in addition is highly interactive.

It is well known that perfect control is obtained in the limit as the controller gain approaches infinity. This can be seen from the conventional feedback structure:

$$y = (I + GK)^{-1}[GKr + G_d d] \tag{11}$$

Clearly, as the controller gain approaches infinity ($\underline{\sigma}(GK) \to \infty$) we get that $y \to r$ and the control error $e = y - r \to 0$. We also have that $y = Gu + G_d d$. Then, as the controller gain approaches infinity and the control error approaches zero, we get that

$$e = y - r \to 0 \quad \Rightarrow \quad u = G^{-1}[r - G_d d] \tag{12}$$

Thus one can implicitly obtain the inverse of the plant as the closed loop controller transfer function, and thereby also obtain perfect control, without an exact plant model. It may therefore appear that cause 5 for imperfect control above is not a fundamental restriction on control performance. However, high gain can only be used to achieve (almost) perfect control at low frequencies, because of points 1 - 4 above. On the other hand, model uncertainty is not very important at high frequencies where points 1 - 4 above make the loop gain low anyway. Thus, the main concern for model uncertainty is at intermediate frequencies, in the crossover region.

2.2.1 Manipulated variable constraints

Manipulated variable constraints can limit the ability to follow setpoint changes and reject disturbances. In the following, we will concentrate our discussion on disturbance rejection, but it is trivial to apply the same reasoning to setpoint following.

[1]IMC controller design for unstable plants require special considerations, see [47] for details.

We will here consider limitations in control performance caused by manipulated variable constraints *alone*, and thus assume that there are no other phenomena that limit achievable control performance for the plant.

Some of the relevant questions that can be asked in this context are:

1. What disturbances can be perfectly rejected?
2. For disturbances too large to be perfectly rejected, what is the minimum achievable offset in the controlled variable?
3. For what disturbances can acceptable control be achieved?

In order to answer these questions quantitatively, we must have a firm idea about what constitutes acceptable control, what the limitations in the manipulated variables are, and what disturbances can be expected. The transfer function matrices $G(s)$ and $G_d(s)$ should be properly scaled to reflect this knowledge, as explained in the introduction.

Question 1 is easiest to answer. We here assume that we have at least as many manipulated variables as controlled variables, otherwise it is clearly impossible to achieve perfect control of all controlled variables. For perfect control (and a setpoint of zero), we have

$$u = -G^{-1}G_d d \tag{13}$$

For the case with more manipulated than controlled variables, the pseudo-inverse has to be used in Eq. (13) above. Perfect control can be achieved provided $|u_i| \leq 1$. Thus, if the sum of the absolute values of the elements in row i of $G^{-1}G_d$ are less than one, perfect control can be achieved for controlled variable i. Large elements of $G^{-1}G_d$ indicate disturbances that are hard to reject, if $|[G^{-1}G_d]_{ij}| > 1$, disturbance j alone can cause imperfect control of controlled variable i.

Questions 2 and 3 are significantly harder to answer. Wolff et al. [64] formulated the corresponding optimization problems. For question 2 this is:

$$\begin{aligned} \max_d(\min_u \quad & \| \quad y \|_\infty) && (14) \\ \text{such that} \quad & \| \quad d \|_\infty \leq 1 \\ & \| \quad u \|_\infty \leq 1 \\ & y \quad = Gu + G_d d && (15) \end{aligned}$$

and for question 3 the corresponding optimization problem is:

$$\max_{d}(\min_{u} \quad \| \quad d \|_\infty) \tag{16}$$

$$\begin{aligned} \text{such that} \quad & \| \quad u \|_\infty \leq 1 \\ & \| \quad y \|_\infty = \| \, Gu + G_d d \, \|_\infty \leq 1 \end{aligned} \tag{17}$$

Here $\| \ x \ \|_\infty$ is the *infinity norm* of the vector x, i.e., the magnitude of the largest vector element. In solving these optimization problems, there is no requirement that there should be at least as many manipulated variables as controlled variables. Unfortunately, the optimization problems in Eqs. (14) and (16) are difficult to solve. Their solution is the subject of [8].

Skogestad and Postlethwaite [61] argue that the most insight is gained by considering one disturbance at the time, and derive the approximate requirement for achieving $\| \ y \ \|_\infty \leq 1$:

$$\sigma_i(G) \geq |U_i^H g_d| - 1 \ \text{ at frequencies where } |U_i^H g_d| > 1 \tag{18}$$

Here $\sigma_i(G)$ is the i'th singular value of G and U_i the corresponding output singular vector, and g_d is the column of G_d corresponding to the disturbance being considered. Because of simplifications in the derivation of Eq. (18), the results may be off by a factor of at mosr $\sqrt{ml}$, where m is the number of outputs and l is the number of manipulated variables. However, the results of Eq. (18) are normally much more accurate. Equation (18) can be used to assess:

1. For which disturbances and which frequencies input constraints may cause problems. This may give ideas to which disturbances should be reduced, for example by redesign or the use of feedforward control.

2. In which input direction i the plant gain i is too small. By looking at the corresponding input singular vector V_i, one can determine which actuators should be redesigned to get more power in certain directions.

3. By looking at the output singular vector U_i, one can determine for which outputs one may have to reduce one's performance requirements.

If , in addition to considering each disturbance separately, one only looks at steady state, the problem of checking whether the input constraints allow for acceptable control becomes much easier. This is because all variables have to be real, and the worst possible disturbance has to be $d = 1$. The problem is then to find

$$u^* = \min_u \| \ u \ \|_\infty \quad \text{such that} \quad \| \ Gu + G_d \cdot 1 \ \|_\infty \leq 1 \tag{19}$$

and the requirement for achieving acceptable control is that $u^* \leq 1$. Equation(19) may be formulated as a linear programming problem, and can hence be readily solved.

It should be noted that the effects of the disturbances are not additive, because of the constraints. Acceptable control may be achieved independently for two disturbances, but their combined effect may still exceed the limits for acceptable control. On the other hand, disturbances may also cancel each other, giving acceptable control for cases where this is not possible for the individual disturbances. In some sense, it is a matter of design philosophy whether disturbances should be considered separately or jointly - it may be argued that several independent disturbances are unlikely to obtain their worst possible value at the same time, and that designing for such a scenario would lead to overdesign of the plant. The opposing view would be that it is imprudent to rely on chance or nature's good will for acceptable plant performance, and that the worst case combination of disturbances therefore should be considered.

2.2.2 Time delays

The effect of deadtime on achievable performance is addressed in [29, 52]. We will here give a brief summary of the results of Holt and Morari [29], who address the effect of achievable performance within the IMC framework. Other factors than time delays that may limit achievable performance are not considered in this section.

A lower bound for the settling time for output i is given by

$$\tau_i = \min_j p_{ij} \tag{20}$$

where p_{ij} is the time delay in element g_{ij}.

Clearly, τ_i is the minimum time necessary for any input to affect output i. It will often not be possible to achieve a settling time of τ_i for all outputs i, but the τ_i's can serve as measures against which a control system can be judged.

The minimum necessary closed loop time delays that are achievable and allow for decoupled control of all outputs, can be shown to be:

$$\rho_j = \max_i(\max(0, \hat{q}_{ij} - \hat{p}_{ij})) \tag{21}$$

where $\hat{q}_{ij}$ is the minimum delay in the denominator of element ij of G^{-1}, and $\hat{p}_{ij}$ is the minimum delay in the numerator of element ij of G^{-1}. Holt and Morari propose that the above ρ_j is viewed as a lower bound on achievable performance, since it will often be possible to achieve faster responses in some of the outputs if dynamic interactions are allowed. Dynamic decoupling is seldom a reasonable requirement. Unless there is large differences in the importance of the outputs, it is reasonable to reduce the offset from setpoint in one output

at the cost of introducing modest offsets in other outputs. If there are large differences in the importance of the outputs, *one-way* decoupling may be justified, whereas Eq. (21) gives the minimum closed loop time delays for *two-way* decoupling.

The only case when the upper and lower bounds on achievable performance due to time delays are equal for all outputs (i.e., $\tau_i = \rho_i \; \forall i$), occurs when the rows and/or columns of G can be rearranged such that for each row in G the element with the smallest time delay is on the main diagonal. The time delays along the main diagonal of the rearranged G will then be the best possible settling times for the corresponding outputs, and decoupled control will be achievable with these settling times.

Increasing time delays may improve controllability!

Holt and Morari argue that whilst it is often difficult to reduce time delays in a plant, it may be possible to increase time delays, for instance by increasing pipe lengths between process units. They give an example of a heat exchanger network, for which a simplified transfer function matrix involving only deadtimes is

$$G(s) = \begin{bmatrix} e^{-6s} & e^{-11s} & e^{-2s} \\ e^{-11s} & 1 & e^{-12s} \\ e^{-8s} & e^{-13s} & 1 \end{bmatrix} \tag{22}$$

and the minimum delays for a dynamically decoupled system are

$$\rho = [e^{-6s}, 1, e^{-4s}] \tag{23}$$

By increasing the deadtime between two heat exchangers (and thereby increasing the time delay in some off-diagonal elements of $G(s)$), the new matrix of delays becomes

$$G(s) = \begin{bmatrix} e^{-6s} & e^{-11s} & e^{-6s} \\ e^{-11s} & 1 & e^{-16s} \\ e^{-8s} & e^{-13s} & 1 \end{bmatrix} \tag{24}$$

and the corresponding minimum delays for a dynamically decoupled system are

$$\rho = [e^{-6s}, 1, 1] \tag{25}$$

showing that faster control is possible with increased delays in som elements. The reason why faster control is possible with increased delays, is that in the original matrix of delays (Eq. (22) element (1,3) has a delay than is 4 units smaller than the delay in element (1,1). In order to achieve a decoupled response, the controller therefore has to delay the control action in input 3 by 4 time units.

Changes in delays may also affect the degree of interaction in a control system, and can therefore affect the control properties of the plant even if we don't require dynamic decoupling. Consider the following example

$$G = \begin{bmatrix} 1 & (1-\epsilon)e^{-\theta s} \\ 1 & 1 \end{bmatrix} \tag{26}$$

As $\epsilon \to 0$ the matrix G approaches singularity at steady state. With $\epsilon = 0.01$, we get a condition number of 399 and an RGA value $\lambda_{11} = 100$. If we are able to choose the deadtime such that g_{12} is -0.99 at the desired bandwidth frequency, we get a condition number of 1.01 and RGA value $\lambda_{11} = 0.5$, and thus a plant that is only moderately interactive at the bandwidth.

Further implications of time delays for decentralized control.
The results on achievable performance in the presence of time delays hold for full (centralized) controllers. If one restricts the controller to be decentralized even the lower bound on settling times in Eq. (21) may not be achievable - and the responses will certainly not be decoupled. Clearly, for decentralized control the minimum time required for a change in setpoint i to affect output i will be the time delay in g_{ii} (assuming inputs and outputs are rearranged to bring the paired elements to the diagonal). If the time delay in g_{ii} is θ_{ii}, an approximate upper bound on the bandwidth of *loop* i is $\omega_{Bi} < 1/\theta_{ii}$ (see e.g. [61]).

2.2.3 Right half plane transmission zeros

In this section we will consider the implications of Right Half Plane (RHP) zeros for feedback control. It is assumed that the reader is familiar with the concept of RHP zeros in SISO systems. Nevertheless, this section initially goes into some details about the implications of RHP zeros in SISO systems, for the following two reasons:

- The individual loops of a decentralized control system are usually required to be stable and perform acceptably on their own. The implications of monovariable RHP zeros are therefore relevant also for multivariable, decentralized control systems.
- It will serve as a background for the presentation Right Half Plane Transmission zeros, the MIMO counterpart of RHP zeros in SISO systems.

The implications of RHP transmission zeros are addressed in the latter part of this section. Although MIMO RHP zeros were discovered some time ago [55], we believe that this phenomenon and it's implications for control are less well known among control practitioners in industry.

This section does not attempt an exhaustive review of the literature on RHP zeros. More details can be found in, e.g., [28, 22, 47, 18, 13, 69]. Also, we will not go into details of how to compute the zeros of a transfer function (matrix), since this can be done by standard software.

Right Half Plane Zeros in SISO systems

Here we consider a SISO transfer function $g(s)$. The zeros of $g(s)$ are the roots of its numerator polynomial, and a zeros is a right half plane zero if its real part is positive. Some of the properties of plants with RHP zeros are:

1. If z_R is a a RHP zero of the plant transfer function $g(s)$, there exists an input $u(t)$ of the form e^{zt} and a set of initial conditions such that the output $y(t) = 0$ for $t = 0$. [41]

2. For a stable plant with an odd number of RHP zeros displays an inverse response to a step change in the input [28].

3. The zeros are invariant under state and output feedback [39].

4. The zeros are the poles of the plant inverse.

Properties 3 and 4 show that RHP zeros pose fundamental problems for feedback systems, and that perfect control is not achievable for plants with RHP zeros.

A more quantitative result on the effect of RHP zeros on feedback control is due to Freudenberg and Looze [22], who found that for a system that is open loop stable, with an RHP zero at $z = x + jy$ (and $x > 0$), the sensitivity function must satisfy

$$\int_{-\infty}^{\infty} \log|S(j\omega)| \frac{x}{x^2 + (\omega - y)^2} d\omega = 0 \tag{27}$$

This means that if $\log|S(j\omega)|$ is plotted as a function of frequency ω, the area under the line $|S| = 1$ (i.e., $\log(|S|) = 0$) must be "balanced" with the area above the line $|S| = 1$. The area above $|S| = 0$ must come mainly at frequencies below $\omega = |z|$, since the weight $x/(x^2 + (\omega - y)^2)$ decreases rapidly for higher frequencies. Therefore, if we want to improve control (make $|S|$ smaller) in a certain frequency range, we must also accept poorer control at other frequencies below $\omega = |z|$.

Thus RHP zeros impose bandwidth limitations on the closed loop system. Skogestad and Postlethwaite [61] use several different design approaches to illustrate that a reasonable, approximate constraint on the open loop crossover frequency ω_c for a real RHP zero at $s = z$ is

$$\omega_c < z/2 \tag{28}$$

RHP zeros in MIMO systems

Several different definitions of zeros of multivariable have been proposed. The definition that has become commonly accepted is that of a *Transmission Zero.*

Definition of Transmission zeros. Transmission zeros are those values of s for which the rank of $G(s)$ drops below its normal value [41].

This definition of transmission zeros show that transmission zeros are relatively uncommon for non-square plants, unless they are associated with a particular subsystem of the plant, such as a particular output (for plants with more inputs than outputs) or a particular input (for plants with more outputs than inputs). This is because, if r is the normal rank of $G(s)$, all minors of order r are unlikely to loose rank at the same value of s.

There are alternative definitions of transmission zeros, (see e.g., [55, 41, 38]). The difference between these definitions are of no practical importance with regard to the control properties of a plant with transmission zeros, although they point to different approaches for calculating transmission zeros.

Note that there is not necessarily any direct connection between zeros in the individual elements of G and transmission zeros in G. Zeros in the individual elements of G do not imply transmission zeros (unless *all* the elements in a row or column of G have a zero for the same value of s), and $G(s)$ may have a transmission zero at a value of s where none of the elements have a zero.

Another difference between monovariable zeros and multivariable transmission zeros, is that a multivariable plant can have poles and zeros for the same value of s.[2] If $G(s)$ have a pole and a transmission zero at the same value for s, one singular value of s will be infinite and another singular value will be zero at this value for s. This can cause problems if one tries to determine the transmission zeros of a square plant by setting the determinant to zero. However, the only transmission zeros that have implications for the achievable control performance of the plant are those in the right half plane. For open loop stable plants, there can be no cancellation of poles and RHP zeros in the determinant, and the RHP transmission zeros are the roots of the numerator of the determinant in the RHP.

Some of the properties of RHP transmission zeros are:

- Let z be a transmission zero of $G(s)$. Then there exists an input of the form $ue^{zt}l(t)$, where u is a complex vector and $l(t)$ is the Heaviside unit step function, and a set of

[2] This can happen even for a minimal realization of the plant transfer function matrix, i.e., even if all states are both controllable and observable.

initial conditions such that $y(t) = 0$ for $t > 0$ [41].

- Consider a square plant $G(s)$ of dimension $n \times n$, and let $n-1$ of the outputs be perfectly controlled by $n-1$ of the inputs[3]. The transmission zeros in $G(s)$ will then appear in the scalar transfer function between the unused input and the uncontrolled output [28] (see also Eq. (32)). Thus, if $G(s)$ has an odd number of RHP transmission zeros, there will be an inverse response from the unused input to the uncontrolled output.
- The transmission zeros are invariant under feedback [39].
- The transmission zeros are the poles of the plant inverse.

The similarities between the properties of monovariable zeros and transmission zeros are striking. The main differences are that the "transmission blocking" property hold for a multivariable as opposed to a monovariable signal, and that perfect control of other outputs is required to obtain an unambiguous manifestation of the inverse response associated with RHP zeros.

Zero direction. If the plant $G(s)$ has a zero at $s = z$, then $G(s)$ looses rank at $s = z$. Thus, if we perform an SVD of $G(z)$, we will find that one singular value is zero. The output singular vector corresponding to the zero singular value is termed the zero output direction, and is here denoted by y_Z. Similarly, the input singular vector corresponding to the zero singular value is termed the zero input direction, and is here denoted by u_Z. Clearly, then we have

$$G(z)u_Z = 0; \qquad y_Z^H G(z) = 0 \tag{29}$$

We will be primarily interested in the zero output direction y_Z. In order to preserve internal stability of the feedback system, the controller K cannot cancel the zero in G, and we get (e.g., [61]).

$$y_Z^H T(z) = y_Z^H G(z)K(z)(I + G(z)K(z))^{-1} = 0 \tag{30}$$

$$y_Z^H S(z) = y_Z^H (I - T(z)) = y_Z^H \tag{31}$$

Thus, not only the location of zeros cannot be changed by feedback control, but also the zero directions are invariant under feedback control. The effect of an RHP transmission zero must therefore appear in (one or more of the) outputs corresponding to nonzero elements in y_Z. This was first observed by Bristol [11], who termed zeros whose zero direction vector y_Z

[3] This assumes that it is possible to obtain perfect control of the $n-1$ outputs using the $n-1$ inputs, which implies that the transfer function matrix from the chosen inputs to the chosen outputs have no RHP transmission zero.

have some elements equal to zero *pinned* zeros, since such zeros are "pinned" to the outputs corresponding to nonzero elements in y_Z. Pinned zeros are quite common in practice. If there is an inverse response or a time delay in the measurement of a plant output, this effect cannot be moved to another output. Similarly, if u_Z contains some zero elements, the zero is "pinned" to the inputs corresponding to the nonzero elements of u_Z.

Moving the effect of an RHP transmission zero to a specific output. Consider a transfer function matrix $G(s)$ that is square and of full rank in most of the complex plane, but has a single RHP transmission zero at $s = z$ and no pole at $s = z$. Assume that the RHP zero at $s = z$ is the only restriction of the performance under feedback control[4]. Let the k'th element of y_Z be nonzero. Then it is possible to obtain "perfect" control for all outputs $j \neq k$, with no steady state offset in output k. One possible choice for the complementary sensitivity function $T(s)$ is:

$$T(s) = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & & & & & \\ \frac{\beta_1 s}{s+z} & \cdots & \frac{\beta_{k-1} s}{s-z} & \frac{-s+z}{s+z} & \frac{\beta_{k+1} s}{s-z} & \cdots & \frac{\beta_n s}{s-z} \\ \vdots & \vdots & & & & & \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \tag{32}$$

where

$$\beta_j = -2\frac{y_{Zj}}{y_{Zk}} \quad \text{for } j \neq k \tag{33}$$

It can easily be verified that $y_Z^H T(z) = 0$. For a complete proof of the above result see [47].

Implications of RHP transmission zeros on the sensitivity function.

Generalizations of the sensitivity integral relation in Eq. (27) to multivariable systems can be found in [13, 69]. One result from Zhou [69] provide a simple illustration that the implications of RHP zeros for MIMO systems are similar to the implications for SISO systems:

Assume that the open loop transfer function $L = GK$ is proper[5]. Then, for any RHP transmission zero $z = x + jy$

$$\int_{\infty}^{\infty} \ln \bar{\sigma}(S(j\omega)) \frac{x}{x^2 + (\omega - y)^2} \geq \pi \ln \bar{\sigma}(S(z)) \tag{34}$$

[4]This implies that $G^{-1}(s)$ is proper and causal.

[5]The technical assumption that must be satisfied is

$$\lim_{R \to \infty} \max_{\phi \in [-\pi/2, \pi/2]} \frac{|\ln \bar{\sigma}(S(Re^{j\phi}))|}{R} = 0$$

Note that since $y_Z^H S(z) = y_Z^H$, $\ln \bar{\sigma}(S(z)) \geq 0$. In the same way as for SISO systems, the frequency range where $\bar{\sigma}(S(j\omega)) > 1$ must be mainly at frequencies $\omega < |z|$.

Alignment between zero directions and disturbances. The closed loop performance will be poor if the direction of the disturbance is aligned with the zero direction. If the plant $G(s)$ has a RHP zero at $s = z$, a requirement for acceptable disturbance rejection is [61]:

$$|y_Z^H g_d(z)| < 1 \tag{35}$$

where g_d is the column of G_d corresponding to the disturbance in consideration.

2.2.4 Limitations imposed by RHP poles

The implications of open loop RHP poles on the closed loop properties of the system will first be presented for SISO systems, and then the generalizations to MIMO systems will be made. It is assumed that the concept of an RHP pole (unstable mode) of a system is familiar to the reader both for the SISO and MIMO cases.

RHP poles in SISO systems. When the system contains RHP poles, Bode's Sensitivity Integral relationship must be modified. In addition to the assumption that the open loop transfer function has at least two more poles than zeros, assume that the open loop transfer function has N_p RHP poles (including multiplicities) at locations p_i. Freudenberg and Looze [22] found that the sensitivity function must satisfy

$$\int_0^\infty \ln |S(j\omega)| d\omega = \pi \sum_{i=1}^{N_p} Re\{p_i\} \tag{36}$$

for the closed loop system to be stable. Thus, the RHP poles increase the "positive area" in Eq. (36) by an amount proportional to the distance of the pole from the imaginary axis.

Another constraint imposed by an RHP pole, is that for internal stability of the control system, we must require $T(p) = 1$ if p is a RHP pole of the open loop system. This, together with the maximum modulus theorem, can be used to show that

$$\| w_T T \|_\infty \geq |w_T(p)| \tag{37}$$

Choosing $w_T = 1$, we find that $\| T \|_\infty \geq 1$, which shows that control is needed to stabilize an unstable plant, since $K = 0$ (no control) implies $T = 0$. Eq. (37) is used by Skogestad and Postlethwaite [61], together with a reasonable weight $w_T(s)$ to argue that for a real RHP pole at $s = p$ one must expect an open loop crossover frequency $\omega_c > 2p$ in order to achieve acceptable control.

RHP poles in MIMO systems. In the same way as for RHP transmission zeros, RHP poles also have *directions*. If $s = p$ is an RHP pole of the plant G, the input and output

pole directions can be found from a singular value decomposition of $G(p)$, and are the input and output singular vectors corresponding to the infinite singular value of $G(p)$. We will be most interested in the *output pole direction*, which we will denote y_P. The SISO constraint $T(p) = 1$ for an RHP pole at $s = p$ for MIMO systems becomes

$$T(p)y_P = y_P \tag{38}$$

Note that a MIMO system may have a pole and a zero at the same value of s. Their directions are then always orthogonal. The alignment of the directions of the RHP poles and RHP transmission zeros are important when considering the combined effects of RHP poles and RHP transmission zeros, as will be shown in the next section.

The Bode Sensitivity Integral relation for MIMO systems with RHP poles is [22]:

$$\int_0^\infty \ln|\det S(j\omega)|d\omega = \pi \sum_{i=1}^{N_p} \mathrm{Re}\{p_i\} \tag{39}$$

A lower bound on the integral of $\bar{\sigma}(S)$ is given in [69], based on a factorization of the plant into a minimum phase part and an allpass part.

2.2.5 Combined effects of RHP poles and RHP zeros

SISO systems. For a SISO system with a RHP zero at $z = x + jy$ and N_p RHP poles at locations p_i, the sensitivity function must satisfy

$$\int_{-\infty}^\infty \ln|S(j\omega)|\frac{x}{x^2 + (\omega - y)^2}d\omega = \pi \ln \prod_{i=1}^{N_p} \left|\frac{z + \bar{p}_i}{z - p_i}\right| \tag{40}$$

In the same way as for open loop stable systems, most of the positive area must be at frequencies below $\omega = |z|$. In addition, the value of the integral is increased by the presence of open loop RHP poles, and increases further at the RHP zero approaches an RHP pole.

Skogestad and Postlethwaite [61] give lower bounds on the peak values of S and T. Let the open loop system have N_p RHP poles at locations p_i, and N_z RHP zeros at locations z_i. For any RHP zero z, define

$$c_1 = \prod_{i=1}^{N_p} \left|\frac{z + \bar{p}_i}{z - p_i}\right| \tag{41}$$

and for and RHP pole p define

$$c_2 = \prod_{i=1}^{N_z} \left|\frac{\bar{z}_i + p}{z_i - p}\right| \tag{42}$$

Then, lower bounds on the peaks of S and T are

$$\| S \|_\infty \geq c_1; \quad \| T \|_\infty \geq c_2 \tag{43}$$

Thus, not only the integral of $\ln S$, but also the peak values of S and T increase as RHP poles and zeros are moved closer.

In theory, any linear plant can be stabilized irrespective of the locations of its RHP poles and zeros. In practice, it is highly desirable that stable controllers are used. Youla et al. [66] showed that a linear plant can be stabilized by a stable controller if and only if every real RHP zero in $G(s)$ lies to the left of an even number (including 0) of real RHP poles. The presence of complex RHP poles or zeros does not affect this result. Thus, if we have a plant $G(s)$ with one real RHP zero at z and one real RHP pole at p, we must require that $z > p$ in order to be able to stabilize the plant with a stable controller. In [61] it is argued that in order to achieve acceptable performance, a more realistic requirement i $z > 4p$.

MIMO systems. For MIMO systems the directions of RHP poles and zeros are important for their combined effects on S and T. A generalization of Eq. (40) to MIMO systems can be found in [69]. Whereas the increase in the integral of $\ln |S(j\omega)|$ depends strongly on the directions of the zero and the poles, the frequency region in which most of this increase in sensitivity occurs depends only on the location of the zero.

The bounds on S and T in Eq. (43) can also be generalized to MIMO systems [61], and this generalization also show the importance of the pole and zero directions. For the case with only one RHP zero z and one RHP pole p, we get that

$$\| S \|_\infty \geq c_1; \qquad \| T \|_\infty \geq c_2 \tag{44}$$

where

$$c_1 = c_2 = \sqrt{\sin^2 \theta + \frac{|z+p|^2}{|z-p|^2} \cos^2 \theta} \tag{45}$$

and θ is the angle between the output directions of the pole and the zero.

3 ANALYSIS OF INTERACTION IN MULTIVARIABLE PLANTS

A fundamental difference between monovariable and multivariable plants is *interaction* in multivariable control systems. It is obvious that changing one plant input in a multivariable plant will affect several plant outputs. It may require a little reflection to understand that this means that the control of one output from a multivariable plant depends on how the other outputs of the plant are controlled. How strong this dependence on the control of the other outputs is, depends on the gains in the different *directions* of the plant, and how these directions are aligned to the plant inputs and outputs. What do we mean by the 'directions' of the plant? Intuitively, the plant inputs and outputs may appear to represent the natural directions of the plant. The eigenvectors of the plant transfer function matrix is another intuitively reasonable definition of the plant directions, in this case the eigenvectors would represent the corresponding gains. However, the most useful concept of plant directions is that resulting from the Singular Value Decomposition (SVD) of the plant transfer function matrix. We expect most readers to be familiar with the SVD, and we have already used it briefly in the preceding section. Details of its computation and mathematical properties can be found in standard textbooks on linear algebra, e.g. Golub and Van Loan [23]. Using the SVD, a transfer function matrix $G(s)$ can for a given value of s be decomposed into

$$G(s) = U(s)\Sigma(s)V^H(s) \tag{46}$$

where the superscript H represents the complex conjugate transpose. For a transfer function matrix with r outputs and c inputs, U has dimensions $r \times r$, Σ has dimensions $r \times c$ and V has dimensions $c \times c$. The column vectors of U are orthogonal and of unit length, and represent the *output directions* of the plant. Likewise, the column vectors of V are orthogonal and of unit length, and represent the *input directions* of the plant. Σ is a diagonal matrix, with the elements ordered by decreasing magnitude along the main diagonal. The elements on the main diagonal of Σ are called the *singular values* of G. The singular values are always real and positive. The i'th singular value is denoted σ_i, the largest singular value is denoted $\bar{\sigma}$, and the smallest singular value is denoted $\underline{\sigma}$. The gain from input direction i (i'th column of V) to output direction i (i'th column of U) is thus given by σ_i. The largest gain in any direction of a multivariable plant is given by $\bar{\sigma}$ and the smallest gain is given by $\underline{\sigma}$. Denoting the eigenvalues of G by λ_i, we have

$$\underline{\sigma}(G) \leq |\lambda_i(G)| \leq \bar{\sigma}(G) \qquad \forall i \tag{47}$$

Some advantages of the SVD over the eigenvalue decomposition for analyzing gains and directionality of multivariable plants are:

- The singular values give more representative information about the gains of the plant than the eigenvalues, as explained above. Only for hermitian matrices ($G = G^H$) are the SVD and eigenvalue decompositions equal. Consider for example the matrix

$$A = \begin{bmatrix} 0 & 100 \\ 0 & 0 \end{bmatrix}. \tag{48}$$

  Both eigenvalues of A are zero, but the largest singular value of A is 100. The largest singular value is the gain from input direction $[0\ 1]^T$ to output direction $[1\ 0]^T$, i.e., the input and output directions are different. In contrast, the eigenvalues only measure gains when the input and output directions are equal.

- The plant directions obtained from the SVD are orthogonal. In contrast, if one wishes to analyze plant gains and plant directionality using the eigenvalue decomposition, one will have to include the angle between the eigenvectors in the analysis.

- The SVD applies directly also for nonsquare plants, i.e., plants where the number of inputs and the number of outputs differ. For a plant transfer function matrix G of dimension $r \times c$, the number of singular values equals the smaller of r and c (assuming G to be of full rank). If $r > c$, columns $c+1$ to r of U then represents combinations of outputs that are unaffected by the inputs. Similarly, if $c > r$, columns $r+1$ to c of V represent combinations of inputs whose effect on the outputs exactly cancel, giving no change in the outputs.

If it had been possible to send in inputs to the plant perfectly aligned with the input directions (from the SVD), and project the measurements on the output directions, multivariable control design would in many cases reduce to a set of single loop designs. Unfortunately, U, Σ and V in general vary with frequency in ways that cannot be described by rational, realizable transfer functions, making it impossible to use the SVD directly in control design. Strictly speaking, the symbol s in Eq. (46) should therefore interpreted as a given complex number, and not as a continuous variable. There are however design approaches that utilize the insight gained from the SVD. One may approximate the SVD at some specific, important frequency with real matrices [37]. In other cases, knowledge about the physical structure of the plant can make it possible to use a slightly modified SVD for controller design in a very direct way [30, 33, 36]. With this introduction to the concept of interactions in multivariable plants, we will next introduce some tools for analyzing interactions.

3.1 The minimum singular value as an indicator of plant controllability

It follows from Eq. (18) that if $\underline{\sigma}(G)$ is small, this can be an indication that disturbances can be impossible to reject, because of input constraints. Similarly, a small $\underline{\sigma}(G)$ implies that

in some directions only small setpoint changes can be made, and that it may not always be possible to follow changes in the optimal operating point. These observations of course rest on the assumption that the plant G is appropriately scaled.

3.2 The Minimized Condition Number.

Consider a case where the plant gains in different directions are significantly different, but for which we have similar performance specifications for all plant directions. We will then want to have similar loop gains in all directions. Thus, we will want a much higher controller gain in the low gain direction of the plant than in the high gain direction. However, it is rarely possible to align the inputs to the plant exactly to the plant directions, for several reasons:

- In general the SVD of the plant cannot be represented by realizable transfer function matrices, as pointed out above.
- It is naive to expect the plant model $G(s)$ used for control design to be exact. For highly interactive plants the directions corresponding to high and low gains may vary significantly even for small variations in the elements of $G(s)$.
- It is also naive to expect the plant inputs to be exactly equal to the controller outputs, because of uncertainties in actuators. This can be particularly detrimental to the alignment of plant inputs to the plant directions.

For these reasons, we will not be able to align the plant inputs perfectly with the plant directions. The high controller gain that we want in the low gain direction of the plant will therefore 'miss', and a component of the high gain of the controller will in reality be in the high gain direction of the plant. Similarly, a component of the low gain direction of the controller will be in the low gain direction of the plant. The result may be that control performance is significantly different in the different directions of the plant, and in severe cases instability may result.

The additional difficulty with multivariable control design arises when the gain in different plant directions are not the same. A natural measure of this additional difficulty caused by the plant interactions is therefore the ratio of the largest to the smallest possible gains in the plant. This is known as the condition number of the plant, and is denoted by γ.

$$\gamma(G(s)) = \bar{\sigma}(G(s))/\underline{\sigma}(G(s)) \tag{49}$$

One problem with the condition number is that it depends on the *scaling* of the plant G. This means that γ depends on the units used for measuring the plant inputs and outputs. This is obviously unreasonable - the condition number could for instance show that it is simpler

to control a plant if pressure is measured in bar than if pressure is measured in Pascal. To avoid this sort of ambiguity, we may use the minimized condition number γ^*, which is found by minimizing the condition number by pre- and postmultiplying G by diagonal, real scaling matrices, D_1 and D_2.

$$\gamma^* = \min_{D_1, D_2} \gamma(D_1 G D_2) \tag{50}$$

The minimized condition number of a transfer function matrix G of dimension $n \times n$ can be calculated from [7, 14]

$$\gamma^*(G(j\omega)) = \inf_{D_1, D_2} \bar{\sigma}^2 \left(\begin{bmatrix} D_2^{-1} & 0 \\ 0 & D_1 \end{bmatrix} \begin{bmatrix} 0 & G(j\omega)^{-1} \\ G(j\omega) & 0 \end{bmatrix} \begin{bmatrix} D_2 & 0 \\ 0 & D_1^{-1} \end{bmatrix} \right) \tag{51}$$

The minimization in Eq. (51) is the same minimization used for calculating the upper bound on the structured singular value (μ). This minimization is convex, and can be performed with commercially available software (e.g., [1]). As indicated in Eq. (51), a new optimization has to be performed for each frequency ω of interest.

Example

Consider a plant with the steady state gain matrix

$$G(0) = \begin{bmatrix} 100 & 0 \\ 0 & 1 \end{bmatrix} \tag{52}$$

Here we get $\gamma(G(0)) = 100$, which would indicate severe interactions at steady state. This is obviously erroneous, since $G(0)$ is diagonal, and thus completely decoupled (non-interacting) at steady state. Of course, it is easy to scale $G(0)$ to minimize $\gamma(G(0))$:

$$D_1 G(0) = \begin{bmatrix} 1 & 0 \\ 0 & 100 \end{bmatrix} \begin{bmatrix} 100 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 100 & 0 \\ 0 & 100 \end{bmatrix} \tag{53}$$

The minimized condition number, $\gamma^*(G(0)) = 1$, indicating a non-interactive plant.

3.3 The Relative Gain Array

The Relative Gain Array (RGA) was introduced by Bristol in 1966 [9], and has since been widely applied in chemical process control for analysis of interactions and for pairing plant inputs and outputs for decentralized control. References to the use of the RGA in the literature are too numerous to cover, some worth mentioning are [10, 44, 58, 24, 62, 59, 31, 32]. The RGA is defined as

$$\Lambda(G(s)) = G(s) \times (G^{-1}(s))^T \tag{54}$$

where $\times$ represents element-by-element multiplication (Hadamard or Schur product). It can be shown (e.g., [24]) that the ij'th element of Λ, λ_{ij}, is the ratio of the open loop gain from

input j to output i when all other outputs are uncontrolled, to the gain from input j to output i when all other gains are perfectly controlled. The RGA therefore measures how the loops interact. Perfect control can of course only be achieved at steady state, and therefore the interpretation of the RGA as the ratio of open loop to closed loop gain has lead many authors to discard the use of the RGA at frequencies other than zero. This is unfortunate, as the RGA as a function of frequency can give useful information and is easily computed.

Some useful mathematical properties of the RGA are:

- It is independent of scaling of the inputs and outputs of the matrix G.
- Rearranging the order of rows and columns of G (corresponding to different pairings of inputs and outputs for decentralized control), only results in the same rearrangement in the order of rows and columns in $\Lambda(G)$.
- All row and column sums of $\Lambda(G)$ equal 1.

In this section we will concentrate on the use of the RGA for analysis of interactions and as an indicator of robustness problems. The use of the RGA for selecting pairings of inputs and outputs for decentralized control will be addressed later.

3.3.1 The RGA and the minimized condition number

Plants with large elements in $G(j\omega)$ will also have a large minimized condition number at the same frequency ω. Nett and Manousiouthakis [48] show that

$$2\max\{\| \Lambda(G(j\omega)) \|_1, \Lambda(G(j\omega)) \|_\infty\} \leq \gamma^*(G(j\omega)) + \frac{1}{\gamma*(G(j\omega))} \tag{55}$$

where $\| \cdot \|_1$ is the induced 1-norm (maximum sum of absolute values of the elements of one column in the matrix), and $\| \cdot \|_\infty$ is the induced infinity norm (maximum sum of absolute values of the elements of one row in the matrix).

Thus, since the RGA is independent of scaling of inputs and outputs, large elements in $\Lambda(G)$ imply a large $\gamma*(G)$, and we know that the plant is ill-conditioned without having to perform the optimization in Eq. (51).

3.3.2 The RGA and individual element uncertainty

Theorem 1 *The (complex) matrix G becomes singular if we make a relative change $-1/\lambda_{ij}$ in it's ijth element, that is, if a single element of G is perturbed from g_{ij} to $g_{Pij} = g_{ij}(1-\frac{1}{\lambda_{ij}})$.*

Theorem 1 was originally proven by [67], and a much simpler proof is given in [31].

Some implications of this theorem for control are:

- Element uncertainty. Consider a plant with transfer function matrix $G(s)$. If the relative uncertainty in an element at a given frequency is larger than $|1/\lambda_{ij}(j\omega)|$ then the plant may have $j\omega$-axis zeros and RHP zeros at this frequency. However, the assumption of individual element uncertainty is often a poor one from a physical point of view because the elements are normally coupled in some way.

- Process identification. Models of multivariable plants $G(s)$ are often obtained by identifying one element at the time, for example, by using step or impulse responses. From Theorem 1 it is clear that such methods for process identification are very unreliable if there are large RGA elements within the bandwidth where the model is intended to be used. Useless models (e.g., with wrong sign of $det(G(0))$ or non-existing RHP zeros) can easily result. Consequently, identification must be combined with physical knowledge if a good multivariable model is desired in such cases.

- Uncertainty in the state matrix. Consider a stable linear system written on state space form; $dx/dt = Ax + \cdots$. Then changing the ijth element in A from a_{ij} to $a_{ij}(1 - 1/\lambda_{ij}(A))$ yields one eigenvalue of A equal to zero. Thus, we may conclude that systems with large RGA-elements of A will become unstable for small relative changes in the elements of A. Note that the RGA only gives the magnitude of the relative perturbation necessary to make one eigenvalue of A equal zero, even smaller perturbations may cause a complex conjugate pair of eigenvalues to cross the imaginary axis.

3.3.3 The RGA and diagonal input uncertainty

We mentioned above that it is naive to expect the plant inputs to be exactly equal to the controller outputs, any control system should be designed to tolerate some uncertainty in the inputs. Let the nominal plant model be $G(s)$, and the true (perturbed) plant be $G_p = G(I+\Delta)$. $\Delta = \text{diag}\{\Delta_i\}$ is a matrix consisting of the relative uncertainty (error) in the gain of each input channel. If an "inverse-based" controller (decoupler) is used, $K(s) = G^{-1}(s)C(s)$, where $C(s)$ is a diagonal matrix, then the true open loop gain G_pK is

$$G_pK = (I + G\Delta G^{-1})C \tag{56}$$

The diagonal elements of $G\Delta G^{-1}$ are directly given by the RGA [59]:

$$(G\Delta G^{-1})_{ii} = \sum_{j=1}^{n} \lambda_{ij}(G)\Delta_j \tag{57}$$

Thus, if the plant has large RGA elements and an inverse-based controller is used, the overall system will be extremely sensitive to input uncertainty.

Control implications. Consider a plant with large RGA-elements in the frequency range of importance for feedback control. A diagonal controller is robust (insensitive) with respect to input uncertainty, but will be unable to compensate for the strong directionality of the plant, even for the nominal plant model. On the other hand, an inverse based controller may yield excellent performance for the nominal plant $G(s)$, but will be very sensitive to input uncertainty. Applied to the true plant $G_p(s)$ the inverse-based controller must therefore be expected to perform poorly.

3.4 Diagonal Dominance

The concept of *diagonal dominance* is used both for analyzing interactions and for design of non-decentralized compensators. The work in this field is based on Gershgorin's theorem:

Theorem 2 *The eigenvalues of a $n \times n$ matrix A are contained within the union of n circles in the complex plane. The centers of these circles are located at*

$$a_{ii}, \qquad i = 1, \cdots, n \tag{58}$$

and the radii of the circles are given by

$$r_i = \sum_{j \neq i} |a_{ij}| \tag{59}$$

Alternatively, the sum of magnitudes of the offdiagonal elements in *column i* can be used for the radius of circle i, which should be obvious since A and A^H have the same eigenvalues. One has to sum either column-wise for all circles or row-wise for all circles to find the radii, the two ways of finding the radii cannot be mixed.

If $|a_{ii}| > r_i \ \forall i$, the matrix A is called "diagonally dominant", either "row dominant" or "column dominant", depending on how the radii are calculated. The concept of diagonal dominance can be used to design compensators such that the compensated plant is diagonally dominant. The interested reader is referred to [56, 42], such techniques are not explained here. Economou and Morari [19] propose the use of the ratio $|g_{ii}|/r_i$ as a measure of interaction in a transfer function matrix G. This is what is known as the IMC Interaction Measure, and can be plotted as a function of frequency. Interaction cannot cause instability if

$$|\tilde{t}_i(j\omega)| < |g_{ii}(j\omega)|/r_i \quad \forall i, \forall \omega \tag{60}$$

Hence, if $|g_{ii}|/r_i > 1 \ \forall \omega$, there is no bandwidth limitation in loop i caused by interactions.

Alternatively one may plot the "Gershgorin bands" of $G(j\omega)$, i.e., plots of the loci of $g_{ii}(j\omega)$ in the complex plane, with circles of radius $r_i(j\omega)$ superimposed. If the Gershgorin

bands include the origin, bandwidth limitations because of interactions may occur. However, it is easier to see from plots of IMC interaction measures for what range of frequency interaction is a problem.

3.4.1 The structured singular value interaction measure

Let $\hat{G}$ be a block diagonal matrix with blocks along the diagonal $\hat{G}_i$ that are equal to the corresponding blocks of the plant G. Similarly, let $K = \text{diag}\{K_i\}$ be a block diagonal matrix of controllers, each controller of compatible dimension with the corresponding diagonal block of $\hat{G}$. Assume that $\hat{G}$ and G contain the same number of RHP poles [6], and let K_i be a stabilizing controller for $\hat{G}_i$, that is, $\hat{T} = \hat{G}K(I + \hat{G}K)^{-1}$ is stable. Grosdidier and Morari (1986) found that the overall system is stable if

$$\bar{\sigma}(\hat{T}(j\omega)) < \mu^{-1}(E(j\omega))\forall\omega \tag{61}$$

where $E = (G - \hat{G})\hat{G}^{-1}$. The value of μ is computed with respect to the structure of $\hat{T}$. The measure $\mu(E)$ is known as the Structured Singular Value Interaction Measure (SSV-IM). The result in Eq. (61) follows easily from the Multivariable Nyquist Theorem and the definition of the structured singular value, by observing that

$$(I + GK) = (I + E\hat{H})(I + \hat{G}K) \tag{62}$$

Eq. (61) provides the tightest possible norm bound on $\hat{T}$, in the sense that if $\bar{\sigma}(\hat{T}) > \mu^{-1}(E)$ then there exists another system $\hat{T}'$ such that $\bar{\sigma}(\hat{T}') = \bar{\sigma}(\hat{T})$ corresponding to an unstable overall system T.

It is clear that if the controller K has integral action, then $\hat{T}(j0) = I$, and in order to guarantee stability with integral action we must require $\mu(E) < 1$. Furthermore, in order to guarantee stability of the overall system, the bandwidths of the individual subsystems must be constrained to the frequency range for which $\mu(E) < 1$.

The SSV-IM is flexible since it does not require the control system to be fully decentralized (single loops only), multivariable subsystems can easily be included in the analysis. However, it provides a single bound that applies to all subsystems. In [25] it is indicated how to use weights to give preference to specific subsystems.

Like other interaction measures that give guarantees for stability of the overall system, the SSV-IM is conservative. It is often possible to design decentralized control systems that perform satisfactorily even though the bound in Eq. (61) is violated.

[6] Normally, this assumption only holds if G is stable.

4 SELECTION OF INPUTS AND OUTPUTS FOR CONTROL

In this section we will give guidelines on how to select inputs and outputs to be used for controlling a plant. We emphasize that there is no substitute for physical understanding of the plant, the engineer needs a clear understanding of what is required of the control system and how the different parts of the plant interact.

In the early stages of plant design, the numbers and locations of measurements and actuators can be changed relatively easily, and the number of decisions a control engineer has to make can be enormous. In the late stages of design, and for plants that have already been constructed, changes to the instrumentation are costly and should if possible be avoided. This difficulty of changing the instrumentation in simplifies the task of the control engineer, in the sense that the number of decisions that has to be made is reduced. On the other hand, the control engineer is then also constrained by the decisions that already have been made (whether these decisions are made consciously or not), and the quality of control that can be obtained may suffer from bad decisions that already have been made. The engineer must then argue that the plant design imposes fundamental and severely restrictive limitations on the control quality that can be obtained (because of, e.g., time delays, RHP transmission zeros or interactions), and that design modifications are therefore required.

Traditionally the control engineer gets involved relatively late in the design of a plant, and many decisions have then been made - often without considering how such decisions affect the control of the plant. Several authors have argued that control should be considered throughout the design project (e.g., [70, 20]), but the control engineer then needs effective tools (in addition to physical understanding) to aid in making the numerous decisions in the early stages of plant design.

The quality of the information about the plant that is available to the control engineer also varies depending on what stage of the design or operational life the plant is in. Below we will first give a qualitative description of the role of the control engineer in the different stages of design and operational life, and thereafter present some tools that can be used to determine which inputs and outputs should be used for control. The qualitative description probably reflects most accurately the typical situation in chemical process control, but the tools presented have general applicability.

4.1 The Early Stages of Design

A rough sketch of the plant, showing only the main components of the plant, is often all that is available at the early stages of design. A steady state model of the plant is also often available at a relatively early stage of the design. In cooperation with the design engineers,

one may use physical understanding to identify variables that are important for plant operation and safety or product quality. Where possible, such variables should be measured directly. If some variables cannot be measured directly, or the available measurement is slow or unreliable, one should search for secondary measurements that are easily measurable that have a strong relationship with the primary variable. A typical example of this is distillation, where temperatures are often used for control although the product compositions are the variable that is relevant for product quality. One may then either use temperatures to estimate product compositions [46], or control the temperatures directly and update the temperature setpoint if the product compositions are found to be off specifications.

One should also consider which variables relating to plant safety need to be continuously controlled and which variables will be left uncontrolled during normal operation. For example, vibration in a pump is normally not controlled continuously, but if the vibration exceeds an alarm limit some remedial action is taken - possibly shutting down the plant. We will subsequently only consider measured variables for which continuous control is found to be necessary, and the term "plant outputs" will only refer to such variables.

For each of the plant outputs that should be controlled independently, there must be at least one plant input (manipulated variable) which has an effect on the plant output. The number of independent plant inputs should allow for assigning at least one plant input to the control of each of the plant outputs, in such a way that a plant input has an effect on the output to which it is assigned. This will ensure that the plant outputs can be controlled independently. Such "degrees of freedom analysis" has been addressed for chemical plants by Ponton [53]. However, many manipulated plant inputs will affect several plant outputs, and we recommend that the final pairing of inputs and outputs is postponed until later in the design, when more detailed information about the plant is available.

4.2 Later Stages of Design

In the later stages of plant design, it is often possible to develop a dynamical model of the plant. This makes it possible to take into account more sophisticated criteria when selecting inputs and outputs for control. The accuracy of the model will depend strongly on the type of plant being designed. For example, for some chemical processes only very crude models exist, whereas fairly accurate models for flexible structures in satellites can be developed from fundamental physical relationships. Paradoxically, the availability of accurate models at the design stage often does not imply that the task of designing a control system becomes easier. If you have access to an accurate model, so will probably also your competitor, and the availability of an accurate model merely translates into stricter performance requirements for the control system.

As the design progresses, more information about the plant becomes available - but on the other hand it also becomes progressively more difficult to make changes to the design. It should therefore be determined as early as possible what types and numbers of measurements and actuators that are needed for control, and where these should be installed. This means that a dynamical model of the plant should be established as early as possible in the design phase, but a very flexible modeling tool must be used to allow for easy modification of the model when changes are made to the design. The modeling tool should also enable extraction of simplified models from the rigorous model, such as linearlized models for controllability analysis and controller design.

4.3 Existing Plants

With existing plants, models can be tuned to match the observed behavior of the plant. Thus relatively accurate models can be made available - at least if identification experiments on the plant can be allowed. On the other hand, modifications to the plant design are now quite costly, even simple design modifications like installing new measurements and actuators usually involves shutting down the plant. In some plants, design modifications are clearly unacceptable or impossible - such as for space based structures. This means that the engineer should try very hard to achieve acceptable control with the plant inputs and outputs that are available. In many cases the number of available inputs and outputs is large even for existing processes, and some guidelines are needed for choosing the inputs and outputs used for control.

4.4 Selection of Plant Outputs for Control

Plant outputs are selected to a large extent based on physical understanding of the plant. Preferably, variables that are important for plant safety and product quality should be chosen as controlled outputs. All states that are not asymptotically stable must affect at least one of the measurements used for control, otherwise it will not be possible to stabilize the state.

Ideally, the measurement of a controlled output should be fast and direct, with little or no time delay or inverse response. This avoids limitations to control performance that are due to the measurements, and are not caused by the plant *per se*. In some cases, it is not possible to obtain fast and direct measurements of an important plant variable. One should then consider obtaining fast estimates of the plant variable using fast secondary measurements.

Care should be taken to ensure that the controlled variables are independent of each other. A trivial example of dependent variables are the mole fractions of the chemical components of a product - since the mole fractions must sum to 1, the mole fractions of all components cannot be determined independently. If the controlled outputs are dependent of each other,

the control system will not behave acceptably. Such inconsistencies are normally avoided if physical insight is used when choosing controlled outputs.

4.5 Selection of Plant Inputs for Control

Physical understanding is invaluable also when selecting plant inputs for control. The selected inputs should have a fast and direct effect on the plant outputs, and any state that is not asymptotically stable must be state controllable with the selected inputs.

The number of plant inputs for control should be equal to or larger than the number of controlled outputs from the plant, if offset-free control of all the controlled outputs is desired. When the plant outputs have been identified, a lower bound on the number of plant inputs for control has thus also been established. On the other hand, the types of plant inputs to use, and the locations of the plant inputs still need to be determined. The number of alternatives of plant inputs for control may therefore be very large at the early stages of plant design. Note that some alternatives may be mutually exclusive, and therefore one needs to ensure that the selected inputs are independent of each other. For example, if two tanks are connected by a pipeline, one cannot use the flow out of the first tank to control the level in the first tank, and at the same time use the flow into the second tank to control the level in the second tank.

4.5.1 The RGA for non-square systems and selection of inputs and outputs

The RGA has been applied to non-square systems [68, 12, 27]. Naturally, the definition in Eq. (54) needs to be modified in order to apply the RGA to non-square systems. For a matrix G of dimension $m \times n$ the non-square RGA is defined as

$$\Lambda(G) = G \times (G^+)^T \tag{63}$$

where G^+ is the pseudoinverse. Similar to the RGA for square systems, the non-square RGA can be interpreted as the ratio of open loop to closed loop gain, but some attention is needed to define *what* closed loop gain is considered, see [12] for details.

Some properties of the non-square RGA are [12]:

- If $m \leq n$ and $rank(G) = m$ then any row of Λ sums to one, if $m \geq n$ and $rank(G) = n$ then any column of Λ sums to one.
- If $m \leq n$ and $rank(G) = m$ then Λ is output scaling independent, if $m \geq n$ and $rank(G) = n$ then Λ is input scaling independent.
- Any permutation of rows or columns of G results in the same permutation of the rows and columns of $\Lambda(G)$.

When choosing plant inputs for control, we prefer inputs which have a strong effect on the outputs. Similarly, when choosing plant outputs for control, we prefer outputs in which the plant variations are easily observed. The Single Input Effectiveness (SIE) and the Single Output Effectiveness (SOE) provide quantitative measures of how well a specific input or a specific output fulfill these criteria.

Let v_j be the projection of a single input u_j on the column space of G. Cao and Biss [12] define the Single Input Effectiveness of input j is the ratio $\| v_j \|_2 / \| u_j \|_2$. Similarly, let z_i be the projection of a single output y_i on the row space of G. Then the Single Output Effectiveness of output i is the ratio $\| z_i \|_2 / \| y_i \|_2$.

The RGA and the input and output effectiveness are related [12, 27] through

$$\eta_j^2 = \sum_{i=1}^{m} \lambda_{ij}; \quad \xi_i^2 = \sum_{j=1}^{n} \lambda_{ij} \tag{64}$$

where η_j is the input effectiveness of input j and ξ_i is the output effectiveness of output i.

For plants with more inputs than outputs, the input effectiveness can be used to discard inputs with low effectiveness, since these inputs will have little effect on the outputs. Conversely, for plants with more outputs than inputs, the output effectiveness can be used to discard outputs with low effectiveness since the variations in these outputs will be small and the outputs with low effectiveness contain little information about the plant G.

4.6 Partial Control

When selecting inputs and outputs for control one can easily come in a situation where the number of outputs for which control is desired differs from the number of plant inputs available for manipulating the plant. In other cases the numbers of inputs and outputs are equal, but the overall plant has undesirable controllability characteristics, such as very strong interactions. Thus, one may come in a situation where one wants to investigate using only subsets of the available inputs and outputs for control. We will use the term "partial control" to denote a control system in which some of the plant outputs are deliberately left permanently uncontrolled. This should not be confused with the situation where a subset of the control loops are taken out of service (switched to manual mode), which is a situation that occurs from time to time in most control systems, e.g., during maintenance.

The need for control can arise from three possible causes:

1. The need for stabilizing an unstable plant.

2. A need for rejecting disturbances.

3. A requirement for setpoint following.

As noted previously, the need for stabilizing an unstable plant results in the requirement that any unstable state must be observable in the set of plant outputs chosen for control and controllable from the set of inputs used for control.

We will here consider more closely how to determine whether disturbance rejection and setpoint following requirements result in a need for controlling a specific output with a specific set of inputs. Consider a case where we have m outputs for which control is desired, and n inputs that can be used for control, and we consider using only k inputs and outputs for control, with $k \leq m$ and $k \leq n$. We then partition the outputs into two complementary subsets y_1 containing $m-k$ outputs and y_2 containing k outputs. Similarly we partition the inputs into two subsets u_1 and u_2, with u_1 containing $n-k$ inputs and u_2 containing k inputs[7]. We want to investigate whether y_1 needs to be controlled when y_2 is controlled using u_2, leaving u_1 unchanged. We then have

$$\begin{aligned} e &= y - Rr = Gu + G_d d - Rr \qquad (65) \\ &\Updownarrow \\ \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} &= \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix} u + \begin{bmatrix} G_{d1} \\ G_{d2} \end{bmatrix} d + \begin{bmatrix} R_1 & 0 \\ 0 & R_2 \end{bmatrix} \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} \qquad (66) \end{aligned}$$

If u_2 is used to control y_2 perfectly, we have that $e_2 = 0$, and we get

$$\begin{aligned} u_2 &= -G_{22}^{-1}[G_{d2}d - R_2 r_2] \qquad (67) \\ &\Downarrow \\ e_1 &= G_{11}u_1 - R_1 r_1 + [G_{d1} - G_{12}G_{22}^{-1}G_{d2}]d + G_{12}G_{22}^{-1}R_2 r_2 \qquad (68) \end{aligned}$$

Obviously, for the offsets e_1 to be small, the setpoint changes for y_1 must be small, i.e. R_1 must be small. Thus, it is natural to choose as uncontrolled outputs the outputs for which no (or small) setpoint changes are anticipated. Assume therefore that $R_1 = 0$, and when u_1 is constant we get that the effect of the disturbances on the uncontrolled outputs is given by $[G_{d1} - G_{12}G_{22}^{-1}G_{d2}]$, and the effect of setpoints for the controlled outputs on the uncontroller outputs is given by $[G_{12}G_{22}^{-1}R_2]$.

Whether disturbances and setpoints should be considered separately or simultaneously is a matter of design philosophy. It may be considered unlikely that several disturbances attain their worst possible values simultaneously. If disturbances and setpoints are considered separately, the requirement for acceptable offsets in e_1 is that all elements of $[G_{d1} - G_{12}G_{22}^{-1}G_{d2}]$

[7] Note that u_1 may be empty, but y_1 must be non-empty for the subsequent analysis to make sense.

and $[G_{12}G_{22}^{-1}R_2]$ should have magnitude less than one. If we consider the combined effects of setpoints and/or disturbances on output y_i in y_1, we must sum the magnitudes of the elements of row i in $[G_{12}G_{22}^{-1}R_2]$ and/or $[G_{d1} - G_{12}G_{22}^{-1}G_{d2}]$.

Clearly, this analysis will depend on the scaling of inputs, outputs, disturbances and setpoints. Furthermore, the results also depend on the choice of the sets u_1 and y_1. That is, the effects of disturbances and setpoints on uncontrolled output y_i depend not only on what inputs are used for control, but also on what other outputs are left uncontrolled. For large systems, there may therefore be a large number of possible choices of u_1 and y_1, and searching for the best choice of u_1 and y_1 may be laborious and time-consuming. In order to alleviate this problem, the input and output effectiveness (see Eq. (64)) can be used as a guide to selecting unused inputs and uncontrolled outputs.

After this preliminary analysis for selecting a set of outputs y_2 to be controlled and a set u_2 of inputs to use for control, the controllability of the resulting plant (G_{22} above) should be analyzed with respect to interactions, input constraints, RHP transmission zeros and time delays.

4.7 Analyzing and Modifying the Selection of Inputs and Outputs

Assume that a preliminary selection of inputs and outputs for control have been made, resulting in a vector of n outputs y controlled by the same number of inputs u. Controllability analysis may show that the candidate sets of inputs and outputs are unacceptable. Here we consider what remedial action can be taken in such cases.

Input constraints. A value of $\| y \|_\infty$ in Eq. (14) larger than one or a small $\underline{\sigma}(G)$ indicates that problems with input constraints may be expected. If $\| y \|_\infty > 1$ in Eq. (14), one should consider

1. The individual outputs in y with unacceptably large offsets. Are the specifications for these outputs too tight? Consider whether the specifications for the outputs can be relaxed with only minor implications for plant safety of product quality.

2. The inputs that reach their constraints. Consider replacing the physical device which sends the input to the plant. For example, this could mean replacing a valve with a larger one.

3. The individual disturbances that contribute strongly to constraining the plant input. In large plants, control actions in one section of the plant can cause disturbances to another plant section. One should in such cases consider changes to the control in the section of the plant where the disturbance originates. For example, if the flow out of a tank is used to control the tank level, fast level control can result in large flowrate

disturbances for the process unit receiving the flow from the tank. One should then consider slower control of the tank level, thus using the tank as a buffer to filter flowrate disturbances to the downstream process unit.

If $\underline{\sigma}(G)$ is small, one should study the corresponding input and output singular vectors. If the output singular vector is closely aligned to a specific output, this may indicate that the specifications for that output are unrealistic, see point 1 above. Similarly, an input singular vector closely aligned to a specific input indicates that a larger input is needed see point 2 above. On the other hand, is the input and output singular vectors corresponding to $\underline{\sigma}(G)$ have significant elements in the direction of more than one input/output, this indicates that the problem with input constraints is truly multivariable in nature. One may then consider changing the sets of inputs and outputs used for control. An illustration can be high purity binary distillation, which has a small $\underline{\sigma}(G)$ in the output direction corresponding to increasing the purity of both products at the same time. Increasing the purity in one of the products is easy if one accepts that the other product then becomes less pure. In practice, many distillation columns are operated with control of only one composition, either the top composition is controlled using the reflux flow, or the bottom composition is controlled using the boilup to the reboiler.

Interactions. If the plant is strongly two-way interactive (as measured by the minimized condition number or the RGA), this may indicate that there are strong couplings between different plant outputs. With a little luck, this can mean that it is not necessary to control all outputs simultaneously, and partial control should be investigated.

Similarly, large interactions can also be caused by the inputs having similar effects on the plant, i.e, the inputs being close to colinear. In this case one can also consider partial control; dropping one or more of the interacting inputs. If partial control is found not to be acceptable, one should try to identify alternative inputs for control.

If one cannot modify the selections of inputs and outputs, one should try to minimize the implications of the interactions. It has been noted previously that interactions are particularly troublesome if the plant model is uncertain. Efforts should therefore be made to reduce the uncertainty about the plant as much as possible. This can involve using high quality actuators and measurements, or using local feedback loops. For example, the flow through a valve may be uncertain because of wear, inaccurate valve position, variations in upstream or downstream pressure, and nonlinear valve characteristics if a linear model is used. Much of this uncertainty can be removed if the flow through the valve is measured and the valve position is used to control the flowrate. The higher level controls must then use the setpoint to the flow control loop as an input for control, instead of using the valve position as an input.

With large interactions, the issue of *what type of controller to use* is particularly important. Multivariable controllers which counteract the interactions in the plant *model*, may be sensitive to any uncertainties in the model. If a multivariable controller is chosen, one should therefore take robustness into account when designing the controller, see e.g. Morari and Zafiriou [47]. Decentralized control is more robust to uncertainties, but will not be able to counteract the interactions even if the model is assumed to be perfect.

Right half plant transmission zeros. One should make an effort to understand how RHP transmission zeros occur. In some cases, the RHP transmission zero is caused by an inverse response in the measurement of the output, and the RHP transmission zero will then be pinned to that specific output. If it is considered important to control the output with the pinned RHP transmission zero, two options are available:

1. Install an improved measurement which does not have any inverse response.

2. If option 1 is not available, one will have to accept that the control performance will suffer. However, analyzing the causes for control offset may give indications to how remedial action can be taken. For instance, feedforward control from disturbances or local control loops in cascade with the main loop may counteract disturbances before they can affect the main controlled output.

If the RHP transmission zero is truly multivariable (not pinned to a specific output), then one can make use of the observation that transmission zeros are unlikely for non-square plants, and add one or more extra inputs to the control problem. Consider the plant $y = [G_1\ G_2]u = Gu$, where $u = [u_1^T\ u_2^T]^T$. We wish to control y using u_1[8], but because of RHP transmission zero(s) in G_1 input vector is augmented with u_2, with u_2 chosen such that G has no RHP transmission zero within the desired bandwidth. One may then to find a compensator K_Z such that GK_Z is square and has no RHP transmission zero within the desired bandwidth.

Example. Consider the plant

$$y(s) = \frac{1}{f(s)} \begin{bmatrix} s+1 & s+4 & s+5 \\ 1 & 2 & 4 \end{bmatrix} u(s) \tag{69}$$

Inputs u_1 and u_2 are the primary inputs chosen for control, with input u_3 added because of an RHP transmission zero. Using *only* inputs 1 and 2, there is a transmission zero at $s = 2$, using inputs 1 and 3 there is a transmission zero at $s = 1/3$, and using inputs 2 and 3 there is a transmission zero at $s = -3$. Considering only controllability, it therefore appears preferable to use only inputs 2 and 3 for control. However, other considerations can make it necessary to minimize the use of u_3, we will here assume that u_3 is much more expensive to

[8] y and u_1 are both assumed to be of dimension n_1.

use than u_1. Therefore u_3 is used only at high frequencies, whereas at frequencies sufficiently below the RHP transmission zero u_1 is used. The result may be termed *parallel control*. This can be done by defining the original inputs as linear combinations of new inputs:

$$\begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = K_U u' = \begin{bmatrix} \frac{1}{\tau s+1} & 0 \\ 0 & 1 \\ \frac{\tau s}{\tau s+1} & 0 \end{bmatrix} \begin{bmatrix} u_1' \\ u_2' \end{bmatrix} \tag{70}$$

Choosing $\tau = 0.8$, we find that GK_U has transmission zeros at -1.59, -1.25, and -0.79, i.e., all in the left half plane. Decentralized controllers may now be designed using u_1' and u_2' as inputs for control. With K_U chosen as in Eq. (70), one should take some care when implementing the controller to avoid pole-zero cancellations between the decentralized controller and K_U, particularly if an integrating controller is used. In Fig. 4, it is shown how to avoid this problem when an ordinary PI controller is used for u_1' ($k_1(s) = k_I \frac{\tau_I s+1}{\tau_I s}$), by combining the PI controller and K_U.

An alternative is to implement the control in a cascaded manner; first use u_3 to control y_1, and then use u_1 to control u_3 to it's optimal value. This alternative implementation has the advantage of being relatively simple to tune on line, first tuning the loop $u_3 - y_1$ to be fast and thereafter tuning $u_1 - u_3$ to be slower. However, with this cascaded implementation, all control of y_1 is lost if u_3 is lost as a manipulated variable. The parallel implementation will retain some (albeit slower) control of y_1 if u_3 is lost.

Time delays. Time delays often occur because of transportation delays in the plant or in measurements. Holt et al. [29] give an example where controllability is improved by increasing one time delay in a multivariable plant (see also Eq. (26)), but normally it is desirable to have the time delays as short as possible. Unfortunately, removing time delays by placing equipment closer together is not always possible. One should then consider feedforward control or local control loops in cascade with the main control loop, as described above for RHP transmission zeros in a single output.

5 INTEGRITY TO CONTROL LOOP FAILURE

Control loops may fail for a variety of reasons, for example

- Measurement (plant output) failure.
- Actuator (plant input) failure or saturation.
- Operator intervention - switching controllers to manual.

The possible reasons for equipment malfunction will depend on both equipment type and design, the operating environment, and how the equipment is installed. Lost communication between controllers and actuators or measurements will also cause control loop failure.

Operators may switch controllers to manual because of poor controller performance, in which case the individual controller or a larger part of the overall control system should be redesigned or retuned. However, there are also valid reasons for operators to switch controllers to manual which do not imply that the control system is performing poorly, such as

- Maintenance of equipment in the control loop. Preferably, it should be possible to do as much as possible of equipment maintenance without having to close down an entire plant.
- Change of operating point. When moving between operating points, plant dynamics may be significantly different from the dynamics at any of the steady state operating points of the plant. Ideally, one would like to design a control system that makes changing the operating point easier for the operators, but this may be difficult to achieve. In many plants, it is therefore common to switch parts of the control system to manual when changing operating points, even if there are only trivial differences between the control systems that are used in the different operating points.
- Startup and shutdown. The control system is often put into service gradually (loop by loop) during startup, and taken out of service gradually during controlled shutdowns.

In addition to outright failure of control loops, the dynamics of the individual loops may also change, due to either changes in the plant (e.g., changes in feedstock or operating point) or changes to the controller tunings.

We would like the control system to be robust to all such changes. Both for outright failure of the loops and for moderate changes in the plant or controller dynamics, we would like the control system to remain stable and the performance to deteriorate in a graceful manner.

The issue of designing decentralized controllers which are robust to changes in the plant dynamics and the controller tuning parameters is addressed in a subsequent section. In this section we will give simple necessary conditions, in terms of the steady state gain matrix $G(0)$, for the existence of controllers yielding a closed loop system with integrity to loop failure. These necessary conditions assume that offset free control is obtained at steady state (implying use of integral action) and that the open loop transfer function GK rolls off at high frequencies. These assumptions are not very restrictive, and the assumption of offset-free control at steady state can in practice be relaxed somewhat, the results will hold provided high gain is used at steady state ($\bar{T}(0) \approx I$).

Failure *detection* is not considered here, it is assumed that the failures are discovered and the corresponding actuator is locked in an acceptable position.

Tolerance to loop failure will depend on the control structure, i.e., how the plant inputs and outputs are connected for control. We will therefore here assume that a pairing of inputs and outputs is chosen, and that the plant transfer function matrix $G(s)$ is rearranged to bring the paired elements on the diagonal.

5.1 Stable Plants

The Relative Gain Array. If a pairing corresponding to a negative steady state RGA element $\lambda_{ii}(G(0))$ is chosen, and integral action is used in the controllers in all the loops , at least one of the following will be true (e.g., [24]):

1. Loop i is unstable by itself, or
2. the system will become unstable if loop i is taken out of service, or
3. the overall system is unstable.

All of these three possibilities for instability are undesirable, and instability of the overall system is clearly unacceptable. However, for systems of dimension larger than 2×2, it may not be possible to choose a pairing of inputs and outputs corresponding to only positive relative gains. In such cases one will either have to try to find new inputs and/or outputs for control, or if this is not possible, ensure that the instability occurs where it can most easily be accepted (usually this means that one single loop, which is unlikely to operate without other loops in operation, is chosen to be unstable).

Note that for cases where only a subset of the loops are in operation, the RGA of the corresponding submatrix of G is also of interest.

The Niederlinski Index. The Niederlinski Index [50], N_I, is defined as

$$N_I = \frac{\det G(0)}{\det \tilde{G}(0)} \tag{71}$$

where $\tilde{G} = \text{diag}\{g_{11}, g_{22}, \cdots, g_{nn}\}$. With a pairing corresponding to a negative N_I is used, and integral action is used in all loops, then

1. at least one of the loops is unstable by itself, or

2. the overall closed loop system is unstable.

For plants of dimension 2×2, the Niederlinski Index and the RGA are equivalent, but for systems of larger dimension they contain different information.

5.2 Unstable Plants

Unstable plants will require feedback control for stabilization, they will obviously not be stable if all control loops are taken out of service. Furthermore, it is essential that the decentralized control structure allows for stabilization of any unstable modes. The results in this section on RGA and N_I for unstable plants are taken from Hovd and Skogestad [32].

Decentralized fixed modes. A plant mode is called a decentralized fixed mode if it cannot be changed by decentralized feedback [63]. Decentralized fixed modes are a structural property of the plant, and depend only on the decentralized control structure used, but are independent of the tuning of the decentralized controllers.

Normally it is not very difficult to avoid decentralized fixed modes; any mode corresponding to a pole in an element on the main diagonal of the plant transfer function matrix $G(s)$ can be moved by decentralized feedback[9].

If it for some reason is not possible to choose pairings of inputs and outputs such that all unstable poles appear in at least one of the elements on the main diagonal of $G(s)$, the most straight forward way to check whether an unstable mode ia a decentralized fixed mode is to try with different static controllers. If the location of an unstable pole is unchanged for all these static controllers, the unstable pole corresponds to a decentralized fixed mode and another pairing of inputs and outputs must be chosen. Any mode which is fixed for static feedback will also be fixed for dynamic feedback. An alternative, more rigorous way of testing for decentralized fixed modes can be found in Lunze [40].

The Relative Gain Array. Let n_G be the number of unstable poles in $G(s)$, and $n_{\hat{G}}$ be the number of unstable poles in $\hat{G}(s) = \text{diag}\{g_{ii}(s), G^{ii}(s)\}$, where $G^{ii}(s)$ is $G(s)$ with row i and column i removed. Note that normally $n_{\hat{G}} > n_G$. If G contains one unstable pole which appears both in g_{ii} and in G^{ii}, then $n_G = 1$ and $n_{\hat{G}} = 2$. The only case when it is possible to have $n_G > n_{\hat{G}}$ is if an unstable pole appears *only* in offdiagonal elements in row i or column i of $G(s)$. In this case it is possible for the unstable pole to correspond to a decentralized

[9] This assumes that the order of the inputs and outputs are rearranged to bring the paired elements of $G(s)$ on the main diagonal.

fixed mode - which obviously should be avoided. For unstable plants, in order to achieve stability of loop i, the closed loop system with loop i out of service, and the overall closed loop system, we must require that

$$\text{sign}\lambda_{ii}(0) = (-1)^{n_{\hat{G}}-n_G} \tag{72}$$

Note that for stable plants we have $n_G = n_{\hat{G}} = 0$, which is consistent with the requirement that λ_{ii} should be positive.

The Niederlinski Index. Let n_G be the number of unstable poles in $G(s)$, and $n_{\tilde{G}}$ be the number of unstable poles in $\tilde{G}(s) = \text{diag}\{g_{ii}(s), g_{22}(s), \cdots, g_{nn}(s)\}$. Note that normally $n_{\tilde{G}} > n_G$. If G contains one unstable pole which appears in all diagonal elements of $G(s)$, then $n_G = 1$ and $n_{\tilde{G}} = n$. In order to achieve stability of all the individual loops and the overall system, we must require that

$$\text{sign}N_I = (-1)^{n_{\tilde{G}}-n_G} \tag{73}$$

Possible drawbacks with these criteria involving the RGA and N_I for unstable plants, are:

- It is not sufficient to know only that the plant is unstable; the number of unstable poles, the multiplicity of unstable poles, and the distribution of the unstable poles in the transfer function matrix must be known.
- Obtaining $G(0)$ for an unstable plant can be more involved for an unstable plant than for a stable plant. It will obviously not be possible to obtain $G(0)$ by performing step responses on the uncontrolled plant.

Both these drawbacks are avoided if the plant model is obtained from a rigorous mechanistic model, and such a model is normally preferable if they can be obtained and validated at reasonable expense.

Example. Consider an unstable plant $G(s)$ with one unstable pole. Let p denote an element of $G(s)$ in which the unstable pole appears, and x denote an element in which the unstable pole does not appear. Assume that after rearranging inputs and outputs to bring the paired elements to the main diagonal, the distribution of elements in $G(s)$ containing the unstable pole can be described by

$$G(*) = \begin{bmatrix} x & x & x \\ p & p & p \\ p & p & p \end{bmatrix} \tag{74}$$

Then we will want $N_I < 0$, $\lambda_{11}(0) > 0$, $\lambda_{22}(0) < 0$, and $\lambda_{33}(0) < 0$. This illustrates that the value of $n_{\hat{G}}$ can depend on which loop is considered.

6 LOOP GAIN REQUIREMENTS FOR SETPOINT FOLLOWING AND DISTURBANCE REJECTION

We have earlier discussed the bandwidth requirements imposed by RHP transmission zeros and poles. Here we will consider bandwidth requirements for setpoint following and disturbance rejection using diagonal control. Throughout this section, it is assumed that a candidate pairing of inputs and outputs have been chosen, and that the order of inputs and outputs of the plant have been rearranged such that the elements of G that correspond to the paired inputs and outputs appear on the main diagonal. It is also assumed that the plant model $G(s)$ is appropriately scaled, as explained in the introduction. For simplicity of the exposition, we will take controller $K(s)$ to be fully decentralized (i.e. $K(s)$ is a diagonal matrix), however the necessary adjustments to apply the analysis to partially decentralized control ($K(s)$ block diagonal) are relatively straight forward.

6.1 Performance Specifications

As a performance specification, we will require that for any setpoint change r_j the offset e_i is bounded:

$$|e_i(i\omega)/r_j(i\omega)| = |[SR]_{ij}(i\omega)| < 1/|w_{ri}(i\omega)|; \quad \forall\omega, \forall i, \forall j \tag{75}$$

Here $w_{ri}(s)$ is a scalar performance weight. For any disturbance d_k we require that

$$|e_i(i\omega)/d_k(i\omega)| = |[SG_d]_{ik}(i\omega)| < 1/|w_{di}(i\omega)|; \quad \forall\omega, \forall i, \forall k \tag{76}$$

Typically, both weights are large at low frequencies where small offset is desired. $|w_{ri}|$ is often about 0.5 at high frequencies to guarantee an amplification of high-frequency noise of 2 or less. Thus we have a number of performance specifications we want to have satisfied simultaneously.

6.2 Bounds on Single-loop Designs

In this section we will use the above definition of performance to obtain bounds on the individual loop transfer functions $g_{ii}k_i$ at low frequencies. At frequencies below the bandwidth ω_B we may usually assume

$$S = (I + GK)^{-1} \approx (GK)^{-1} \tag{77}$$

We thus have

$$e = -SRr + SG_d d \approx -K^{-1}G^{-1}Rr + K^{-1}G^{-1}G_d d \quad (78)$$

$$= -(\tilde{G}K)^{-1}\tilde{G}G^{-1}Rr + (\tilde{G}K)^{-1}\tilde{G}G^{-1}G_d d, \quad \omega < \omega_B \quad (79)$$

where $\Gamma = \tilde{G}G^{-1}$ is known as the Performance Relative Gain Array (PRGA), and ΓG_d is known as the Closed Loop Disturbance Gain (CLDG). The steady state PRGA was introduced by Grosdidier [26], whereas Hovd and Skogestad [31] first demonstrated how to use the PRGA and CLDG as functions of frequency. This presentation differs from that of [31] in that the diagonal scaling matrix R for the setpoints is used, to make explicit how different scalings can be used for offsets and setpoints. The elements of Γ are denoted by γ_{ij}, and the elements of ΓG_d are denoted by δ_{ik}. The step from (78) to (79) requires that the diagonal elements of G are nonzero. We have proven the following theorem:

Theorem 3 *For plants with nonzero diagonal elements in $G(s)$, and at frequencies $\omega < \omega_B$ where (77) holds, the performance specifications (75) and (76) are satisfied iff*

$$|g_{ii}k_i(j\omega)| > |\gamma_{ij}R_{jj}w_{ri}(j\omega)|; \quad \forall\omega < \omega_B, \forall i, \forall j \quad (80)$$

$$|g_{ii}k_i(j\omega)| > |\delta_{ij}w_{di}(j\omega)|; \quad \forall\omega < \omega_B, \forall i, \forall k \quad (81)$$

For a given choice of pairings, Theorem 3 provides lower bounds on the individual loop gains to achieve nominal performance. We get one bound on the loop gain $g_{ii}k_i$ for each setpoint j and each disturbance k. The bounds may be difficult to satisfy if γ_{ij} or δ_{ik} are large. A plot of $|\gamma_{ij}R_{jj}(j\omega)|$ as a function of frequency will give useful information about for which input-output pairs we can expect interactions. A plot of $|\delta_{ij}(j\omega)|$ will give useful information about which disturbances are difficult to reject.

Comparison with all loops open. To get a better physical interpretation of the PRGA and CLDG consider the response e_i to a setpoint change r_j and disturbance d_k when all the other loops are *open*. We get

$$e_i = -(1 + g_{ii}k_i)^{-1}R_{jj}r_j + (1 + g_{ii}k_i)^{-1}g_{dik}d_k \quad (82)$$

When all the loops are *closed* simultaneously and we assume $\tilde{S} \approx (\tilde{G}K)^{-1}$ we get

$$e \approx -\tilde{S}\Gamma Rr + \tilde{S}\Gamma G_d d; \quad \omega < \omega_B \quad (83)$$

or

$$e_i \approx -(1 + g_{ii}k_i)^{-1}\gamma_{ij}R_{jj}r_j + (1 + g_{ii}k_i)^{-1}\delta_{ik}d_k \quad (84)$$

Comparing (82) and (84) we see that for a setpoint change r_i in loop i the performance relative gain, γ_{ii}, gives the approximate change in offset caused by closing all the loops. In addition, $\gamma_{ij}R_{jj}$ gives the effect of setpoint change r_j on output e_i when all loops are closed. That is, for $\omega < \omega_B$, we have $s_{ij}/\tilde{s}_{ij} \approx \gamma_{ij}$, and γ_{ij} is thus a measure of performance degradation at low frequencies. Similarly, we see that δ_{ik} is the approximate gain from disturbance d_k to offset e_i when all loops are closed, which explains why ΓG_d is called the closed loop disturbance gain.

6.3 Comparisons Between the RGA and the PRGA

The PRGA is closely related to the RGA, as their names suggest. The diagonal elements of the PRGA equal the diagonal elements of the RGA, but the off-diagonal elements generally differ. Some disadvantages of the PRGA relative to the RGA are:

1. The off-diagonal elements of the PRGA depend on the scaling of the outputs, but are independent of the scaling of the inputs. The RGA is independent of scaling.

2. The PRGA depends on the chosen pairing, and needs to be recomputed for every pairing under consideration. In contrast, the RGA for a new pairing can be found by simply permuting the RGA matrix for the original pairing[10]. The RGA therefore only needs to be computed once.

An advantage of the PRGA over the RGA is that the PRGA gives information about one-way interactions, whereas the RGA only contains information about two-way interactions. For example, the RGA of a triangular matrix is the identity matrix, but severe one-way interactions may nevertheless be present.

6.4 The PRGA and the CLDG in the Bandwidth Region

It is apparent that the approximation in Eq. (77) that the PRGA and CLDG are based on, breaks down in the bandwidth region. Nevertheless, experience shows that it is preferable to chose a pairing corresponding to a PRGA (and hence also an RGA) element that is close to unity in the bandwidth region. We will explain this in two different ways, first by considering closed loop stability and then by considering closed loop performance.

For *stability*, observe that the sensitivity function can be factorized as

$$S = (I + \tilde{S}(\Gamma - I))^{-1}\tilde{S}\Gamma \tag{85}$$

[10]To bring the paired elements in G to the main diagonal for a new pairing, G must have its rows and columns permuted. Permuting the rows and columns of G brings about the same permutations in the RGA matrix.

Assume that the individual loops have been tuned to be stable($\tilde{S}$ stable) and that both G and $\tilde{G}$ are stable and with no zeros in the right half plane. It then follows that the overall system will be stable provided $(I + \tilde{S}(\Gamma - I))^{-1}$ is stable. Here $\tilde{S}(\Gamma - I)$ is stable, and it then follows from the spectral radius stability condition (see e.g. [61]) that the overall system is stable if

$$\rho(\tilde{S}(\Gamma - I)) < 1 \ \forall \omega \tag{86}$$

where ρ denotes the spectral radius, i.e., the magnitude of the largest eigenvalue. At low frequencies, this condition is usually satisfied because $\tilde{S}$ is small. At higher frequencies where the elements of $\tilde{S}$ approach and possibly exceed one in magnitude, Eq. (86) may be satisfied if $G(j\omega)$ is close to triangular, since $(\Gamma - I)$ and hence $\tilde{S}(\Gamma - I)$ is then close to triangular with diagonal elements close to zero. The eigenvalues of $\tilde{S}(\Gamma - I)(j\omega)$ are then also close to zero, Eq. (86) is satisfied, and we have stability of S.

This provides a theoretical justification for choosing a pairing which gives a Γ close to triangular, with diagonal elements close to one in the bandwidth region. This corresponds to choosing a pairing with $\Lambda \approx I$ in the bandwidth region.

For closed loop *performance*, we ideally want all off-diagonal PRGA elements and all CLDG elements to be small. The reason for this is that the loop gains can only decrease at a limited rate around the bandwidth frequency - particularly if stability of the individual loops is desired. The work of Bode [5] tells us that the faster we decrease loop gain with increasing frequency, the more negative the phase of the open loop transfer function will be. It is well known from classical single - loop control theory that in order to preserve stability of loop i, the phase of $g_{ii}(j\omega)k_i(j\omega)$ must be larger than $-180°$ at the frequency ω_B where $|g_{ii}(j\omega_B)k_i(j\omega_B)| = 1$. Thus, the gradient of the loop gain in the Bode magnitude plot ($\log 10|g_{ii}k_i|$ vs. $\log 10\omega$) must be larger than -2/decade. In practice a gradient closer to -1/decade would be desirable in order to have sufficient phase margin at ω_B.

Thus, the accuracy of the approximation in Eq. (77) will improve gradually as one moves from frequency ω_B to lower frequencies, it will not be totally off at one frequency and close to perfect at a frequency only slightly lower. If an element of the CLDG or an off-diagonal element of the PRGA is large in the frequency range approaching the bandwidth, it is therefore likely that the approximation in Eq. (77) is sufficiently accurate to indicate problems with performance in this frequency region even though the loop gain is not very high. Although this argument cannot be used to reject pairings giving PRGA or CLDG elements of magnitude 2 in the bandwidth region (although magnitudes of 1 or less would be preferable), a PRGA or CLDG of magnitude 10 or more in the bandwidth region is a clear indication that problems with performance can be expected.

What if one is unable to choose a pairing giving small CLDG's or PRGA's in the band-

width region? Some suggestions are:

- For large $|\gamma_{ij}R_{jj}|$, use feed forward from setpoint j to loop i. Alternatively, setpoint j can be low pass filtered.
- For large $|\delta_{ik}|$, use feed forward from disturbance k to loop i. If disturbance k cannot be measured or calculated from other measurements, one may consider changes to the plant, for example installation of buffer tanks between unit processes in a plant. The CLDG may be used to estimate the needed holdup of such buffer tanks.

7 TUNING OF DECENTRALIZED CONTROLLERS

We have stated that one main advantage of decentralized control is the relative ease with which it can be tuned online. Nevertheless, for large, interactive plants the online tuning can be greatly simplified by having reasonable initial guesses for the tuning parameters. We will therefore consider design methodologies for decentralized control in this section. There exists several synthesis methods for the design of multivariable controllers, i.e., for synthesis of H_2- or H_∞-optimal controllers. However, these synthesis methods cannot accommodate any requirement for a specific structure for the controller. Indeed, the H_2- or H_∞-optimal decentralized controller is known to have an infinite number of states [57].

In the absence of any closed form solution to the design of an optimal decentralized controller, some pragmatic approaches to the design of decentralized controllers have evolved:

- Independent design [60, 35].
- Sequential design [43, 4, 49, 34].
- Simultaneous design using parameter optimization.

Below we will briefly discuss each of these three approaches.

7.1 Independent Design

Independent design was introduced by Skogestad and Morari [60], within the H_∞/μ framework. With this approach, bounds on the sensitivity and complementary sensitivity functions of the individual loops are found. Provided all controllers satisfy the bounds, the overall system will be stable and satisfy the specified performance criteria.

With the independent design of Skogestad and Morari, any type of controller can be used in the individual loops, as long as the bounds are satisfied. Although this gives a lot of flexibility to the design, it also results in the design often being conservative. This conservatism can be reduced by choosing a specific parametrization of the controllers in the individual loops, as demonstrated in [35].

Robustness to uncertainties in the plant model can easily be incorporated into independent design. Allowable ranges of variation for controller parameters can also be found. Tolerance to measurement or actuator failure can be ensured by performing independent designs for subsystems of lower dimension, provided the following assumptions are made

1. Failures are detected and the controllers in the corresponding loops are taken out of service.

2. In the case of a failure it is acceptable to give up control of the output(s) of the loop(s) in which the actuator(s) or measurement(s) has failed.

7.2 Sequential Design

Sequential design was introduced in the control literature by Mayne [43], but it is probably fair to say that it has always been the most common way of designing decentralized controllers in industry. In sequential design, the controllers in the individual loops are designed one at the time. When designing the controller for a specific loop, it is assumed that the controllers that have already been designed are in service.

One possible drawback with sequential design is that the result can depend on the order in which the individual controllers are designed. A heuristic rule is to design (and close) first the loops which have to be fast. The argument for this rule is that the fast loops are less affected by the control action in the slow loops than vice versa. This argument often holds, but there are some exceptions where there are strong one-way interactions from the slow loops to the fast loops. The PRGA and CLDG can give valuable information about the bandwidth requirements for the individual loops. In Hovd and Skogestad [34] such information is used to approximate the effect of the loops that are still open on the loops that are closed.

When one individual controller is designed, and the corresponding loop closed, this may cause unacceptable performance in a loop that has been closed previously. In such cases it will be necessary to redesign the controller in the loop with unacceptable performance.

It is normal to require that the system is stable after designing each individual controller. If this requirement is fulfilled, and no controller has to be redesigned, sequential design automatically provides a limited degree of failure tolerance. The system will remain stable if loops are closed in the same order as they were designed, or if loops are opened in te reverse order. This limited degree of failure tolerance can be useful during startup or shutdown, when it is common to bring loops into or out of service one at the time. However, this type of failure tolerance is of less help in the case of actuator or measurement failure (or saturation), which cannot be assumed to occur in any specified order.

We therefore have the following guidelines to the order of designing the individual controllers:

- Design first the controllers in the fast loops.
- If loops have to be put into operation in a specific order during plant startup, or taken out of operation in a specific order during shutdown, use this information to determine the order of designing the controllers.
- Individual elements or subsystems of the plant may have right half plane zeros that are

not RHP transmission zeros of the overall plant. If such RHP zeros in subsystems impose unacceptable bandwidth limitations in some loops, this problem should be avoided by designing controllers for these loops at a later stage in the design, or by changing the pairing of inputs and outputs. Beware that such changes in the order of designing controllers will make the overall system sensitive to failure of some of the loops.

It can be problematic to take account of robustness with respect to model uncertainty when performing sequential design. A practical way of approaching this problem is to design for robustness of the subsystem under control at each step in the design. However, this does not necessarily result in an overall design with very good robustness properties. Chiu and Arkun [15] circumvents this problem by formulating each controller step as an independent design problem for the loops that remain open. Thereby the conservatism of the independent design method is introduced also into sequential design, and in order to be successful Chiu and Arkun needs the independent design procedure to be feasible in the first step.

7.3 Simultaneous Design with Parameter Optimization

Conceptually, this approach to the design of decentralized controllers is the simplest. Parametrizations of the individual controllers are chosen à priori, and some criterion reflecting the control specifications is optimized with respect to the controller parameters. Problems with local minima may occur, since the optimization problem is not necessarily convex. The resulting decentralized controller can anyway only be optimal for the parametrization used.

Robustness with respect to model uncertainty may be achieved with parameter optimization, if the criterion that is optimized takes robustness into account. On the other hand, control performance need not be acceptable even with modest changes in controller parameters, and the method does not address failure tolerance, which must be checked separately. The parameter optimization method provides no guidelines for how to achieve failure tolerance if this is not achieved with the original design.

7.4 Special Case: Decentralized Controller Design for 2×2 Systems

For the special case of decentralized control of plants of dimension 2×2, Balchen (e.g. [3]) presents a graphical design procedure where the tradeoff between the design of the two loops is very clear. This procedure is based on the Multivariable Nyquist Theorem. Assuming that the plant is stable in open loop, the Multivariable Nyquist Theorem states that the closed loop system will be stable provided the map under the Nyquist D-contour of $\det(I + G(s)K(s))$ does not encircle the origin[11].

[11] The Multivariable Nyquist Theorem is the same as the classical monovariable Nyquist stability theorem except that one counts the number of encirclements of $\det(I + G(s)K(s))$ around the origin instead of the encirclements of $g(s)k(s)$ around the point -1.

Recall that for a decentralized controller $K(s)$ we have $(I+G(s)K(s)) = (I+E_H(s)\tilde{T}(s))(I+\tilde{G}(s)K(s))$, where $E_H(s) = (G(s) - \tilde{G}(s))\tilde{G}^{-1}(s)$. Hence, if the open loop system and the individual loops are stable, we can consider only the encirclements around the origin of $\det(I + E_H(s)\tilde{T}(s))$. For a plant G of dimension 2×2, we then get

$$\det(I + E_H\tilde{T}) = (1 + \frac{g_{12}g_{21}}{g_{11}g_{22}}\tilde{t}_1\tilde{t}_2) = (1 + Y\tilde{t}_1\tilde{t}_2) \tag{87}$$

Here Y is known as the *Rijnsdorp Interaction Measure* [54, 2]. Instead of plotting $Y(s)\tilde{t}(s)_1\tilde{t}(s)_2$ and checking for encirclements of the point -1, Balchen checks for "encirclements" of $\tilde{t}(s)_1\tilde{t}(s)_2$ around $-1/Y(s)$. This way, $\tilde{t}(s)_1\tilde{t}(s)_2$, which depends on the controller, is separated from $-1/Y(s)$ which is controller independent. The tradeoff between the two loops is clear since it is only the product of the complementary sensitivity functions for the two loops which is plotted. This method can therefore be helpful when modifying an unacceptable initial design. Unfortunately, the method does not generalize to systems with more control loops. Full details and examples on the use of this method can be found in Balchen and Mummé [3].

8 CONCLUSION

In the introduction we started off by stating that traditional single-loop control is clearly incomplete for many control problems, which are by nature multivariable. After first describing the typical structure of control systems for large plants, and discussed the reasons why structure is imposed on control systems, we addressed some fundamental issues which are independent of the choice control structure:

- Fundamental limitations to what is achievable with feedback control.
- The presence of *interactions* in multivariable plants, and its implications for the choice of control structure.
- How to determine what variables should be controlled, and what plant inputs should be used for control. In some applications this is obvious, whereas in other applications this is far from the case. The chemical processing industries, in which the authors are most familiar, have many control problems which fall into the latter category.

The latter sections are focused on issues specific to decentralized control, which is extensively used in the regulatory control level of large plants. issues covered in these sections are:

- Control system integrity. How to ensure that the behavior of the control system deteriorates gracefully when a part of the control system becomes inactive - in particular, stability should be maintained. The relative ease of designing of integrity in a decentralized control system is an important reason for the popularity of decentralized control at the regulatory control level.
- What manipulated variable should be used to control a specific controlled variable. Pairing of controlled and manipulated variables can be critical for the performance and integrity of a decentralized control system.
- Approaches to the tuning of decentralized controllers. Mathematical controller synthesis techniques cannot handle a requirement for a specific controller structure, but some practical approaches to controller tuning have evolved instead. The relative merits of these approaches have been discussed.

Despite the incompleteness of classical single-loop theory for multivariable control problems, decentralized control is likely to remain an important part of the control system for large plants. We hope this chapter have filled some of the holes in the theory, and that it will make it easier for the reader to understand what can be achieved with feedback control, and to design (at least the regulatory layer of) a control system.

References

[1] G. Balas, J. C. Doyle, K. Glover, A. Packard and R. Smith, *μ-analysis and Synthesis Toolbox. Users Guide*. MUSYN Inc., Minneapolis, Minnesota, USA (1991).

[2] J. G. Balchen, *Lecture Notes*, Norwegian Institute of Technology (in Norwegian), (1958).

[3] J. G. Balchen and K. I. Mummé, *Process Control. Structures and Applications*. Van Nostrand Reinholdt, New York, USA (1988).

[4] D. S. Bernstein, *Int. J. Control*, **46** (1987), pp. 1569 - 1577.

[5] H. W. Bode, *Bell System Tech. J.*, **19** (1940), pp. 421 - 454.

[6] H. W. Bode, *Network Analysis and Feedback Amplifier Design* (D. Van Nostrand Company, New York, USA, 1945).

[7] R. D. Braatz and M. Morari. Submitted to *Numerische Mathematik*, (1991).

[8] R. D. Braatz and M. Hovd *Efficient Computation of Disturbance Rejection Measures*, in preparation (1996).

[9] E. H. Bristol, *IEEE Trans. Autom. Control*, **AC-11** (1966), pp. 133-134.

[10] E. H. Bristol, Paper at the *AIChE Annual Meeting*, Chicago, Illinois, USA (1978).

[11] E. H. Bristol, Presented at *Joint American Control Conference*, (1981).

[12] Y. Cao and D. Biss, To appear in *Comp. chem. Engng.* (1996).

[13] J. Chen, *IEEE Trans. Automatic Control*, **AC- 40** (1995), pp.1700 - 1716.

[14] J. Chen, J. S. Freudenberg, C. N. Nett, *Automatica*, **30** (1994), pp. 1029 - 1035.

[15] M.-S. Chiu and Y. Arkun, *Automatica*, **28** (1992), pp. 997-1001.

[16] J. C. Doyle, J. E. Wall and G. Stein, *Proc. IEEE Conference on Decision and Control*, Orlando, Florida, USA, pp. 629 - 636 (1982).

[17] J. C. Doyle, *IEE Proc., Pt. D*, **129** (1982), pp. 242 - 250.

[18] J. C. Doyle, B. Francis and A. Tannenbaum, *Feedback Control Theory*, Macmillan Publishing Company (1992).

[19] C. G. Economou and M. Morari, *Ind. Eng. Chem. Proc. Des. Dev.*, **25** (1986), pp. 411-419.

[20] A. S. Foss, *AIChE Journal*, **19** (1973), pp. 209-214.

[21] J. S. Freudenberg and D. P. Looze, *IEEE Trans. Autom. Control*, AC-30 (1985), pp. 555-565.

[22] J. S. Freudenberg and D. P. Looze, *Frequency Domain Properties of Scalar and Multivariable Feedback Systems*, Lecture Notes in Control and Information Sciences 104, Springer Verlag, Berlin, Germany (1988).

[23] G. H. Golub and C. F. Van Loan,*Matrix Computations*, John Hopkins University Press, Baltimore, Maryland, USA (1983).

[24] P. Grosdidier, M. Morari and B. R. Holt, *Ind. Eng. Chem. Fundam.*, **24** (1985), pp. 221 - 235.

[25] P. Grosdidier and M. Morari, *Automatica*, **22** (1986), pp. 309 - 319.

[26] P. Grosdidier, Comp. chem. Engng, **14** (1990), pp. 687 - 689.

[27] K. Havre and S. Skogestad, Paper presented at *ESCAPE VI*, Rhodes, Greece, (1996).

[28] B. R. Holt and M. Morari, *Chem. Eng. Sci.*, **40** (1985), pp. 59 - 74.

[29] B. R. Holt and M. Morari, Chem. Eng. Sci., **40** (1985), pp. 1229-1237.

[30] Hovd, M., *Studies on Control Structure Selection and Design of Robust Decentralized and SVD Controllers*, Dr. Ing. Thesis, University of Trondheim - NTH (1992).

[31] M. Hovd and S. Skogestad, *Automatica*, **28** (1992), pp. 989 - 996.

[32] M. Hovd and S. Skogestad, *Ind. Eng. Chem. Res.*, **33** (1994), pp. 2134 - 2139.

[33] M. Hovd and S. Skogestad, *Automatica*, **30** (1994), pp. 957 - 973.

[34] M. Hovd and S. Skogestad, *Automatica*, **30** (1994), pp. 1601 - 1607.

[35] M. Hovd and S. Skogestad. *J. Proc. Cont.*, **3** (1993), pp. 43 - 51.

[36] M. Hovd, R. D. Braatz and S. Skogestad, *Proc. American Control Conference*, 1994.

[37] Y. S. Hung and A. G. J. MacFarlane, *Multivariable Feedback: A Quasi-Classical Approach*, Lencture Notes in Control and Information Sciences, Springer Verlag, Berlin, Germany (1982).

[38] T. Kailath, *Linear Systems*, Prentice Hall Inc., Englewood Cliffs, N. J., USA (1980).

[39] Kwakernaak and Sivan, *Linear Optimal Control Systems*, Wiley, New York, USA, (1972).

[40] J. Lunze, *Feedback Control of Large-Scale Systems*, Prentice Hall, New York, N. Y., USA (1992).

[41] A. MacFarlane and N. Karcanias, *Int. J. Control*, **24** (1976).

[42] J . M. Maciejowski, *Multivariable Feedback Design*, Addison-Wesley, Wokingham, England (1989).

[43] Mayne, D. Q., *Automatica*, **9** (1973), pp. 201 - 207.

[44] T. J. McAvoy, *Interaction Analysis*, ISA Monograph Series, ISA, Research Triangle Park, Maryland, USA (1983).

[45] A. L. Mees, *Systems and Control Letters*, **1** (1982), pp. 155- 158.

[46] T. Mejdell, *Estimators for Product Compositions in Distillation Columns*, Dr. Ing. Thesis, University of Trondheim - NTH (1990).

[47] M. Morari and E. Zafiriou, *Robust Process Control*, Prentice Hall, Englewood Cliffs, New Jersey, USA (1989).

[48] C. N. Nett and V. Manousiouthakis, *IEEE Trans. Autom. Control*, **AC-32** (1987), pp. 405 - 407.

[49] C. N. Nett and J. A. Uthgenannt, *Automatica*, **24** (1988), pp. 261 - 265.

[50] A. Niederlinski, *Automatica*, **7** (1971), pp. 691 - 701.

[51] A. Packard. *What's new with μ.* PhD Thesis, University of California, Berkeley (1987).

[52] J. D. Perkins and M. P. F. Wong, *Chem. Eng. Res. Des.*, **63** (1985), pp. 358 - 362.

[53] J. Ponton, *Chem. Eng. Sci.*, **49** (1994), pp. 2089 - 2095.

[54] J. E. Rijnsdorp, *Automatica*, **1** (1965), pp. 15-28.

[55] H. Rosenbrock, *State-space and Multivariable Theory*, Nelson, London, England (1970).

[56] H. Rosenbrock, *Computer Aided Control System Design*, Academic Press, London (1974).

[57] N. R. Sandell jr., P. Varaiya, M. Athans and M. C. Safonov, *IEEE Trans. Autom. Control*, **AC-23** (1978), pp. 108 - 128.

[58] F. G. Shinskey, *Distillation Control*, 2nd ed., McGraw Hill, New York (1984).

[59] S. Skogestad and M. Morari, *Ind. Eng. Chem. Res.*, **26** (1987), pp. 2323 - 2330.

[60] S. Skogestad and M. Morari, *Automatica*, **25** (1989), pp. 119 - 125.

[61] S. Skogestad and I. Postlethwaite, *Multivariable Feedback Control. Analysis and design*, Wiley, New York, USA, (1996).

[62] J. Slaby and I. Rinard, AIChE National Meeting, Miami, FL, USA, (1986).

[63] S.-H. Wang and E. J. Davison, *IEEE Trans. Autom. Control*, **AC-18** (1973), pp. 473-478.

[64] E. A. Wolff, S. Skogestad, M. Hovd and K. W. Mathiesen, *Preprints IFAC Workshop on Integration of Process Design and Process Control*, London, England (1992).

[65] E. A. Wolff, *Studies on Control of Integrated Plants*, Dr. Ing. Thesis, Univ. of Trondheim - NTH, 1994.

[66] D. C. Youla, J. J. Bongiorno jr. and C. N. Lu, *Automatica*, **10** (1974) pp. 159 - 173.

[67] C.-C. Yu and W. L. Luyben, *Ind. Eng. Chem. Res.*, **26** (1987), pp. 1043 - 1045.

[68] J.-W. Chang and C.-C. Yu, *Chem. Eng. Sci.*, **45** (1990), pp. 1309 - 1323

[69] K. Zhou, J. C. Doyle and K. Glover, *Robust and Optimal Control*, Prentice Hall Inc., Englewood Cliffs, N. J., USA (1995).

[70] J. G. Ziegler and N. B. Nichols, *Trans. ASME*, **65** (1943), pp. 433-444.

A THE STRUCTURED SINGULAR VALUE

The Structured Singular Value (usually denoted μ or SSV), was introduced by Doyle (1982) as a tool for analyzing the robustness of a control system with respect to uncertainties in the model of the plant. Both robust stability and robust performance problems can be addressed using the SSV, and it provides a flexible framework for specifying performance requirements. Iterative procedures exist for designing controllers which optimizes robust stability or robust performance within this framework (e.g. [1]).

We will here give a very brief introduction to the structured singular value, since it is a tool that is used both for controllability analysis and design of decentralized controllers. There exists a substantial literature on this subject, and more information can be found in e.g. [16, 17, 51, 47, 69, 61]. The basic idea within the structured singular value framework is to accept that no model of a physical system is perfect. One therefore attempts to model the uncertainties in the model, i.e., its location, structure and magnitude. In Fig. 5 is an example of a control system for which there is uncertainty in the plant inputs and the plant outputs, represented by the perturbation blocks Δ_I and Δ_O, respectively. The weights W_I and W_O are frequency-dependent and normalize the maximum magnitude of Δ_I and Δ_O to unity. The individual perturbation blocks can can be restricted to have a certain structure. For instance, individual inputs and outputs normally do not affect each other, therefore Δ_I and Δ_O can be assumed to be diagonal.

Any block diagram with uncertainties represented by perturbation blocks can be rearranged into the $M-\Delta$ structure in Fig. 6, if external inputs and outputs are neglected. In Fig. 6, Δ is a block diagonal matrix with the perturbation blocks of the original block diagram on the diagonal, and M contains all the other blocks in the original block diagram (plant, controller and weights). For the specific case in Fig. 5, we have that

$$\Delta = \mathrm{diag}\{\Delta_I, \Delta_O\}; \qquad M = \begin{bmatrix} -W_I K G (I+KG)^{-1} & -W_I K (I+GK)^{-1} \\ W_O G (I+KG)^{-1} & -W_O G K (I+GK)^{-1} \end{bmatrix} \tag{88}$$

Provided M is stable and Δ is norm bounded and stable (stable perturbation blocks), It can be shown that the overall system is stable provided $\det(I - M\Delta) \neq 0 \ \forall \Delta, \ \forall \omega$. The structured singular value, μ, is defined such that

$$\mu_\Delta^{-1} = \min_\delta \{\delta | \det(I - M\Delta) = 0 \qquad \text{for some } \Delta, \bar{\sigma}(\Delta) \leq \delta\} \tag{89}$$

The subscript Δ in μ_Δ emphasize that the value of μ depends on the structure of Δ. The perturbation matrix Δ contains structure on two levels, firstly, it is a block diagonal matrix of perturbation blocks, secondly, each perturbation block within Δ may themselves be structured.

If weights are used to normalize the maximum value of the largest singular value of Δ to unity ($\bar{\sigma}(\Delta) = 1$) at all frequencies, like in Fig. 5, the system will remain stable for any allowable perturbation Δ provided $\sup_\omega \mu_\Delta(M) < 1$. To simplify notation, we use "$\mu(M)$" in the meaning $\sup_\omega \mu_\Delta(M)$.

It is possible to calculate $\mu(M)$ exactly only in a few special cases, but reasonably tight upper and lower bounds are readily available. Some useful properties of μ are:

$$\rho(M) \leq \mu(M) \leq \bar{\sigma}(M) \tag{90}$$

$$\mu(M) \leq \bar{\sigma}(D_l M D_r^{-1}) \tag{91}$$

Eq. (90) holds for any complex valued perturbation Δ, but μ may be lower than the lower bound if the perturbations are constrained to be real. D_l and D_r are real positive matrices with a structure such that $D_r^{-1}\Delta D_l = \Delta$. If all blocks in Δ are square, $D_l = D_r$. The upper bound on the value of μ in Eq. (91) is usually quite tight [17], and minimizing this upper bound can be formulated as a convex optimization problem, which means that it is computationally tractable.

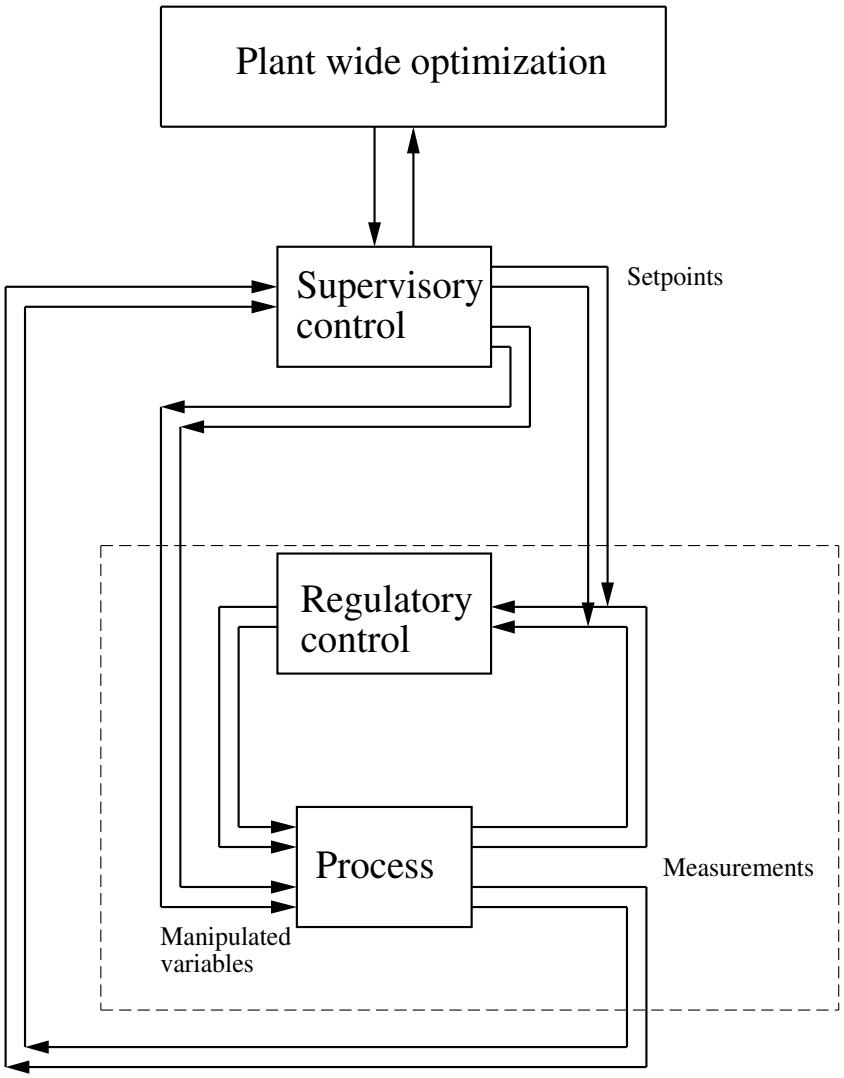

Figure 1: Hierarchical structure of a control system of a typical chemical plant.

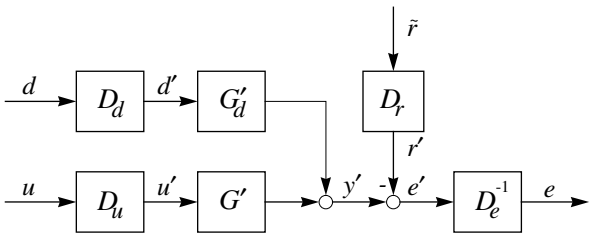

Figure 2: Scaling of variables.

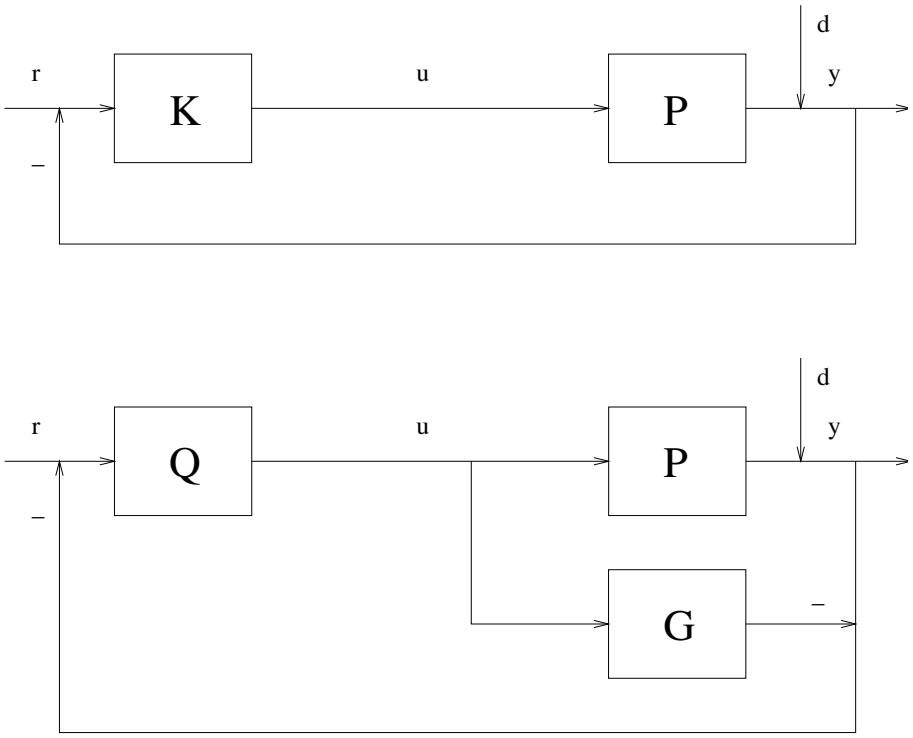

Figure 3: Conventional and Internal Model Control feedback structures.

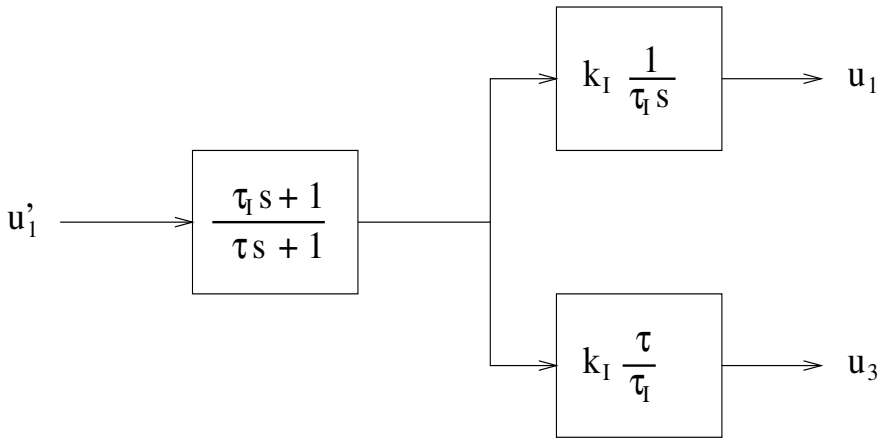

Figure 4: Implementation of integrating controller for parallel control. u_1 is used at low frequencies, and u_3 is used at high frequencies.

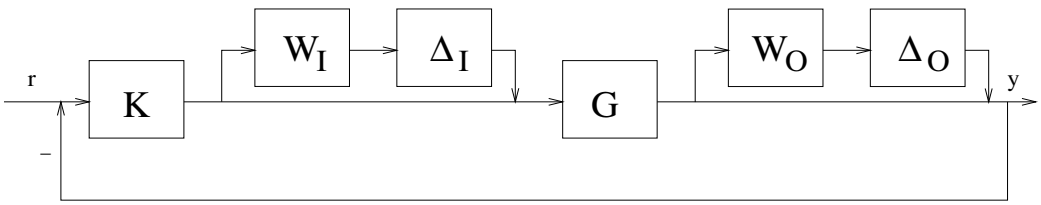

Figure 5: Control system with uncertainty in the plant inputs and plant outputs.

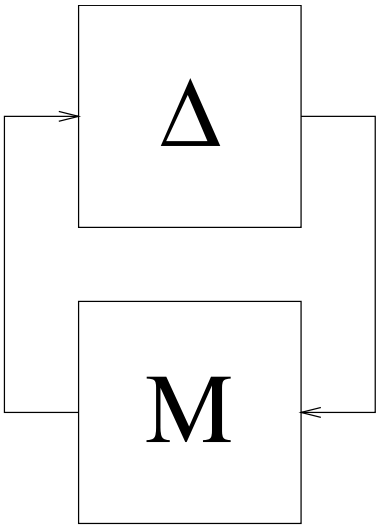

Figure 6: The $M - \Delta$ structure, with the uncertainties collected in Δ and M the "generalized plant".