Presented at ACC'90, San Diego, May 23-25, 1990

Use of Frequency-Dependent RGA for Control Structure Selection

Sigurd Skogestad Morten Hovd

Chemical Engineering
Norwegian Institute of Technology (NTH)
N-7034 Trondheim, Norway

Abstract

The paper reviews existing and presents some new results for the frequency- dependent RGA. It is shown how frequency-dependent plots of the relative gains, λ, and the closed-loop disturbance gains, δ, can be used to evaluate the achievable performance (controllability) of a plant under decentralized control. These controller-independent measures give constraints on the design of the individual loops.

1 Introduction

The relative gain array (RGA) has found widespread use as a measure of interaction and as a tool for control structure selection for single-loop controllers. It was first introduced by Bristol(1966), and Shinskey (1967, 1984) and several other authors has demonstrated its practical application. Important advantages with the RGA is that it depends on the plant model only and that it is scaling independent. It is straightforward to generalize the RGA from single-loop controllers to block-diagonal controllers by introducing the block relative gain (BRG) (Manousiouthakis et al, 1986), and most of the results presented in this paper may be generalized in such a manner. However, to simplify the presentation, and because single-loop controllers are most common in practice, we shall consider only the RGA in this paper.

Most authors have confined themselves to use the RGA at steady state, and a thourough review of the use and interpretation of the steady-state RGA is given by Grosdidier (1985). A frequency-dependent interaction measure Δ, which is equivalent to the RGA for 2×2 systems, was introduced by Balchen (1958) and is discussed for $n \times n$ systems in Balchen and Mumme (1988). Balchen also gives some performance interpretation to his measure. The RGA was extended to higher frequencies for 2×2 system by Witcher and McAvoy (1977) and for larger systems by McAvoy (1983). Their definition is consistent with Bristols original definition of the RGA and is used in this paper. Most authors consider mainly the magnitude of the RGA-elements as a function of frequency, but Balchen (1958) and Slaby and Rinard (1986) also discuss use of the phase angle.

Other definitions of a dynamic RGA have also been proposed. Witcher and McAvoy (1977) proposed a time domain definition of the RGA, as did Tung and Edgar (1981). Arkun (1987, 1988) has proposed measures (DBRG and Relative Sensitivity) which include the controller. Balchen and Mumme (1988) also include the controller in their measure Δ. However, one then looses one of the main advantages of the RGA which is that it depends on the plant model only. These alternative definitions are not considered in this paper.

A measure related to the RGA is the relative disturbance gain, RDG, introduced by Stanley et al. (1985). It was given a performance interpretation and extended to other frequencies by Skogestad and Morari (1987a). It is also scaling independent and depends on the model of the plant and its disturbances.

The objective of this paper is to derive measures which may be assist the engineer in selecting the best control structure and in designing single- loop controllers such that the overall system has acceptable performance. We shall prove that the frequency-dependent RGA and RDG are very useful tools in this respect. A number of interesting properties are already known about the RGA. We shall give a review of some of the most inportant ones, which we believe are significant for engineering applications. Furthermore, we shall present some new results, for example, one that relates the RGA and RHP-zeros.

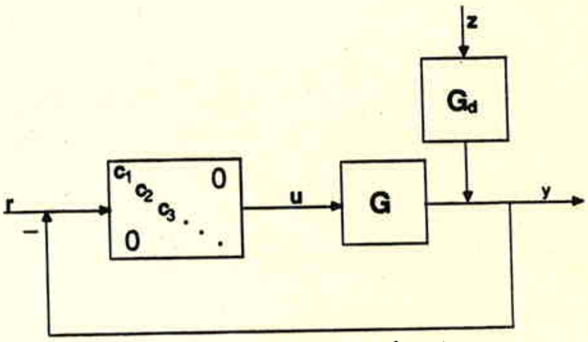

Fig. 1. Block diagram of control system.

2 Definitions and Notation

2.1 Notation

The notation follows that of Skogestad and Morari (1989). The controller $C(s)$ is diagonal with entries $c_i(s)$ (see Fig.1). This implies that after the variable pairing has been determined, the order of the elements in y and u has been arranged so that the transfer function matrix G has the elements corresponding to the paired variables on the main diagonal. The matrix consisting of the diagonal elements of $G(s)$ is denoted $\tilde{G}$. Let $y(s)$ denote the response for the overall system when all loops are closed. $\tilde{y}(s)$ denotes the response of the individual subsystems, that is, $\tilde{y}_i(s)$ is the response when loop i is closed and the other loops are open. The sensitivity function S and complementary sensitivity function H for the overall system are defined by

$$e(s) = -S(s)r(s) + S(s)G_d(s)d(s); \quad S = (I + GC)^{-1} \quad (1)$$

$$y(s) = H(s)r(s); \quad H = GC(I+GC)^{-1} \tag{2}$$

We also have $S + H = I$. G is assumed to be a $n \times n$ square matrix, but G_d may be nonsquare. The bandwidth of the system, ω_B is defined as the frequency where the asymptote of $\bar{\sigma}(S(j\omega))$ crosses one. The freuqency range near ω_B will in the paper be denoted "crossover region" or sometimes simply "high frequency". The closed-loop transfer functions for the individual loops may be collected in the diagonal matrices $\bar{S}$ and $\bar{H}$:

$$\bar{e}(s) = -\bar{S}(s)r(s) + \bar{S}(s)G_d(s)d(s) \tag{3}$$

$$\bar{S} = (I + \bar{G}C)^{-1} = diag\{\bar{s}_i\}; \quad \bar{s}_i = (1 + g_{ii}c_i)^{-1} \tag{4}$$

$$\bar{y}(s) = \bar{H}(s)r(s); \quad \bar{H} = \bar{G}C(I+\bar{G}C)^{-1} \tag{5}$$

Note that the elements in $\bar{S}$ and $\bar{H}$ are not equal to the diagonal elements of S and H. The following relations apply

$$S = \bar{S}(I - E_S\bar{S})^{-1}\bar{G}G^{-1}; \qquad E_S = (G - \bar{G})G^{-1} \tag{6}$$

$$H = G\bar{G}^{-1}\bar{H}(I + E_H\bar{H})^{-1}; \qquad E_H = (G - \bar{G})\bar{G}^{-1} \tag{7}$$

$$S = \bar{S}(I + E_H\bar{H})^{-1} \tag{8}$$

$$(H - \bar{H})\bar{H}^{-1} = -SE_H = (I - \bar{S}E_S)^{-1}\bar{S}E_S \tag{9}$$

2.2 RGA

Consider a $n \times n$ plant $G(s)$.

$$y(s) = G(s)u(s) \tag{10}$$

When all other outputs are uncontrolled, the gain from input u_j to output y_i is $g_{ij}(s)$. Solving Eq. (10) for u, it can be seen that the gain from u_j to y_i with all the other y's perfectly controlled is $1/[G^{-1}(s)]_{ji}$. The relative gain is the ratio of these two gains. Thus a matrix of relative gains, the RGA matrix, can be computed using the formula

$$\Lambda(s) = G(s) \times [G^{-1}]^T(s) \tag{11}$$

where the $\times$ symbol denotes element by element multiplication (Hadamard product). The inverse $G^{-1}(s)$ may be non-proper or non-causal, and a physical interpretation in terms of perfect control is of course not strictly meaningful except at steady-state. This has caused many authors to discard use of a dynamic RGA, or to restrict its use to plants with no RHP-zeros (Manousiouthakis et al, 1986). This is unfortunate as the dynamic RGA as defined above proves to have a number of useful properties. Furthermore, we shall mainly consider the $\Lambda(s)$ as a function of frequency, $s = j\omega$, and in this case $\Lambda(j\omega)$ may be computed for any plant G provided it has no $j\omega$-axis zeros.

The RGA matrix as defined above has some interesting algebraic properties (eg., Grosdidier et al., 1985):

a) It is scaling independent (ie., independent of units chosen for u and y)

b) All row and column sums equal one

c) Any relative pertubation in a element of G results in the same pertubation in the RGA matrix Λ.

These properties are easily proven from the following expressions for the individual elements of Λ

$$\lambda_{ij}(s) = g_{ij}(s)[G^{-1}(s)]_{ji} \tag{12}$$

or

$$\lambda_{ij}(s) = (-1)^{i+j}\frac{g_{ij}(s)det(G^{ij}(s))}{det(G(s))} \tag{13}$$

Here G^{ij} denotes the matrix G with subsystem ij removed, that is, row i and column j is deleted.

2.3 RDG

For disturbances, an analogue to the Relative Gain Array is found in the Relative Disturbance Gain (RDG) introduced by Stanley et al. (1985). For a particular disturbance z_k, the RDG is defined for each loop i as the ratio of the change in u_i for perfect disturbance rejection, to the change in u_i needed for complete disturbance rejection in the corresponding output y_i when all the other manipulated variables are kept constant. A matrix of relative disturbance gains can be calculated from (Skogestad and Morari, 1987a):

$$RDG = \frac{G^{-1}G_d}{\bar{G}^{-1}G_d} \tag{14}$$

The division in this case denotes element by element division. Here G_d is the open loop transfer matrix from disturbances to outputs, and $\bar{G}$ consists of the diagonal elements of G. The RDG matrix is scaling independent. Individual elements of the RDG matrix are given by

$$\beta_{ik} = \frac{[G^{-1}G_d]_{ik}}{g_{dik}/g_{ii}} \tag{15}$$

Note that the RDG has to be recomputed whenever another choice of pairings is selected, whereas the RGA need only be rearranged in accordance with the rearrangement of G.

3 Use of steady state RGA

3.1 The RGA and integral action

A well established rule for pairing inputs and outputs is that if integral action is required, pairings corresponding to negative steady state relative gains should be avoided.

Theorem 1 *Consider a control system using single-loop controllers with integral action. A pairing of outputs and manipulated inputs corresponding to a negative steady state relative gain will give a closed loop system with at least one of the following properties:*

(a) The closed loop system is unstable.

(b) The loop with the negative relative gain is unstable by itself.

(c) The closed loop system is unstable if the loop with the negative relative gain is removed.

The proof is given by Grosdidier et al. (1985) and is based on using (13) at steady state, combined with the fact that integral action with positive feedback leads to instability.

4 Use of frequency-dependent RGA

4.1 The RGA and right half plane zeros

Bristol (1966) claimed in his original paper that there was a relationship between RHP-zeros and negative values of $\lambda_{ii}(0)$, but Grosdidier et al (1985) showed with a counterexample that this is not true. However, as we shall see there proves to be a relationship if we assume that the loops have been paired such that $\lambda_{ii}(\infty)$ is positive.

Theorem 2 *Assume* $\lim_{s\to\infty}\lambda_{ij}(s)$ *is finite and different from zero. Let* $g_{ij}(s)det(G^{ij}(s))$ *have* z_{0n} *zeros at the origin,* z_{Rn}

zeros in the right half plane, p_{0n} poles at the origin and p_{Rn} poles in the right half plane. Similarly, let $det(G(s))$ have z_{0d} zeros at the origin, z_{Rd} zeros in the right half plane, p_{0d} poles at the origin and p_{Rd} poles in the right half plane. Define $z_0 = z_{0n} - z_{0d}$, $p_0 = p_{0n} - p_{0d}$, $z_R = z_{Rn} - z_{Rd}$, and $p_R = p_{Rn} - p_{Rd}$. Then the net change in the phase of $\lambda_{ij}(j\omega)$ as the frequency goes from 0 to ∞ is $\frac{\pi}{2}(p_0 - z_0) + \pi(p_R - z_R)$.

Proof: Appendix. If all elements of $G(s)$ are stable (all poles in the closed left half plane), then any net change in phase must have been caused by a different number of RHP zeros in $g_{ij}(s)det(G^{ij}(s))$ and $det(G(s))$. The direction of the phase change will then tell whether the numerator or denominator of (13) has the most RHP zeros. The theorem is useful, for example, if the frequency-response is known, but not the plant realization such that the zeros can be computed. The following Corollary is even more useful since it only requires knowledge about the diagonal RGA-elements at $\omega = 0$ and $\omega = \infty$ which may be avaialble even when the detailed dynamics are unknown.

Corollary 1 *Assume $\lim_{s\to\infty} \lambda_{ij}(s)$ is finite and different from zero. Consider a transfer matrix with stable elements and no zeros or poles at $s = 0$. If $\lambda_{ij}(j\infty)$ and $\lambda_{ij}(0)$ have different signs then at least one of the following must be true:*
a) $g_{ij}(s)$ has a RHP zero.
b) $G(s)$ has a RHP transmission zero.
c) $G^{ij}(s)$ (ie., the subsystem with input j and output i removed) has a RHP transmission zero.

All of these options may be detrimental for the control of the system. However, note that there may RHP-zeros present even if the RGA elements do not change sign. For example, adding a time delay or RHP-zero to an individual input or output channel will not effect the RGA as it may simply be viewed as a kind of scaling. In most cases the pairings are chosen such that $\lambda_{ii}(\infty)$ is postive (usually close to 1, see pairing rule 4 below) and this confirms Bristols claim that negative RGA-elements imply presence of RHP-zeros. In the case when the process does contain zeros or ploes at $s = 0$ the Corallary still applies if $\lambda_{ij}(0)$ is corrected by adding $\frac{\pi}{2}(p_0 - z_0)$ to the phase angle.

Example 1. Consider a plant

$$G(s) = \frac{1}{\tau s + 1}\begin{pmatrix} s+1 & s+4 \\ 1 & 2 \end{pmatrix} \quad (16)$$

We have $\lambda_{11}(\infty) = 2$ and $\lambda_{11}(0) = -1$. Since none of the diagonal elements have RHP-zeros we conclude from Corollary 1 that $G(s)$ must have a RHP-zero. This is indeed confirmed as $G(s)$ has a transmission zero at $s = 2$.

4.2 RGA and individual element uncertainty

Systems with large RGA values are sensitive to small relative errors in the individual elements of the system transfer matrix $G(s)$.

Theorem 3 *The (complex) matrix G becomes singular if we make a relative change $-1/\lambda_{ij}$ in its ij-th element, that is, if a single element in G is perturbed from g_{ij} to $g_{pij} = g_{ij}(1 - \frac{1}{\lambda_{ij}})$.*

(We have stated the theorem as a matrix property, but it will of course also apply to $G(s)$.) This is actually a quite amazing algebraic property of the RGA which seems to be largely unknown. The theorem was originally presented by Yu and Luyben (1987). However, their proof is somewhat difficult to follow, and a much simpler proof is presented here.
Proof. Let $G_p(s)$ denote $G(s)$ with g_{pij} substituted for g_{ij}. Using (13), we find by expanding the determinant of $G_p(s)$ by row i or column j that

$$det(G_p) = det(G) - \frac{det(G)}{(-1)^{i+j}det(G^{ij})}(-1)^{i+j}det(G^{ij}) = 0 \quad (17)$$

Example 2. Consider the matrix

$$A = \begin{pmatrix} 3 & 9 & 5 & 1 \\ 4 & 2 & 7 & 6 \\ 1 & 1 & 8 & 7 \\ 5 & 2 & 4 & 0 \end{pmatrix} \quad (18)$$

This matrix is non-singular as $\det(A) = 634$. The 2,4-element of the RGA is $\lambda_{24}(A) = 2.5836$. Thus the matrix becomes singular if $a_{24} = 6$ is perturbed to $a_{p24} = 6(1 - 1/2.5836) = 3.6777$. This is indeed confirmed since the resulting matrix has $\det(A_p) = 0$.

Theorem 3 is primarily an important algebraic property of the RGA, but it also has some important control implications:

1) The RGA-matrix $\Lambda(j\omega)$ is a direct measure of sensitivity to element-by-element uncertainty. 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. This is of course detrimental for control pefor-mance. However, as noted by Skogestad and Morari (1987b) the assumption of element-by-element uncertainty is usually poor from a physical point of view because the elements are always coupled in some way. The importance of Theorem 3 as a "proof" of why large RGA-elements imply control problems is therefore not as obvious as it may first seem. However, for process indentification the result is definitely useful as shown next.

2) 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 3 it is clear this method will most likely give meaningless results (eg., wrong sign of $det(G(0))$ or non-existing RHP- zeros) if there are large RGA-elements within the bandwidth where the model is intended to be used. Consequently, identification must be combined with first principles modelling if a good multivaraible model is desired in such cases.

4.3 RGA and diagonal input uncertainty

We stressed above that the element-by-element uncertainty generally is a poor uncertainty description. However, a kind of uncertainty that is always present, and which often limits achievable performance, is input uncertainty. We shall in this section allow $C(s)$ to be a full multivariable controller. Let the nominal plant model be G, and the true (perturbed) plant be $G_p = G(I + \Delta)$. Δ is a diagonal matrix consisting of the relative uncertainty (error) in the gain of each input channel. The true open loop gain G_pC can then be written in terms of the nominal loop gain GC and an "error term" $G\Delta C$.

$$G_pC = GC + G\Delta C \quad (19)$$

If a diagonal controller $C(s)$ is used then we simply get $G_pC = GC(I + \Delta)$ and there is no particular sensitivity to this uncertainty. On the other hand, with an inverse-based controller,

$C(s) = G^{-1}(s)K(s)$, where $K(s)$ is a diagonal matrix we get $G_p(s)C(s) = (I + G(s)\Delta G^{-1}(s))K(s)$. Here the diagonal elements of the error term prove to be a function of the RGA (Skogestad and Morari, 1987b)

$$(G\Delta G^{-1})_{ii} = \Sigma_{j=1}^{n} \lambda_{ij}(G)\Delta_j \quad (20)$$

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 couplings (as expressed by the large RGA- elements) and will yield poor performance (even nominally). On the other hand, an inverse-based controller which corrects for the interactions may yield excellent nominal performance, but will be very sensitive to input uncertainty and will not yield robust performance. The physical reason for the problems with the inverse-based controller is that the controller tries to apply large input signals in certain directions to match weak directions in the plant. The input uncertainty changes these directions and ruins the desired match. In addition, stability problems are also expected for the inverse-based controller. In summary, plants with large RGA-elements around the crossover- frequency are fundamentally difficult to control, and inverse-based controllers should never be used for such plants.

5 RGA and RDG and control performance

Design objectives. When model uncertainty is not explicitly taken into account the following nominal (N) design objectives may be considered for a decentralized control system:
1) Performance (NP): Performance of overall system
2) Stability (NS): Stability of overall system
3) Loop performance (NLP): Performance of individual loop
4) Loop stability (NLS): Stability of individual loop
5) Subsystem performance (NSP) and stability (NSS): Behavior of remaining system when some loops are removed (or possibly detuned)

In this paper we consider only NP, NS and NLS. For 2×2 systems objective 5 is a special case of objectives 3 and 4, but not otherwise.

5.1 Performance Relationships

Performance requirements. Assume that G and G_d have been scaled such that 1) the expected disturbances, z_k are less than one at all frequencies, and 2) the outputs y_i are such that the expected setpoint changes, r_j, are less than one. As a NP performance specification we shall require for any disurbance at a frequency ω that the offset $|e_i(j\omega)| < 1/|w_{di}(j\omega)|$ where $w_{di}(s)$ is a scalar performance weight, and for setpoints changes that $|e_i(j\omega)| < 1/|w_{ri}(j\omega)|$. Typically, both weights $|w_{di}(j\omega)|$ and $|w_{ri}(j\omega)|$ are large at low frequencies where small offset is desired. w_{ri} is usually about 0.5 at high frequencies to guarantee an amplification of high-frequency noise of less than 2. Thus we have a number of performance specifications we want satisfied simultaneously. In process control the requirements for set-point tracking are often lax and performance is mainly determined by disturbance rejection.

Setpoints. The closed loop relationship between setpoints and offset is $e(s) = -S(s)r(s)$. For $\omega < \omega_B$ we have $(I + G(s)C(s))^{-1} \approx (G(s)C(s))^{-1}$. From the definition of the RGA we have $[G^{-1}(s)]_{ij} = \lambda_{ji}(s)/g_{ji}(s)$, and the offset in output i to a setpoint change for output j becomes

$$e_i(s) = -[S(s)]_{ij} r_j(s) \approx -\frac{\lambda_{ji}(s)}{g_{ji}(s)c_i(s)} r_j(s); \quad \omega < \omega_B \quad (21)$$

For loop i we note that $|\lambda_{ii}|$ gives the approximate increase in offset caused by closing the other loops. For acceptable setpoint tracking it is required that

$$|g_{ji}c_i(j\omega)| > |\lambda_{ji} w_{ri}(j\omega)|; \quad \omega < \omega_B \quad (22)$$

We see that a large value of $|\lambda_{ji}|$ indicates poor response in output i to a setpoint change r_j.

Disturbances. The closed loop relationship between disturbances and outputs is $e(s) = (I + G(s)C(s))^{-1} G_d(s)z(s)$. For $\omega < \omega_B$ we have $(I + G(s)C(s))^{-1} \approx (G(s)C(s))^{-1}$. We get from the definition of the RDG that $[G(s)^{-1}G_d(s)]_{ik} = \beta_{ik}(s)g_{dik}(s) / g_{ii}(s)$. Consequently, we have for each output e_i and each disturbance z_k:

$$e_i(s) = [S(s)G_d(s)]_{ik} z_k(s) \approx \frac{\beta_{ik}(s)g_{dik}(s)}{g_{ii}(s)c_i(s)} z_k(s); \quad \omega < \omega_B \quad (23)$$

For loop i and disturbance k we note that $|\beta_{ik}|$ gives the approximate increase in offset caused by closing the other loops. For good disturbance rejection, we require $|e_i(j\omega)w_{di}(j\omega)| < 1$, that is

$$|g_{ii}c_i(j\omega)| > |\beta_{ik} g_{dik} w_{di}(j\omega)|; \quad \omega < \omega_B \quad (24)$$

We see that a large value of $|\beta_{ik} g_{dik}|$ indicates poor response in output i to a disturbance d_k. The important product $\delta_{ik}(s) = \beta_{ik}(s)g_{dik}(s)$ is henceforth termed the Closed Loop Disturbance Gain (CLDG). A plot of $|\delta_{ik}(j\omega)|$ will give useful information about which disturbances k are difficult to reject. If $|\delta_{ik}|$ is large, a poor response in loop i to disturbance k may be expected.

5.2 Comparison with previous work.

Mathematically, the performance specification used above is $\|[W_r S \quad W_d S G_d](j\omega)\|_{max} < 1, \forall\omega$ where the *max*-norm used spacially (channels) is the largest element in the matrix (sometimes called the spacial ∞-norm). $W_r = diag\{w_{ri}\}$ and $W_d = diag$ $\{w_{di}\}$ are diagonal matrices specifying the desired performance in in each output. This performance specification is very similar to the usual H_∞-norm, but in the latter case the induced 2-norm is used spacially. Consider the special case where $W_r = W_d = W_P$ and we have the H_∞ performance specification

$$\bar{\sigma}(W_P[S \quad SG_d](j\omega)) < 1, \forall\omega \quad (25)$$

Skogestad and Morari (1989) have shown how one from the NP-condition (25) may derive the tightest possible bounds on the individual loops, for example, in terms of bounds on $|h_i|$, $|s_i|$ or $|g_{ii}c_i|$. These results are very powerful, but unfortunately the same bound is used for all loops, and this may be conservative. It is possible to derive different and less conservative bounds by introducing additional adjustable parameters ("weights"), but it is not all obvious how this should be done *a priori* (see Nett and Uhtgenannt, 1988). Using the *max*-norm for the matrix as in (22) and (24) makes it is much simpler to derive tight bounds on the individual loops.

5.3 Implications for controller design

Above we derived bounds on the designs of the individual loops

$$|g_{ii}c_i(j\omega)| > b_i(\omega) \quad (26)$$

which when satisfied yield performance of the overall system (NP). For setpoint following (Eq.22) the bound $b_i(\omega)$ is given by the relative gains, λ_{ii}, and for disturbance rejection (Eq.(24) by the closed-loop disrurbance gains, δ_{ik}. It is desirable that the bound b_i be as small as possible because a large $b(\omega)_i$ requires a large bandwidth in loop i. Since our design approach requires stability of the individual loops (NLP), this may be impossible if $g_{ii}(s)$ contains time delays, neglected or uncertain dynamics, or rhp-zeros, which limits the achievable bandwidth, ω^*_{Bi}, for loop i.

5.4 Implications for control structure selection (loop pairing)

Note that all the above equations apply to a particular choice of pairings. However, since a rearrangenent of G gives the same rearragement in the RGA matrix, the following is clear from (22) for an input-output pair ji:

Pairing Rule 1. Avoid parings with large RGA-values $|\lambda_{ij}|$ (in particular at frequencies near cross-over).

Pairing Rule 2. Prefer pairings where $g_{ij}(s)$ puts minimal restrictions on the achievable bandwidth ω^*_{Bi}.

Rule 2 implies that one should prefer loops with minimum-phase behavior and without uncertain high-order uncertain dynamics, and is consistent with common engineering practice (eg., Balchen and Mumme, 1988, p.48). Rule 1 is also consistent with common design practice. However, note that (22) actually seems to imply that one should prefer pairings with as small values of $|\lambda_{ij}|$ as possible. This is not consistent with the usual rule of preferring pairing with λ_{ij} close to 1. The reason is that (22) is a performance (NP) condition only and does not guarantee stability (NS) of the overall system. Note that the loop gain $|g_ic_{ii}|$ increases approximately with a factor $|1/\lambda_{ii}|$ as the other loops are closed, and intuitively it seems obvious that small values of $|\lambda_{ii}|$ may create stability problems. These issues are discussed in some more detail below. The comment in Rule 1 regarding frequencies near crossover should be stressed since a number of authors indicate that large values of the RGA at steady-state is not acceptable (eg., Shinskey, 1984). Since there is generally no limitation on large loop gains at steady-state, Eq.(22) does not require small $\lambda_{ii}(0)$ for a pairing to be acceptable. Indeed, a study on control of distillation columns (Skogestad et al., 1990) has confirmed that even pairings corresponding to infinite $\lambda_{ii}(0)$ may be acceptable provided it approaches 1 at high frequencies. However, we might at this point add one rule which does apply to the steady-state RGA (recall Theorem 1):

Pairing Rule 3. Avoid pairings with $\lambda_{ij}(0) < 0$.

To use δ_{ik} for control structure selection one has to compute a separate RDG for each choice. Since there are $n!$ possible single-loop alternatives for a $n \times n$ plant this is quite a formidable task for large systems. One cannot use a large value of δ_{ik} to eliminate a *particular* loop pairing. However, one may conclude that

Pairing Rule 4. A control structure (an entire *set* of loops) should be avoided if it has large values of $|\delta_{ik}|$ at high frequencies, and in particular if the achievable bandwidth ω^*_{Bi} for the corresponding loop i is restricted (because of $g_{ii}(s)$, see rule 2).

5.5 Overall Stability (NS)

The relationships for performance derived above requires stability (NS) as a prerequisite. Since the stability of the individual loops (NLS) does not gurantee overall stability (NS) this issue has to adressed separately. Actually, most of the papers on interactions measures (eg., Grosdidier and Morari, 1986a) address this issue. Assuming G and $\tilde{G}$ have the same number of unstable poles and using Eq.(8), Grosdidier and Morari (1986) derive the following "small gain" stability condition:

$$\text{NS if } \rho(E_H\bar{H}(j\omega)) < 1 \forall\omega \quad (27)$$

Grosdidier and Morari used this condition to derive the tightest possible bound on the singular value of the individual designs: $\bar{\sigma}(\bar{H}) < 1/\mu(E_H)$ where $\mu(E_H)$ is the SSV interaction measure. However, here we are looking for a relationship between the NS-condition (27) and the RGA. There does not seem to be any simple relationship between the elements in E_H and Λ. One exception is the 2×2 case where we have (we are here considering the more general case of two blocks)

$$E_H\bar{H} = \begin{pmatrix} 0 & G_{12}G_{11}^{-1}\bar{H}_1 \\ G_{21}G_{22}^{-1}\bar{H}_2 & 0 \end{pmatrix} \quad (28)$$

Use the fact that $\rho\begin{pmatrix} 0 & A \\ B & 0 \end{pmatrix} = \sqrt{\rho(AB)}$ to derive

$$\text{NS} \quad \text{if} \quad \rho(G_{12}G_{11}^{-1}\bar{H}_1G_{21}G_{22}^{-1}H_2(j\omega)) < 1\forall\omega \quad (29)$$

Note that this condition may be weakened by using singular values to rederive Nett and Uthgenannt's (1988) condition for the optimally scaled SSV interaction measure. Our derivation is clearly much simpler. For scalar blocks the stability condition becomes

$$\text{NS} \quad \text{if} \quad |\bar{h}_1\bar{h}_2| < |\frac{g_{11}g_{22}}{g_{12}g_{21}}| = |\frac{\lambda_{11}}{1-\lambda_{11}}| \quad \forall\omega \quad (30)$$

The bound is obviously trivially satisfied if $\lambda_{11} = 1$. At low frequencies the product $|\bar{h}_1\bar{h}_2|$ is approximately 1, but it may have a peak value of $p > 1$ around crossover. At low frequencies (30) is then satisfied if $0.5 < \lambda_{11} < \infty$. At crossover the requirement is $\frac{p}{p+1} < \lambda_{11} < \frac{p}{p-1}$, for example, with $p = 2$ we have $0.67 < \lambda_{11} < 2$. Note that (30) may be conservative as it neglects phase information: The tightest NS-condition is that the overall system has no RHP-poles, where the pole polynomial is $1 - \bar{h}_1\bar{h}_2g_{12}g_{21}/(g_{11}g_{22})(s) = 0$ (Balchen, 1958). For example, it may be shown for 2×2 systems with $\bar{h}_1(0) = \bar{h}_2(0) = 1$ that the lower bound on $\lambda_{11}(0)$ should be 0 rather than 0.5. However, at higher frequencies the phase of $\bar{h}_1\bar{h}_2$ may take on almost any value and the above results give justification to the following rule:

Pairing rule 5. Prefer pairings with $\lambda_{ij}(j\omega)$ close to 1, in particular, for ω in the crossover region.

One implication of pairing rule 5 when applied to high frequencies is to prefer pairing on elements $g_{ij}(s)$ with a low-order transfer function ("pair on variables "close" to each other"). Again, this is in agreement with engineering practice.

Note that the diagonal elements in the matrix E_S are equal to $(1-\lambda_{ii})$. Nett (eg. Minto and Nett, 1989) has used this fact together with Eq.(9) to show that large changes in $(H -$

$\bar{H})\bar{H}^{-1}$ will result if $|\lambda_{ii}-1|$ is large at high frequencies. These results provide further justification for rule 5 above. Again, we want to stress the importance of the high-frequency region. For example, McAvoy (1983) reports on control of a head-box in paper manufacturing. The chosen pairing corresponds to $\lambda_{ii}=0$ at steady state. However, at high frequencies $\lambda_{ii}\approx 1$ and control proved successful.

6 Example

In order to demonstrate the use of the frequency dependent RGA and RDG for evaluation of expected control performance and control structure selection, a binary distillation column with 40 theoretical trays plus a total condenser is considered. This is the same example as studied by Skogestad et al. (1988), but we use a much more rigorous model with both composition and liquid dynamics. Using model reduction, the number of states in the model was reduced from 82 to 22. Disturbances in feed flowrate F (z_1) and feed composition z_F (z_2), are included in the model. The LV configuration is used, that is, the manipulated inputs are reflux L (u_1) and boilup V (u_2). Outputs are the product compositions y_D (y_1) and x_B (y_2). The model then becomes

$$\begin{pmatrix} dy_1 \\ dy_2 \end{pmatrix} = G(s)\begin{pmatrix} du_1 \\ du_2 \end{pmatrix} + G_d\begin{pmatrix} dz_1 \\ dz_2 \end{pmatrix} \qquad (31)$$

The disturbances and outputs have been scaled such that a magnitude of 1 corresponds to a change in F of 30%, a change in z_F of 20%, and a change in x_B and y_D of 0.01 molefraction units.

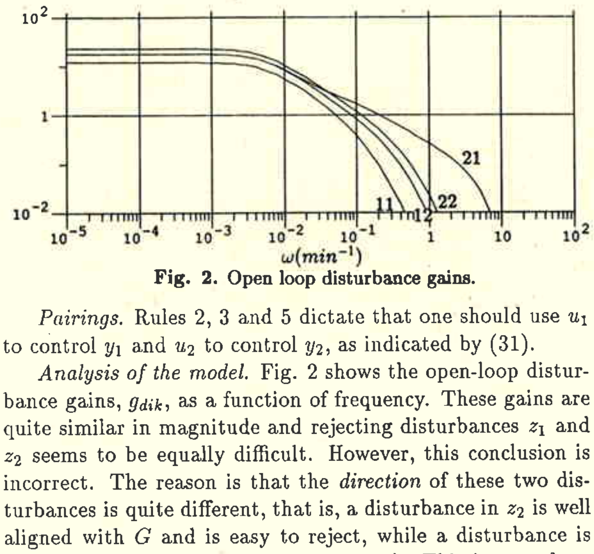

Fig. 2. Open loop disturbance gains.

Pairings. Rules 2, 3 and 5 dictate that one should use u_1 to control y_1 and u_2 to control y_2, as indicated by (31).

Analysis of the model. Fig. 2 shows the open-loop disturbance gains, g_{dik}, as a function of frequency. These gains are quite similar in magnitude and rejecting disturbances z_1 and z_2 seems to be equally difficult. However, this conclusion is incorrect. The reason is that the *direction* of these two disturbances is quite different, that is, a disturbance in z_2 is well aligned with G and is easy to reject, while a disturbance is z_1 is not (Skogestad and Morari, 1987a). This is seen from Fig. 3 where the closed-loop disturbance gains, δ_{i2}, for z_2 are seen to be much smaller than δ_{i1} for z_1. The relative gains for the loops are also included in Fig.3 (note thet $\lambda_{11}=\lambda_{22}$ for 2×2 plants). We see that rejection of disturbance z_1 (as indicated by $|\delta_{i1}|$) and setpoint following (as indicated by $|\lambda_{ii}|$) put similar bounds on the loop gain $|g_{ii}c_i|$. Assuming that the performance requirement around crossover corresponds to performance weights $|w_d(j\omega_B)|\approx|w_r(j\omega_B)|\approx 1$ we find that the bandwidth requirement for both loops is about 0.5 min^{-1}.

Observed control performance. To check the validity of the above results we designed single-loop PI controllers by

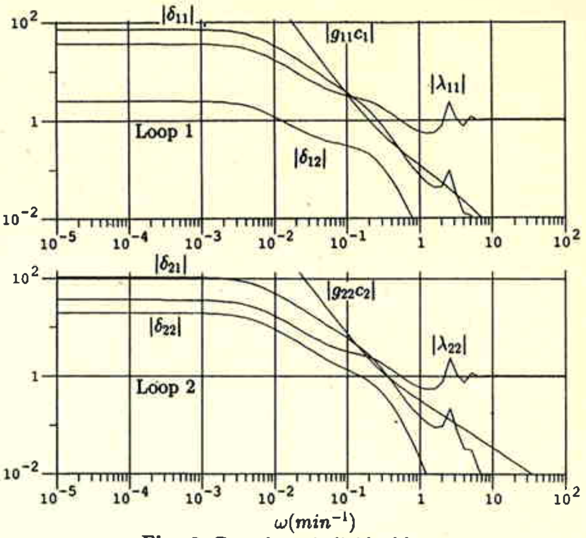

Fig. 3. Bounds on individual loops.

optimizing robust performance using (25) as the performance specification.

The loop gains, $|g_{ii}c_i|$ with these controllers are shown in Fig. 3. The loop gains are seen to be larger than the closed-loop disturbance gains δ_{ik} at all frequencies up to the crossover. Closed-loop simulations with these controllers are shown for disturbance 1 and 2 in Fig. 4 and 5. These simulations confirm that disturbance 2 is much easier rejected than disturbance 1. Note from Fig. 3 that δ_{2k}'s for loop 2 are higher than the δ_{1k}'s for loop 1. This indicates poorer control of y_2 than of y_1, but this is not confirmed by the simulation in Fig.4. The reason is that the controllers are designed such that $|g_{22}c_2|>|g_{11}c_1|$, that is, loop 2 is faster, and this compensates for the larger disturbance gain in loop 2.

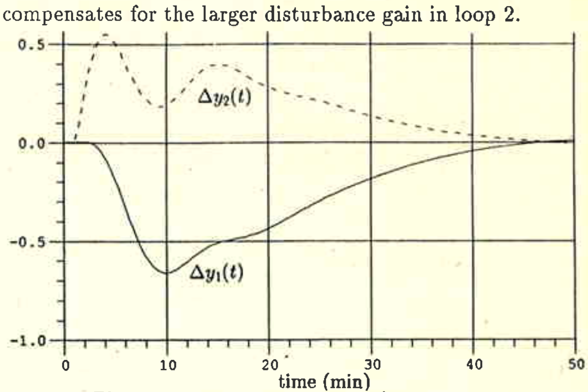

Fig. 4. Response in $\Delta y_1(t)$ (solid line) and $\Delta y_2(t)$ (dotted line) to z_1 (a 30% increase in feed rate).

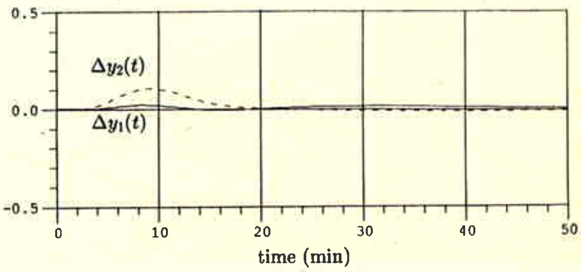

Fig. 5. Response in $\Delta y_1(t)$ (solid line) and $\Delta y_2(t)$ (dotted line) to z_2 (a 20% increase in the fraction of the more volatile component in the feed).

In summary, there is an excellent correlation between the analysis based on Fig. 3 and the simulations. This is not surprising when one considers Fig. 6 which shows that accuracy of the approximation $[S(s)G_d(s)]_{ik} \approx \frac{\delta_{ik}(s)}{g_{ii}(s)c_i(s)}$ which was used to derive Eq.(24) and which formed the basis for the analysis in Fig.3. The approximation is very good at low frequencies, but as expected poorer at frequencies around the closed loop bandwidth. From Fig. 6 we see that the actual disturbance rejection in this frequency range is better than the approximation (however, for other examples this may be different). In particular, the deviation is large for $ik = 12$, and this explains why the effect of z_2 on y_1 in Fig. 5 is even smaller than expected from Fig. 3.

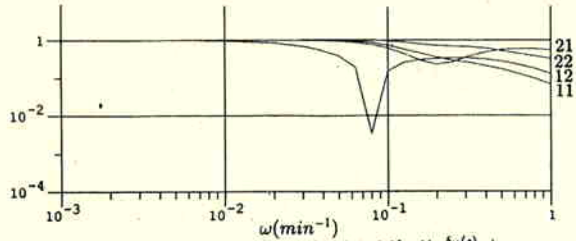

Fig. 6. The magnitude of $[S(s)G_d(s)]_{ik}/(\frac{\delta_{ik}(s)}{g_{ii}(s)c_i(s)})$

7 Conclusions

In the paper we have derived bounds on the designs of the individual loops. These bounds depend on the model of the process only, that is, are independent of the controller. This means that frequency- dependent plots of λ_{ii} and δ_{ik} may be used to evaluate the achievable closed-loop performance (controllability) under decentralized control. Plants with small values of these measures are preferred. Furthermore, the values of δ_{ik} may tell the engineer which disturbance k will be most difficult to handle using feedback control. This may pinpoint the need for using feedforward control, or for modifying the process. For example, in process control adding a feed buffer tank will dampen the effect of disturbances in feed flowrate and composition. Plots of $\delta(j\omega)$ may be used to tell if a tank is necessary and what holdup (residence time) would be needed.

Nomenclature (also see Section 2 and Fig. 1)

$e = y - r$ - vector of offsets
g_{ij} - ij'th element of G
g_{dij} - ij'th element of G_d
G^{ij} - G with row i and column j removed
r - vector of reference outputs (setpoints)
u - vector of manipulated inputs
w_{di} - performance weight for disturbance rejection in loop i.
w_{ri} - performance weight for setpoint following in loop i.
y - vector of outputs
z - vector of disturbances
β_{ik} - ik'th element of RDG matrix
$\delta_{ik} = [G^{-1}G_d]_{ik}/g_{ii} = \beta_{ik}g_{dik}$ - Closed Loop Disturbance Gain.
Λ - RGA matrix
λ_{ij} - ij'th element of Λ
$\mu(A)$ - structured singular value of matrix A
ω - frequency
ω_B - closed loop bandwidth
$\rho(A)$ - spectral radius of matrix A
$\bar{\sigma}(A)$ - maximum singular value (spectral norm) of matrix A

Subscripts
i - index for outputs or loops
j - index for manipulated inputs or setpoints
k - index for disturbances

References

Arkun, Y., 1987, "Dynamic block relative gain and its connection with the performance and stability of decentralized structures", *Int. J. Control*, **46**, 4, 1187-1193.

Arkun, Y., 1988, "Relative Sensitivity: A Dynamic Closed-Loop Interaction Measure and Design Tool", *AIChE Journal*, **34**, 4, 672-675.

Balchen, J.G., 1958, *Lecture notes*, Norwegian Institute of Technology, Trondheim, Norway (In Norwegian).

Balchen, J. G. and Mumme, K. , 1988 *Process Control. Structures and Applications*, Van Nostrand Reinhold, New York.

Bristol, E.H., 1966, "On a new measure of interactions for multivariable process control", *IEEE Trans. Automat. Control*, **AC-11**, 133-134.

Grosdidier, P., Morari, M. and Holt, B.R., 1985, "Closed-Loop Properties from Steady-State Gain Information", *Ind. Eng. Chem. Fundam.*, **24**, 221-235.

Grosdidier, P., and Morari, M., 1986, "Interaction Measures for Systems Under Decentralized Control", *Automatica*, **22**, 3, 309-319.

Manousiouthakis, V., Savage, R. and Arkun, Y., 1986, "Synthesis of Decentralized Process Control Structures Using the Consept of Block Relative Gain", *AIChE Journal*, **32**, 6, 991-1003.

McAvoy, T.J., 1983, *Interaction Analysis. Principles and Applications*, ISA Monograph Series 6, Triangle Park.

Minto, K. D. and Nett, C.N., 1989, A Quantitative Approach to the Selection and Partitioning of Measurements and Manipulations for the Control of Complex Systems." Presented at ACC 1989, Pittsburgh, PA.

Nett, C.N. and Uthgenannt, J.A., 1988, "An explicit formula and optimal weight for the 2-block structured singular value interaction measure", *Automatica*, **24**, 2, 261-265.

Shinskey, F.G., 1967, *Process Control Systems*, McGraw-Hill, New York.

Shinskey, F.G., 1984, *Distillation Control*, 2nd Edition, McGraw Hill, New York.

Skogestad, S. and Morari, M., 1987a, "Effect of Disturbance Directions on Closed Loop Perfomance", *Ind. Eng. Chem. Res.*, **26**, 2029-2035.

Skogestad, S. and Morari, M., 1987b, "Implications of Large RGA Elements on Control Performance", *Ind. Eng. Chem. Res.*, **26**, 11, 2323-2330.

Skogestad, S., Morari, M. and Doyle, J.C, 1988, "Robust Control of Ill- Conditioned Plants: High Purity Distillation", *IEEE Trans. Autom. Control*, **33**, 12, 1092-1105.

Skogestad, S. and Morari, M., 1989, "Robust Performance of Decentralized Control Systems by Independent Designs", *Automatica*, **25**, 1, 119-125.

Skogestad, S., Lundström, P. and Jacobsen, E.W., 1990, "Selecting the best distillation control configuration", *AIChE Journal*.

Slaby, J. and Rinard, I.H., 1986, "A complete interpretation of the dynamic relative gain array", Presented at AIChE Annual Meeting, Miami Beach.

Stanley,G., Marino-Galarraga, M., McAvoy, T.J., 1985, "Shortcut operability analysis. 1. The relative disturbance gain. *Ind. Eng. Chem. Process Des. Dev.*, **24**, 4, 1181-1188.

Tung, L.S. and Edgar, T.F., 1981, "Analysis of Control-Output Interactions in Dynamic Systems", *AIChE Journal*, **27**, 4, 690-693.

Witcher, M.F. and McAvoy, T.J., 1977, "Interacting Control Systems: Steady State and Dynamic Measurement of interacion", *ISA Transactions*, **16**, 3, 35-41.

Yu, C.-C. and Luyben, W.L., 1987, "Robustness with Respect to Integral Controllability", *Ind. Eng. Chem. Res.*, **26**, 1043-1045.

Appendix. Proof of Theorem 2. Consider Eq.(13) as a function of frequency, ie., let $s = j\omega$. Since $\lim_{s\to\infty} \lambda_{ij}(s)$ is finite and different from zero, Eq.13 may be written as a fraction of two polynomials in s where the numerator polyomial and denominator polynomial are of the same order. The phase change in $\lambda_{ij}(j\omega)$ as ω goes from 0 to ∞ must then be caused by RHP-poles or zeros in $g_{ij}(s)$, $det(G^{ij}(s))$ or $det(G(s))$ and the theorem follows.