

NOTICE WARNING CONCERNING COPYRIGHT RESTRICTIONS:

The copyright law of the United States (title 17, U.S. Code) governs the making of photocopies or other reproductions of copyrighted material. Any copying of this document without permission of its author may be prohibited by law.

OPERABILITY, RESILIENCY, AND FLEXIBILITY -
PROCESS DESIGN OBJECTIVES FOR A CHANGING WORLD

by

I.E. Grossmann & M. Morari

December, 1983

DRC-06-55-83

OPERABILITY, RESILIENCY AND FLEXIBILITY -
PROCESS DESIGN OBJECTIVES FOR A CHANGING WORLD

by

Ignacio E. Grossmann
Chemical Engineering Department
Carnegie Mellon University
Pittsburgh, Pa 15213

Manfred Morari
Chemical Engineering Department
University of Wisconsin
Madison, WI 53706

ABSTRACT

Chemical plants have to be designed based on uncertain thermodynamic, kinetic and equipment performance correlations, market forecasts, raw material and product prices. They have to be able to adjust to changes in operating conditions, disturbance levels, product specifications, product distribution and demand, and they should be able to tolerate equipment malfunction without leading to serious safety hazards. The general term "operability" will be used to describe the ability of the plant to perform satisfactorily under conditions different from the nominal design conditions. Current industrial practice accounts for operability at the design stage in an ad hoc fashion through empirical overdesign factors and by introducing large storage and surge tanks for raw materials, products and intermediates. It is shown here that the heuristic approach is not only often costly and ineffective but that it can have an adverse

UNIVERSITY LIBRARIES
CARNegie-MELLON UNIVERSITY
PITTSBURGH, PENNSYLVANIA 15213

effect: a design modification intended to improve operability can actually make it worse. Systematic methods to include operability as a design objective are reviewed and directions for future research in this area are given. Application examples demonstrate the utility of the suggested approaches.

The best part of our knowledge is that which teaches us where knowledge leaves off and ignorance begins

Oliver Wendell Holmes

INTRODUCTION

It is not uncommon that in our function as engineering educators we encounter students with highly developed scientific and mathematical skills which are unable to solve relatively simple engineering design problems. The difficulties arise from the fact that these problems tend to be more loosely defined; they require a series of assumptions to be made before an answer can be obtained with reasonable effort in an acceptable time period. Because of the uncertainties inherent in the problem formulation there is rarely only one "correct" or "best" solution which adds further complications and confusion. The skills required from the student for "design" are quite the opposite of what is stressed in the early training when exact scientific reasoning is emphasized.

Even after the early education period one of the main challenges in engineering design remains that decisions have to be made based on limited knowledge. For example, the engineer has to deal with the uncertainty in the kinetic data, the thermodynamic correlations, the projected equipment performance, the feedstock quality, the product price and demand etc. More important when designing a chemical plant it is generally either not possible or not desirable to remove all the uncertainties involved. For example, it is not possible to predict prices with certainty even with an unlimited marketing research budget. Or, for example, it might be economically unsound to try to obtain more accurate performance correlations for distillation column trays.

This discussion makes clear that it is very hard to provide an unambiguous definition of the concept of "optimal design". The

Improvements obtained in a given economic objective function by involved optimization techniques might become insignificant in the engineering sense when compared to the uncertainties involved in the problem formulation. Furthermore some of the more complex design objectives like safety, reliability and operability are often equally or more important than the economic objective (e.g. return on investment) but much more difficult to quantify.

In somewhat oversimplified terms one could say that this paper is about uncertainty in engineering design. How much of it constitutes a healthy dose in a particular situation, how one can cope with it and how one can plan for it through the appropriate design; or in the words of management guru Peter Drucker, what "we have to do today to be ready for an uncertain tomorrow". We want to show that engineering rules of thumb can fail quite miserably, what better techniques are available, what their drawbacks are and where future research should be directed. In order to do that we first have to establish a common vocabulary with clear definitions of all the terms involved.

Synthesis vs. Analysis: By synthesis we mean the integration of processing units into a system (plant, control system, etc.) such that it has specified properties. By analysis we mean the examination of an existing systems¹ properties.

Structure vs. Parameters: A system can be specified by its structure and its design parameters. The design parameters which correspond to sizes, flowrates, pressures, temperatures, etc. can generally take on all real values within specified bounds. The structure (presence or absence of system parts and their interconnections)*can also be described by design parameters. However, these design parameters are restricted to the integer values "0" and "1", where "0" denotes the absence and "1" the presence of a certain system part or interconnection.

Design vs. Control: In design, decisions are made once and for all before a plant is constructed. In control, decisions are made continuously during the operation of the plant.

We will use the general term operability to describe the ability of the plant to perform satisfactorily under conditions different from the nominal design conditions. The major objectives that are to be achieved in the operability of a chemical plant include the following:

- a) Feasibility of steady-state operation for a range of different feed conditions and plant parameter variations
- b) Fast and smooth changeover and recovery from process disturbances
- c) Safe and reliable operation despite equipment failures
- d) Easy start-up and shut-down.

The first two objectives deal with the satisfactory performance of the plant during periods of "normal" plant operation, whereas the last two objectives are concerned with the plant performance during "abnormal" operation. This distinction is of course somewhat arbitrary, but it reflects the expected time of operation: most of the time the plant will be under "normal" operation, whereas the occurrence of "abnormal" operation is much less frequent. This paper will deal exclusively with the first two objectives (a) and (b), and the purpose will be to present how these objectives can actually be addressed at the design stage.

The attributes that denote the first two objectives of operability will be denoted in this paper as Flexibility and Resiliency. The dictionary defines resiliency as the "power of recovery after strain". In the context of process design we mean by it the ability of the plant to tolerate and to recover from undesirable changes and upsets. For example, the plant can tolerate parameter variations and it can easily recover from process disturbances in a fast and smooth manner. We will refer to the former quality as "static" resiliency when only steady state operation is considered, and we will refer to the latter quality as "dynamic resiliency". * ^

The dictionary defines flexibility as the ability to readily adjust to meet the requirements of changing conditions. For example, a flexible plant can be adapted to different feedstocks, product specifications or process conditions. In summary, the main difference is that resiliency refers to the maintenance of satisfactory performance despite adverse conditions while flexibility is the ability to handle alternate (desirable) operating conditions. Needless to say the distinction between resiliency and flexibility is not always clear cut. However, the emphasis in resiliency is on the

dynamic operation of the plant, whereas the emphasis in flexibility is on the steady-state operation.

Armed with these definitions we can now proceed with four motivating examples which should demonstrate the practical importance of operability, and the failure of simple minded heuristic rules to incorporate operability as one of the design objectives.

SOME MOTIVATING EXAMPLES

Overdesign or underdesign for resiliency and flexibility?

The conventional procedure for introducing resiliency in a chemical plant is to use empirical overdesign. That is, a nominal or "conservative" basis is selected for designing and optimizing the plant. Empirical factors are then applied to the sizes of equipment and extra units are also often introduced. However, although this empirical procedure will in general add resiliency and flexibility of operation to a plant, it has the following drawbacks:

1. Not much insight is gained on the actual degree of flexibility that is obtained in the chemical plant*
2. Conditions that give rise to infeasible operation may not be detected due to the fact that the interactions among the different units in the process are not explicitly taken into account.
3. The resulting overdesigned plant may not operate efficiently and may not be optimal from an economic viewpoint.

In order to illustrate some of these drawbacks, and in particular the problem of overlooking effects of interactions, consider the example of the heat exchanger network shown in Fig. 1. Note that in this case the outlet temperatures of streams H and C2 have been specified in the form of inequalities: stream H must be cooled down to at least 410K, while stream C2 must be heated up to at least 430K.

Assume that the areas of exchangers 1 and 2 are sized with the nominal values of heat transfer coefficients $U_i = U_2^s 800 \text{ W/m}^2$, and that the resulting areas are oversized by 20%. If such a design were implemented in practice the following situation might occur:

Suppose that U_i is 20% higher than the nominal value while U_2 is 20% lower. For such a case, as is shown in Fig. 2, the exit

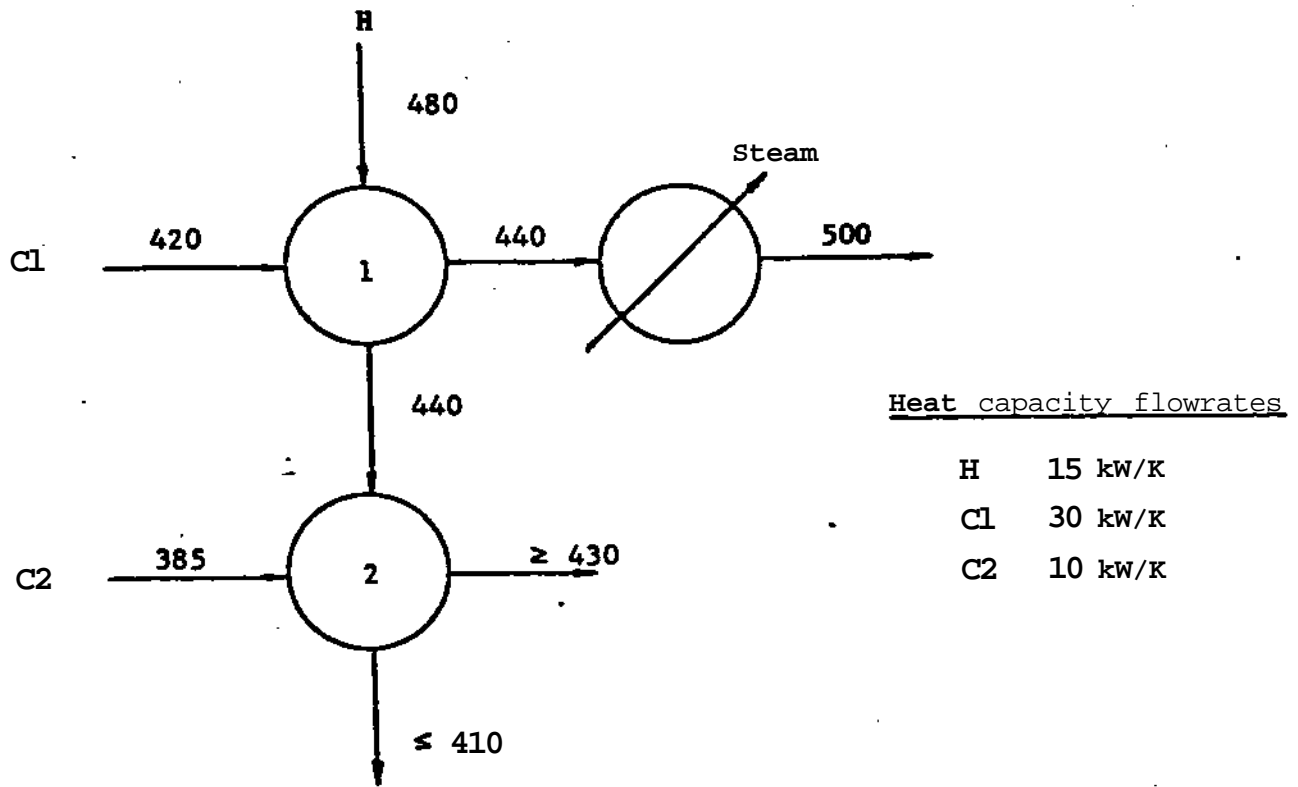


Fig. 1 Heat exchanger network for example 1

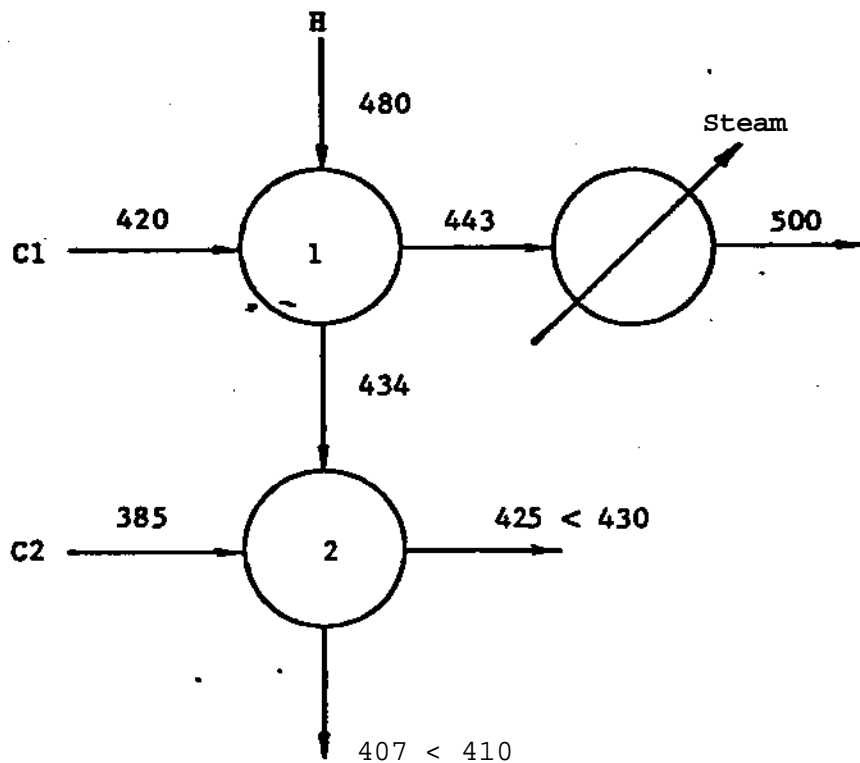


Fig. 2. Performance of network with 20%L oversizing of exchangers 1 and 2 when heat transfer coefficients are + 20% and - 20% respectively

temperature of stream H from exchanger 1 would drop from the expected 440K down to 434K due to the larger transfer coefficient. However, with this temperature change the temperature driving force in exchanger 2 is reduced, which when coupled with the lower transfer coefficient causes the outlet temperature of C1 from this exchanger to be 425K, or 5K below the minimum temperature that was specified. Therefore, for the above cited realization of transfer coefficients the network exhibits infeasible operation since it violates the temperature specification.

It should be noted that this network design satisfies the temperature specifications when both heat transfer coefficients are 20% lower than the nominal values, which intuitively would be regarded as the "worst" condition. This example illustrates then the danger of overlooking interactions when using empirical overdesign. Furthermore, it shows that identifying "worst" conditions for feasible operation may not always be obvious from intuition. This observation will be elaborated on further in the next section.

Another point of interest in the example is related to the choice of areas such that temperature specifications are not violated for any deviation within $\pm 20\%$ of the nominal values of U_1 and U_2 . For instance, if one were to insist in oversizing the area of exchanger 1 by 20%, one would find that the area of exchanger 2 would have to be oversized by 108%! On the other hand, if one were to oversize exchanger 2 by 23%, one would find that the first exchanger would not have to be oversized, but rather it would have to be undersized by 16%! This then shows that the choice of a resilient design which in addition is economically optimal, may not be quite obvious in general. Hence, the need for a systematic treatment of resiliency and flexibility in process design should be evident.

What constitutes a "worst" operating condition?

Traditional industrial practice generates resilient systems by designing them for what are perceived to be "extreme" operating conditions. Naturally, if these extremes are selected properly the system will perform satisfactorily for the whole range of expected situations. The following example is meant to demonstrate that the

proper selection of "extremes" is far from trivial and that seemingly logical choices can lead to extremely poor systems. For the problem data in Table 1 the network shown in Fig. 3A was designed. There are no other designs with a smaller number of heat transfer units, the approach temperatures fall nowhere below 10°C and therefore this structure is likely to be close to optimal economically. It is known that the heat capacity flowrate of stream HI can be as large as 1.85 at times. The natural approach of the design engineer would be to test his design for this extreme condition. The test reveals that the network structure performs satisfactorily also at this flowrate (Fig. 3B). It appears then logical to expect that the structure can handle all flowrates in the range between 1 and 1.85. Figure 3C reveals that this is not the case. Even if exchanger 1 had an infinite area, for a flowrate of 1.359 the outlet temperature of HI cannot be decreased below 71° . With a reasonable approach temperature difference of 10° (Fig. 3D) the minimum attainable outlet temperature for HI is 102.2° , corresponding to a target violation of 52° . In particular if HI were the feed stream to a reactor this design error could have serious consequences.

By switching the cooler from H2 to HI the network can be made flexible (Fig. 3E). In all exchangers the approach temperatures exceed 10°C over the whole range of flowrate variations $1 < \text{WHI} < 1.85$ and therefore the capital costs remain reasonable. The example shows that flexibility can be reached not through additional exchangers or excessive oversizing but rather by a proper redesign of the network structure.

Let us also look at the slightly modified problem where the inlet temperature of stream C2 is increased to 120°C ($T^{\text{sc}2} = 120^{\circ}\text{C}$). The network structure used in Fig. 3A can be demonstrated to suffer from the same deficiencies as previously. The flexible structure is shown in Fig. 3F. It involves only three heat exchangers while the other one had four. Selecting networks with a larger number of transfer units does not only increase capital costs but can lead to a decrease in flexibility. Flexibility cannot be accomplished by ad hoc addition of equipment but by systematic design techniques based on a thorough understanding of the physico-mathematical problem.

Stream No.	w(kW/°C)	T ^S (°C)	T ^T (°C)
HI	1	310	50
H2	2	450	280
C1	3	40	120
C2	2	115	290

Table 1: Stream data for Example 2

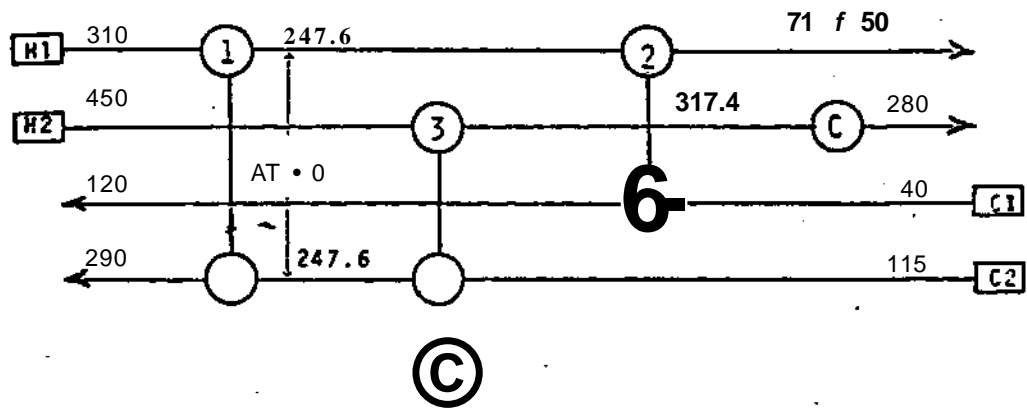
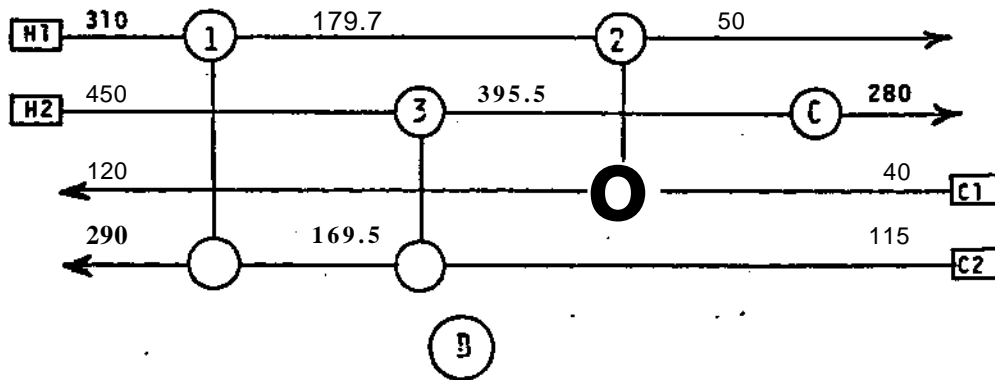
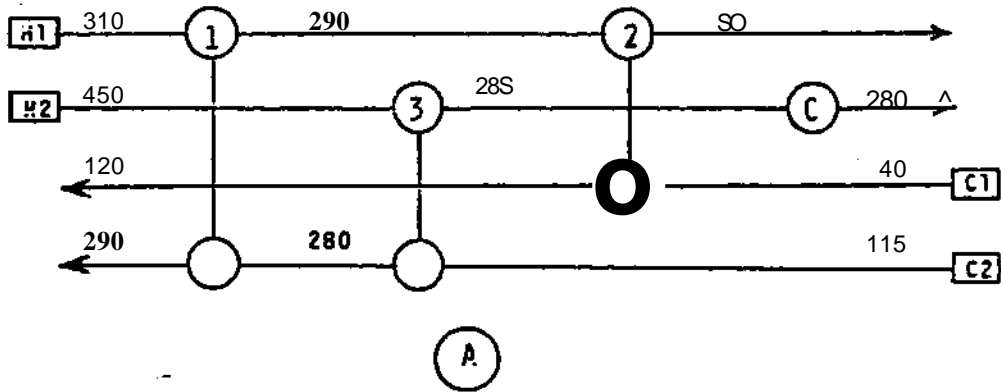


Fig. 3.A,B,C: Stream temperatures in heat exchanger network for example 2. A, B: Extreme flowrates of stream H1. C: Temperature target violation by 21° despite exchanger with infinite area.

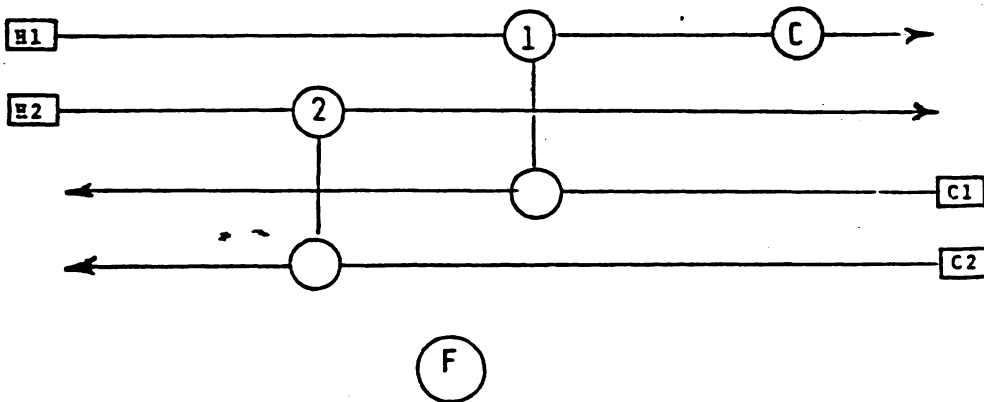
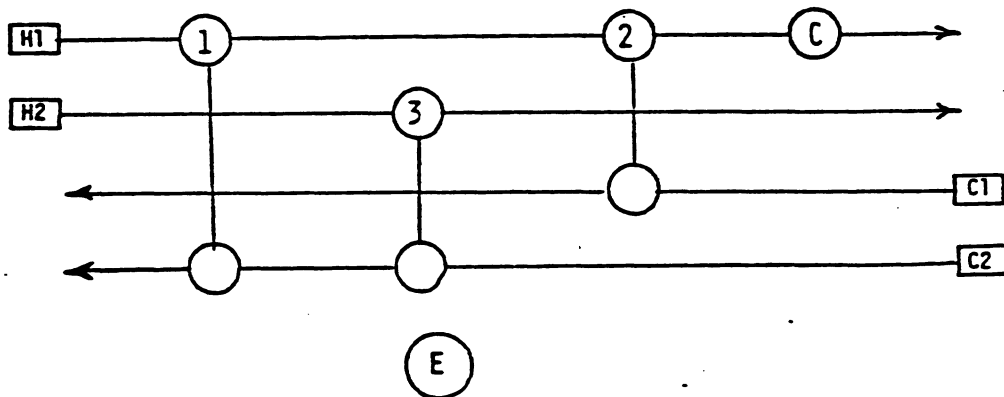
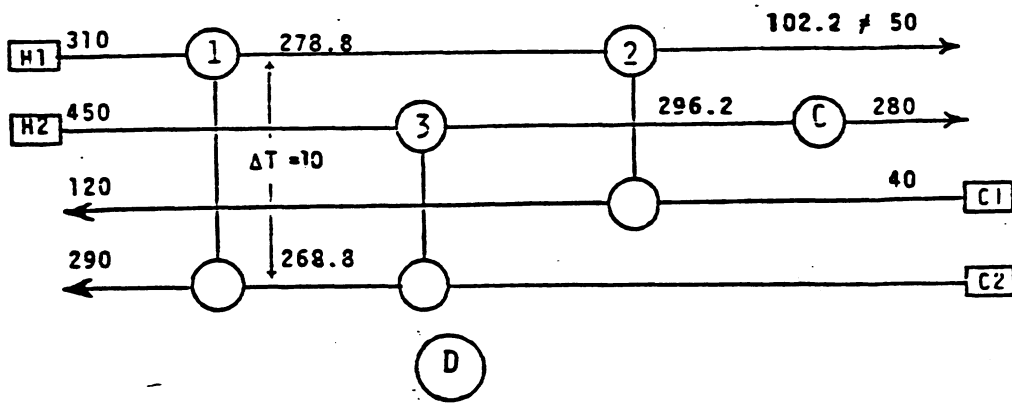


Fig. 3.D,E,F: Stream temperatures in heat exchanger network for example 2. D: Temperature target violation by 52° with exchanger of reasonable size. E: Flexible network structure. F: Flexible network structure for modified example 2 ($T_{C2}^S = 120^\circ\text{C}$).

Are long dead times bad for dynamic resilience?

The dynamic behavior of even quite complicated chemical engineering systems like distillation columns or reactors can usually be approximated well by first order plus dead time models. When dead times dominate the dynamics, as is often the case, they can exert severe limitations on the response time and therefore on the dynamic resilience of a process. Consequently they are of central importance in a resilience analysis.

In the single-input-single-output (SISO) case the detrimental effect of dead times is clearly proportional to their magnitude. Confronted with a number of possible SISO designs when dead times dominate, the design engineer generally chooses the system with the smallest delay. If dead times are significant but other aspects of the dynamics are also important, then dead times can serve as the basis of a first rough screening procedure where those designs whose dead times are significantly larger are removed from consideration. A more detailed analysis can then be carried out to select the proper design from those remaining.

In the multi-input-multi-output (MIMO) case the analysis of the effect of dead time on dynamic resilience is significantly more complicated. Let us consider the following example: It is desired to control the outlet temperatures of the streams 2, 3 and 4 of the heat exchanger network shown in Fig. 4. The heat exchangers are assumed to be distributed throughout a plant and the transport lag between heat exchangers dominates the dynamics. The network is controlled by the bypasses on heat exchangers 2, 4 and 5 with the transport lag between heat exchangers indicated in Fig. 4. It can be shown (Holt, Morari, 1983) that for a dynamically decoupled system the best servo response, i.e. output response to set point changes, is described by the transfer matrix

$$G = \text{diag}(e^{-6s}, 1, e^{-4s})$$

That is the stream 2 temperature responds to a set point change after 6 minutes, stream 3 immediately and stream 4 after 4 minutes. It is not surprising that by decreasing some of the delays between the exchangers the response can be made faster.

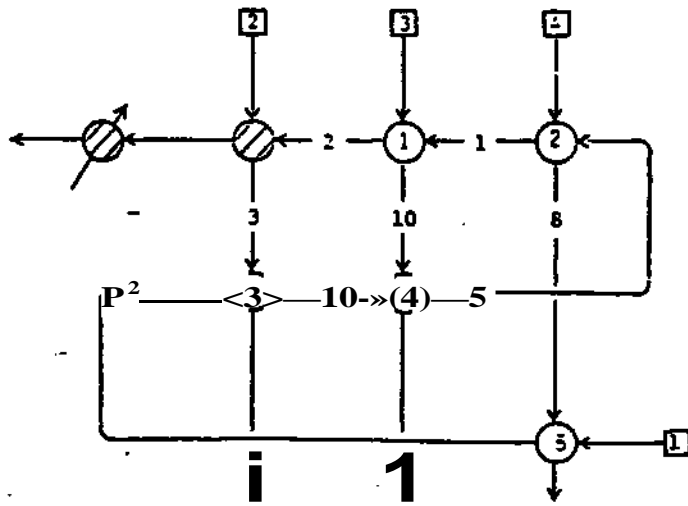


Fig. 4. Heat exchanger network with transport lag (Example 3).

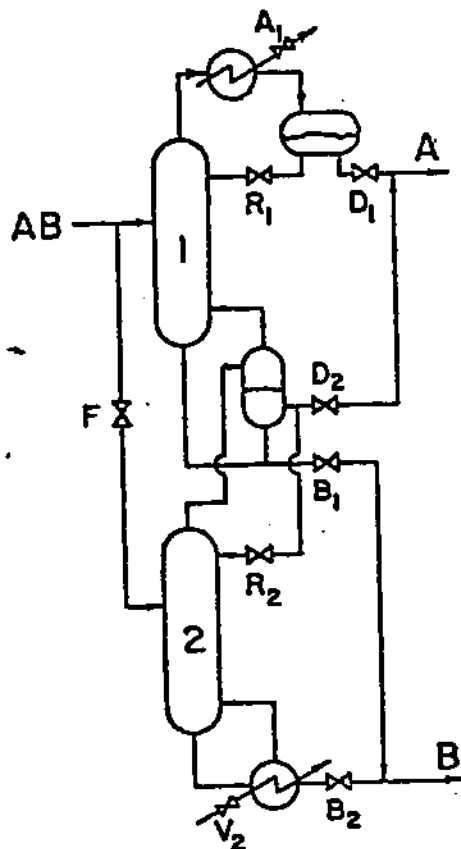


Fig. 5. Thermally coupled distillation columns (Example 4).

It may not always be possible to decrease dead times but in many cases it is possible to increase them. Contrary to the SISO case and normal intuition this can lead to improved resilience in the MIMO case. Continuing the example suppose it is not physically possible to shorten the distance between units but it is possible to increase the transport lag. Consider increasing the lag between heat exchangers 10 and 11 from 2 to 6 minutes. Then it can be shown (Holt, Morari, 1983) that the best servo response is described by the transfer matrix

$$G = \text{diag}(e^{-6s}, 1, 1)$$

Thus increasing a time delay has resulted in a significant improvement of the best achievable performance in a dynamically decoupled system.

Again, it is clear that simple rules of thumb are unable to explain the effect of design changes on resilience and more rigorous techniques are called for.

How resilient are controlled systems to plant parameter variations?

In order to compensate for disturbances and to speed up the adjustment of the plant to new operating conditions or in other words, to make the plant more resilient, automatic controllers are employed. The tuning of the controllers is always based on a model of the process, albeit sometimes a grossly simplified one. The models are invariably inaccurate because of identification problems and time varying system characteristics. It is desirable that the control performance be insensitive to modelling errors, such that the required modelling effort can be kept to a minimum and frequent retuning of the controllers can be avoided. Clearly the sensitivity is a function of the control system design, but as we will demonstrate next, even more so of the system itself.

Consider the system of coupled distillation columns {20 stages each) shown in Fig. 5 which is used to separate a 70% methanol/water mixture into a 99% methanol distillate and a 0.1% methanol bottom product (Example 4). The detailed model and all the parameters are reported by Lenhoff and Morari (1982). We will investigate two different control structures for these columns.

	Manipulated Var.	Fixed Var.
Structure 1	R_1, V_2	F, R_2
Structure 2	F_1, V_2	D, R_f

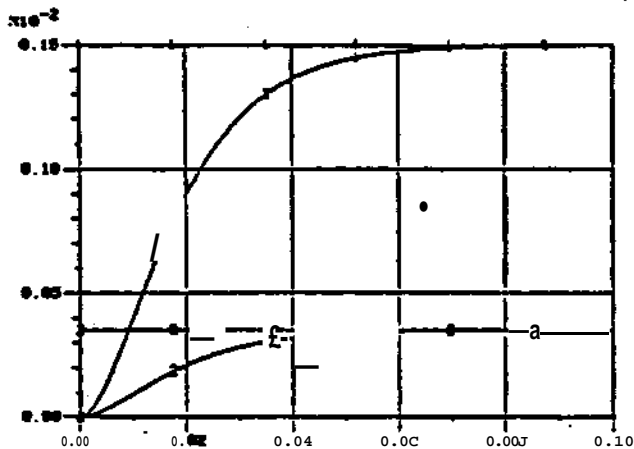
Here the "manipulated variables" are the two valves used for composition control and the "fixed variables" are not used at all for control. The other variables shown in Fig. 5 are employed in loops maintaining the mass balance in the column. These loops are assumed "fast" and are not included in the model.

For both structures 1 and 2 the multivariable composition controllers were tuned based on the linear model to yield the response to set point changes shown in Fig. 6A. The question is now how well the controllers would work for the two systems if the real plant were different from the model. To mimic a "real" plant time delays in length equal to about 8% of the dominant open loop time constants were introduced into the models. The performance of the controllers on the "real" plant is shown in Fig. 6B and C. Even for this unrealistically small modelling error the performance deteriorates significantly for structure 1 and much less for structure 2. For a slightly larger time delay or gain error the system with structure 1 would become unstable. Naturally, any designer will opt for structure 2 which promises to allow a much simpler control system design and to require less modelling effort. Instead of the control structure we could have varied other design parameters and similar effects on the sensitivity could have been observed (Saboo, 1982).

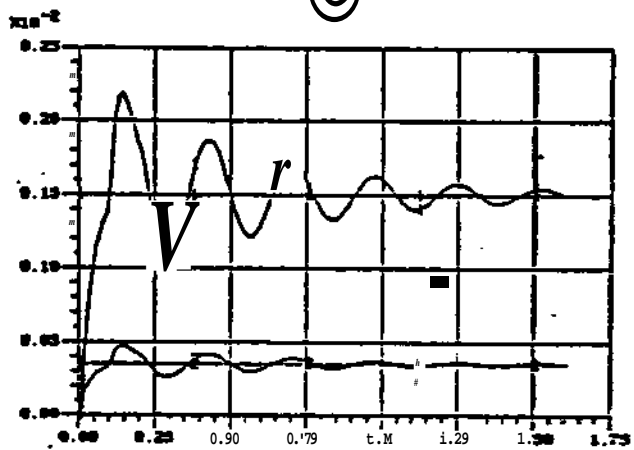
The example demonstrates that design decisions can have a very pronounced effect on the dynamic resilience of a plant. Therefore, it would be highly desirable to have a reliable criterion to assess the dynamic resilience at the design stage, which does not require extensive simulation runs. This is especially important since - in the authors¹ opinion at least - the observed sensitivity differences do not seem obvious on physical grounds and no heuristic rules suggest themselves.

The moral from the examples

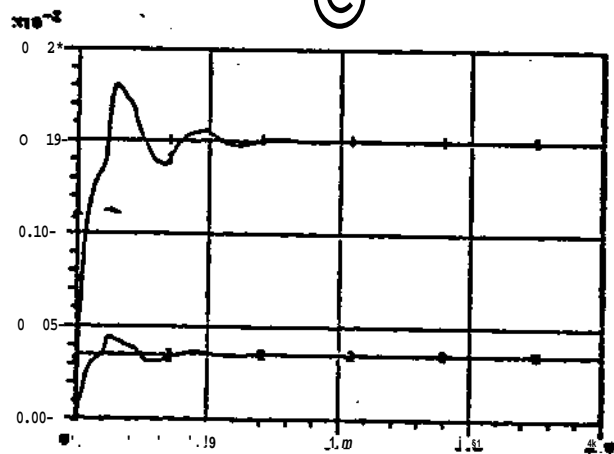
- 1) The oversizing of existing and the addition of new units into a process is not only costly but can lead to a decrease



(C)



(C)



(C)

Fig. 6. Overhead and bottom product composition response to set point changes vs. time. A: Structures 1 & 2 without modelling error. B: Structure 1 with modelling error. C: Structure 2 with modelling error.

in flexibility and resilience.

- 2) What constitutes a "worst" operating condition on which to base a conservative design is impossible to determine for a complex plant with many interacting pieces of equipment without a systematic analysis tool.
- 3) Longer dead times can sometimes improve a plants dynamic resilience.
- 4) Design changes can have very pronounced but difficult to predict effects on the sensitivity of the performance of a controlled system to modelling errors and thus on the dynamic resilience.

Much progress has been made over the last few years toward the understanding of these counterintuitive phenomena and the foundations have been laid for a framework that will allow operability considerations to become an integral part of the design process.

FLEXIBILITY AND STATIC RESILIENCY

Problem definition

A first step in incorporating operability considerations at the design stage is to provide an adequate treatment of operational flexibility or static resiliency. As mentioned in the introduction section, these attributes are mainly concerned with the problem of ensuring that a plant is able to handle a number of different steady-state conditions during periods of normal operation. For example, this would involve the capability of processing different feedstocks, producing different products, operating at various capacity levels or -at a variety of process conditions. In other words, the basic concern in flexibility or static resiliency is to ensure feasible steady-state operation of the plant not only for a single nominal condition, but rather for a whole range of conditions that may be encountered in the operation.

Since the dynamic behavior of the process will be neglected in this section it is impossible to distinguish mathematically between resiliency and flexibility; in both cases the plant has to cope with parameter variations in the steady-state. These parameters involve uncertainties in internal process conditions such as catalyst activity

and heat transfer coefficients. Alternatively, the parameters involve uncertainties in external process variations such as feed or ambient conditions. Because our emphasis will be on the mathematical formulation we will use for simplicity the term flexibility.

It is important to point out that design decisions related to selecting the process configuration, equipment sizes and mode of operation, all have an impact in determining the flexibility of a process. This was clearly demonstrated in the preceding examples. However, the impact will in general be much greater at the synthesis stage where the process configuration is selected. Furthermore, since for the flexibility of the process to be "optimal" requires also that the advantages of flexibility be balanced mainly in relation to its cost, flexibility in design requires that it be incorporated early in the synthesis stage as well as in the more detailed stages of design. This clearly requires the development of a variety of analysis and optimization tools which have to be based on a solid foundation that captures the basic nature of the flexibility problem, which is on establishing the existence of feasible regions of operation.

For most design applications flexibility in chemical processes is determined through the allowable variations of a vector of uncertain parameters p . In the case of conventional design procedures these parameters are usually treated as fixed nominal values, and typically they correspond to feed or ambient conditions, or process parameters such as reaction constants, transfer coefficients and other physical properties. Since the values of these uncertain parameters can normally be expected to change widely during the plant operation, it is a major design objective to ensure that the chemical plant has the required flexibility to operate over a given range of parameter values.

A substantial number of methods have been reported in the literature for dealing with parameter uncertainties in process design. These methods have as a major objective to optimize a given flow sheet configuration while introducing flexibility according to some specified criteria or strategies. The methods consider the design problem as given by the optimization problem

$$\begin{aligned} \min \quad & C(d,u,x,p) \\ \text{s.t.} \quad & h(d,u,x,p) = 0 \\ & g(d,u,x,p) < 0 \end{aligned} \tag{1}$$

where C is an economic objective function, h and g are vectors of equalities and inequalities that define the performance and specifications of the design; and d , u , x are the vectors of design, control and state variables for the process. The basic difference in the methods lies on how the effect of the vector of uncertain parameters p is taken into account for introducing flexibility. A recent review of these methods is given in Grossmann et al. (1982), and Table 2 lists a selected number of contributions. However, rather than discussing in detail the relative advantages of these methods, it would seem more appropriate for the purpose of this paper to discuss some of the main issues that are involved in the synthesis and design of flexible chemical processes.

I. Information on the uncertain parameters The first important question in flexibility is on the kind of information that is available on the uncertain parameters. Clearly, it is the task of the designer to decide first as to what are the particular parameters in the design that should be treated as uncertain. Since conceptually these parameters can be regarded as random variables, their probability of realization would be given by a distribution function. However, the difficulty in practice is that these distribution functions are normally not available at the design stage since no measurements can be made to infer them. Furthermore, although one could conceptually assign economic penalties for those parameter realizations which cause violations in the design specifications, it has to be recognized that accurate knowledge on penalties is also normally not available. Therefore, the minimum amount of information that can be expected is the nominal parameter value p^N , as well as its expected range which is specified in the form of lower and upper bounds

$$P^L < P < P^U \tag{2}$$

Clearly, the actual value of these bounds may be somewhat arbitrary since in general they have to be provided by the designer. However, these bounds could in principle be derived from a distribution function if it were available, so that they would represent confidence limits. In either case the parameter bounds in (2) are to a great extent meaningful and easy to interpret since they can be used to define the parameter ranges for which it is desired to guarantee feasible operation. In this way the designer has the capability of specifying explicitly the regions of operation which are of interest in a flexible design. On the other hand, it is recognized that since flexibility implicitly defines a probability for feasible operation, the designer may have to provide some subjective distribution function in order to define the expected economic performance of the design.

Finally, it should also be noted that in general the uncertain parameters will not necessarily be independent, in which case they will typically be related by algebraic relationships which very often can be expressed in terms of a subset of independent parameters.

II. Specification of flexibility requirements The flexibility of a design is determined by its capability to meet constraints and specifications for a range of conditions. However, it is clear that in practice not necessarily all constraints will have to be satisfied exactly when considering a variety of operating conditions in a plant. In general, there will be on the one hand "hard" constraints which cannot or should not be violated under any circumstances, and on the other hand there will be "soft" constraints which can be violated to some degree without affecting significantly the performance of the system. An example of the former type of constraints would be safety constraints or product specifications, whereas examples of the latter type would be specifications on minimum temperature approaches or maximum outlet temperatures of cooling water which can often be relaxed to some extent. One possible approach to handle the two types of constraints would be to enforce both the "hard" and "soft" constraints at the nominal point, but only the "hard" constraints for the other parameter values. However, whatever type of approach is used to handle these constraints, it is ultimately

the designer who has to decide which constraints in a design should be strictly enforced. Therefore, when one refers to the flexibility of a design, one has to realize that this is a relative concept and by no means a universal attribute. Flexibility merely reflects the capability of feasible operation with respect to the desired goals that are set by the designer, and which are expressed explicitly in the form of constraints that must be satisfied.

III. Flexibility problems The three basic types of problems that would seem to be most relevant for synthesizing and designing flexible processes are the three following:

a) Flexibility Analysis. The first subproblem addressed here would be on how to test feasibility of operation of a design given specified bounds of the uncertain parameters. This would help to assert that the design has the required flexibility, or else, it would allow to identify those parameter values that lead to infeasible operation. The second subproblem which is more general would be to measure the inherent flexibility in a design. For example, this could involve the computation of a scalar index that would reveal the size of the parameter space over which feasible operation can be attained. This would allow, for instance, the evaluation of flexibility for different process configurations at the synthesis stage, and also to identify parameter values which limit the flexibility in each design. Swaney and Grossmann (1983) have recently proposed an index of flexibility for this purpose, and Morari (1983) has also proposed an index for measuring the flexibility and resiliency in heat exchanger networks.

b) Optimal Synthesis and Design with Fixed Degree of Flexibility. The problem addressed here would be to obtain a minimum cost design which is feasible to operate over a prespecified parameter range. Most of the work shown in Table 2 has concentrated on this type of problem for the case of fixed flowsheet configurations. For the synthesis case the main work that has been published is by Morari and coworkers for maximum energy recovery networks (Marselle, Morari, Rudd, 1982; Saboo, Morari, 1983). It should be pointed out that the major challenge in this class of problems not only lies in optimizing the economics of the process, but also in obtaining a design for which

Table 2

Methods reported in the literature for optimal design
with uncertain parameters

<u>Authors</u>	<u>Design Strategy</u>
Kittrel and Watson (1966)	Min. expected cost
Wen and Chang (1968)	Min. expected cost and/or max. change in cost function
Avriel and Wilde (1969)	Two-stage programming and per- manently feasible as applied to geometric programming
Weisman and Holzman (1972)	Min. expected cost with penalties for constraint violations
Watanabe, Nishixmura and Matsubara (1973)	Min. combination expected cost and maximum probable cost
Takamatsu, Hashimoto and Shioya (1973)	Min. deviation of cost from nominal point while satisfying linearized constraints
Nashida, Ichikava and Tazaki (1974)	Minimax strategy
Freeman and Gaddy (1975)	Min. expected cost for given level of dependability
Dittmar and Hartmann (1978)	Min. deviation of cost from nominal point while satisfying linearized constraints
Johns, Marketos and Rippin (1978)	Multiperiod two-stage programming
Grossmann and Sargent (1978)	Two-stage programming with feasibil- ity constraint
Malik and Hughes (1979)	Two-stage programming
Halemane and Grossmann (1983)	Two-stage programming with feasibil- ity constraint

feasible operation can be guaranteed for the specified parameter range. This clearly involves as an important subproblem the flexibility analysis.

c) Design with Optimal Degree of Flexibility. This is a generalization of the previous problem as it is concerned with establishing proper trade-offs, mainly between flexibility and the economics of the process. Due to the common lack of information on penalties for constraint violations a suitable approach would be to develop trade-off or pareto-optimal curves that can help the designer to decide on what is an "optimal" degree of flexibility, either when developing the structure of the process or when sizing the equipment. Swaney and Grossmann (1983) provide a framework to accomplish this objective, but much work remains to be done in this area.

IV. Design strategy Since in flexibility the main concern is to ensure feasible steady-state operation for a variety of conditions, a very important question is on the type of plant operation that should be anticipated at the design stage. Clearly, chemical plants have a number of variables that can be adjusted or manipulated during the operation (e.g. flows, temperatures, pressures). These variables, which can be regarded as control variables, represent degrees of freedom that ought to be considered not only for optimal operation, but also to attain feasible operation for the different parameter realizations. Although this might seem a trivial point, it is interesting to note that only few of the methods listed in Table 2 address explicitly this problem.

For instance methods based on minimizing the expected cost

$$\begin{aligned} \min_{d,u} \quad & E\{C(d,u,x,p)\} \\ \text{s.t.} \quad & h(d,u,x,p) = 0 \\ & g(d,u,x,p) < 0 \end{aligned} \tag{3}$$

assume that a single choice of the control variables u is made for all the parameter realizations.

Similarly, in the case of methods based on the minimax strategy

$$\begin{aligned} \min_{d,u} \max_p C(d,u,x,p) \\ h(d,u,x,p) = 0 \\ g(d,u,x,p) < 0 \\ pL < p < pU \end{aligned} \tag{4}$$

the single choice of the control variable u is for the "worst" economic outcome in the plant. In this way these two approaches fail to account for the fact that control variables can be adjusted during operation. Furthermore, the actual methods that have been proposed do not guarantee feasible operation for the range of parameters considered.

A much more suitable strategy for flexibility is the two-stage programming strategy

$$\begin{aligned} \min_d E\{\min_{p,u} C(d,u,x,p)\} \\ h(d,u,x,p) = 0 \\ g(d,u,x,p) < 0 \end{aligned} \tag{5}$$

in which it is assumed that the control variables u are adjusted for every parameter realization to achieve optimal operation. This is clearly a more ambitious strategy, but it is more realistic in that it anticipates more closely the way in which chemical plants are actually operated. It should be noted, however, that from the methods based on this approach (see Table 2) only the ones proposed by Grossmann and Sargent (1978) and yalemane and Grossmann (1983) have as an explicit objective feasible operation for the selected parameter range by proper manipulation of the control variables.

In summary, as one can see from this section the type of problems and assumptions involved in flexibility give rise to rather challenging research questions for developing useful analysis and synthesis tools. However, as will be shown in the next sections, a number of ideas has emerged in the area of flexibility over the last few years which have helped to gain some fundamental understanding and insight into the nature of this problem.

Parametric region of feasibility

For a design in which the process configuration and equipment sizes are given, the steady-state performance of the system can be represented by a system of nonlinear equations

$$h(d_f u, x_f p) = 0 \quad (6)$$

where h is the vector of equations consisting of heat, material balances and design equations

d is the vector of design variables which define the equipment sizes

u is the vector of control variables

x is the vector of state variables

p is the vector of uncertain parameters

It should be noted that the control variables u do not necessarily have to correspond to variables that can be physically manipulated in the plant, but rather they represent a suitable selection of degrees of freedom in (6) when the design variables and uncertain parameters have fixed values. In other words, the only requirement is that the control variables u be selected such that the system of equations in (6) is solvable for the state variables x , given fixed values for the vectors d and p .

The feasibility requirements of the system are specified through the vector of inequalities,

$$g(d, u, x, p) < 0 \quad (7)$$

which define product specifications, allowable ranges for state or control variables or other types of physical constraints that should hold in the process.

Since for a fixed design and parameter values, the state variables can be expressed from Eq. (6) as an implicit function of u

$$h(d, u, x, p) = 0 \quad \bullet \quad x = x(d, u, p) \quad (8)$$

the inequalities in (7) can be expressed in a reduced form as

$$g(d, u, x(d, u, p)) = f(d, u, p) < 0 \quad (9)$$

In this way the Inequalities, $f(d,u,p) < 0$, determine the feasibility or infeasibility of steady-state operation for a chosen control u , when for a given design d the plant operates at the parameter value p . Note that since the control variables represent degrees of freedom that may be adjusted to suit the prevailing conditions, feasibility of operation for a given d and p , requires only the existence of some control u for which the constraints can be satisfied, i.e. $f(d,u,p) < 0$.

In order to determine the actual set of parameter values p for which feasible steady-state operation can be attained, the following parametric region of feasibility can be defined,

$$R = \{p \mid \exists u \mid f(d,u,p) < 0\} \quad (10)$$

This region R defines the set of parameter values p for which control variables u exist such that the reduced inequalities in (9) can be satisfied. This region provides then the basic information on the flexibility of operation of a given design. An example of this region is depicted in Fig. 7.

A computationally more convenient, but equivalent form for defining the feasibility region R has been shown by Swaney and Grossmann (1983) to be given by

$$R = \{p \mid U(d,p) < 0\} \quad (11)$$

where

$$U(d,p) = \min_u \max_{j \in J} \{f_j(d,u,p)\} \quad (12)$$

which alternatively can be written as the nonlinear program

$$\begin{aligned} t(d,p) = \min_{u,a} a \\ \text{s.t. } a > f_j(d,u,p) \quad j \in J \end{aligned} \quad (13)$$

where J is the index set for the constraints, and a is a scalar variable.

This function $U(d,p)$ provides a quantitative measure of feasibility ($U < 0$) or infeasibility ($U > 0$) for the chosen design d

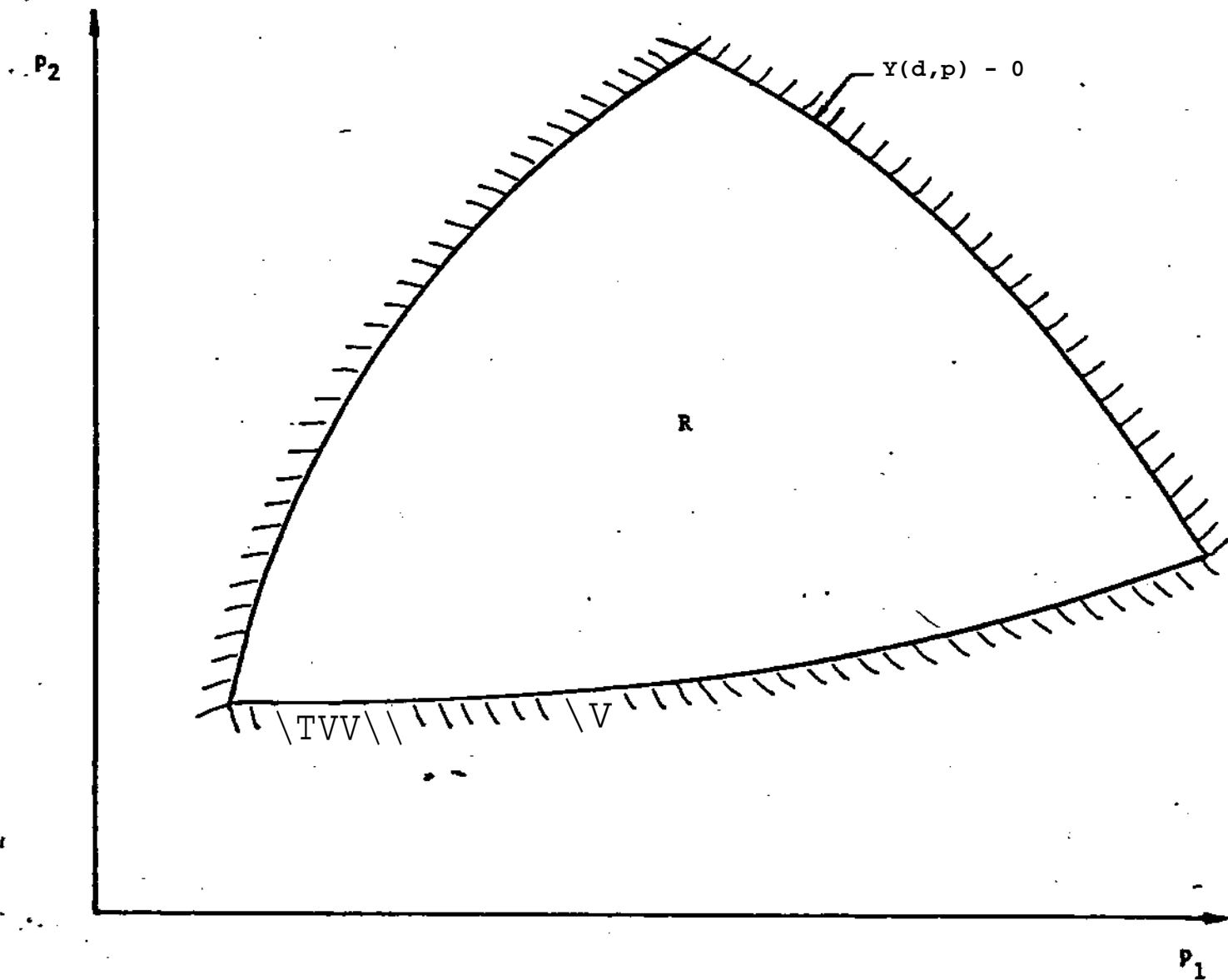


Fig. 7. Parametric feasible region of operation R for a fixed design.

at the parameter p , since it determines the values of the control variables for which the maximum constraint j is minimized. The significance of this function $*(d,p)$ is that it provides a systematic way for defining the region R , and in particular its boundary which is given by the parameters p for which $*(d,p) = 0$ as is shown in Fig. 7.

It is also interesting to note that if the gradients of the constraints f_j are linearly independent the number of active constraints in (13) is $\dim(u) + 1$ (Swaney and Grossmann, 1983). Therefore, in order to solve for the function $t(d,p)$ all that is required in general is to solve a system of nonlinear equations for the appropriate set of active constraints in (13).

Flexibility Analysis

The definition of the feasible region R as given in (10) or (11) provides the conceptual framework that is required for analyzing the feasibility of operation for the specified set of bounded parameters,

$$P - (p_{lP} < p < p^u) \quad (14)$$

which for simplicity in the presentation will be assumed to be independent.

As indicated previously, the set P can be interpreted as the desired parameter range for feasible steady-state operation specified by the designer. Figure 8 illustrates the case when the set P , which is a rectangle, is feasible for the region R since it is totally contained within that region. On the other hand, Fig. 9 shows an example where the rectangle P is infeasible since part of the parameter points in P lies outside from the feasible region R .

When analyzing the flexibility of a given design it is also of interest to determine the maximum feasible parameter set that a given design can handle. In order to accomplish this task, assume that we define the family of parameter sets

$$P(\delta) = \{p | p_N - \delta A_{p-} < p < p_N + \delta A_{p+}\} \quad (15)$$

where $A_{p-} = p_N - p_L$, $A_{p+} = p_U - p_N$, p_N is a feasible nominal parameter point and δ is a scalar variable. The sets $P(\delta)$ can be interpreted as

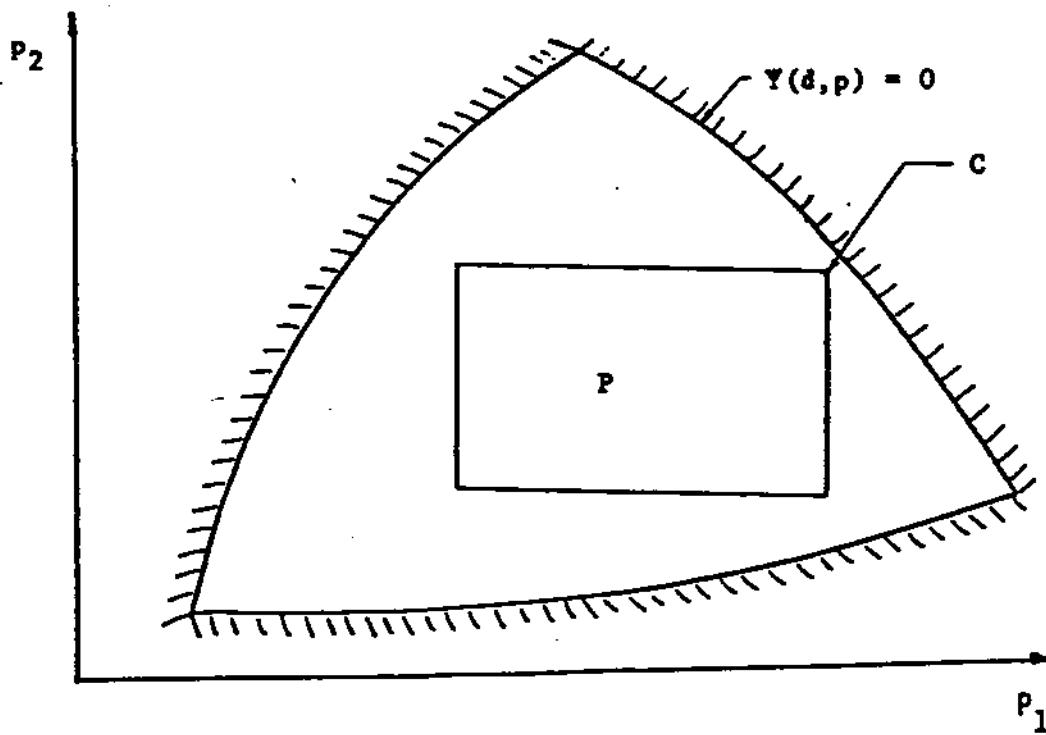


Fig. 8. Feasible parameter set P .

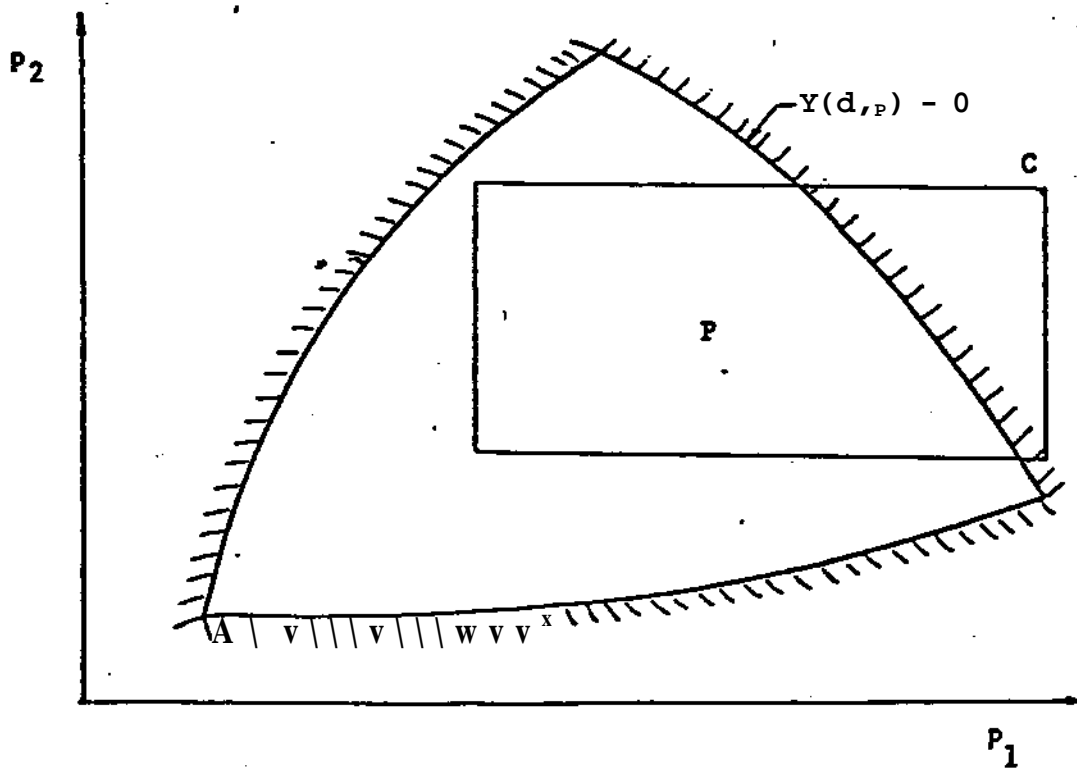


Fig. 9. Infeasible parameter set P .

a family of hyper-rectangles of different sizes that are expanded around the nominal point (Fig. 10). The sides of the rectangles are proportional to the expected parameter deviations Δp^+ , Δp^- , and to the scalar δ which defines through (15) the actual size of the parameter set. Note that if the scalar $\delta = 1$, then $P(1) = P$; i.e. we have the rectangle defined by (14). On the other hand if $0 < \delta < 1$, the hyper-rectangle $P(\delta)$ is a subset of the specified parameter set P .

The motivation in defining the family of parameter sets in (15) is that it provides a way for quantitatively measuring flexibility in a given design. That is, by determining the maximum parameter set $P(\delta)$ that can be inscribed within the feasible region R , it is possible to define the scalar flexibility index F as

$$\begin{aligned} F &= \max \delta \\ \text{s.t. } P(\delta) &\subseteq R \end{aligned} \quad (16)$$

The index F defines then through (15) the size of the maximum hyper-rectangle $P(F)$ that can be inscribed in the feasible region R . Furthermore, this flexibility index defines through (16) the actual parameter bounds

$$\begin{aligned} \hat{p}^L &< p < \hat{p}^U \\ \hat{p}^L &= p^N - F\Delta p^- , \quad \hat{p}^U = p^N + F\Delta p^+ \end{aligned} \quad (17)$$

for which feasible steady-state operation can be guaranteed.

Note that a design featuring a flexibility index $F > 1$ exceeds the specified bounds for feasible operation. On the other hand a flexibility index $0 < F < 1$ implies that the design can only operate within a maximum fraction F of any of the expected deviations. The example in Fig. 11 shows a rectangle that defines the flexibility index F .

Critical points for flexibility

A very important concept when analyzing the flexibility of a design is the notion of critical parameter points. Qualitatively, these points can be interpreted as "worst" conditions for feasible operation. Clearly if these points could be predicted a priori, they

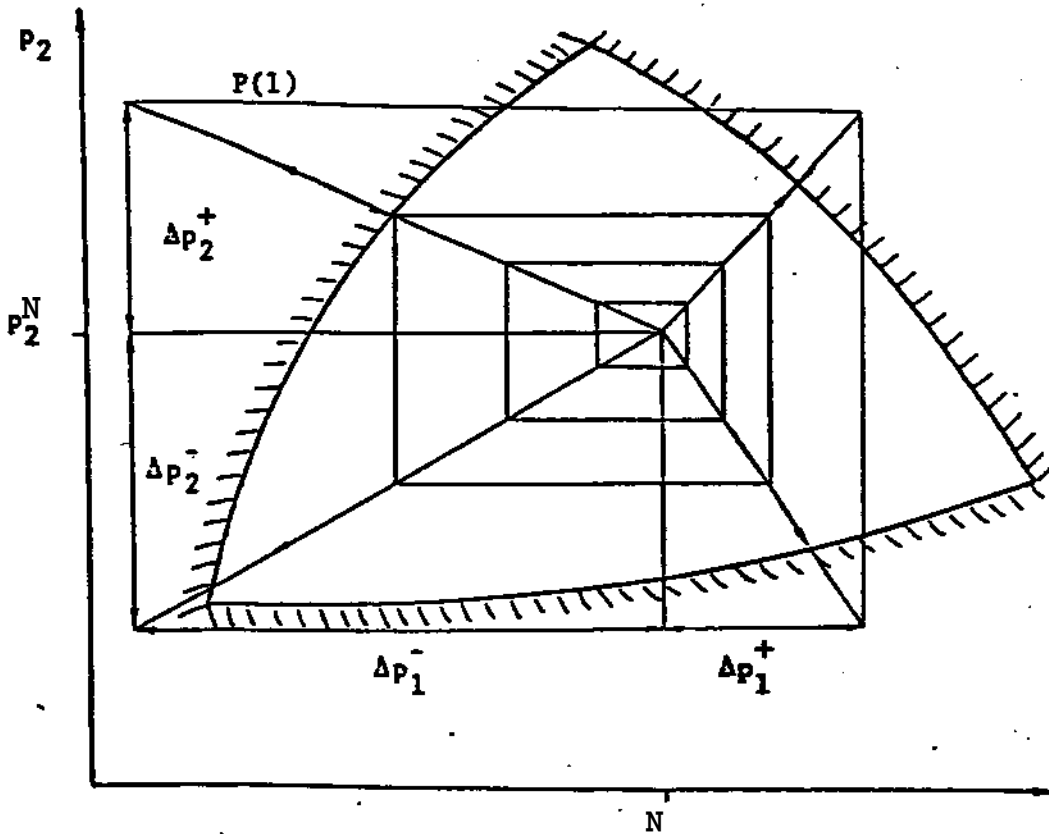


Fig. 10. Parameter sets $P(6)$ expanded from nominal point p^N .

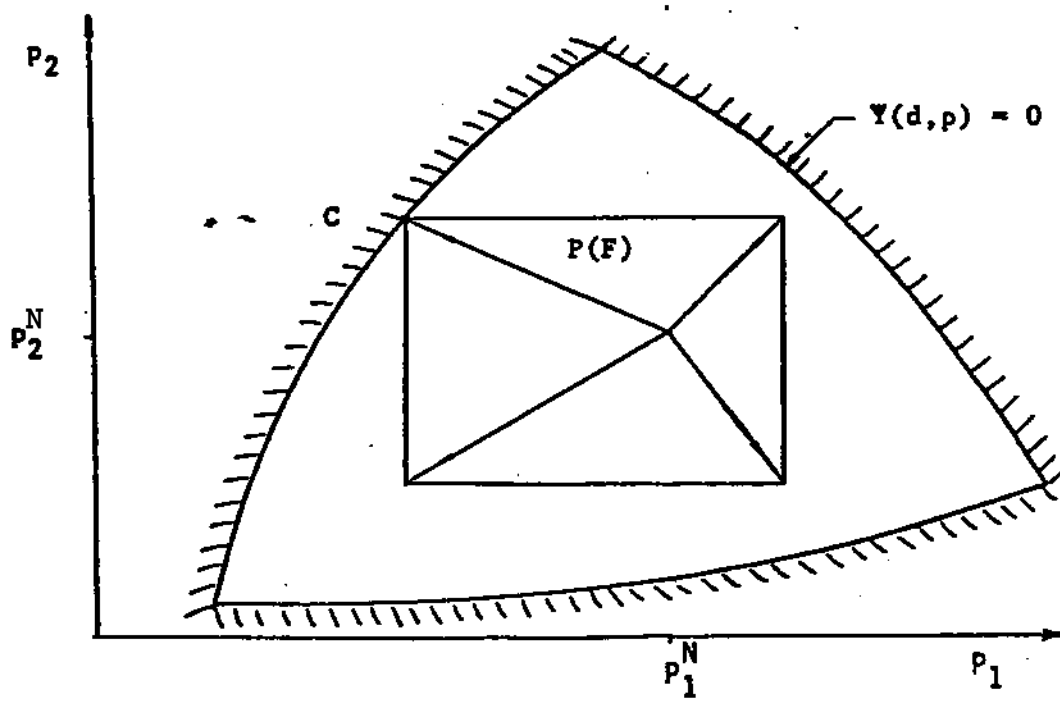


Fig. 11. Maximum feasible parameter set $P(F)$

could simplify considerably the design of flexible processes. However in general it is a nontrivial problem to correctly predict critical points in a design. Therefore a good understanding is required on how critical points arise in design. Formally, the critical points, p^c , can be defined as those points for which the feasibility function $l(d,p)$ attains a global maximum over a given parameter set $P(6)$; that is

$$p^c = \arg\{ \max_{p \in P(6)} *(d,p) \} \quad (18)$$

According to this definition, points C in Figs. 8, 9 and 11 correspond to critical points. Also, as shown in Fig. 12, the critical points need not be unique since different parameter values may attain the global maximum value of $*(d,p)$, which in this figure has zero value.

The importance of identifying critical points in a design is that they have the following properties:

- a) If the critical points are feasible (i.e. $l(d,p^c) < 0$), they guarantee also feasible operation for all other $p \in P(6)$. This simply follows from the fact that in (18) by the definition of global maximum, $*(d,p) < *(d,p^c)$ for $p \neq p^c$, $p \in P(6)$.
- b) If the critical points lie on the boundary of the region R (i.e. $t(d,p^c) = 0$), then p^c represents a parameter point that limits the flexibility in the design, since then the hyper-rectangle $P(6)$ cannot be expanded further within the region R for any $\delta^1 > 0$ (e.g. see Fig. 11).
- c) If $*(d,p^c) > 0$, it follows from (12) that the critical parameter point p^c represents the parameter point in $P(6)$ for which there are maximum constraint violations.

Although physical intuition can often predict correctly the location of critical points, interactions in a process may lead to rather unexpected values as was shown in the two heat recovery network problems at the beginning of the paper. In these two examples the critical points did not correspond to the intuitive "worst" conditions, namely, the lowest heat transfer coefficients for example

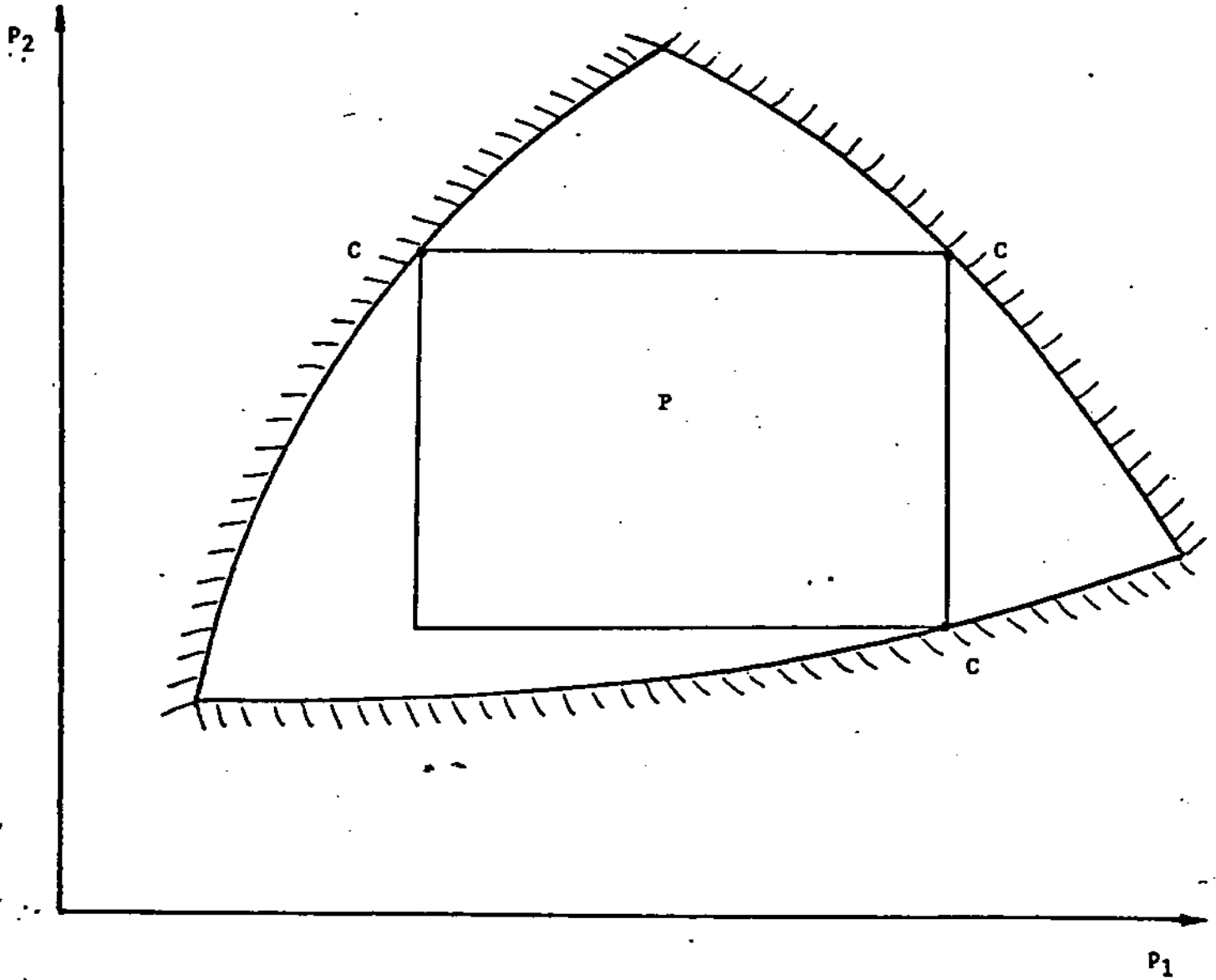


Fig. 12. Region with multiple critical points

1 and the highest heat capacity flowrate for example 2. In the former example the critical point was defined by having one transfer coefficient at the upper bound and the other at the lower bound. In the latter example the critical point was not even an extreme point but rather at an intermediate value. Before discussing how one can identify systematically critical points in a design, it is useful to consider first how they arise in the general formulations of flexibility problems.

Formulation of flexibility problems

We will consider in this section how the three main types of flexibility problems can be formulated mathematically so that they have a form that is amenable for numerical solution.

a) Flexibility analysis The flexibility index F can in principle be determined by the formulation in (16). However, the constraint of that problem $P(6) \in R$, cannot be handled readily by standard numerical optimization procedures since this constraint imposes feasibility conditions for the infinite number of parameter points in $P(5)$. Therefore, to formulate this constraint in a more convenient form the following equivalence can be established,

$$\begin{aligned}
 P(6) \in R &\Leftrightarrow \forall p \in P(6) [\exists u | f(d,u,p) < 0] \\
 &\Leftrightarrow \forall p \in P(6) [* (d,p) < 0] \\
 &\Leftrightarrow \max_{p \in P(6)} * (d,p) < 0 \qquad (19) \\
 &\Leftrightarrow \max_{p \in P(6)} \min_u \max_{j \in J} f_j(d,u,p) < 0
 \end{aligned}$$

by treating the \max and \min operators as global operators.

With the equivalence in (19), problem (16) can be formulated as

$$\begin{aligned}
 F &= \max_{u, \delta} \delta \\
 \text{s.t. } &\max_{p \in P(6)} \min_u \max_{j \in J} f_j(d,u,p) < 0 \\
 P(6) &= \{ p | p^N - \delta A p^- < p < p^N + \delta A p^+ \}
 \end{aligned} \tag{P1}$$

Note that in (P1) the max-min-max constraint imposes the feasibility condition since from (19) it can be seen that it simply

states that the maximum of the function $*(d,p)$ taken over all $p \in P(6)$ should be non-positive to ensure that the rectangle $P(6)$ is indeed contained in the region R . Also, the solution to this constraint yields the critical point p^c for which feasibility of the constraints must hold.

For the case when the flexibility analysis consists in testing feasibility for a given parameter set P ($F = 1$), problem (PI) reduces to

$$\max_{p \in P} \min_u \max_{j \in J} f_j(d,u,p) = \max_{p \in P(1)} *(d,p) \quad (P2)$$

In this case when $+(d,p^c) < 0$ feasibility is confirmed, and when $t(d,p^c) > 0$ infeasible operation is detected at the critical point.

b) Optimal design with fixed degree of flexibility In this problem the objective is to obtain an optimal and feasible design for a specified set of parameters P (or equivalently for a flexibility index $F = 1$).

Assuming that the control variables are adjusted for both feasible and optimal operation depending on the parameter realization p_f the optimal expected cost of operation is given by

$$\bar{C} = E_{p \in P} \{ \min_u C(d,u,p) | f(d,u,p) < 0 \} \quad (20)$$

However, since the design variables d must in this case be selected optimally (to minimize \bar{C}), and so as to guarantee feasible operation ($P \in R$), the problem corresponds to the two-stage programming problem

$$\begin{aligned} \min_{d \in P} E \{ \min_u C(d,u,p) | f(d,u,p) < 0 \} \\ \text{s.t. } \max_{p \in P} \min_u \max_{j \in J} f_j(d,u,p) < 0 \end{aligned} \quad (P3)$$

$$P = \{ p | p^L < P < P^U \}$$

Note that in this formulation by including the max-min-max constraint the design variables d are selected in such a way so as to ensure that

feasible operation can be guaranteed, provided that the control variables are adjusted to achieve both feasible and optimal operation for every parameter $p \in P$.

c) Design with optimal degree of flexibility This case can be regarded as a generalization of problem (P3) in which the simultaneous objectives are to minimize the cost and to maximize flexibility, while ensuring feasible steady-state operation over the parameter set $P(\delta)$ that is to be determined. This problem then leads to the bi-criterion optimization problem

$$\begin{aligned} \min_{d, \delta} \quad & E \{ \min_u C(d, u, p) \mid f(d, u, p) < 0 \} \\ \max_{d, \delta} \quad & \delta \end{aligned} \tag{P4}$$

$$\text{s.t.} \quad \max_{p \in P(\delta)} \min_u \max_{j \in J} f_j(d, u, p) < 0$$

$$P(\delta) = \{ p \mid p^N - \delta \Delta p^- < p < p^N + \delta \Delta p^+ \}$$

which as is well known defines not a single optimal solution, but rather an infinite number of pareto-optimal or trade-off solutions as shown in Fig. 13. Note that when $\delta = 1$, (P4) reduces to problem (P3), while fixing the design variables d and eliminating the first objective function reduces (P4) to problem (P1).

General approaches for solving flexibility problems

The formulations presented in the previous section have in common that they correspond to nonlinear infinite programming problems (see Fiacco and Kortanek, 1983) since they require that the infinite number of constraints contained in the specified parameter range be satisfied. The direct solution to these problems (P1, P2, P3 and P4) poses in general a formidable problem since the max-min-max constraint usually involves a non-differentiable global optimization problem (see Grossmann et al., 1982).

The only general algorithm for solving this class of problems has been proposed recently by Polak (1982). The main idea in his method

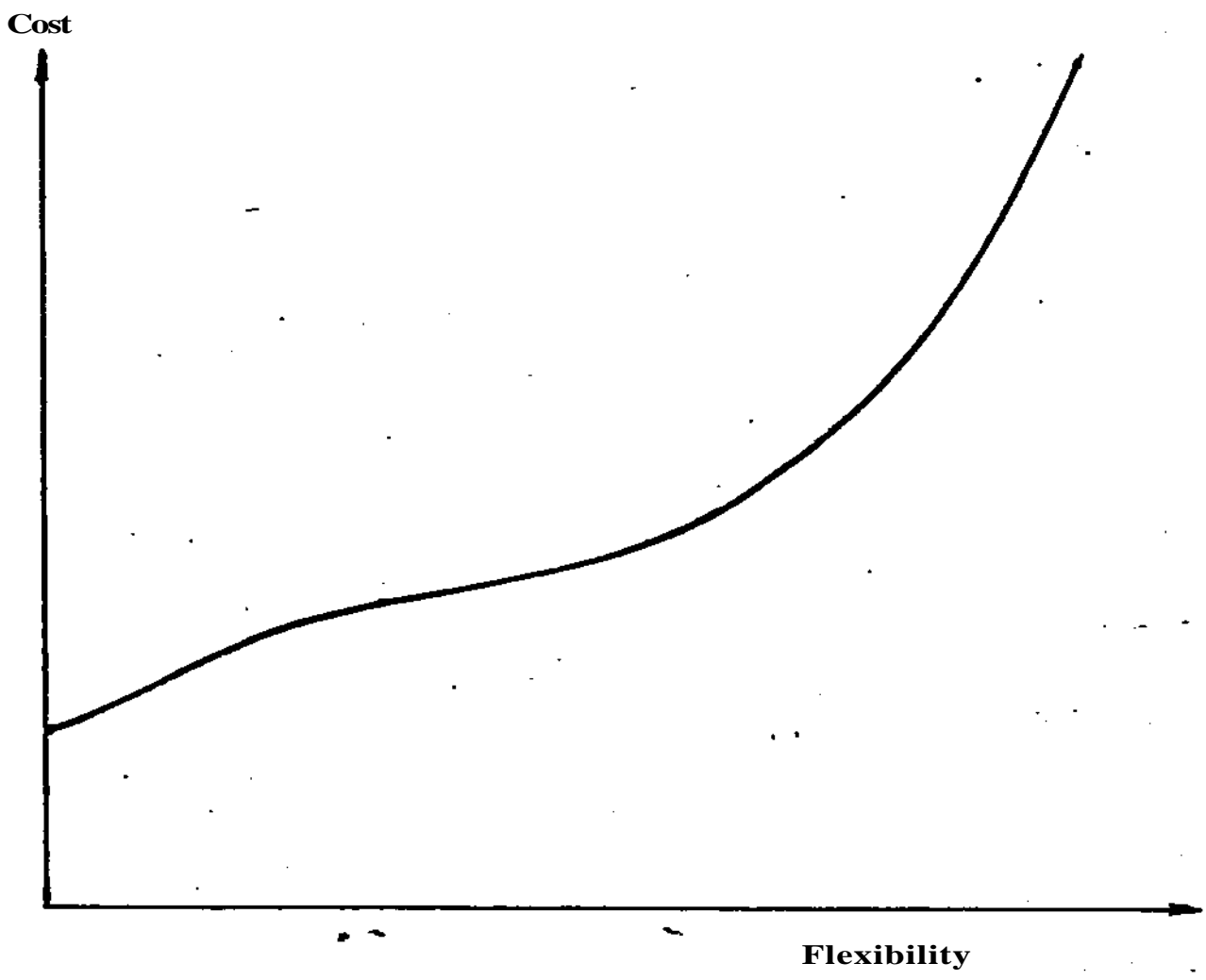


Fig. 13. Trade-off curve for cost vs. flexibility

is to construct outer-approximations of the feasible region by discretizing the parameter sets P or $P(\delta)$ which allow the problems to be solved as a sequence of nonlinear programming problems. Although the advantage in Polak's method is that it does not assume any particular location for the critical points, the drawback is that the method is computationally very expensive as it requires the solution of global optimization subproblems in order to ensure feasibility of the constraints.

An alternate approach for solving flexibility problems is to assume that the critical points correspond to vertices or extreme values of the parameter sets P or $P(\delta)$ (e.g. see Halemane and Grossmann, 1982; Marselle et al., 1982). This has the advantage of reducing the infinite dimensional problem into one of finite dimensions since feasibility must then only be ensured at the vertices. This number, however, can still be very large, particularly for a large number of parameters n_p , since the number of vertices is given by 2^{n_p} . For instance if $n_p = 10$ the number of vertices is 1024, and for $n_p = 20$ the number is 1,048,576! Despite this limitation, from a computational viewpoint this assumption simplifies the global optimization problem, since in this case one is also assuming that the global maximum of the feasibility function $\psi(d,p)$ lies at one of the vertices. However, the drawback is that the property that critical points correspond to vertices does not necessarily apply for any arbitrary type of constraint functions $f_j(d,u,p)$ $j \in J$. As has been shown by Swaney and Grossmann (1983), a sufficient condition for the property to hold is that the feasibility region R in (10) must be one-dimensional convex. That is, as shown in Fig. 14, in this type of region the convexity condition must only hold for points that are parallel to the coordinate directions (e.g. line A-B in Fig. 14). It should be noted that this class of regions is not necessarily convex as it covers some types of nonconvex regions. Swaney and Grossmann (1983) have also shown that a region can be guaranteed to be one-dimensional convex if the constraint functions $f_j(d,u,p)$ are jointly one-dimensional quasi-convex in p and quasi-convex in u . This is a more general condition than requiring that the constraint functions be jointly convex in p and u .

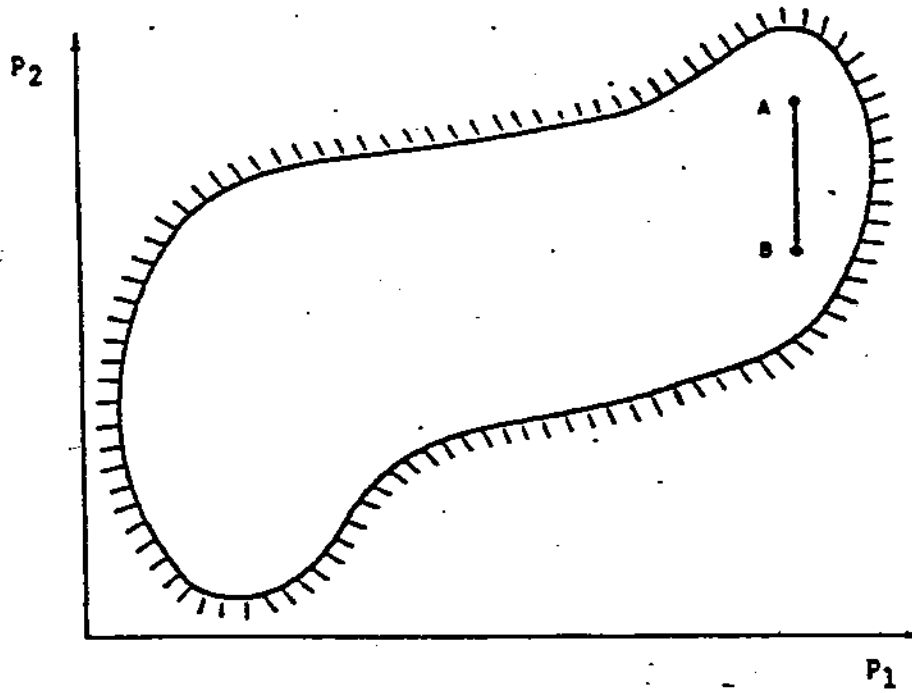


Fig. 14. One-dimensional convex region

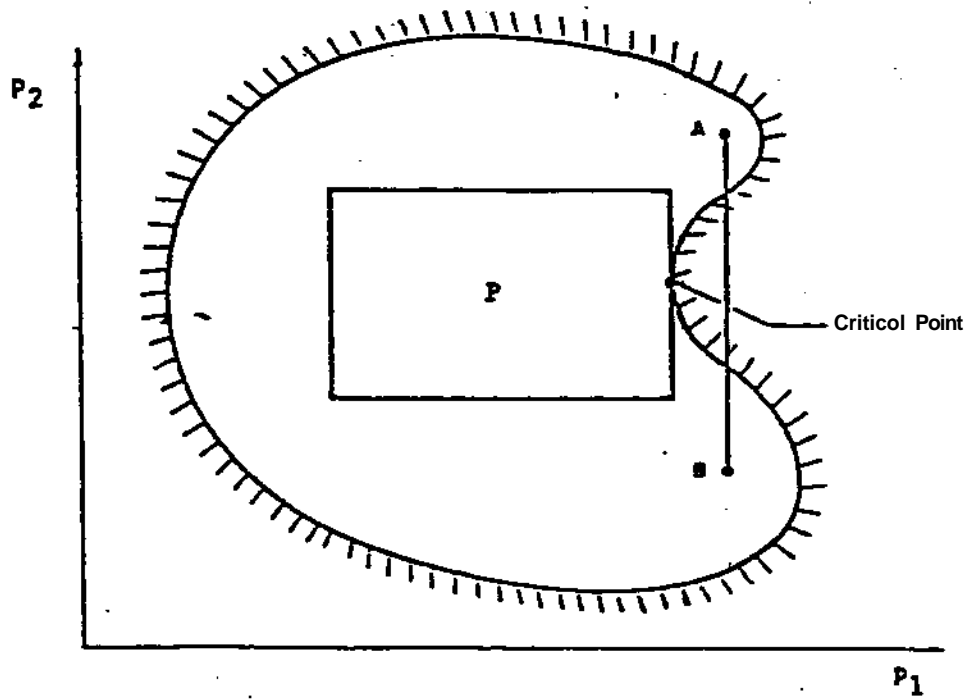


Fig. 15. Nonconvex region with near-vertex critical point

Unfortunately in practice, unless a particular problem has a special mathematical structure, it is not possible in general to verify that critical points will correspond to vertices. Although physical intuition would tend to support this conjecture for a very large number of cases, as was shown in the second example of the heat recovery network, one cannot always expect the critical point to be a vertex. This situation may occur for some types of non-convex feasible regions such as the one depicted in Fig. 15 where the critical point lies at one of the faces of the rectangle.

Practical algorithms

Despite possible exceptions to the assumption that critical points correspond to vertices, it would still seem to be worthwhile to develop algorithms based on this assumption since the scope for computational efficiency is much greater in this case. A detailed description of these methods can be found in Grossmann et al. (1982) and Saboo and Morari (1983). This section will outline only very briefly the main idea behind the methods.

In the case of the flexibility index, problem (PI) can be decomposed in subproblems that determine the maximum parameter deviation along the vertex direction $(\Delta p)_k$, $k \in V$, where V is the index set for all the vertices. These subproblems have the form

$$\begin{aligned} \max_{u, \delta_k} \quad & \bar{g}_k \\ \text{s.t.} \quad & f(d, u, p) < 0 \\ & p = p^N + \delta_k (\Delta p)_k \end{aligned} \tag{21}$$

which can be shown (see Swaney and Grossmann, 1983) to define points on the boundary of the feasible region R . The value of F is then simply given by $F = \min_{k \in V} \{f_c\}$. However, in order to avoid solving (21) explicitly for each vertex k , an enumeration procedure has been proposed by Swaney and Grossmann (1983) which tests with problem (13) the feasibility of vertices with the current upper bound \bar{g} . If a

vertex is found to be infeasible, it is solved with (21) to update the value of S. Also, Swaney and Grossmann (1983) have developed an alternate procedure that assumes monotonic constraints with which only a small number of subproblems need to be solved out of the total number of vertex points. Both procedures can also detect under some conditions whether the critical point is not a vertex, in which case a local maximization procedure is used to find a non-vertex critical point.

For the problem of optimal design with fixed degree of flexibility (P3), Halemane and Grossmann (1983) have proposed an algorithm that considers a finite set P_j of N_j parameter points in successive iterations $j = 1, 2, \dots$. These points include vertices that are estimates of critical points, as well as other parameter values that provide a suitable weighting for the cost function. With this discretization problem (P3) reduces to:

$$\min_{d, u^1, \dots, u^{N_j}} \sum_{i=1}^{N_j} w_i C(d, u^i, p^i) \quad (22)$$

$$\text{s.t. } f(d, u^{1 \wedge 1}) < 0 \quad i = 1, \dots, N_j$$

where w_i correspond to the probabilities of realization of the selected parameters $p \in P_j$. Rather than considering all vertices of P in the set P_j to ensure feasibility, the algorithm solves problem (22) iteratively by augmenting the set of P_j with those vertices that were not included in the set P_{j-1} , but were found to have the largest infeasibility by solving the subproblem in (13) for all vertices. The algorithm makes use of the method suggested by Grossmann and Sargent (1978) to generate the initial vertices in the set P_i . The procedure consists in analyzing the gradients of individual constraints in order to determine the vertex for which each constraint is maximized. The resulting set of vertices is then merged in a smaller set of predicted critical points, with which very often problem (22) needs to be solved only once.

It should be noted that problem (22) has the structure of a multi-period design problem where the weights W_j can be interpreted as lengths of time of each time period i in which the plant operates under the parameter p^i . The projection-restriction strategy proposed by Grossmann and Halemane (1982) and implemented in the computer package FLEXPACK (Avidan and Grossmann, 1983) can be used to solve the multi-period problem very efficiently since its computational effort tends to vary only linearly with the number of points N_j .

As for the solution of the problem of design with optimal degree of flexibility (P4), one approach would just simply be to solve it for different values of θ for generating the trade-off curve. Since this requires solving a sequence of problems (P3) this approach tends to be computationally expensive. Work is needed to develop a more efficient procedure for this problem.

On the synthesis of flexible processes the only problem that has been studied systematically is the one of resilient heat recovery networks. Marselle et al. (1982) consider the flowrates and inlet temperatures of n process streams as uncertain parameters that are specified within given lower and upper bounds. The objective is then, to determine a configuration of a heat recovery network that is feasible to operate and attains maximum energy recovery for the specified range of parameters. Out of the 4^n vertices that are possible, Marselle et al. (1982) identified four predicted critical vertices that are shown in Table 3. Note that these vertices correspond to four physical situations: maximum total heat exchanged, maximum heating and cooling duties, maximum heat transfer area. The proposed synthesis procedure consists in deriving a network structure that is feasible and attains maximum energy recovery for the four vertices. The actual network structure is derived by combining networks for each of the four vertices. Very often the resulting network structure will be feasible for all the parameter points in the specified range. However, this may not be true in general as was shown in the second example at the beginning of this paper which involved an uncertainty in the flowrate of one of the streams.

Recently, Saboo and Morari (1983) have identified a class of network problems for which feasibility can rigorously be guaranteed if

Table 3

Predicted critical vertices for resilient
heat recovery networks

<u>Case</u>	<u>Stream Type</u>	<u>Inlet Temperature</u>	<u>Flowrate</u>	<u>Description</u>
A	Hot	U	U	Maximum heat exchanged
	Cold	L	U	
B	Hot	U	U	Maximum cooling
	Cold	U	L	
C	Hot	L	L	Maximum heating
	Cold	L	Q	
D	Hot	L	D	Min AT Max area
	Cold	U	U	

L - lower bound, U - upper bound

feasibility holds at the vertices of the uncertain parameters* In particular, the following assumptions must hold for that class of problems:

- a) Heat capacities are temperature independent.
- b) No boiling/condensing streams.
- c) Only uncertainties in the inlet temperatures are considered.
- d) The networks do not feature stream splitting.
- e) The range of pinch point variations is defined by the inlet temperature of the same processing stream.

In practice, however, these assumptions can be relaxed somewhat. Although the assumptions are rather restrictive they illustrate the great difficulty involved in the problem of guaranteeing the existence of regions of feasible operation.

Finally, Papoulias and Grossmann (1983) have developed a synthesis procedure based on mixed-integer linear programming for designing flexible utility systems. They specify a finite number of different demands that the utility system must provide in a sequence of time periods. By formulating the problem as a multi-period mixed-integer problem a minimum cost structure of the utility system is obtained which is feasible for each time period. Although this synthesis problem is not as general as the one where demands are specified as uncertain parameters, it has the interesting feature that the structure is obtained automatically through the solution of a mathematical programming approach.

Examples

In order to illustrate some of the ideas on flexibility three examples will be discussed. The first one is taken from Swaney and Grossmann (1983), the second from Halemane and Grossmann (1983) and the third from Saboo and Morari (1983).

Example 5: In the system shown in Fig. 16 a pump must transport liquid at flowrate m from its source at pressure P_i through a pipe run to its destination at pressure P_2^* . The actual pressure P_z must remain within a tolerance ± 20 kPa of the desired pressure P_2^* . Both the flowrate m and the pressure P_2^* are uncertain parameters. The nominal value of m is 10 kg/s with expected deviations of +2 and -5

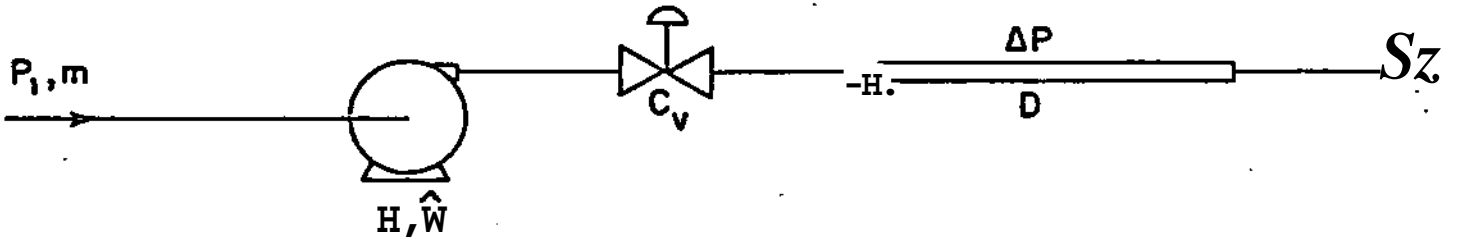


Fig. 16. Pipeline example problem

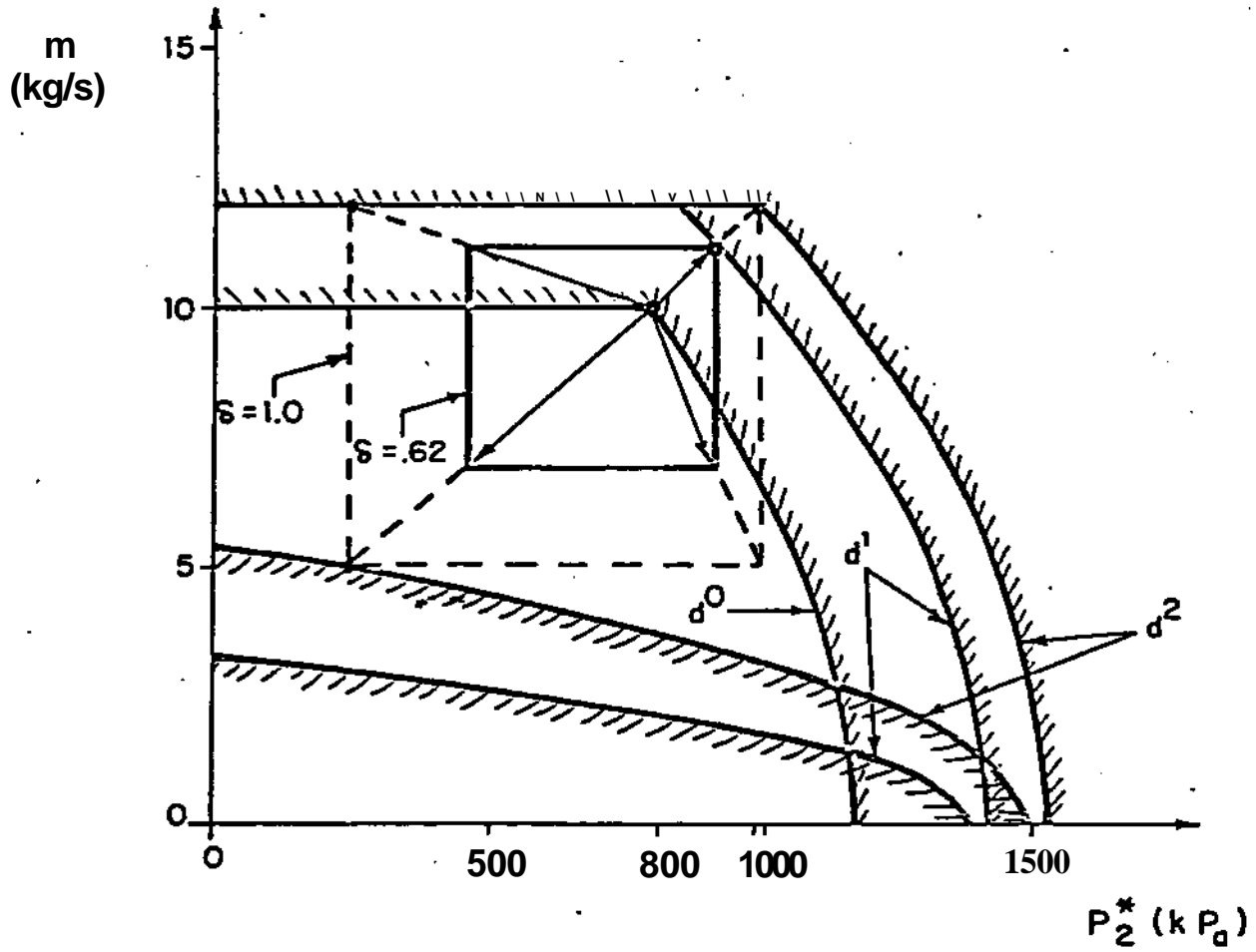


Fig. 17. Regions of feasible operation for designs d^0 , d^1 and d^2 .

kg/s* The nominal value for $P2^*$ is 800 kPa with expected deviations of +200 and -550 kPa; P^A is fixed at 100 kPa. The design variables to be selected are the pipe diameter D , the pump head H , the driver power tf , and the control valve size C_v^{max} ; the control variable is the valve coefficient c_y which can be adjusted with different valve openings.

The resulting feasible regions R corresponding to three different proposed designs are depicted in Fig. 17. Note that these regions are one-dimensional convex. The first design $\langle P$ is a design which has been optimized at the nominal parameter point. Since this point lies at the boundary of the feasible region R for $\langle P$, no rectangle of finite size can be expanded, and therefore the flexibility index for this design is zero. Design d^1 has been oversized by increasing the pump head by 230 kPa and by sizing the driver power H at the expected high value for $m = 12$ kg/s. The index of flexibility for d^1 is illustrated in Fig. 17 by the rectangle inscribed within the d^* region. The critical point for this design lies at the vertex which simultaneously maximizes $P2^*$ and m , for which $F = 0.62$. Note that in spite of the chosen oversize allowances the condition $P2^* = 1000$ kPa, $m = 12$ kg/s remains infeasible. Design d^2 shown in Fig. 17 is one for which the flexibility targets are exactly met, i.e. $F = 1$, at minimum cost. Note that in this case the region R is modified in such a way so as to exactly accommodate the rectangle whose sides correspond to the specified expected deviations. Also note that in this design, three vertices of the rectangle are critical points. Finally, Fig. 18 shows the trade-off curve between the annualized cost and flexibility for this system. As can be seen from this curve, for values of flexibility that lie within 0 and 1.3 only a moderate linear increase is experienced in the cost. However, for flexibility values greater than 1.3 a rather sharp increase in the cost is experienced since the system becomes more inefficient to operate. Analysis of this trade-off curve gives the required insight to the designer to select an appropriate point in the curve, and so establish the degree of flexibility considered to be optimal.

Example 6: The heat exchanger network 4SP1 of Lee et al. (1970) with outlet temperatures specified as inequalities is considered. The network involves two hot and two cold streams and

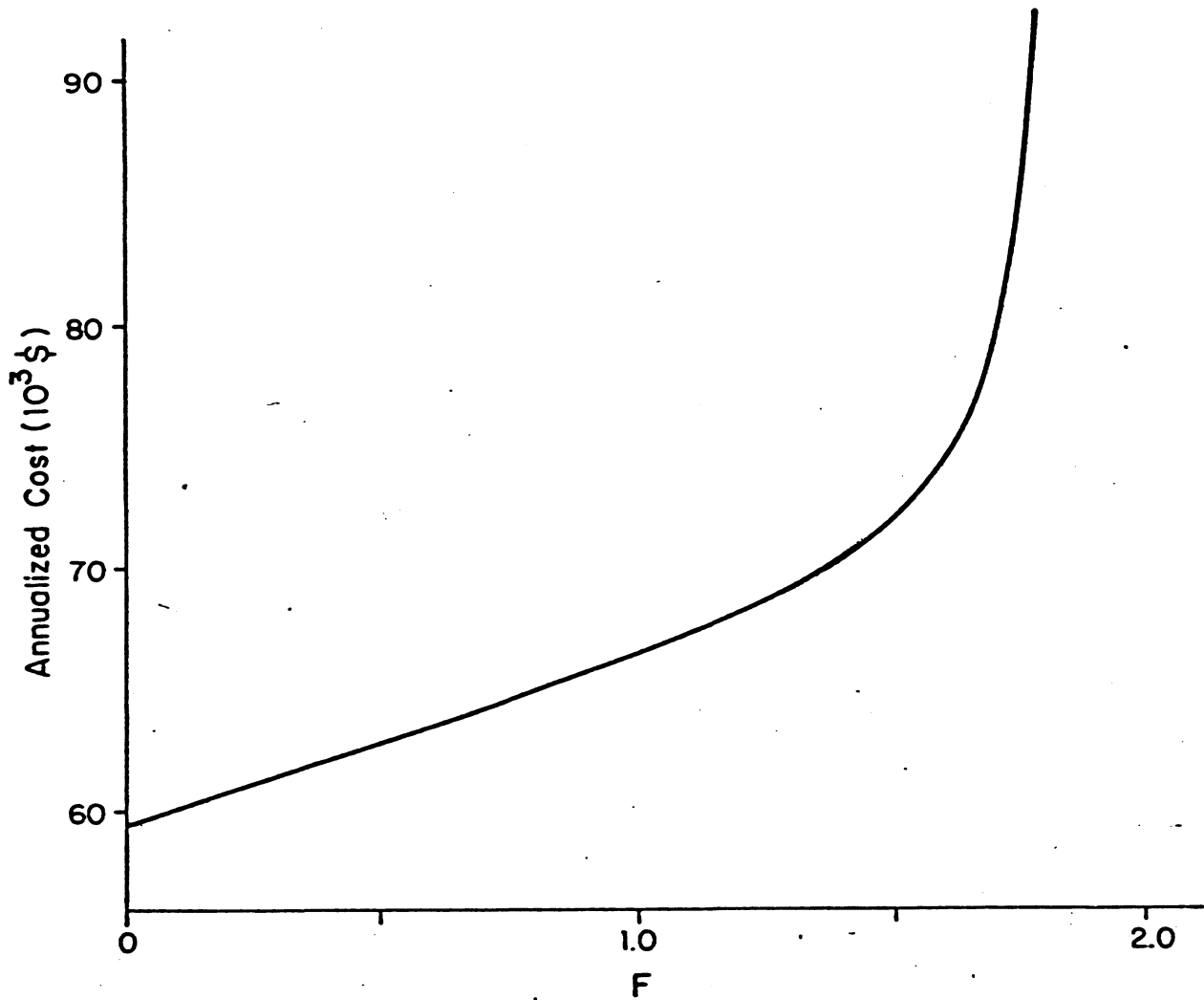


Fig. 18. Trade-off curve of cost vs. flexibility index for pipeline problem.

five exchangers as shown In Fig. 19. The overall heat transfer coefficients are considered as uncertain parameters with expected deviations of $\pm 20\%$ with respect to their nominal values. The problem consists In selecting the five areas so that irrespective of the actual values of the heat transfer coefficients (with $\pm 20\%$ range), the specifications on the outlet temperatures are satisfied by suitable choice of the cooling water outlet temperature $l\backslash s$ and the steam temperature $T13$.

Table 4 gives the initial set of predicted critical vertices considered for design, which were obtained by analyzing the signs of gradients of constraints as suggested by Grossmam and Sargent (1978); the nominal point Is included to provide an adequate weighted cost function. Note also that the vertex where the five transfer coefficients lie at their lower bounds (intuitively the "worst" condition) is not included. The optimal design of the network corresponding to these set of points was obtained by solving the corresponding multi-period design problem. The resulting design was found to be feasible for the 32 vertices. It is Interesting to note that 24 of these vertices are actually critical points since they attain the maximum value of $+(d,p) = 0$. This result is to be expected since the optimization procedure will have the tendency of adjusting the feasible region so that its boundary touches the parameter set on as many vertices as possible.

The actual areas that were obtained are shown in Table 5 which also shows the areas that are obtained when the network is optimized at only the nominal joint. Clearly the striking feature is that the overdesigns that are predicted are quite different for each exchanger; for instance, exchanger 5 has been oversized by 64.2% whereas exchanger 4 has been oversized by only 8.4%. However, more interesting Is the fact that exchanger 2 is actually being undersized by 7.7%. Physically, the explanation of this is that when the transfer coefficient in exchanger 2 takes the upper bound value, the outlet temperature of H2 drops to a point which rauld make the heat exchange infeasible in exchanger 3. Therefore, to avoid this situation the area of exchanger 2 must be undersized. This result shows the effectiveness of the procedure by Halemane and

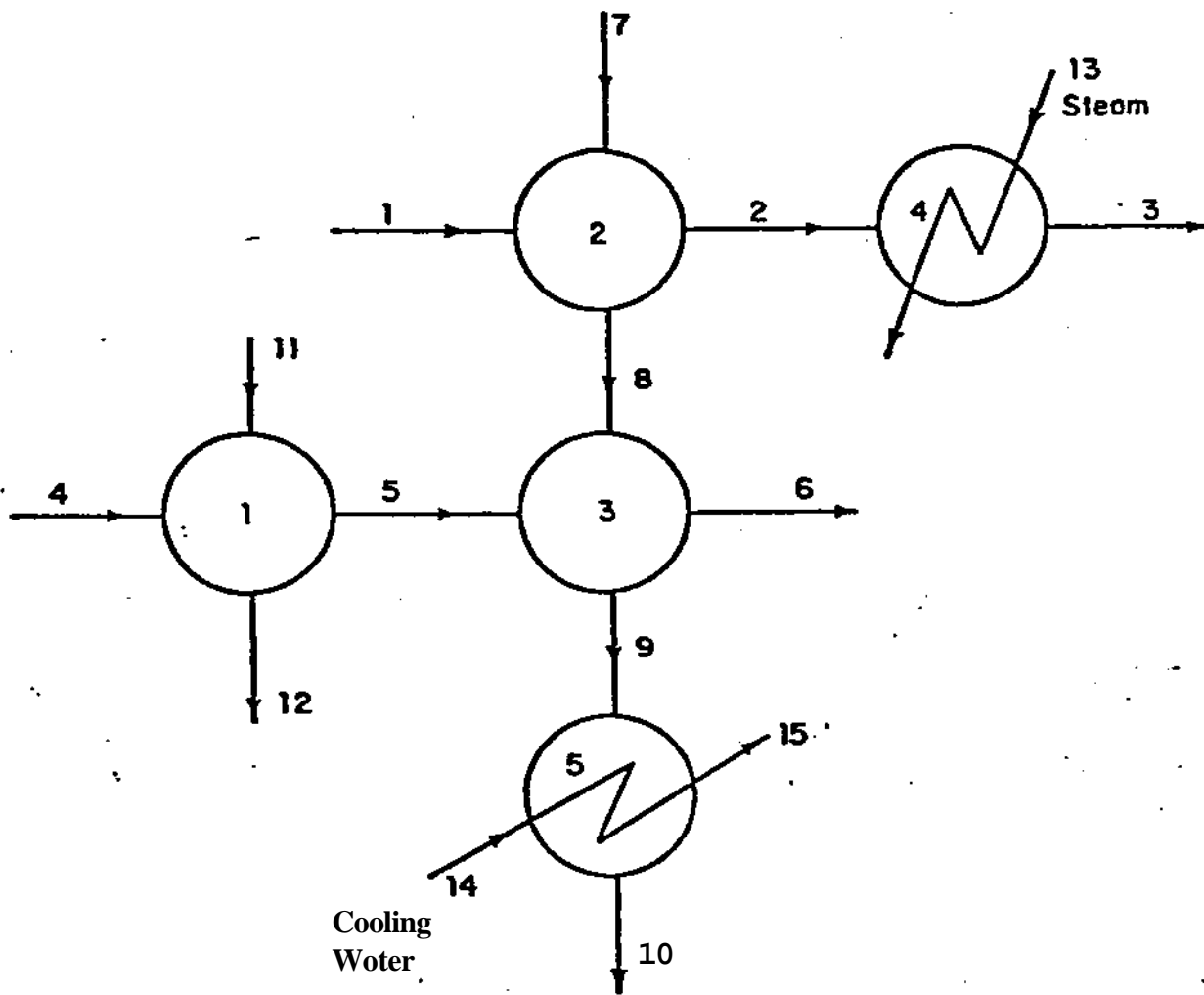


Fig. 19 Heat exchanger network 1

Table 4

Parameter values considered for heat exchanger network
(Example 6)

Heat transfer coefficients					
	D_1	U_2	U_3	U_4	U_5
1	N	N	N	N	N
2	U	U	U	U	N
3	L	U	U	U	N
A	U	L	L	L	L
5	L	U	L	U	N

N-nominal, L-lower bound, U-upper bound

Table 5

Comparison of optimal areas predicted for
heat exchanger network (Example 6)

	ϵ_1	Areas (m ²)				Expected cost (\$/yr)
		A_j	A_3	A_4	A_5	
H&G Algorithm	30.8	65	45.6	3.9	2.8	11,758
Nominal design	24.6	70.4	40.6	3.6	1.7	9,959
Z Overdesign	25	-7.7	12.2	8.4	64.2	

Grossmann (1983) in accounting for the interactions that may lead to infeasible operation in a processing system.

Example 7: The objective is to synthesize a heat exchanger network for the problem specified in Table 6 such that arbitrary inlet temperature variations up to $\pm 10K$ and up to $\pm 15K$ can be tolerated while maintaining maximum energy recovery. It can be shown (Saboo & Morari, 1983) that for this problem all assumptions stated above are satisfied which ensure that feasibility at the vertices is sufficient for feasibility over the whole operating range. In particular for all possible temperature variations the pinchpoint is defined by the inlet temperature of stream H2. Saboo and Morari developed the program RESHEX, which - among other features - provides an efficient means for checking the feasibility at all the vertices. A network structure capable of providing maximum energy recovery at the nominal operating conditions with the minimum number of exchangers is shown in Fig. 20A. For inlet temperature variations up to $\pm 10 K$ the only vertex at which the approach temperature constraint of $AT_{min} = 10K$ is violated is Case C (cf. Tab. 3) (Fig. 20A). For all practical purposes the network is therefore structurally resilient for this temperature range. However care has to be given to the area design to account for the small approach temperature which occurs for Case C. Figure 20B & C show that Cases B and C for $\pm 20K$ are physically impossible with this structure. Based on recommendations for modifications generated by the program RESHEX the structure in Fig. 20D 4 E was obtained which is structurally resilient for a range of $\pm 20K$.

DYNAMIC RESILIENCE

Problem definition and general terminology

Clearly the primary requirement for plant operability is steady state resilience and flexibility. But these qualifications are of little help when the transients of the process moving from one operating condition to another or the actions taken to compensate for a disturbance, are exceedingly slow resulting in large amounts of off-specification product. It is then very likely that at the next level of the screening process a design candidate with less favorable static but more attractive dynamic resilience characteristics will be

Table 6
Stream data for Example 7

Stream No*	w(lcW/K)	T ^S (K)	T ^T (K)
H1	8	450	380
H2	13	510	410
C1	8	355	445
C2	6	390	530

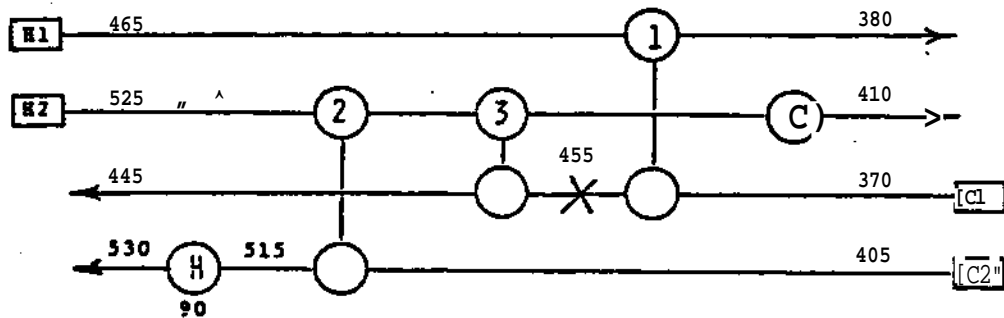
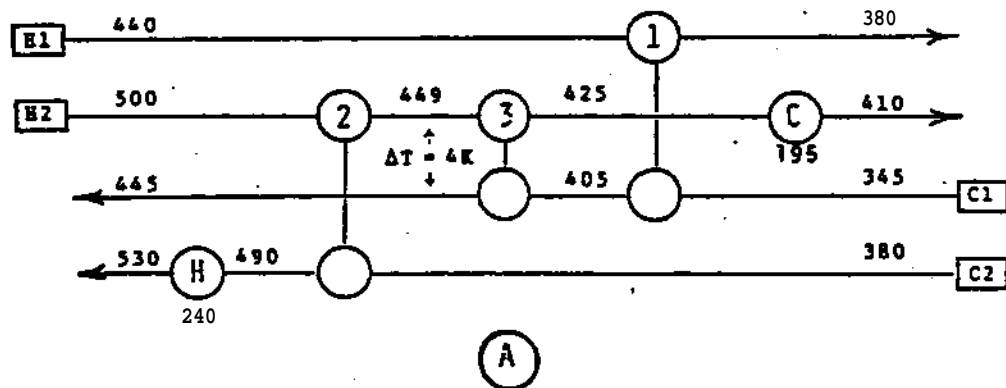


Fig. 20.A,B: Heat exchanger networks featuring maximum energy recovery generated for Example 7. A: Case C with $\Delta T = -10$ K. B: Case B with $\Delta T = +15$ K. Situation B is physically impossible because temperature crossovers occur.

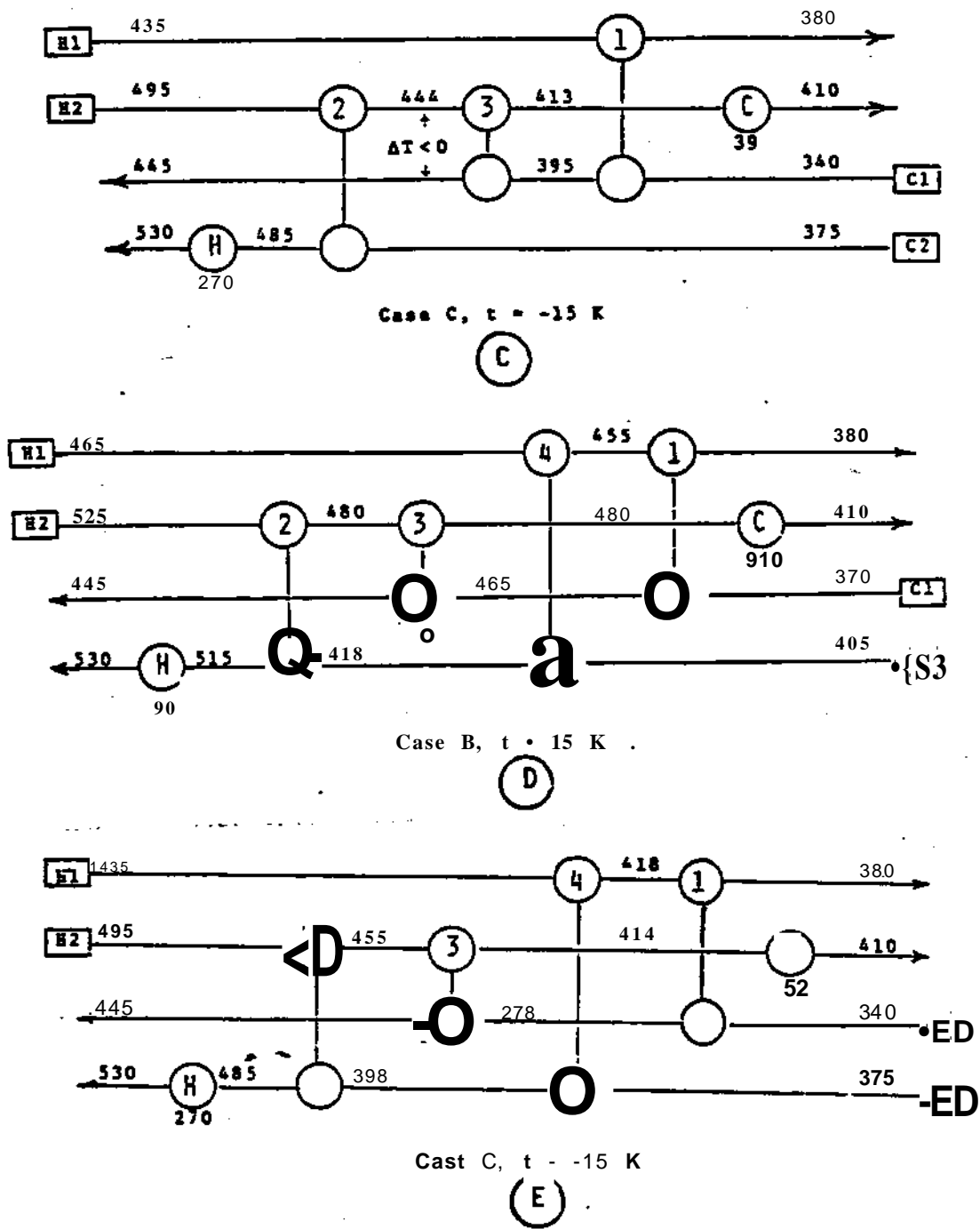


Fig.20.C,D,E: Heat exchanger networks featuring maximum energy recovery generated for Example 7. C: Case C with $\Delta T = -15$ K. D: Case B with $\Delta T = +15$ K. E: Case C with $\Delta T = -15$ K. Situation C is physically impossible because temperature crossovers occur.

preferred. The quantitative evaluation of "dynamic resilience" is much more difficult than that of its static counterpart. Though it could be cast into the form of a variational problem with an integral square error objective function and though the trade-offs could be judged conveniently using a vector valued objective function (Lenhoff & Morari, 1982), it is essentially impossible to relate these numbers to features which are observed on the real plant by the operators. Therefore this approach is rightfully frowned upon by the practitioner. A further difficulty arises from the fact that a judgement of the dynamic behavior without including a controller in the analysis appears to have as little meaning as including a specific controller which might very well bias the results: a different controller could lead to a very different conclusion on the dynamic resilience of a design.

The situation can be compared quite accurately to the problem of judging the performance of a car. Obviously the performance depends on the driver (controller) but clearly there are also properties inherent in the car, e.g. power, which determine its dynamic characteristics independent of the skill of the driver. In practical use the best performance will be realized by a car which has not only good inherent dynamic characteristics but also one whose performance is not extremely sensitive to our driving skills or - expressed differently - to our lack of detailed knowledge about the car's properties. For example, a high powered car is not much use if it is difficult to handle.

Translating these arguments into the area of process design we are looking for a method to judge the inherent dynamic characteristics of a plant independent of the installed controller and a technique to quantify the sensitivity of the performance to uncertainty. To obtain good performance for a "sensitive" plant a good model and a good controller are needed. A good model is often very difficult to obtain and therefore sensitive plants should be avoided.

System description

In order to evaluate the dynamic resilience in a process it will be assumed that both the configuration and equipment sizes at the

plant are given. Furthermore, although no specific controller type will be assumed, it is clear that a given set of controlled and manipulated variables must be specified. In general there might be a number of different choices for the manipulated variables in a process as was illustrated in example 4 at the beginning of this paper.

Because of the complexity of the problem and because in control we are usually concerned about "small" deviations from steady state operating points, we restrict our discussion to linear systems. Furthermore, we assume that the uncertainty is described in the frequency domain as a "region" around the nominal model in which the real plant lies. More specifically, the input/output model is

$$y(s) = G(s) u(s) + p(s) \quad (23)$$

where $y(s)$ is the vector of controlled variables, $u(s)$ the vector of manipulated variables, $p(s)$ the vector of unmeasured disturbances affecting the outputs and $G(s)$ is the transfer matrix. This matrix can be obtained by linearizing the performance equations of the process at the nominal operating conditions. Apart from the additive uncertainty expressed through $p(s)$ there is uncertainty in the model itself.

Uncertainty description for SISO systems A convenient way to describe model uncertainty for single-input-single-output (SISO) systems ($\dim u = \dim y = 1$) is

$$g(s) = g(s)(1 + f(s)) \quad (24)$$

where $g(s)$ denotes the plant and $g(s)$ its model and $f(s)$ is bound by

$$|\Re\{f(i\omega)\}| < \bar{f}(\omega) \quad (25)$$

Here $\bar{f}(\omega)$ is a function defined on the positive reals. Combining (24) and (25) we find

$$\frac{|g(i\omega) - \bar{g}(i\omega)|}{|\bar{g}(i\omega)|} < \bar{\lambda}(\omega) \quad (26)$$

On a Nyquist diagram (26) describes an uncertainty bound swept by circles of radius $\bar{\lambda}(\omega)|\bar{g}(i\omega)|$ within which the plant lies (Fig. 21). In all practical situations the model $\bar{g}(s)$ is of lower order than the plant $g(s)$. Therefore $|g(i\omega)|$ will vanish faster at high frequencies than $|\bar{g}(i\omega)|$ and typically

$$\lim_{\omega \rightarrow \infty} \bar{\lambda}(\omega) = 1 \quad (27)$$

This form of uncertainty description has become quite standard in the control literature but it is not necessarily the most convenient in every situation. As we will show next it can be extended to multivariable systems but some information on the location of the uncertainty is lost in the process. Before we discuss this extension we have to introduce the notion of gain for multi-input-multi-output (MIMO) systems.

Vector and matrix norms In the SISO case the gain is simply the amplitude ratio as a function of frequency. For MIMO systems the situation is somewhat more complicated because the magnitude of the output of a system does not only depend on the magnitude but also on the direction of the input. We will follow here the ideas by MacFarlane & Scott-Jones (1979). Taking norms on both sides of the equation describing the system

$$y(i\omega) = G(i\omega)u(i\omega) \quad (28)$$

we obtain

$$\begin{aligned} \|y(i\omega)\| &= \|G(i\omega) u(i\omega)\| \\ &< \|G(i\omega)\| \|u(i\omega)\| \end{aligned}$$

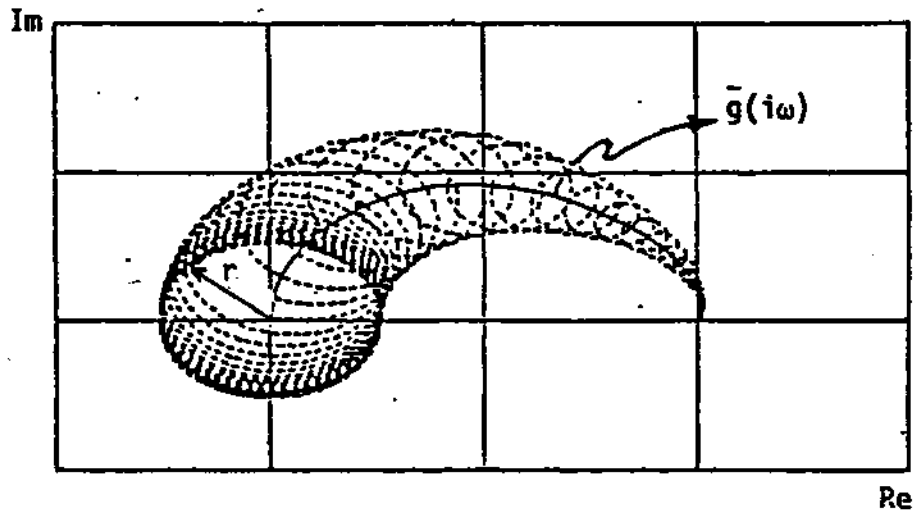


Fig. 21. Uncertainty band around model $\bar{g}(i\omega)$ within which the real plant $g(i\omega)$ lies.

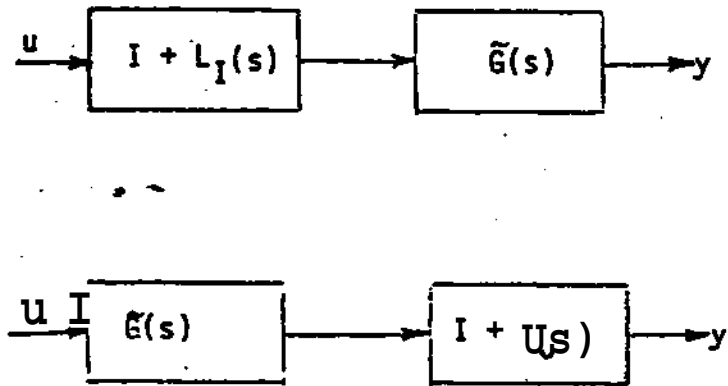


Fig. 22. Multiplicative input and output uncertainties for a multivariable system.

$$\text{or } \frac{\|y(i)\|}{\|u(i)\|} < \|G\| \quad (29)$$

Employing the Euclidean vector norm a compatible matrix norm is the spectral norm

$$\|G\| = \max_i \lambda_i^{1/2}(G^*G) \quad (30)$$

The square roots of the eigenvalues of G^*G (* denotes complex conjugate transpose) are called the singular values (Kiema S Laub, 1980) or, in the control context, the principal gains (MacFarlane & Scott-Jones, 1979) of G . We will employ the notation

$$\sigma_M(G) = \lambda_{\max}^{1/2}(G^*G)$$

$$\sigma_m(G) = \lambda_{\min}^{1/2}(G^*G)$$

It can be shown that

$$\sigma_m(G) \|u\| < \|Gu\| < \sigma_M(G) \|u\| \quad (31)$$

Thus the maximum singular value is a natural definition of gain for multivariable systems. Singular value plots play much the same role for MIMO systems as Bode plots for SISO systems. As is clear from the definition, for square systems the number of singular values is equal to the number of inputs (outputs). The minimum and maximum singular value provide bounds (31) on the stretching action exerted by the system on a particular input vector.

Uncertainty description for MIMO systems As shown in Fig. 22 the multivariable multiplicative uncertainties can act either on the inputs (Lj) or the outputs (LQ)

$$G(s) = \tilde{G}(s)(I + L(s)) \tag{32}$$

$$G(s) = (I + L_0(s))\tilde{G}(s)$$

$$\|\tilde{G}^{-1}(G - \tilde{G})\| < \epsilon_1(\omega) \tag{33}$$

$$\|(G - \tilde{G})\tilde{G}^{-1}\| < \epsilon_0(\omega)$$

where ϵ_1, ϵ_0 are scalar functions defined on the positive reals. These functions do not allow to distinguish between uncertainty localized in one element and uncertainty "spread" over all elements. However, this might not be so disadvantageous at the design stage where better uncertainty information is rarely available.

The Internal Model Control (IMC) structure

The classic feedback structure (Fig. 23) makes an identification of the system inherent characteristics affecting dynamic resilience quite complicated. This analysis is greatly facilitated when the IMC structure (Fig. 24) (Garcia, Morari, 1982) is employed. Obviously, through simple block diagram manipulations the IMC structure and the classic feedback structure can be made equivalent

$$G_c = C(I + \tilde{G} C H) \tag{34}$$

$$C = G_c(I - \tilde{G} G_c)^{-1} \tag{35}$$

but the IMC structure makes certain results more transparent as we will see shortly. The closed loop relationship is

$$y = G(I + G_c(G - \tilde{G}))^{-1} G_c(y_s - p) + p \tag{36}$$

For $\tilde{G} = G$ this reduces to

$$y = G G_c(y_s - p) + p \tag{37}$$

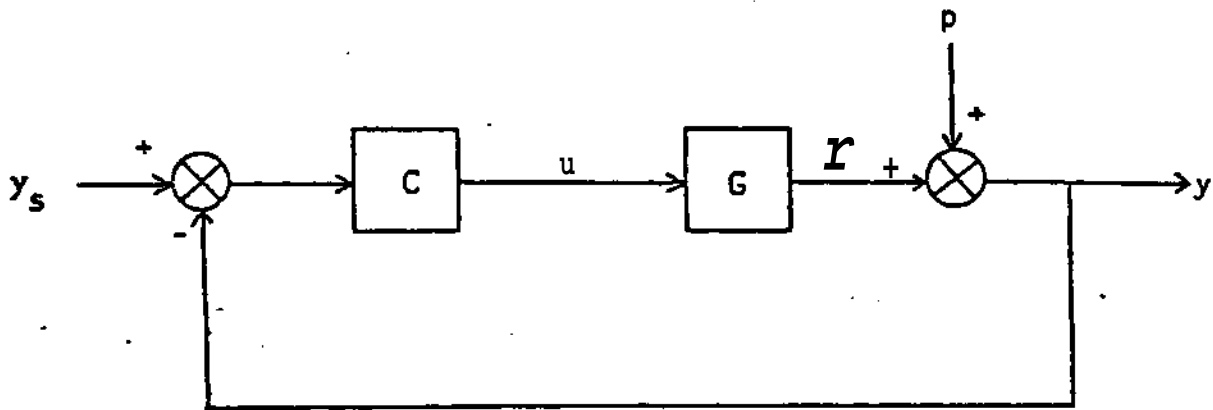


Fig. 23. Classic Feedback Structure.

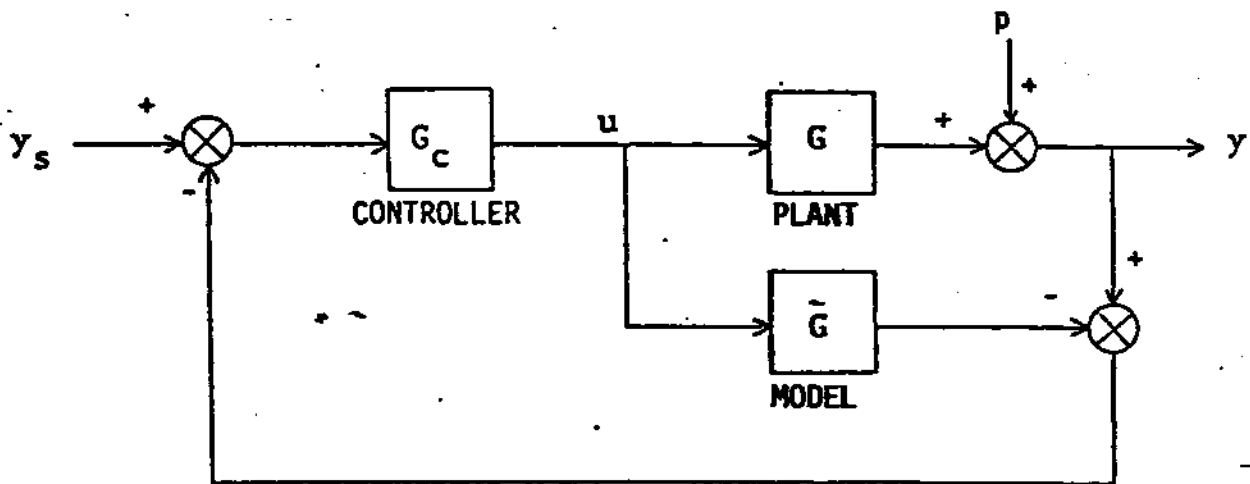


Fig. 24. The Internal Model Control Structure.

i.e. it becomes equivalent to an open-loop controller where the disturbances are measured. Note that G_c allows to specify the structure of the closed loop response. For example, when GG_c is diagonal the response will be decoupled. The following three properties follow directly from (36).

Property 1: Assume that the model is perfect ($\tilde{G} = G$). The system is closed loop stable if the plant G and the controller G_c are stable.

Property 2: Let the controller be the right inverse of the plant model $fi(G_c \text{ } G^{*1})$ and assume that the system is closed loop stable. Then $y(t) = y_s(t) + f^{o^f}$ all times t and all disturbances p .

Property 3: Let the steady state controller gain be the right inverse of the steady state model gain, i.e. $G_c(0) = \tilde{G}^{-*}$ (OK Further assume that the system is closed loop stable and that the set points $y_s(t)$ and the disturbances $p(t)$ are asymptotically constant. Then the system will have no offset:

$$\lim_{t \rightarrow \infty} y(t) = \bar{y}_s$$

where $\lim_{t \rightarrow \infty} y_s(t) = \bar{y}_s$ constant.

Fundamental limitations of dynamic resilience

We can now look at the resilience assessment in the light of the IMC framework. We will analyze the reasons why the "perfect" controller ($G_c = \tilde{G}^{-*}$) can usually not be implemented. Because the controller is simply the process inverse we obtain from this analysis of the controller a direct indication of the system characteristics which limit resilience.

The controller G_c is to be selected "close" to \tilde{G}^{-*1} subject to the constraints that it be causal, realizable (to allow physical implementation) and stable (to guarantee closed loop stability according to Property 1).

G^{-1} is not causal and/or unstable if G has one of the following properties.

- 1) G involves time delays. Inversion of time delays can result in

noncausal expressions (prediction) which can not be implemented
 2) G has zeros in the right half plane (RHP). Inversion leads to poles in the RHP and therefore an unstable G_c

A transfer function containing either delays or RHP zeros is commonly called non-minimum phase (NMP). In the presence of NMP elements G is factored into an invertible part G_L and a noninvertible part \tilde{G}_+

$$\tilde{G} = \tilde{G}_+ \tilde{G}_- \quad (38)$$

such that $\tilde{G}_+(0) = I$ and L^{-1} is realizable and stable. G_c is then chosen as

$$G_c = \tilde{G}_-^{-1} L^{-1} \quad (39)$$

The factorization is clearly not unique but from (36) we note that for $\tilde{G} = G$, G_+ is the closed loop transfer function

$$y = G_c G (y_s - p) + p \quad (40)$$

$$y^s = M y_s - p + P \quad (41)$$

The noninvertible part of the system G_+ expresses the "best" achievable performance by a system and is therefore a direct measure of its dynamic resiliency. Except for SISO systems it is generally not possible to define a unique factor G_+ . Clearly G_+ should be "close" to unity but the designer often has some freedom in giving preference to certain outputs. We will now discuss in more detail the options available for choosing G_+ and finally elaborate on the other limitation to control quality, namely the constraints on the manipulated variables and the model uncertainty.

I. Time delays: The aspects of the time delay factorization have been discussed in detail by Holt and Morari (1983). In general a trade-off between the speed of the closed loop response and decoupling is possible. For example three possible factorizations for

$$G = \begin{bmatrix} 0 & e^{-2s} \\ -e^{-2s} & 1 \end{bmatrix}$$

are

$$G_+^1 = \begin{bmatrix} e^{-4s} & 0 \\ 0 & e^{-2s} \end{bmatrix}; \quad G_+^2 = \begin{bmatrix} e^{-2s} & 0 \\ (1-e^{-2s}) & e^{-2s} \end{bmatrix}$$

$$G_+^3 = \begin{bmatrix} e^{-2s} & e^{-2s}(1-e^{-2s}) \\ 0 & 1 \end{bmatrix}$$

G indicates that output 1 can react only after two time intervals, output 2 can react immediately. These numbers are a lower bound on the response time but they are not an indication of the actual settling time. If both outputs are equally important and decoupling is chosen, G_+^* provides an upper bound on the settling time. This is verified by G_+^2 , where preference is given to the first output which settles in minimum time (cf. G), at the cost of decoupling and a maximum settling time for the second output (cf. G_+^1). Analogously, in G_+^3 preference is given to the second output. These upper and lower bounds on the settling time serve as measures of resiliency for a multivariable system involving time delays. Holt & Morari (1983) have shown that a diagonal G_+ which renders G^{-1} causal is "optimal"¹¹ if and only if the rows and columns of G can be rearranged such that the smallest time delay of each row is on the diagonal. For example, the Wood & Berry (1973) distillation column has the transfer matrix

$$G(s) = \begin{bmatrix} \frac{12.8 e^{-s}}{16.7s+1} & \frac{-18.9 e^{-3s}}{21s+1} \\ \frac{6.6 e^{-7s}}{10.9s+1} & \frac{-19.4 e^{-3s}}{14.4s+1} \end{bmatrix} \quad (42)$$

Here the smallest time delays are on the diagonal and therefore $G_+ = \text{diag}(e^{-s}, e^{-3s})$ is "optimal". The lower and upper bounds on the settling time coincide.

We are now in a position to explain the counterintuitive result presented earlier which indicated that lengthening dead times can improve the dynamic resiliency. For the original network (Fig. 4) the simplified transfer function involving only dead times is

$$G(s) = \begin{bmatrix} e^{-6s} & e^{-11s} & e^{-2s} \\ e^{-11s} & 1 & e^{-12s} \\ e^{-8s} & e^{-13s} & 1 \end{bmatrix} \quad (43)$$

and the diagonal factor is

$$G_+ = \text{diag}(e^{-6s}, 1, e^{-4s}) \quad (44)$$

If the lag between heat exchangers 3 and 5 is increased from 2 to 6 minutes the transfer function of dead times becomes

$$G(s) = \begin{bmatrix} e^{-6s} & e^{-11s} & e^{-6s} \\ e^{-11s} & 1 & e^{-16s} \\ e^{-8s} & e^{-13s} & 1 \end{bmatrix} \quad (45)$$

and the diagonal factor is

$$G_+ = \text{diag}(e^{-6s}, 1, 1) \quad (46)$$

which is an improvement over (44)•

II, Right Half Plane zeros: The issues concerning the factorization RHP zeros are discussed by Holt and Morari (1984) for transfer matrices not involving time delays. The question here is how to choose G_+ such that G_-^{-1} is stable. Two major results were obtained.

Result 1: Let the SISO system $g(s)$ have a single RHP zero at $s=z$, then the Integral Square Error (ISE) to a step change is $2/z$ and this error is obtained from the factor $g_+(s) = (-s+z)/(s+z)$.

Result 2: Let the MIMO system $G(s)$ have a single RHP zero at z , then, in general, the "bad" effect of a RHP zero can be localized to any particular output,

$$G_+ = \begin{bmatrix} 1 & & & 0 & & \\ & \dots & & & & \\ & & 1 & & & \\ x & & x & \frac{-s+z}{s+z} & x & \dots & x \\ & & & 1 & & \dots & \\ & 0 & & & & & 1 \end{bmatrix} \quad (47)$$

where all the off-diagonal elements are zero except in the row which contains the RHP zero.

For example, consider the system

$$G(s) = \frac{1}{s+1} \begin{bmatrix} 1 & 1 \\ \mathbf{Us} & 2 \end{bmatrix}$$

which has a zero at $s = 1/2$. Three possible factorizations are shown below together with the ISE resulting from a unit step change in both set points

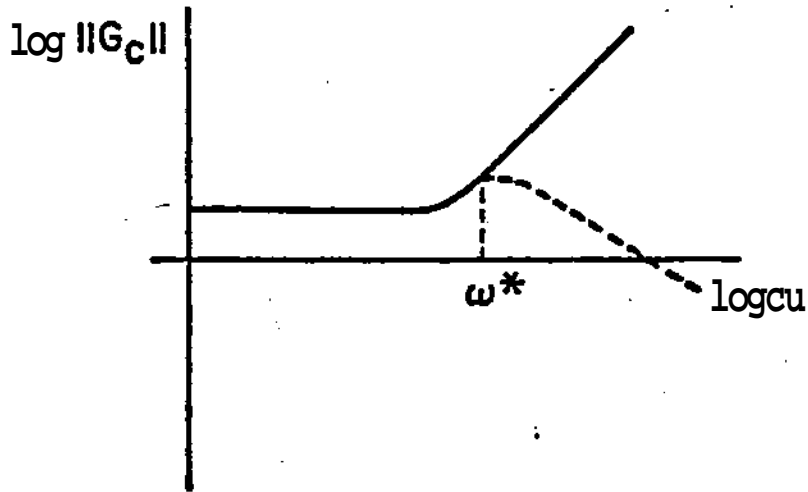


Fig. 25. Practical controller (---) starts to deviate from ideal controller (—) at frequency ω^* .

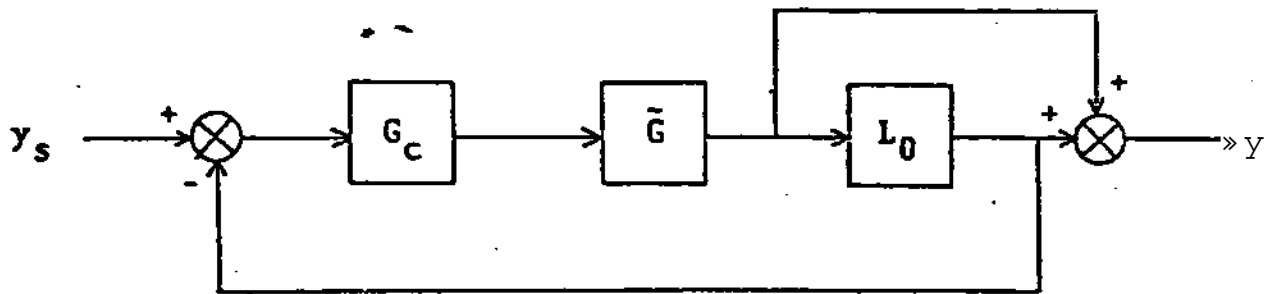


Fig. 26. IMC structure for multiplicative model uncertainties occurring at the model output.

Therefore, in order to satisfy the constraint

$$l_{ui} < i_{ui_{max}} \quad (49)$$

we should require

$$i_{y_s} - p_i < IG^{-1} i_{-1} l_{uimax}$$

$$\text{or } i_{y_s} - p_i < cWG) l_{ui}^x \quad (50)$$

In practice, depending on the frequency spectrum of the disturbances it is very well possible that the bound (50) is exceeded. This does not imply that the system will become unstable but simply that the closed loop performance deteriorates because the "perfect" controller cannot function properly due to the saturation of the manipulated variable. One way to quantify this effect is to envision a "practical" controller which deviates from the "perfect" controller and for which the manipulated variables do not saturate. We see from (48) that the plot of $\text{onf!}(G)$ shows the effective controller gain of the "perfect" controller as a function of frequency. If $G(s)$ is strictly proper (this is in principle true for all physical systems) $G^{-1}(s)$ is improper, which means that $\text{onf!}(G)$ becomes infinite as $\omega \rightarrow \infty$. A "practical" controller will have to depart from the perfect controller at high frequencies as is shown in Fig. 25. The departure point which is generally close to the corner frequency gives the bandwidth w^* over which perfect control is possible, w^* is a simple way to measure the effect of the manipulated variable constraints on closed loop performance and can be used as a tool for resilience assessment.

IV. Model uncertainty In the previous analysis the availability of a perfect model was assumed ($G = \hat{G}$). If this is indeed the case and if measurement noise is negligible, as we have assumed throughout, the resilience analysis carried out to this point is all that is needed. In practice model uncertainties and nonlinearities will always be present. In the following we will investigate the effect of model/plant mismatch on control quality.

In Fig. 26 use was made of (32) to redraw the IMC diagram of Fig. 24. Assuming the uncertainty description (33) it can be shown that a

necessary and sufficient condition for "robustness", (i.e. closed loop stability under plant variations), is that the loop gain must always be less than 1:

$$|G_c(i\omega)| |G(i\omega)| < 1/\lambda(\omega) \quad (51)$$

If we select $G_c = \tilde{G}^{-1}$ (51) becomes

$$\lambda(\omega) < \frac{\sigma_m(\tilde{G})}{\sigma_M(\tilde{G})} < 1 \quad (52)$$

This implies that the system is only closed loop stable if the uncertainty radius $\lambda(\omega)$ never exceeds 1. For any practical process control problem $\lambda(\omega)$ will grow beyond 1 for high frequencies. This forces us to detune the controller and to give up performance for robustness. Therefore let the controller have the form

$$G_c(s) = e^{-ts} F(s) \quad (53)$$

with $F(0) = I$

where $F(s)$ is a dynamic compensator, a "filter". Its purpose is to lower the $|G_c|$ at high frequencies to make the system robust against model uncertainties. The function of the filter is best understood by substituting (53) into (51)

$$Y(\omega) \leq \frac{\sigma_m(\tilde{G})}{\sigma_M(\tilde{G})} |G(i\omega)| < \frac{1}{\sigma_M(F(s))} \quad (54)$$

Note that $Y(\omega) > 1$ and that this ratio can be either unbounded or bounded as $\omega \rightarrow \infty$. Ideally we want for good closed loop performance $F(s) = I$ because for $\tilde{G} = G$, $y = FG + (y_s - p) + p$. The robustness requirement (54) forces $F(s)$ to be "small" especially at high frequencies when $\lambda(\omega)$ is large.

For SISO systems $Y = 1$ always and (54) becomes

$$|F(i\omega)| < \frac{1}{\gamma(\omega)}$$

(55)

Thus for SISO systems the restriction on the filter norm depends only on the model uncertainty $J_t(\omega)$ but not on the model itself. In the MV case the restriction becomes tighter ($\gamma > 1$) and depends on properties of the model. $\gamma(0)$ is a measure of singularity of $\tilde{G}(0)$ and is usually called the condition number. In the extreme case that \tilde{G} does not have full row rank γ becomes infinity and system stability can only be guaranteed when $F(s) \neq 0$ or in other words, by opening the feedback loop. The larger γ , the more sensitive the control performance is to a possible model/plant mismatch. Therefore we will call $\gamma(\omega)$ the sensitivity function of the system, γ is a system inherent property which limits control quality independent of the employed controller. It is therefore a convenient tool to judge the dynamic resilience of alternate designs. If we assume that $f(\omega)$ is similar for the systems under comparison, those where $\gamma(\omega)$ is small over a wider range of frequencies are preferable.

One objection commonly invoked against the use of the condition number is that it is strongly scale dependent. That is, if inputs and outputs are measured in different physical units, an entirely different γ can result. This argument is correct. However, it does not invalidate γ as a sensitivity measure. Rather, it points out that γ becomes only meaningful after scaling. \tilde{G} should be scaled to minimize γ such that the bound (54) on the filter gain is least conservative. Though an optimum scaling procedure is not available, simple suboptimal rules are available in the numerical analysis literature.

Example

The main purpose here is to demonstrate that the theoretical developments have practical significance. The last one of the motivating examples discussed in the introduction, the thermally coupled distillation columns, still awaits an explanation. When equipped with two different control structures they displayed a strikingly different performance sensitivity to modelling errors. It would be hoped that this kind of behavior is detectable by the newly

introduced techniques. The models were found not to contain time-delays or RHP zeros. The singular values and the sensitivity functions obtained after scaling (Figs. 27 & 28) are quite revealing. $Y(0)$ is about three orders of magnitude larger for Structure 1 than for Structure 2; thus the high sensitivity is not surprising. Furthermore $\text{onri}_n(0)$ for Structure 1 is larger than for Structure 2 by a factor of 30. This is reflected in the smaller steady state excursion of the manipulated variables shown in Fig. 29. In conclusion, the singular values and the sensitivity function are good indicators of closed loop sensitivity and of the effects of constraints on closed loop performance. Therefore they can form an important tool in dynamic resilience analysis.

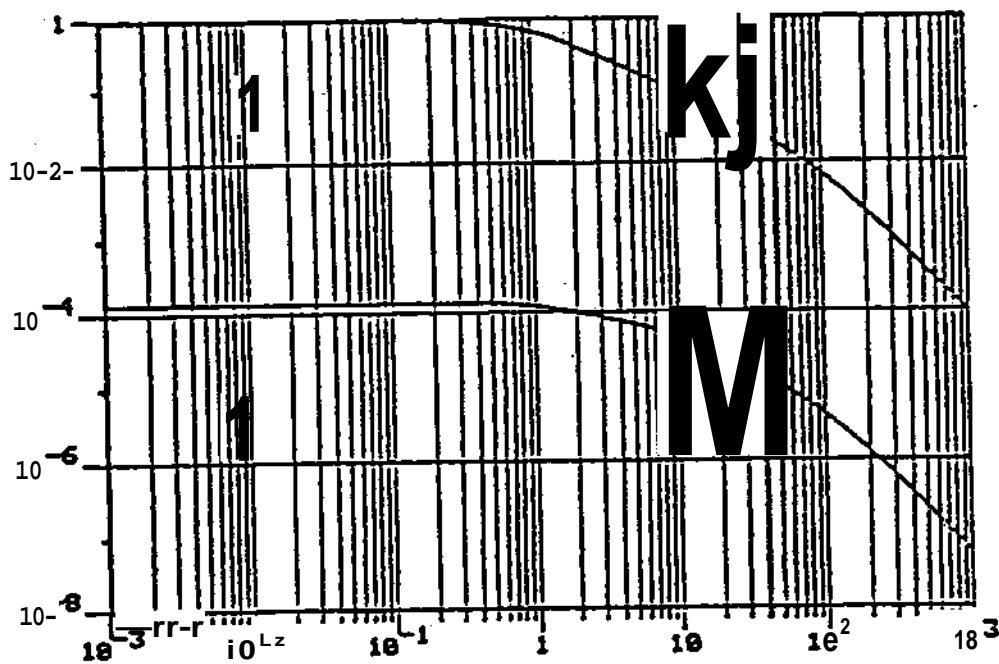
CONCLUSIONS

The main motivation behind the work presented in this paper lies in recognizing the fact that designing chemical plants for optimum economic performance or energy efficiency at nominal design conditions is usually not sufficient for guaranteeing successful designs. The objective of ensuring good operability characteristics is often of equal or greater importance due to uncertainties and changing conditions that are normally faced during plant operation.

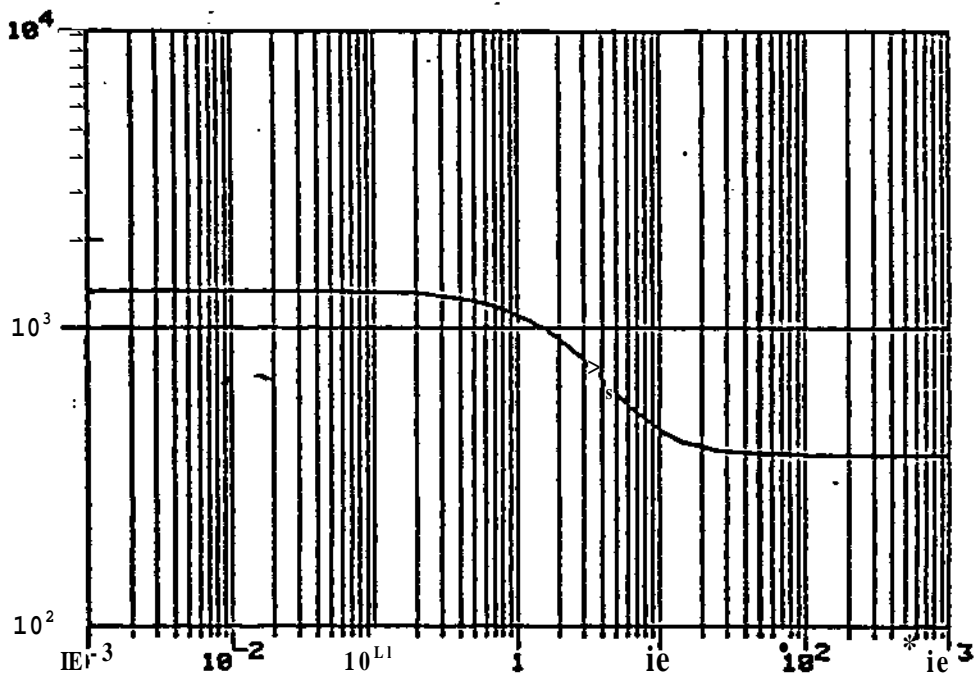
The fact that it is not always a trivial problem to incorporate properly the objective of operability in design has been demonstrated clearly with several example problems in which intuition and heuristics failed miserably. The common ideas of oversizing for flexibility, identifying "obvious" worst conditions for feasible operation, and avoiding long dead times for dynamic resiliency proved to be all incorrect in these example problems. Furthermore, the importance of selecting proper process configurations and equipment sizes to achieve flexibility, as well as the impact of design changes on the sensitivity of dynamic resilience were also established.

This paper has presented a rigorous framework for handling systematically the objectives of flexibility and dynamic resilience which are major components in plant operability. On flexibility the following fundamental concepts were introduced:

- Definition of parametric region of feasible operation
- An index of flexibility that provides a scalar measure of this

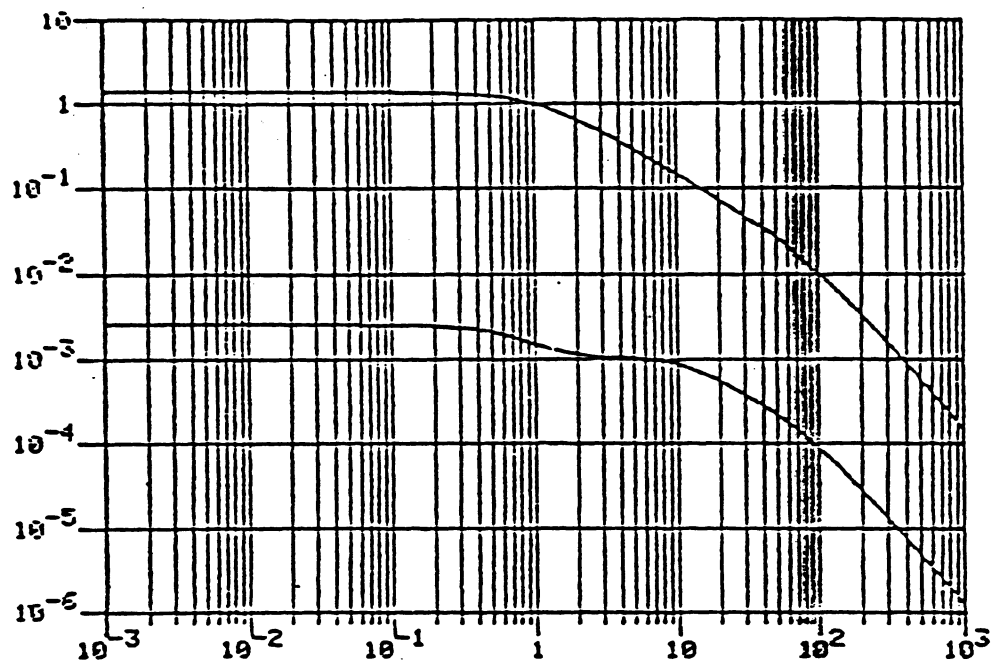


e

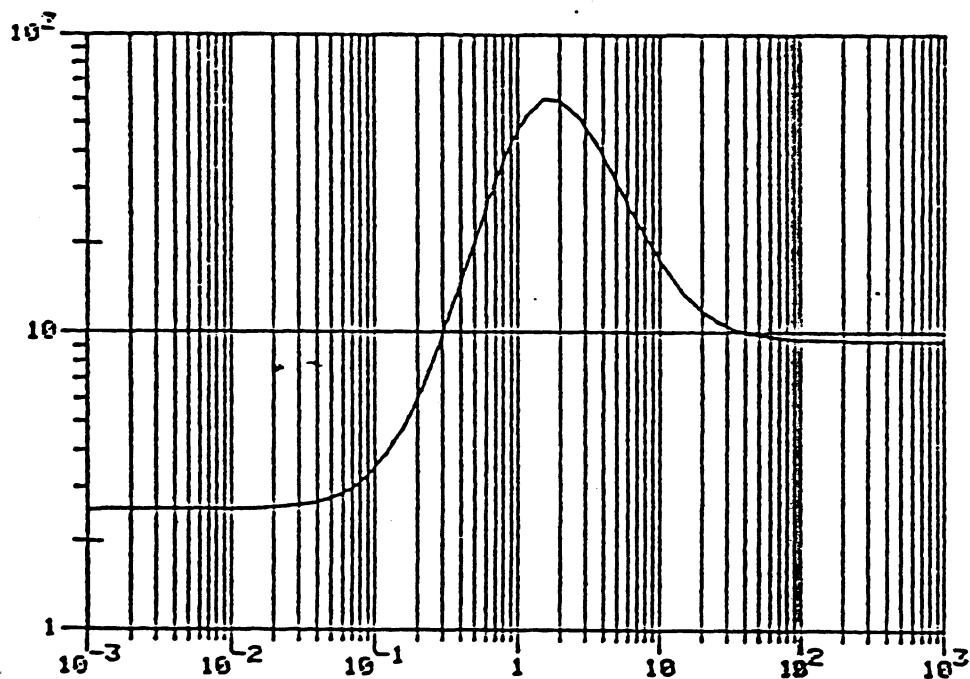


B

Fig. 27. Singular values (A) and sensitivity function (B) for distillation column with structure 1.

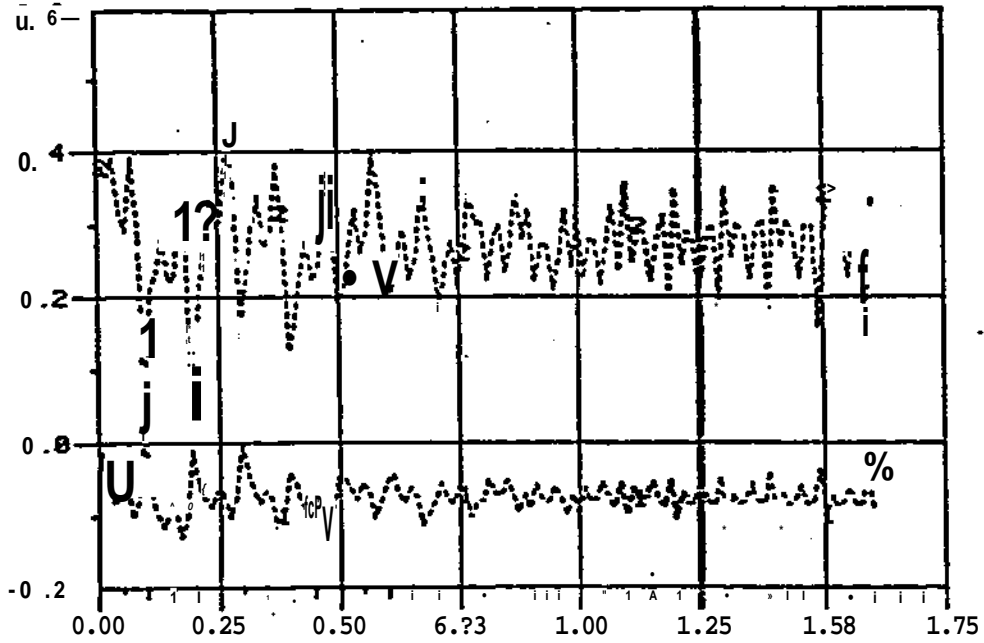


A

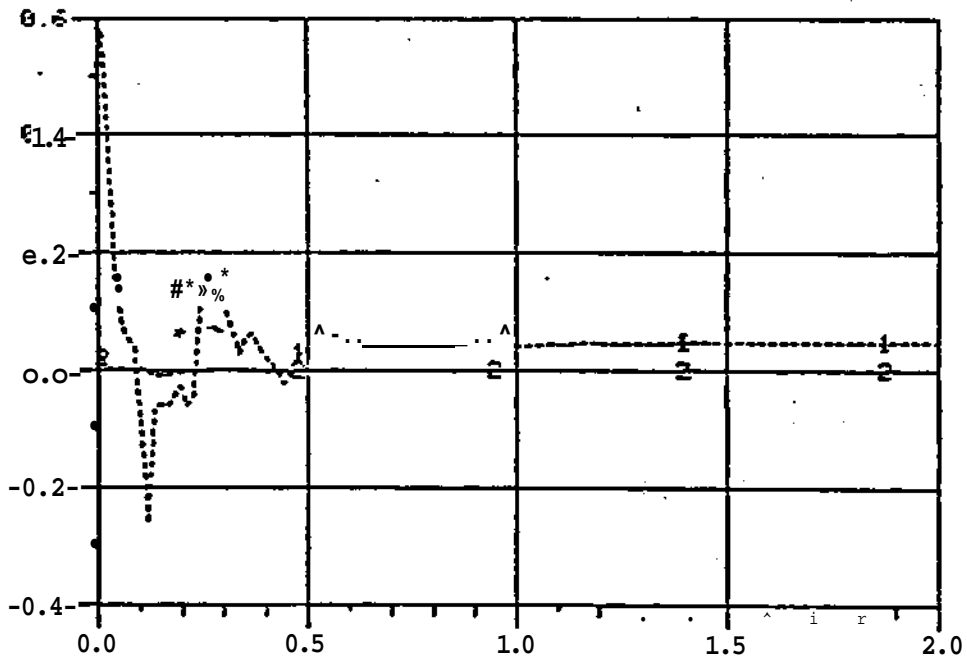


B

Fig. 28. Singular values (A) and sensitivity function (B) for distillation column with structure 2.



©



©

Fig. 29, Excursion of the manipulated variables for response shown in Fig. 6B (A) and Fig. 6C d).

region

- Definition and properties of critical parameter points for feasible operation

These concepts provide useful insights and a clear basis for understanding the nature of the problems involved in flexibility.

Three major subproblems were also formulated for incorporating flexibility in design: flexibility analysis, optimal design with fixed degree of flexibility and design with optimal degree of flexibility. It was shown that the solution of these subproblems can be simplified considerably for the case when critical points correspond to vertices of the uncertain parameters. The rigorous conditions under which this property holds true have been established.

In recent years considerable progress has been made in the development of efficient algorithms and procedures for solving various types of flexibility problems; these include,

- Computation of flexibility index through efficient vertex enumeration schemes
- Projection-restriction strategy for designing plants under multi-period operation
- Iterative multi-period method for optimal design with fixed degree of flexibility
- Synthesis of resilient heat exchanger networks

It is clear, however, that a number of important issues still remain to be answered. Among the more challenging questions, we can cite the following:

- Location of critical points. Ideally one would like to develop procedures that do not necessarily assume the critical points to be vertices. However, even when this assumption is made it would be desirable to develop procedures that only need to examine a small number of vertices for finding critical points.
- Synthesis of flexible processes. Most of the work so far has been directed to designs with fixed flowsheet configurations. However, as has been shown in resilient heat exchanger networks the selection of proper configurations has a great impact on the flexibility

of these networks. Therefore, it is clear that for other types of processes there is a great incentive to develop procedures that can account for flexibility at the synthesis stage.

- Trade-offs for cost versus flexibility. Very often increased flexibility in a process implies larger capital investment. Although this trade-off problem can be formulated conceptually as a bi-criterion optimization problem, efficient and meaningful strategies are still required to establish these trade-offs, both at the synthesis stage and at the stage of equipment sizing.

On the objective of dynamic resilience, it was shown that this property can be attributed to three characteristics inherent in the system:

- Nonminimum phase elements
- Constraints on the control action
- Sensitivity/Robustness

Methods for assessing the effect of each one of these quantitatively were presented. Depending on the system and the expected set point changes and/or disturbances one or the other can dominate. No attempts were made to combine these characteristics into a scalar objective function. The philosophy of the approach is to provide the design and control engineers with a more rigorous basis for their decision making rather than to take the decisions out of their hands.

For a variety of reasons the new framework presented shows high promise to become a standard industrial tool in the near future:

- Modelling:

For the assessment of the sensitivity and the effect of the manipulated variable constraints on the performance only frequency response data are needed. These data can be easily obtained from complex dynamic models by pulse testing. On the other hand, the steady state gains calculated from an available static model could be augmented by time constants and delays estimated from experience. Yes, even the steady state data themselves are clearly sufficient to obtain $Y(0)$ and $\omega_m(0)$, good initial indicators of dynamic resilience. Thus the modelling requirements are very flexible.

- Fundamental Rigor:

The methodology presented here assesses only the fundamental limitations to control quality which are inherent in the system itself and not those imposed by the control system.

- **Intuitive Appeal:**

All the results derived for MIMO systems are natural extensions of the heuristics which have been available for SISO systems for years.

Despite all the progress quite a few questions remain unresolved:

- The accurate computation of the RHP zeros is a problem for which no reliable numerical procedure exists at present in particular when the system contains time delays.
- The type of $G+$ factorization employed affects the condition number. It is not clear how G should be factored from the point of view of sensitivity.
- All the results were derived assuming that the open loop system is stable. An extension to open-loop unstable systems appears non-trivial.
- The method used for describing model uncertainty leads to "clear" theoretical results but might be overly conservative when information on the structure of the uncertainty is available.
- The new technique for dynamic resilience revolves around the idea of a "perfect" controller and the performance will deteriorate from that predicted if, for example, a set of single loop PI controllers is implemented instead. A way has to be found to establish the performance deterioration associated with a controller simplification.
- It is also important to point out that the linear analysis is somewhat restrictive because we know from experience the problems caused by nonlinearities which are typical for chemical processes. Though a complete nonlinear analysis lies far away in the future, a better - namely nonlinear - method to assess the effect of constraints on dynamic resilience would be desirable.

Finally, to conclude this paper, we have attempted to present here a unified treatment for operability as a process design objective. We realize that this effort marks only the beginning of a research area which is both intellectually challenging and of

practical significance. We hope that this paper will motivate researchers in academia and industry to work on many of the problems that still remain to be solved in this area.

Acknowledgement: Support from the National Science Foundation (CPE-8115022 and CPE-8121665) and the Department of Energy (DOE contract DE-AC02-80ER10645) is gratefully acknowledged.

References

- Avidan, A. I. and I. E. Grossmann, "FLEXPACK - * Computer Package for Optimal Multiperiod Design"¹¹, Proceedings for the 3rd International Congress on Computers and Chemical Engineering (Paris), Vol. II, C-158 (1983).
- Avriel, M., and D. J. Wilde, "Engineering Design under Uncertainty", Ind. Eng. Chem. (Proc. Des. Dev.), £, 124-131 (1969).
- Dittmar, R., and K. Hartmann, "Calculation of Optimal Design Margins for Compensation of Parameter Uncertainty", Chem. Eng. Sci., 31, 563-568 (1976).
- Fiacco, A. V., and K. O. Kortanek (Eds), Semi-Infinite Programming and Applications, Springer-Verlag, Heidelberg (1983).
- Frank, P. M., "Entwurf von Regelkreisen mit vorgeschriebenem Verhalten", G. Braun Verlag, Karlsruhe (1974).
- Freeman, R. A., and J. L. Gaddy, "Quantitative Overdesign of Chemical Processes", A.I.Ch.E. J., 21, 436-440 (1975).
- Grossmann, I. E., and K. P. Halemane, "A Decomposition Strategy for Designing Flexible Chemical Plants", A.I.Ch.E. J., 28, 686 (1982).
- Grossmann, I. E., K. P. Halemane and R. E. Swaney, "Optimization Strategies for Flexible Chemical Processes", Proceedings of the International Symposium on Process Systems Engineering (Kyoto), 239 (1982).
- Grossmann, I. E., and R. W. H. Sargent, "Optimum Design of Chemical Plants with Uncertain Parameters", A.I.Ch.E. J., 24, 1021-1028 (1978).
- Halemane, K. P., and I. E. Grossmann, "Optimal Process Design under Uncertainty", A.I.Ch.E. J., 2£, 425 (1983).
- Holt, B. R., and M. Morari, "Design of Resilient Processing Plants - The Effect of Deadtime on Dynamic Resilience"¹, Chem. Eng. Sci., submitted (1983).
- Holt, B. R., and M. Morari, "Design of Resilient Processing Plants - The Effect of Mght-Half-Plane Zeros on Dynaoc Resilience", in preparation (1984).
- Johns, W. R., G. Marketos and D. W. T. Rippin, "TSe Optimal Design of Chemical Plant to Meet Time-Varying Demands in the Presence of Technical and Commercial Uncertainty", Trans. Inst. Chem. Eng., 5£, 249-257 (1978).
- Kittrel, J. R., and C. C. Watson, "Don't Overdesifn Process Equipment"¹¹, Chem. Eng. Prog., 62, 79-83 (196§).
- Klema, V. C, and A. J. Laub, "The Singular Value Decomposition: Its Computation and Some Applications"¹¹, IEEE Trans. Autom. Contr., AC-25, 164 (1980).
- Lee, K. F., A. H. Masso and D. F. Rudd, "Branch aid Bound Synthesis of Integrated Process Designs", Ind. Eng. Chem. Proc. Des. Dev., 13, 209 (1974).

- Lenhoff, A. M., and M. Morari, "Design of Resilient Processing Plants. I. Process Design under Consideration of Dynamic Aspects", Chem. Eng. Sci., 37, 245 (1982).
- MacFarlane, A. G. J., and D. F. A. Scott-Jones, "Vector Gain", Int. J. Contr., 29, 65 (1979).
- Malik, R. K., and R. R. Hughes, "Optimal Design of Flexible Chemical Processes", Comp. Chem. Eng., 3, 473-485 (1979).
- Marselle, D. F., M. Morari, and D. F. Rudd, "Design of Resilient Processing Plants II: Design and Control of Energy Management Systems", Chem. Eng. Sci., 37, 259 (1982).
- Morari, M., "Flexibility and Resiliency of Process Systems", Proceedings of the International Symposium on Process Systems Engineering (Kyoto), 223 (1982).
- Nishida, N., A. Ichikawa and E. Tazaki, "Synthesis of Optimal Process Systems with Uncertainty", Ind. Eng. Chem. Proc. Des. Dev., 13, 209-214 (1974).
- Papoulias, S. A., and I. E. Grossmann, "Optimal Synthesis of Flexible Utility Systems", Proceedings of the 18th Intersociety of Energy Conversion Conference (Orlando) (1983).
- Polak, E., "An Implementable Algorithm for Optimal Design Centering, Tolerancing, and Tuning Problem", JOTA, 37, 45 (1982).
- Saboo, A. K., ChE 770 Term Project, Chemical Eng. Dept., U. of Wisconsin (1982).
- Saboo, A. K., and M. Morari, "Design of Resilient Processing Plants - III. Some New Results on Heat Exchanger Network Synthesis", Chem. Eng. Sci., in press (1983).
- Swaney, R. E., and I. E. Grossmann, "An Index for Operational Flexibility in Chemical Process Design. Part I: Formulation and Theory", submitted for publication (1983).
- Swaney, R. E., and I. E. Grossmann, "An Index for Operational Flexibility in Chemical Process Design. Part II: An Efficient Vertex Enumeration Algorithm", manuscript in preparation (1983).
- Takamatsu, T., I. Hashimoto and S. Shioya, "On the Design Margin for Process System with Parameter Uncertainty", J. of Chem. Eng. of Japan, 6, 453-457 (1973).
- Watanabe, N., Y. Nishimura and M. Matsubara, "Optimal Design of Chemical Processes Involving Parameter Uncertainty", Chem. Eng. Sci., 28, 905-913 (1973).
- Weisman, J., and A. G. Holzman, "Optimal Process System Design under Conditions of Risk", Ind. Eng. Chem. Proc. Des. Dev., 11, 386-397 (1972).
- Wen, C. Y., and T. M. Chang, "Optimal Design of Systems Involving Parameter Uncertainty", Ind. Eng. Chem. Proc. Des. Dev., 7, 49-53 (1968).
- Wood, R. K., and W. M. Berry, "Terminal Composition Control of Binary Distillation Column", Chem. Eng. Sci., 28, 1707 (1973).