OPERABILITY AND CONTROL IN PROCESS SYNTHESIS AND DESIGN

Steve Walsh and John Perkins

Centre for Process Systems Engineering
Imperial College of Science, Technology and Medicine
London, England

This chapter presents an integrated approach to process design and control in which operability and controller design are considered at the same time as process design. The aim of the approach presented is to assist designers in achieving a good trade-off between cost and

ADVANCES IN CHEMICAL ENGINEERING, VOL. 23

risk, uncertainties and disturbances notwithstanding. We address the following questions:

- *Can a particular design or structure be made to satisfy its operating constraints despite uncertainty and disturbances?*
- *What is the impact of dealing with disturbances and uncertainty on the economics of the processing system considered?*

We propose a range of tests that can be applied in a hierarchical manner to give increasingly accurate performance estimates as the number of options being considered is reduced and the design approaches its final form.

The key design tools discussed are:

- *a robust dynamic optimization package*
- *a screening method for evaluating the effect of delays on achievable performance*
- *an optimization method for dealing with parametric uncertainty.*

The application of this approach to chemical waste water treatment is also discussed extensively.

I. Introduction

Operability and control tend to be considered near the end of process design and synthesis. A process may be designed by one group of people, then another design group is handed the process flowsheet and required to add on a control scheme to ensure that it operates successfully. Changes in the process design at this stage are usually expensive both in resources and in time. This approach has long been recognized as far from ideal, but the problem of how to do things better has not yet been resolved satisfactorily.

Integration of operability and control considerations into early-stage process design can be facilitated by organizational measures. Process control experts can be brought into the early-stage design team to identify and investigate possible operational difficulties. Process engineers can receive additional training to enhance their understanding of operational and control issues. Such organizational measures need to be supported by appropriate design procedures and tools if they are to deliver their full potential benefit.

Operability encompasses the ability of the system to cope with uncertainty and disturbances and also with issues of reliability and maintenance. In this contribution, we restrict our discussion to methods of dealing with uncertainty and disturbances. We address the following questions:

- Can a particular design or structure be made to satisfy its operating constraints despite uncertainty and disturbances?
- What is the impact of dealing with disturbances and uncertainty on the economics of the processing system considered?

It should be noted that these questions do not fit comfortably within any of the conventional categories of operability analysis, such as controllability (regulatory capability about a steady state), switchability (ability to switch between operating modes), flexibility (ability to accommodate uncertainty at steady state), and robust control (ability to maintain stability and performance of a control scheme despite perturbations in the characteristics of the controlled system). Rather, we attempt to develop an integrated design approach that allows us to consider operability issues on a par with economic issues, thus permitting design and synthesis decisions to be made within a common framework.

An important strand of previous work addressing the issues of dynamics and uncertainty in process and control system design is the development of design guidelines based on a mixture of experience, theory, and intuition. Among the notable contributors to such guidelines are P. Buckley, F. G. Shinskey, and W. L. Luyben. We value these contributions and believe they repay careful study. However, because such qualitative guidelines do not place operability issues on a par with process design issues, they do not allow integrated consideration of process and control decisions. Our approach exploits advances in optimization methods and computing capabilities in order to achieve this.

Engineers are required to produce designs that will operate safely and legally at all times while providing an adequate return on investment. They must meet these requirements despite incomplete knowledge of the process and equipment characteristics and the operating demands on the plant. Success or failure will depend on how well the designers manage the trade-off between risk and return on investment and how well they anticipate potential operational problems. Integrated design approaches aim to provide tools, procedures, and environments to aid engineers in tackling these issues in a systematic, well-coordinated manner.

The design problem described above is a multifaceted, multiobjective, incompletely defined problem and as such requires the application of creative intelligence. As no computer yet exhibits such creativity, any integrated design approach should be built around a human project team. Computer-based techniques should be designed to assist such a team and should not attempt to carry out tasks that humans are better able to do. The challenge in integrated design is to find a good "division of labor" between human and computer and an appropriate way of interfacing between the two.

With these considerations in mind, we have chosen the following approximation to the integrated design problem. The integrated design objective is to

maximize return on investment while ensuring that performance constraints are satisfied for all possible plant parameters and disturbances.

This approximation involves a number of nontrivial assumptions.

- The performance requirements can be represented either by constraints or by a contribution to the operating costs. In some cases, this is straightforward—equipment and material and energy inputs can be included in an economic objective; constraints for which violation leads to plant shutdown are well approximated by a hard constraint. In other cases, the choice of representation is more difficult—excessive quality variations may lead to complete loss of the market for a product (constraint) or reduced price for the product (objective).
- Adequate models must be available to allow evaluation of the effect of design decisions on performance. This can be very difficult to achieve, particularly if we desire to characterize a wide range of options in depth at the outset of a design. We assume that the project team acts to refine and augment the models required for performance prediction as the design progresses. The models can take any form, from correlation curves through differential-algebraic equations.
- Uncertainty and disturbances can be described in terms of mathematical constraints defining a finite set of *bounded* regions for the allowable values of the uncertain parameters of the model and the parameters defining the disturbances. If uncertainty or disturbances were unbounded, it would not make sense to try to ensure satisfaction of performance requirements for all possible plant parameters and disturbances. If the uncertainty cannot be related mathematically to model parameters, the model cannot adequately predict the effect of uncertainty on performance. The simplest form of description arises when the model is developed so that the uncertainty and disturbances can be mapped to independent, bounded variations on model parameters. This last stage is not essential to the method, but it does fit many process engineering problems and allows particularly efficient optimization methods to be deployed. Some parameter variations are naturally bounded; e.g., feed properties and measurement errors should be bounded by the quality specification of the supplier. Other parameter variations require a mixture of judgment and experiment to define, e.g., kinetic parameters.

Matching the above approximation to the original design problem requires the intervention of the project team to define

- objectives and constraints to approximate performance requirements
- models to link these to the choice of design variables
- bounds on uncertain variables to manage the trade-off between cost and risk

Problem definition and solution will usually be an iterative process. Each solution of the approximate problem can be used to provide estimates of cost and to indicate the critical values of the uncertain parameters. This information allows the project team to trade off risk and cost at a very high level. In working with industrial project teams, we have found them to be quite comfortable with defining problems in this way and with exploring the interactions between problem definition and the cost of the resulting design.

If good information is available on the probability distribution of the uncertain parameters, it becomes possible in principle to design to a specified probability of infeasibility (Straub and Grossmann, 1990). This would provide an alternative method of trading off cost and risk. The algorithms to deal with this case are significantly more complex than those for dealing with bounded uncertainty as it is necessary to identify the complete boundaries of the feasible region in order to evaluate the probability of the uncertain parameters lying within the feasible region. Because of the difficulty of solving this problem for general nonlinear systems and the difficulty of obtaining reliable statistical data on the variation of the uncertain parameters, we have not pursued this approach.

In considering the solution of the approximate problem, a number of points are evident. It is highly desirable to tackle the problem with a hierarchy of methods and models. The design process can be likened to a mixture of broad and deep searches over the "space" of design options. The search has two main mechanisms:

- narrowing the search domain by eliminating sets of options
- deepening the search by picking likely "winners"

Eliminating options is assisted by simple criteria or tests that can identify a large number of options as unviable. Deepening the search can be assisted by more heuristic measures as the aim is usually to establish a viable base-case design rapidly, without necessarily discarding other options. The appropriate criteria to use and the balance between search methods is clearly problem-dependent.

There are a number of general techniques suggested by the problem formulation. At the most detailed level of design, the design parameters need to be optimized in relation to performance criteria based on a nonlinear dynamic model. This points to a need for effective tools for dynamic optimization. At a more preliminary level in a hierarchy of techniques, it might be useful to evaluate steady-state performance or to carry out tests on achievable dynamic performance to eliminate infeasible options. Appropriate screening techniques are therefore needed. All these methods can use nominal models for initial analysis, but a full analysis should be based on design with uncertainty.

The rest of this chapter is structured as follows. The next section considers general techniques for use in the integrated design approach proposed: design with uncertainty, screening tools for disturbance rejection, and dynamic opti-

mization. Then a particular problem domain—chemical waste water neutralization—is introduced. A procedure is presented for coordinating the techniques in an integrated design framework for this problem area. Key modeling issues for this problem domain are discussed. A generic example and four industrial case studies are then presented, illustrating different aspects of the use of this procedure. We conclude the chapter by reviewing our experience on the problem domain discussed and highlighting some directions for further extension of the proposed integrated design approach.

II. General Techniques

In the previous section, an approach to synthesis and design taking account of uncertainty and the impact of process disturbances on performance was outlined. In this section, a set of tools to support the approach will be presented. First, we examine techniques for the design of plant for which uncertainties in design data are explicitly represented. These techniques are computationally demanding; currently, they are most suitable at the later stages of design when the most promising designs are being verified and finalized. In the early stages, where rapid screening of a large number of options is required, less computationally demanding tools are more appropriate. A technique based on the ability to reject process disturbances is presented for the rapid evaluation of design alternatives. Finally, since the aim is to evaluate the dynamic behavior of the process, and since optimization is a key component of our approach, algorithms for solving optimal control problems (i.e., optimization problems subject to differential equation constraints) are discussed.

A. Design with Uncertainty

Some approaches to design with uncertainty are reviewed in order to identify the appropriate basic approach and relevant techniques from previous work. A new algorithm is then presented which builds on previous work and attempts to exploit the characteristics of the design problems.

1. Review

As noted by Grossmann *et al.* (1983), the problem of design with uncertainty is not well-defined and many different approaches exist. As discussed above, we are interested in solving problems in which the uncertainty is assumed to be defined by bounds on model characteristics or parameters and in which it is

assumed that the constraints must be satisfied at all times. Attention will therefore be restricted to techniques to ensure feasibility for all possible realizations of the uncertainty (worst-case design). This review considers only optimization-based methods that are potentially applicable to general nonlinear systems. Algorithms designed for linear or bilinear systems or involving manipulation of explicit objectives and constraints are therefore not covered.

Two approaches to worst-case design that have been applied successfully to engineering design problems with uncertainty are those of Grossmann *et al.* (1983; Swaneyand Grossmann, 1985a, b; Grossmann and Floudas, 1987) and Polak *et al.* (Polak, 1982; Tits, 1985; Polak and Stimler, 1988; Mayne *et al.*, 1990). The problem tackled by both approaches is how to deal with an effectively infinite number of constraints (i.e., constraints must be satisfied for the infinite number of parameter realizations that must be considered), which makes the design optimizations *semiinfinite*. In both cases the approach adopted is to approximate the continuous uncertain parameter space, $v \in V$, by a discrete set, $v \in V_i$ and to update this set until it gives a design for which no constraint violation can be found. This is known as an outer-approximation algorithm, as the constraints associated with $v \in V_i$ define a feasible region which contains (i.e., is an outer approximation to) the feasible region associated with $v \in V$. This is illustrated in Fig. 1, which shows how three scenarios of the uncertain parameters give an approximation to the true feasible region.

Grossmann *et al.* consider the set V to be defined explicitly by bounds on the uncertain parameters. Polak *et al.* allow for general constraints on the values of V subject to the requirement that V is made up of the union of a finite number

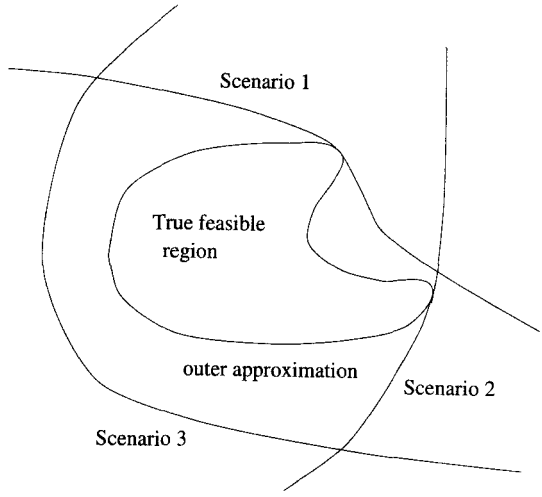


FIG. 1. Effect of outer approximation.

of compact subsets. While many engineering problems fit the simpler case of independent bounded parameter variation, given appropriate models, this approach is not universally appropriate. For example, a feed may have a quality specification on total impurity level as well as specifications on levels of individual impurities. This would give rise to an uncertainty specification of a limit on the sum of the uncertain parameters defining the impurities. We focus below on the simpler case, but comment on extensions necessary to handle the more general case.

The general structure of an outer-approximation algorithm for worst-case design is as follows

1: Choose an initial set $V_i (i = 0)$ to approximate V

2: Carry out a design so that the constraints are satisfied for all $v \in V_i$.
If this stage fails then the problem is infeasible.

3: Find the maximum constraint violation over the set V, c^*, and the corresponding value of v, v^*
If $c^* \leq 0$ then solution found
Else $V_{i+1} = V_i \cup v^*$, $i = i + 1$, goto step 2

The methods differ mainly in step 3.

Grossmann *et al.* (1983) formulate the worst-case design problem as

$$\begin{aligned} &\min_{p \in P} \; E_{v \in V} \{ \min_{o \in O} J(p, v, o) | c(p, v, o) \leq 0 \} \\ &\text{s.t.} \quad \forall v \in V \{ \exists o \in O \langle \forall k \in K [c_k(p, v, o) \leq 0] \rangle \} \end{aligned} \tag{1}$$

where J is the objective function to be minimized, p is a vector of design variables, v is a vector of uncertain variables, c_k is the kth element of the constraint vector c, and o is a vector of operating variables that may be adjusted to reduce costs and to maintain feasibility in the light of the value of v. The set V is assumed to be a polyhedron defined by simple bounds on elements of v. The infinite constraint may be reformulated as

$$\max_{v \in V} \min_{o \in O} \max_{k \in K} [c_k(p, v, o)] \leq 0 \tag{2}$$

The design variables are chosen to allow constraint satisfaction for *all* the uncertain variable values while the operating variables are adjusted for *each* value of the uncertain variables. If operating variables are eliminated, the problem simplifies to

$$\begin{aligned} &\min_{p \in P} \; E_{v \in V} \{ J(p, v) \} \\ &\text{s.t.} \quad \forall v \in V \{ \forall k \in K [c_k(p, v) \leq 0] \} \end{aligned} \tag{3}$$

The formulation without operating variables is qualitatively easier to solve as one level of optimization is eliminated. More specifically, finding v^* for the optimization problem defined by Eq. (2) is a nondifferentiable global optimization problem that is extremely difficult to solve rigorously in the general case. It is therefore important to consider the pros and cons of using operating variables carefully.

The motivation for including operating variables is that certain variables may be adjusted during plant commissioning or operation to give improved performance in the light of the actual plant behavior. Requiring such variables to be chosen so as to accommodate all possible uncertain variables, i.e., as design variables, introduces an element of conservatism into the design. This is particularly the case if the operating variables include process inputs which would be adjusted by a control scheme to maintain satisfaction of constraints and if the uncertain parameters are fixed but unknown rather than variable. On the other hand, including operating variables in the problem formulation assumes an *ideal* adaptation of the operating variables to *all* the uncertain variables, which may actually vary over time. This problem will generally be optimistic as the actual "operator" will have only partial knowledge of some of the uncertain parameters and will adapt the operation in a nonideal way. A design generated using optimization of operating variables will therefore *usually* be infeasible for some values of $v \in V$. Failure of a design problem with operating variables indicates that no control scheme for adjusting the operating variables can achieve feasible operation for all the uncertain parameters. Success of such a design problem does not imply that an *implementable* control scheme exists which can achieve feasible operation. Operating variables, therefore, should not be used in determining design parameters, although they may be useful in certain screening tests if the resulting problem can be solved efficiently (see Section II.B).

To go beyond the potential conservatism of having all the variables as design variables and the probable optimism of using operating variables, it is necessary to include the adaptation mechanism (control scheme) within the model. If desired, the parameters controlling this adaptation can be made design variables. Grossmann *et al.* (1983) consider that this "would make the problem virtually unmanageable" at the design stage. In the integrated design context, which involves consideration of process *and* control system design, it is certainly appropriate. More generally, including basic control information may be accomplished simply by requiring that certain variables remain at their setpoints, which may be added to the design variables, and eliminating the operating variables, o, using the extra equality constraints. The key formulation for design with uncertainty in relation to the integrated design approach is that without operating variables. Solving general problems with operating variables is not necessary in the context of our approach to integrated design. However, the penalty paid for adopting a particular adaptation scheme is that we obtain no information

from the optimization as to what improvements could be made by a different scheme.

In replacing operating variables by a control scheme with associated design variables, the designer has moved from a test for which "no" means *no* and "yes" means *maybe* to a test for which "yes" means *yes* and "no" means *maybe*. The two approaches are clearly complementary. The operating variable approach is best suited to screening the potential feasibility of a design (we want "no" to mean *no*) or setting targets for design, while it is only appropriate for direct use in design if the designer is confident that approximately ideal adaptation can be achieved in practice. The embedded control scheme approach is best suited to use in design, particularly when it is necessary or desirable to design the control scheme itself, while it is ideally suited for screening only if one can parameterize and optimize the best controller one might conceivably implement. In the new algorithm and examples presented below, we do not make use of operating variables.

It is possible to solve the general formulation (including operating variables) rigorously for certain special cases. The algorithms developed for this purpose are of interest as they include techniques that are useful for developing a general algorithm.

An algorithm for solving the general problem under the assumption that the worst-case uncertain parameters lie at vertices of the parameter space, V, is given by Grossmann *et al.* (1983). This is presented below following the general outer-approximation algorithm structure given above.

1: $V_i(i = 0)$ is selected based on the sign of gradients of individual constraints with respect to v. Positive sign indicates that the parameter would maximize that particular constraint at its upper limit if the constraint is monotonic.

2: Solve for new design parameters, p^*, which minimize the expected cost subject to the constraints associated with each $v_j \in V_i$

$$\min_{p, o_j, j=1, \ldots, N_i} \sum_{j=1}^{N_i} w_j J(p, o_j, v_j)$$

$$\text{s.t. } c(p, o_j, v_j) \leq 0, \qquad j = 1, \ldots, N_i$$

(The weighted sum is used by Grossmann *et al.* to approximate the expectation, N_i is the dimension of V_i.)

3: For each vertex, v^v, in V choose o to minimize the maximum constraint violation for the new design parameters.

$$\min_{o \in O} \{c_{\max} | c_{\max} \geq c_k(p, v_j, o), \forall k\}$$

(It is not necessary to carry out the minimization exhaustively as the minimization for a particular vertex can stop once $c_{\max} < 0$.)

Choose v^* as the vertex giving the largest value of $c_{\max}$

If $c_{\max}(v^*) \leq 0$ then solution found
Else $V_{i+1} = V_i \cup v^*$, $i = i + 1$, goto step 2

The fundamental limitation of this approach is that it *assumes* the worst-case parameters are always at a vertex. This requires convexity properties that will not be satisfied in all the problems of interest. A *sufficient* condition for a vertex solution (Swaney and Grossmann, 1985a) is that all $c_k(p, v, o)$ are jointly quasi-convex in o and one-dimensional quasi-convex in v. One-dimensional quasi-convexity implies that

$$\begin{gathered} \max(f(x_1), f(x_2)) \geq f(\alpha x_1 + (1 - \alpha)x_2) \\ \forall \alpha \in [0, 1], \forall \alpha_2 \in \mathcal{R}, \forall x_1, x_2 = x_1 + \alpha_2 e_i \end{gathered} \tag{4}$$

where e_i is a vector with ith element 1 and all other elements 0. A nonvertex solution requires a maximum in v, for some $\min_o c_k$, which is not at a bound of V. This is precluded by the above conditions. Nonvertex solutions may occur for the design problems of interest. Two examples illustrating this are given below.

1. If the time between two step disturbances is an element of v, then the worst case is not necessarily at either bound of this variable.
2. A sinusoidal disturbance with uncertain frequency will tend to have a worst case near the resonant frequency of the control system design which will not, in general, lie at a limit of the uncertainty range.

However, many elements of v can be expected to have a worst case at a bound, e.g., measurement bias, flow, concentration, buffering, and reaction rates. For the general case, the vertex assumption cannot be relied upon but provides a basis for useful heuristics in the search for a solution.

Even with the assumption of vertex constraint maximizers, vertex enumerations with a large number of parameters can be very time-consuming. Algorithms for efficient exploration of the vertices and a proposed "flexibility index" are presented in Grossmann *et al.* (1983) and developed further in later papers (Swaney and Grossmann, 1985a, b). These are discussed below.

The flexibility index, F is defined by

$$\begin{gathered} F = \max \delta \\ \text{s.t.} \max_{v \in V_\delta} \min_{o \in O} \max_{k \in K} c_k(p, v, o) \leq 0 \\ V_\delta = \{v | (v_0 - \delta \Delta v^-) \leq v \leq (v_0 + \delta \Delta v^+)\} \end{gathered} \tag{5}$$

where v_0 is the nominal value of v. Therefore, F is the factor by which a polyhedron representing nominal variability or uncertainty of the parameters can be

expanded without leading to constraint violation for any parameter contained within it. An equivalent representation more useful for solution is

$$F = \min_{\tilde{v}\in\tilde{V}} \max_{\delta,o} \bar{\delta}$$

$$\text{s.t. } c(p, v, o) \le 0 \qquad \forall \delta \in [0, \bar{\delta}] \tag{6}$$

$$v = v_0 + \delta\tilde{v}$$

$$\tilde{V} = \{\tilde{v} | -\Delta v^- \le \tilde{v} \le \Delta v^+\}$$

The condition $\forall\delta \in [0, \bar{\delta}]$ is *assumed* to be satisfied in the solution methods presented if $\bar{\delta}$ gives a feasible point giving the simplified problem

$$F = \min_{\tilde{v}\in\tilde{V}} \max_{\delta,o} \delta$$

$$\text{s.t. } c(p, v, o) \le 0$$

$$v = v_0 + \delta\tilde{v} \tag{7}$$

$$\tilde{V} = \{\tilde{v} | -\Delta v^- \le \tilde{v} \le \Delta v^+\}$$

This index could be used for trading off risk and cost but such indices oversimplify a complex tradeoff. The mapping between F and risk is not direct. Some parameters may never violate their bounds, others may have a significant probability of doing so. The use of F treats all parameters as having a uniform likelihood of violating their bounds, which is not generally appropriate.

The calculation methods for the flexibility index discussed below can be applied to step 3 of the worst-case design optimization procedure to maximize constraint violation instead of minimizing δ.

Under the vertex solution assumption, the flexibility index can be calculated by evaluating the maximum δ along each vertex direction and taking the minimum of the results. This approach becomes computationally impractical for more than about 15–20 parameters, so two procedures are presented which give upper bounds on F more efficiently (a vertex search method and a branch-and-bound method). Of the two methods, the vertex search method is found to be more efficient on the examples considered and therefore seems the best candidate for use to get an approximate solution for the worst-case vertex where appropriate. The vertex search procedure is given below.

1: Set $\rho^{\max}$ (see step 3) and choose an initial value of v

2: Update v using $\text{sign}(\tilde{v}_i) = -\text{sign}(\partial\delta/\partial v_i)$
until either δ fails to decrease as predicted or
the method predicts a vertex already examined

3: For each $c_k < 0$ compute the maximum increase Δc_k in the constraint, based on the constraint gradients, and identify the corresponding vertex
Compute the projected fractional change $\rho = -\Delta c_k / c_k$
If the maximum value of ρ over all the constraints is less than ρ^{max} or no new vertices are identified then STOP
Else select the new vertex value of v giving the maximum value of ρ and goto 2:

ρ^{max} governs the thoroughness of the local search. $\rho^{max} = 1$ corresponds to examining a vertex if the linearized constraints at the present vertex indicate that some currently inactive constraint might become active. $\rho^{max} = 0.5$ corresponds to examining a vertex if the linearized constraints at the present vertex indicate that the constraint will move halfway to becoming active.

Swaney and Grossmann (1985b) note that the assumption of vertex solutions can be relaxed somewhat by carrying out a local search from the solution vertex if a descent direction exists and give one example where this approach identified a nonvertex worst point.

Grossmann and Floudas (1987) present a complementary approach to evaluating the flexibility index, F. This approach introduces integer conditions which for linear constraints allow prediction of the subset of constraints that will be active at the solution, based on the gradients of the constraints with respect to the operating variables, o. These conditions can be directly incorporated into a mixed-integer linear program (MILP). For nonlinear constraints, a modified approach is presented to decompose the problem into a series of smaller nonlinear programs (NLPs) corresponding to the predicted active constraint sets. In the absence of operating variables, this approach reduces to solving the same number of NLPs as the number of inequality constraints, with the inequality constraints replaced in each case by a single equality constraint. For general problems the method requires that the constraints be monotonic in o for all v. Global solutions to the NLP subproblems are guaranteed if the active constraints are jointly quasi-concave in o and v and strictly quasi-convex in o for fixed v. This approach seems to have little merit for the problems of interest for which the required mathematical properties are unlikely to be met.

Polak (1982) gives a different algorithm for solving the general worst-case design problem that attempts to avoid reliance on special convexity or concavity properties. The allowable values of v are given by a set of inequalities defining a compact subset rather than by simple bounds. Tits (1985) notes an error in this algorithm and this method appears to have been abandoned. Tits suggests using a vertex assumption or using local searches from all or a subset of the vertices at each iteration to improve the likelihood of a global maximum. This is not far from the algorithm of Grossmann *et al.*

In a more recent paper Polak and Stimler (1988) note that

> To date, the use of semiinfinite methods in worst-case control system design with parametric uncertainty has been extremely limited because of the above

> mentioned computational problem of evaluating the maximum of the constraints.

The paper also states that

> The most general semiinfinite optimisation problems that are solvable by existing algorithms are of the form
>
> $$\min_p \{J(p) | c(p) \leq 0;\ \max_{v^i \in V^i} c^i(p, v^i) \leq 0,\ i \in \{1, ..., I\}\}$$

This confirms the difficulty of solving problems with operating variables, o, in a rigorous manner. Note that V^i represents the ith subset of the uncertain parameter space and not an outer approximation set (V_i).

The algorithms used for solving this problem are based on the application of general global optimization methods, such as grid enumeration, for solving the constraint maximization problem without operating variables, o. When applying general global optimization methods to the constraint maximization problem, it does not matter greatly whether the set V is defined in terms of simple bounds, linear inequalities, or nonlinear inequalities. So long as checking whether or not an element v is within V is computationally cheap, the effort to deal with the different cases will not vary much. However, as the definition of V becomes more complex, it becomes more difficult to find efficient approximations to the global optimization problem or special cases for which more efficient methods can be applied. For the case of linear inequalities, the vertices of the linear constraints provide a natural extension of exploring the vertices defined by the bounds. Linear equality constraints on elements of v can be used to eliminate uncertain variables and are therefore beneficial rather than detrimental.

Polak and Stimler (1988) develop approximations to some worst-case controller design problems which are more amenable to solution. This is accomplished by replacing the constraint

$$c(p, v) \leq 0 \qquad \forall v \in V \tag{8}$$

by

$$\tilde{c}(p, v) \leq 0 \qquad \forall v \in V \tag{9}$$

where

$$c(p, v) \leq \tilde{c}(p, v) \qquad \forall v \in V \tag{10}$$

This operation is referred to as majorization and allows simpler constraints to be used, facilitating solution at the expense of reducing the feasible space for p. This method is not readily applicable to nonlinear dynamic optimization, where there is no formula for the constraints; but trading conservatism in design for ease of solution can be achieved for the problems of interest by a different strategy discussed in Section II.A.2.

A more recent paper (Mayne *et al.*, 1990) suggests a technique for reducing the number of global maximizations required in outer approximation methods by tracking local minima previously identified as global minima until they cease to give constraint violations. They recommend gradually increasing the effort deployed on constraint maximization at each iteration, provided a constraint violation is obtained.

None of the methods reviewed provides a complete solution to the type of problems of interest. They suffer either from making assumptions that are not likely to be satisfied or from relying on a computationally unrealistic degree of brute force. This review has, however, highlighted a number of methods that could usefully be exploited in an algorithm for worst-case design.

1. Exploiting the heuristic that many local minimizers of constraints lie at vertices of the parameter space through vertex searches and local searches from promising vertices.
2. Constructing suitable approximations to the original problem to make it easier to solve while still giving meaningful solutions.
3. Increasing the effort employed on constraint maximization as the optimization progresses.

These ideas are used below in developing a new algorithm.

2. *A New Algorithm for Worst-Case Design*

This section presents a new algorithm for design with uncertainty. This algorithm does not represent a radical departure from previous approaches, but rather represents a new *variant* of the basic outer approximation algorithm discussed above. The key feature is the use of approximate global optimization to give reduced reliance on convexity properties compared to the algorithm of Grossmann *et al.* discussed above, while not relying on the rigorous but expensive global optimization methods favored by Polak *et al.* The algorithm seeks to reduce computational cost by relying on local search methods in initial iterations and by introducing a systematic overdesign factor to the design iterations.

The discussion below is supplemented by a pseudocode listing in Appendix A.

a. General Observations. No operating variables are included in the optimization formulation handled by this algorithm. Such adaptation of system variables as is possible is assumed to be embedded in the model via the control scheme so that no operating variables are necessary in the optimization problem. This gives the optimization formulation

$$\min_{p} \{J(p, V) | \max_{v \in V} \max_{k \in K} [c_k(p, v)] \leq 0\} \tag{11}$$

where $J(p, V)$ represents an objective function that depends on the design variables and the set of uncertain variables. This encompasses nominal cost, expected cost, and maximum cost. Maximum cost

$$J(p, V) = \epsilon | J(p, v) \leq \epsilon \qquad \forall v \in V \tag{12}$$

or a weighted cost

$$J(p, V) = \sum w_j J(p, v_j), \qquad v_j \in V_i \tag{13}$$

provide two readily implementable forms of $J(p, V)$. The weights are defined so that one particular model (usually a nominal model) can be given a fixed weight between 0 and 1 and the other active models are given equal weight to make $\sum w_j$ equal to 1.

The algorithm presented could be adapted to the case with operating variables. The main problems for extension to this case are as follows:

1. The local search procedures are more likely to suffer problems due to discontinuities either from changes in the identified local minima with respect to o (discontinuity in value or first derivative) or from changes in the active set of constraints at a minimum with respect to o (discontinuity in first derivative). This is likely to cause failure to converge to a local maximum with respect to v in at least some cases.
2. The computation time increases as an optimization replaces a function evaluation.
3. The design problem is of much larger dimension than p if the dimensions of o and the set V_i are large.

The uncertain and variable parameters are assumed to lie within a polyhedron defined by simple bounds on the variables ($v \in V$). The advantage of this description is that the boundaries of the uncertain parameter space, V, are easily identified and many constraint maximizers can be expected to lie at the vertices of the uncertain parameter space. Similar advantages could be obtained for more general uncertainty descriptions such as linear inequalities, although the implementation would be more complex. The extension to uncertain parameters lying in the union of multiple bounded sets can be made straightforwardly but has not been implemented. For more general representations of V, the heuristic of exploring vertices has no clear meaning and would have to be dropped, with correspondingly greater emphasis on general global optimization techniques.

Vertex constraint maximizers are not *assumed*, although they are expected to be common. Both local and global searches into the interior of V may be carried out as part of the search procedure, the effort expended in finding a new constraint maximum being bounded by user-defined variables.

In order to obtain a solution in reasonable time, it is extremely important that the design algorithm not take too many iterations to identify an outer approximation adequate to force a feasible design. One method of tackling this is to use the method of Mayne *et al.* (1990), in which previous local maximizers are

tracked following the update of design variables and both the original and updated maxima are included in the design set. This is not particularly well suited to the case in which many maximizers are expected to be at vertices of V, as is typical of engineering problems. It is, however, possible to develop an alternative, heuristic method to trade off number of iterations against accuracy of solution.

The method used involves building a projection factor, ϵ_p, into the model. Setting $\epsilon_p = 0$ corresponds to the actual problem definition. $\epsilon_p > 0$ corresponds to a performance specification that is more demanding than the actual specification, with the difficulty of the problem increasing with increasing ϵ_p. For each set of uncertain parameters (scenario) that is to be added to the scenarios considered for design, setting ϵ_p to a value greater than zero should make it more difficult to find a satisfactory design. The resulting design, which is able to accommodate the selected scenarios with the chosen, nonzero ϵ_p, should then be more likely to be feasible for all $v \in V$ than a design based on the same scenarios and $\epsilon_p = 0$. This characteristic may allow convergence to a feasible solution of the actual problem using fewer iterations of the outer approximation algorithm than if ϵ_p were set to zero in the design problems. Faster convergence will, however, be at the expense of a degree of overdesign. The final design will give an objective function value lying somewhere between the cost of a design that gives guaranteed performance for the original specification ($\epsilon_p = 0$ is used in constraint maximization to check feasibility) and the cost of a design that gives guaranteed performance for the chosen, nonzero ϵ_p (used in the design optimization).

The appropriateness of this approach depends on being able to find a sensible definition of ϵ_p. In engineering problems, performance is typically limited by the maximum disturbance magnitude or the maximum or minimum plant throughput. The design problems might therefore be solved with throughput variations or disturbances scaled up by a factor of $1 + \epsilon_p$ compared to the actual scenarios being considered. Feasibility would then be checked by maximizing constraint violation with the actual throughput or disturbance conditions. The projection factor ϵ_p may be large on initial evaluation of designs (where speed is at a premium) and be reduced for a more precise optimization of the final design (where accuracy is at a premium).

For this heuristic strategy to be appropriate, the key conditions are as follows:

1. There should be a sensible way of defining ϵ_p, e.g., a fractional increase of disturbance amplitude or throughput.
2. Adaptation of the design variables to increased ϵ_p should improve performance over most or all of the uncertain parameter set, V, e.g., through increased steady-state offset from the active constraints or increased process capacity.

The first condition is straightforward and usually trivial. The second condition is necessary to exclude cases for which increased ϵ_p pushes the design in a direction which expands the feasible region locally but creates new constraint violations elsewhere in V, and hence gives no net benefit. Although this condition cannot be verified *a priori*, the heuristic was found to work effectively in the design examples (Section V), indicating that this condition is adequately satisfied in the problems examined.

The new worst-case design algorithm is discussed further below following the general outer-approximation algorithm structure; constraint maximization, initialization, and multi-model design.

b. Constraint Maximization. The constraint maximization algorithm makes use of the heuristic that the maximizing values of the uncertain variables will often lie on the bounds. It also uses the idea of progressively increasing the amount of effort applied to the constraint maximization as the search proceeds. This progressive increase of effort is implemented by stopping the search if a constraint violation is found by the time the previously applied depth of search has been completed. If not, the search proceeds further, subject to a specified maximum depth of search, until a constraint violation is found and the effort required is recorded to set the minimum effort in the next constraint maximization.

1. In the early iterations of the design, local vertex searches as used by Grossmann and Swaney provide an efficient way to generate new maximizers and push the design toward the required robustness. Local searches from a subset of the vertices may be used to try to identify nonvertex maximizers. $nloc1_{\max}$ is set to limit the number of local searches used to supplement the local vertex search. $\rho_{\min}^{\max}$ defines the maximum depth of the local vertex search. See Appendix A for further details.
2. If the local search methods are unsuccessful on any iteration, global search procedures are activated for use in all subsequent iterations. In *global* constraint maximization, vertex enumeration may be used (if the designer has selected this by setting *ivert* = 1). A multi-start random search may be used in which the random search is biased toward the vertices by specifying an *a priori* probability for each variable lying on a bound at the worst case (*pvert*). When an increased value for the maximum constraint is found by the random search, a new local search is initiated. The number of random points examined is limited to $nrand_{\max}$. The number of local searches allowed during the global phase of the search is limited to $nloc2_{\max}$. Local searches are initiated whenever a global search procedure identifies an increased value of $c_{\max}$ for which $\partial c_{\max}/\partial v$ indicates an ascent direction within V. The use of *pvert* was found to be essential for an effective random search as nonvertex maximizers usually have only

one or two values away from their bounds and these can often be identified by *a priori* physical argument.

3. Both local and global searches stop either when the computational effort employed exceeds the maximum used in previous iterations and a constraint violation has been identified or when the specified maximum effort has been employed unsuccessfully. This ensures monotonically increasing search effort as the outer approximation improves and the design converges.

This approach provides the flexibility to trade off confidence in final solution against the constraints of finite computational power, making extensive use of appropriate heuristics to improve efficiency of solution. The philosophy in applying the heuristics is to use cheaply obtainable constraint maximizers to push the design toward its final form before applying the more expensive procedures.

In carrying out local searches for

$$\max_{v \in V} \max_{k \in K} c_k(p, v) \tag{14}$$

it is desirable to find the solution in a single NLP problem while avoiding the nondifferentiability of the optimization problem above. A differentiable approximation to $\max_{k \in K} c_k(p, v)$ of the form

$$c_{\max} = \epsilon \left| \log\left(\sum_{k=1}^{N} \exp\left(\frac{c_k(p, v) - \epsilon}{c_{\text{sig}k}} \right) \right) \right| = 0 \tag{15}$$

can be used. This approach is preferred because

$$\max_{v \in V} c_{\max} \tag{16}$$

requires the solution of fewer NLP problems for a local search than the more obvious approach to avoid nondifferentiability

$$\max_{k \in K} \max_{v \in V} c_k(v, x) \tag{17}$$

When $\epsilon_k > 0$, the algorithm must decide whether to add the original model or the projected model to the outer approximation set V_k. If applying ϵ_p gives an increased value of $c_{\max}$, the projected model is added; otherwise, the unprojected model is added.

c. Initialization. Initialization includes defining the sets V and P. Set V is defined in terms of simple bounds on the design parameters; P may be defined by a mixture of simple bounds and more general constraints on p as this does not imply any increase in complexity. These constraints are simply added to the outer approximation constraints for each design iteration.

The initialization must also define the parameters controlling the maximum effort to be applied to the optimization. At one extreme, the search could be limited to a crude vertex search by excluding vertex enumeration (*ivert* = 0) and random searches ($nrand_{\max} = 0$) and setting $\rho_{\min}^{\max} \geq 1$ to limit the depth of the vertex search. At the other extreme, vertices could be exhaustively enumerated (*ivert* = 1) and extensive random searches carried out ($nrand_{\max} > 2^{n_v}$, where n_v is the number of uncertain parameters). The initial vertex exploration can be supplemented by up to $nloc1_{\max}$ local searches and the global searches can be supplemented by up to $nloc2_{\max}$ local searches. Setting $nloc1_{\max}$ and $nloc2_{\max}$ to about 10 effectively removes any constraint on the use of local searches.

pvert sets the probability of the worst-case parameter value being at an upper or lower bound. Unless there is a physical reason to expect a worst case that is *not* at an extreme, this should be set in the range .8–.999.

ϵ_p must be set to control the degree of precision used in solving the problem. Values of .05–.2 are typical.

Initialization must provide an initial outer approximation, V_0, on which to base the design. The selection of the *nmods* parameter set(s) making up V_0 may be based on prior judgment or analysis (first *nset* elements of V_0) or it may be based on a simplified constraint maximization algorithm (the remaining elements). Grossmann *et al.* (1983) use the gradients of the constraints at a *nominal point* to make a prediction of the set of vertices maximizing the individual constraints. An initial application of a local vertex search algorithm (see Appendix A) to each constraint seems more appropriate. Only distinct vertex maximizers corresponding to violated constraints are included in the initial set, as there is a high computational cost for including unnecessary points. If $\epsilon_p > 0$, then *either* the projected *or* the original point identified by the constraint maximization is included in V_0, depending on which gives the larger violation of the constraint. Including both points will usually give no benefit while doubling the computational effort in the first design iteration.

d. Multimodel Design. In the multimodel design stage, it is desirable to carry out the design with the minimal set of models that define the optimum while not dropping and reidentifying models that do affect the solution. The approach adopted is to predict the "active set" of models, design with this active set, and check the solution against all models identified; then if any additional model indicates constraint violations, add it to the predicted active set and rerun the optimization. The active set used is formed from all models that have been active for at least one of the previous three design iterations or have previously had to be added back at the end of a design optimization. Although quite heuristic, this strategy appears reasonable.

The use of the "ϵ_p" heuristic can potentially slow convergence if the unprojected constraint maximizers included in the multimodel design become active again as the design progresses. This behavior would not be expected in typical problems—nor has it actually been observed—but it is guarded against by checking the unprojected models at the same time as other inactive models and adding them to the design set if necessary.

e. Some Practical Considerations. It is important to emphasize that, despite the measures introduced to make the worst-case design algorithm more efficient, it is not practical to deploy the global optimization capabilities using more than about 10 uncertain parameters for dynamic problems or about 15 for steady-state problems. This means that the engineer must exercise judgment as to the key parameters to be given the most rigorous treatment. Engineering judgment may identify some variables that are expected to make all performance measures worst at one extreme of their range or groups of parameters that have an essentially equivalent effect. This may allow elimination of many uncertain parameters on a fairly rigorous basis. Other parameters whose influence is unclear but judged to be minor may be frozen at nominal values if necessary to get a problem of reasonable size. Uncertain cost parameters, which appear only in the objective and not in the model equations or constraints, may be conveniently dealt with by evaluating the expected objective over these parameters and eliminating them from the optimization level.

B. Screening Tools for Disturbance Rejection

It is often important to get a rough solution to a design problem quickly. This may be critical in the early stages of a project, when many alternatives are being considered and when the problem definition is still changing rapidly. Under these circumstances, detailed modeling and analysis may be too slow and too costly to be effective, but key design decisions still have to be made. Screening methods are required which allow rapid assessment of alternative designs. This section reviews some screening methods proposed in the literature and presents an improved method for analyzing the effect of delays on achievable disturbance rejection.

Within an integrated design procedure, steady-state modeling and optimization may be used as a screening tool within the overall process of developing a design that can meet the performance requirements at all times and over all uncertain parameters. In the second Shell Process Control Workshop (Prett *et al.*, 1990), Campo *et al.* use steady-state design with uncertainty as a screening tool for linear systems. They use the method of Grossmann *et al.* discussed

above. Our modifications of this method are used for this purpose in the design procedure presented below.

Here, we focus on screening tools, addressing dynamic performance in response to disturbances and uncertainty.

1. Review

Methods relevant to assessing the effect of dynamics may be divided into two groups: controllability indicators and design validation tests. Controllability indicators may be used to direct designers away from options that are likely to be difficult to control but have an inherent ambiguity when weighed against economic considerations. Design validation tests should determine whether any design within a group of possible designs might achieve the required performance, without actually having to select a working design. If such tests can be carried out efficiently, they are particularly useful as they are unambiguous.

a. Controllability Indicators. Controllability indicators provide a *qualitative* indication of how difficult a plant will be to control; they cannot, however, easily be related to the economics and operating constraints for a particular problem. This means that the controllability indicators cannot be weighed on the same scales as the process economics. Within this framework, a choice between a plant with good steady-state economics but poor controllability indicators and one with inferior steady-state economics but better controllability indicators cannot be made without a *judgment* of how much the better controllability indicators are worth. In addition, there are multiple controllability indicators which are likely to give contradictory messages for any given plant. Finally, controllability indicators are usually based on linear models of the system to be controlled, leaving the effect of the nonlinearity of the true plant and process unexamined. Controllability indicators can play a useful role in highlighting when difficulties might arise, but they cannot be used to rigorously eliminate design options. Their main use in this context is to provide the designer with information that may be of assistance in finding a base-case solution quickly. A review of controllability indicators is presented in Morari (1992).

b. Design Validation. The design validation problem is to determine whether performance requirements can be met without actually carrying out a full design.

The controller validation problem of determining whether any controller exists which will allow a particular plant design to meet its performance requirements under process disturbances and uncertainty has been given particular attention. The *Fundamental Process Control* methodology (Prett and Garcia, 1988) addresses this validation problem as an integral part of the approach to

control system design advocated. Their approach is of general interest as it represents an attempt in the control area to deal with uncertainty and constraints systematically. It is emphasized that a problem specification for a control system design, using their approach, should contain quantitatively defined objectives and constraints together with a process model and an uncertainty description. The design problem is divided into two stages—validation and analysis. The validation stage attempts (1) to determine whether performance requirements can be satisfied irrespective of controller structure and type, and (2) to allow the performance indices to be refined if necessary before moving on to the analysis of the worst-case performance of particular control system designs. Such an objective is clearly desirable, so it is important to consider both some specific problems with their formulation and some general problems in separating validation from analysis.

Prett and Garcia (1988) pose the validation problem as a discrete time linear optimal control problem under uncertainty. The uncertainty is defined by simple bounds, giving a polyhedral set of uncertain parameters V. For this problem, certain forms of uncertainty, e.g., in gains only, together with a quadratic performance index can be shown to satisfy the convexity requirements for the worst-case parameters to lie at vertices of V. This allows the algorithm of Grossmann *et al.*, based on examination only of vertices of V, to be applied (see Section II.A.1). The mathematical formulation is

$$\min_{\Delta u(k)} \max_{v \in V} \{\|y(k+1) - y_s(k+1)\|^2_{W_y^T W_y} + \|\Delta u(k)\|^2_{W_u^T W_u}\} \tag{18}$$

subject to bounds on the outputs y, the control moves $\Delta u(k)$, and the control values $u(k)$. Here, y_s is a vector of target values for the outputs, and W_y and W_u are weighting matrices for deviations from target values and control moves, respectively. This formulation implies that the controller has no knowledge of the uncertain parameters, i.e., that the controller operates completely blindly. This assumption can cause this validation problem to give very pessimistic results.

Consider, for example, a typical stable system, $y(s) = (k1e^{-k2s}/(1 + k3s))\, u(s)$, which is known exactly except for an uncertainty in $k1$. It is required that the control bring the system from a given initial condition to target steady state at some arbitrary time in the future in the absence of disturbances. Each possible input sequence, $u(k)$, achieves this for one and only one value of $k1$. The validation problem formulation above will therefore indicate that no controller exists satisfying this problem specification. In fact, any PI controller giving a stable closed-loop system will achieve the performance specified.

The problem of performing the validation irrespective of the control system is confirmed by Garcia in the *Second Shell Process Control Workshop* (Prett *et al.*, 1990):

> In a nutshell, since the optimization problem . . . does not consider any controller explicitly, it searches for a *fixed* sequence of moves into the future that meets all performance criteria for all plants in the uncertainty description. This can be shown to be impossible for trivial cases. . . . Our current thinking is leaning towards solving the *design analysis* problem . . . for the most comprehensive controller that the designer can implement (e.g. DMC) and use this controller for validating design decisions.

The analysis problem for DMC itself is as yet unsolved. Garcia also notes the limitations of the use of linear models:

> Most likely, if a controllability problem arises it is probably due to nonlinearities. In such cases a linear process representation may not be sufficient to allow the solution of the problem requiring the use of a more detailed or sophisticated model. Therefore, the procedure should include the option of finding a nonlinear description and solving the nonlinear control problem under uncertainties.

Leaving aside the emphasis on DMC and the limitation of the design to control, this is consistent with the design approach used here (see Section II.A).

A possible method of salvaging the validation problem is to solve *both* a minmax (Eq. 18) and a maxmin optimal control problem:

$$\max_{v \in V} \min_{\Delta u(k)} \{\|y(k+1) - y_s(k+1)\|^2_{W_y^T W_y} + \|\Delta u(k)\|^2_{W_u^T W_u}\} \tag{19}$$

If the maxmin analysis indicates the problem to be infeasible, it is indeed infeasible as there exists some v for which no input sequence can be found to satisfy the performance specification. If the minmax problem indicates feasibility, there is some *fixed* input sequence which satisfies the constraints for all v.

For an optimal control problem, the complementary maxmin problem is equivalent to the controller having perfect knowledge of the plant parameters and disturbances, including *future* disturbances. This formulation could be termed the "crystal ball" approach to control. For many problems, the crystal ball control will be successful, but this says very little about whether any *realizable* control system exists that can meet the performance specification.

The most likely outcome of solving the complementary maxmin/minmax problems is that the maxmin problem is feasible (for each v there exists an input sequence that can satisfy the constraints) and the minmax is infeasible (there is no single input sequence that can satisfy the constraints for all v). This is unfortunate, as this outcome is the least informative as to whether a realizable controller exists.

For example, optimizing control moves for disturbance rejection in a blending system subject to step disturbances of variable magnitude would give perfect disturbance rejection for a maxmin/perfect-knowledge formulation and a poor

(possibly infeasible) disturbance rejection for a minmax/no-knowledge formulation. This arises as disturbances can be canceled at the system inlet by adjusting flows to maintain the correct blend; thus the blending system is essentially a mechanism for compensating for model uncertainty, particularly uncertainty regarding the disturbances. The merit of a particular blending system lies in how effectively it compensates for lack of knowledge and facilitates extraction of knowledge. The mathematically convenient extremes of no knowledge and perfect knowledge do not provide good bounds on performance for this type of problem. This means that useful results are unlikely unless the mechanism for knowledge extraction is embedded in the optimization by implementing a control system in the model used to evaluate performance. Unfortunately, this destroys the desirable separation between design validation and design analysis.

An additional point with regard to the use of combined maxmin/minmax optimal control formulations is that they give rise to problems of comparable or greater difficulty to carrying out a worst-case design (Section II.A) with a specified control system. Given that the bounds they produce for performance are generally loose as discussed above, they do not seem to be an efficient use of resources in a design procedure.

2. *A Screening Test for Disturbance Rejection in Nonlinear Processes Subject to Time Delays*

Time delays are a feature of many process systems. Where present, they imply a fundamental limitation on achievable performance that no feedback controller, however sophisticated, may overcome. In this section, a screening test for analyzing achievable disturbance rejection in dynamic systems involving time delays is presented. The test is based on calculating the minimum time before a feedback control system can begin to counteract a disturbance and testing whether the open-loop response violates a constraint before this time has elapsed. Passing this test is a necessary condition for the existence of a controller that can meet the specified disturbance rejection requirements. A variation of this test can be used to calculate the fraction of the disturbances that can be rejected. This provides a measure of controllability that can be used to judge whether an implementable control scheme is likely to be successful. The method includes consideration of process and disturbance dynamics and can be applied to uncertain multivariable nonlinear systems.

Previous process controllability work on analyzing the effect of delays on achievable control performance has centered on defining and computing the effective delay(s) in multivariable systems. This work has generated some useful controllability measures, which are reviewed below, but does not in itself provide an answer to the key question of whether a particular disturbance can

potentially be rejected by control action so as to avoid constraint violations. This section presents a method for extending the earlier controllability measures to answer this question directly and unambiguously.

a. Review of Previous Work. Previous work on controllability analysis of systems with delays has centered on calculating the effective delay associated with multivariable systems. Holt and Morari (1985) present several measures of the effective delay. The first measure is the set of minimum delays from any input to each output. This provides an obvious lower bound on the time for control action to reach each output. In this paper they note that this bound may not always be achievable with a stable causal controller. The minimum delays to make a change in each output without disturbing any other output are computed to provide an achievable upper bound on the minimum delay associated with each output. Perkins and Wong (1985) provide an alternative approach based on functional controllability analysis, computing the minimum delay before an arbitrary trajectory can be imposed on each output independently. This corresponds to the maximum of the upper bound delays computed by Holt and Morari and has the virtue of providing a single measure of the effective delay. Holt and Morari demonstrate that increasing delays in the process dynamics can improve the achievable *decoupled* response subject to delays.

These measures of delay all provide useful indicators of the effective delay, but they do not in themselves indicate whether the effect of the delay prevents a disturbance from being rejected before causing constraint violations. In Holt and Morari's analysis, the disturbances are assumed to appear as steps on the outputs, making the question of whether the disturbance causes constraint violation trivial. In practice, disturbances are often well approximated by steps, but the effect of the step usually propagates dynamically through part of the process before affecting the outputs. This means that the effect of the process dynamics in attenuating the disturbance should be included in order to assess disturbance rejection.

It should also be noted that the analyses of delay times discussed above do not distinguish between feedback and feedforward control. In practice, disturbances may often be measured with a smaller delay than the delay between the disturbance and any of the constrained outputs. This allows feedforward control to make a contribution to achievable control performance which should be considered explicitly.

A common performance estimation method in "classical" single-loop feedback controller design is to check the open-loop disturbance rejection of a system up to the point in time at which the controller action is assumed to take effect. If constraints are violated during this time, the controller cannot prevent the violation. The use of this test can be traced back to Velguth and Anderson (1954).

They consider a single disturbance, a fixed linear process model (series lags), and a heuristic estimate of the controller response time based on the sum of delays and minor lags in the control loop. This estimate of the delay is based on trying to approximate the peak in the response of a proportional-integral-derivative (PID) controller, and a number of variants (including computing the response time by using the Ziegler–Nichols rules for computing effective delay) have appeared, e.g., the work of McMillan (1983). This analysis directly addresses achievable disturbance rejection, but does so in a heuristic manner which does not provide a rigorous bound on performance, even for the special case of PID controllers.

The analysis presented below draws on both the areas of previous work discussed above to develop general and direct methods of determining whether the effect of delays prevents control action from meeting constraints.

b. Analysis of Delay Effects on Disturbance Rejection. The basic technique in the methods discussed below is to compute the open-loop response to the disturbance over the interval from disturbance onset to the minimum delay for effective controller response elapsing. Any constraint violations up to this time indicate that even "ideal delay-limited control" is not adequate. Success with ideal delay-limited control is a necessary, but not sufficient condition for success with an implementable feedback controller such as PI (see Fig. 2).

The response of the ideal delay-limited controller is not specified, but is assumed to be able to prevent constraint violations after the minimum delay has elapsed.

This test is particularly useful for disturbances such as steps and pulses because their full impact is felt immediately. However, the test is not restricted to these disturbances and may also be used to consider the initial effect of sinusoids, ramps, or other disturbances.

It should be noted that if the open-loop response is evaluated by numerical integration, there is no requirement for linearity of the process model. Convenient analytical solutions can be found for some linear systems, such as series stirred tanks with instantaneous reaction and constant flow.

The discussion below covers

1. the definition of the minimum delay
2. the formulation of feasibility tests for disturbance rejection
3. the formulation and interpretation of a controllability measure
4. the simplified analysis for series stirred tanks.

c. Definition of the Minimum Delay. For a single-input/single-output system with a single disturbance entry point, the time between a disturbance occurring and the controller response *beginning* to take effect, t_d, is made up of the time

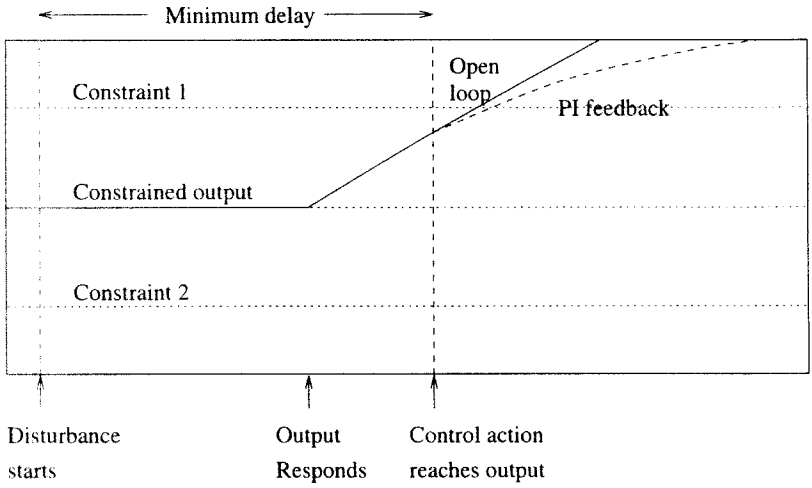


FIG. 2. Ideal and PI control.

delay between the disturbance and the measured value ($t_{d_{d,y}}$) and the delay between the manipulated variable and the measured value ($t_{d_{u,y}}$):

$$t_d = t_{d_{d,y}} + t_{d_{u,y}} \tag{20}$$

The main complication for multivariable systems with a single disturbance entry point or multiple independent disturbance entry points is generalizing t_d for each constrained output. For a general multivariable controller with inputs u_i and outputs y_j the generalization, for output $y_{\bar{j}}$, is

$$t_{d\bar{j}} = \min_j \{t_{d_{d,y_j}}\} + \min_i \{t_{d_{u,y_{\bar{j}}}}\} \tag{21}$$

This expression for the minimum delay from control action to the constrained output $y_{\bar{j}}$, $\min_i\{t_{d_{u,y_{\bar{j}}}}\}$, corresponds to Holt and Morari's lower bound on the minimum delay. Use of their upper-bound values is not appropriate in general, as they are based on a requirement for decoupling which may not be necessary or even desirable for disturbance rejection. If decoupling is assumed to be an additional performance requirement, the upper-bound delays could be substituted for those used. It should be noted that, in the absence of a decoupling requirement, t_d and hence the predicted disturbance rejection, can never be improved by increasing any process delay.

Situations will arise in which paths with minimal delay are technically present but are not realistically useful. For example, pressure variations in a distillation

column may accompany disturbances affecting composition, but do not readily provide a means of countering such a disturbance. In general, only certain manipulated variables, $u_i|i \in I_j^s$, will have a strong enough effect on y_j to be able to counter the disturbance. Similarly, the disturbance can usefully only be detected by using a subset of the measured variables, $y_j|j \in J_d$. These restrictions combine to eliminate weak connections which would form no practical control feedback paths and prevent unrealistic solutions. In assessing the effect of control structure, it is necessary to impose a particular control structure in which only certain inputs, $u_i|i \in I_j^c$, are connected to certain measurements. Combining these restrictions gives

$$t_{d_j^-} = \min_{j \in J_d} \{t_{d_{d,y_j}} + \min_{i \in I_j^s \cap I_j^c} \{t_{d_{u_i,y_j}}\}\} \tag{22}$$

The difference between feedforward and feedback control can be conveniently represented by using an additional *unconstrained* measurement with the appropriate delay in relation to the disturbance to represent the feedforward measurement. The effective delay will usually be less for feedforward than for feedback control, giving an improved bound on disturbance rejection.

d. Feasibility Tests. For a nominal process model, the test for feasibility of disturbance rejection is

$$\exists p \in P \qquad \text{s.t.} \quad \max_{k \in K} \max_{t \in [0, t_{d_k}]} \{c_k(p, t)\} \leq 0 \tag{23}$$

where t_{d_k} is the minimum delay associated with the kth constraint c_k, and p is a vector of design parameters lying in the set P.

In general, there will be uncertain or variable properties to consider and it will be required that the constraints be satisfied for all combinations of the corresponding uncertain parameters. This uncertainty is assumed to be parameterized by a parameter vector v, the values of which lie within a polyhedron V. v may include both parameters defining the possible disturbances and model parameters affecting the initial open-loop response, such as biases on measurements used for control.

Assuming that the effect of control in steady state is defined by a set of controller setpoints and fixed values for a subset of uncontrolled inputs, it is appropriate to use a worst-case design formulation in which a single choice of p is required to accommodate all possible values of the uncertain parameters. In other words, the only steady-state adaptation to uncertainty is that implicit in the choice of controller setpoints. This is potentially conservative; however, it is a fair approximation to industrial practice, where minimizing the need for

operator intervention and control system adjustment is often desirable in normal operation. The feasibility test then becomes

$$\exists p \in P \qquad \text{s.t.} \max_{v \in V} \max_{k \in K} \max_{t \in [0, t_{d_k}]} \{c_k(p, v, t)\} \leq 0 \tag{24}$$

$$V = \{v \mid -\Delta v^- \leq v \leq \Delta v^+\}$$

A more general formulation allows certain variables (o), known as operating variables, to be adjusted in response to the variation in the uncertain parameters. As discussed in Section II.A, this formulation is more difficult to analyze and will often give inappropriately optimistic results. The formulation with operating variables is, however, given below for completeness.

$$\exists p \in P \qquad \text{s.t.} \max_{v \in V} \max_{o \in O} \max_{k \in K} \max_{t \in [0, t_{d_k}]} \{c_k(p, v, o, t)\} \leq 0 \tag{25}$$

$$V = \{v \mid -\Delta v^- \leq v \leq \Delta v^+\}$$

If desired, an objective function, min $J(p,V)$, could be combined with any of the above constraint formulations to obtain the optimal feasible value of p. For example, p might include capacities of equipment and setpoints for the outputs, y, and the objective $J(p,V)$ might be annualized cost. The optimized cost would then give a rough estimate of the effect of process disturbances and dynamics (and uncertainty) on the optimal process cost.

Optimization problems based on ideal disturbance rejection with delays, as above, are much easier to solve than the corresponding problem involving design of a specific controller for the following reasons:

1. The timespan to be simulated is much less than would be required for controller design.
2. A controller need not be included in the model, thus reducing model complexity and the number of design variables.
3. The number of uncertain parameters required to model uncertainty is reduced as some parameters, such as those associated with measurement dynamics, do not affect the open-loop response.

These problems have the additional advantage that they provide a result that bounds the performance of a set of controllers rather than just a particular controller considered in a full design. The "ideal delay-limited control" analysis has been found in the case studies (Section V) to provide a useful bound on the performance achieved with implementable controllers.

In some problems, the delays will themselves be functions of the parameters p and v. This introduces a potential problem for the optimization in that, even

if each delay is a smooth function of the parameters, the path determining the minimum delay may change, introducing a discontinuity. This *might* require the use of nonsmooth optimization methods. As noted by Gill *et al.* (1981), smooth algorithms may often be successful on problems such as those with occasional discontinuities and should be tried first.

The optimization problems above may be used to provide a rigorous bound on the achievable performance if delays are calculated as discussed in Section II.B.2. If it is desired to obtain a heuristic estimate of performance rather than a rigorous bound, the control response delays, $t_{d_{u_i,y_j}}$, within control loops could be replaced by heuristic values such as $\pi/2\omega_n$, where ω_n is the frequency (rad/s) at which the phase shift of the dynamic response between the input and output is π radians. This heuristic corresponds to making $t_d = t_u/4$, where t_u is the natural (ultimate) period of the loop (i.e., the period of oscillation of an open-loop stable system when the feedback gain is just sufficient to sustain an oscillation). The heuristic delay, $t_u/4$, is nearly equivalent to the rigorous method for process dynamics consisting of a small pure delay and a large first-order lag. An example presented in Section V.A.3 confirms this heuristic to be effective in considering PI control of pH in a stirred tank.

e. Controllability Measure. The feasibility test is useful in itself, but a positive or negative result raises additional questions.

1. If the test indicates that control performance may be adequate for several different processes/control schemes, which of these should be pursued in detail first?
2. If the test indicates that feedback control would fail but feedforward control might succeed, what can be said about the likely success of a combined scheme?

One way of addressing these questions is to determine what fraction of the disturbance, δ_f, can be handled by particular schemes assuming ideal delay-limited control. The likelihood of success for alternative schemes can then be estimated from the value of this fraction. Combined feedforward/feedback control is unlikely to be successful unless the relative feedforward error due to measurement and modeling errors is less than the calculated δ_f for feedback control. That is, if δ_f without feedforward control is $x\%$, the feedforward must be accurate to within $\pm x\%$ of value for the combined control to have a chance of success. In considering otherwise comparable design alternatives, the alternative with the larger δ_f should be pursued first. The disturbance fraction therefore represents a useful controllability measure to guide further design effort.

To calculate the disturbance fraction, it is only necessary to include it in the model and apply the optimization formulations discussed above using $J(p, V) = -\delta_f$ as an objective:

$$\min_{p \in P, \delta_f} -\delta_f \quad \text{s.t.} \quad \max_{v \in V} \max_{k \in K} \max_{t \in [0, t_{d_k}]} \{c_k(p, v, t, \delta_f)\} \leq 0$$
$$v = \{v \mid -\Delta v^- \leq v \leq \Delta v^+\} \tag{26}$$

f. Simplified Analysis for Series CSTRs. Although general problems require optimization of a nonlinear dynamic model as discussed above, the analysis can be greatly simplified for some special cases. The case of particular interest for the problems considered later is that of continuous-flow stirred-tank reactors (CSTRs) in series. In this case, it is desired to add reagent so as to keep variations in the net concentration of effluent and reagent, *cnet*, at the exit of the last tank below a certain level, δ_{cnet}, in the face of step disturbances in the inlet concentration of magnitude Δ_{cnet}. This objective can be expressed as a required disturbance attenuation, δ_c, where

$$\delta_c = \frac{\delta_{cnet}}{\Delta_{cnet}} \tag{27}$$

If the reagent flow is much less than the effluent flow and the disturbance is symmetric (equally severe with respect to the upper and lower constraints), the equation

$$\delta_c = \frac{Fr_h - Fr_l}{2\ \Delta Fr} \tag{28}$$

can be used, where ΔFr is the change in reagent flow required to cancel the disturbance and Fr_h and Fr_l are the maximum and minimum reagent flows to remain within the exit limits.

It can be shown that the unit step response of n tanks in series with mixing lag $t_{c_{\text{mix}}}$ and no delays is given by

$$y(t) = 1 - e^{-t/t_{c_{\text{mix}}}} \sum_{i=0}^{n-1} (t/t_{c_{\text{mix}}})^i / i! \tag{29}$$

Given the minimum delay t'_d in the control response relative to the disturbance response at the treatment system exit ($t_d = t_d - t_{d_d,y}$, where y is the exit measurement), the best possible disturbance attenuation, δ_a, is given by

$$\delta_a = 1 - e^{-t'_d/t_{c_{\text{mix}}}} \sum_{i=0}^{n-1} (t'_d/t_{c_{\text{mix}}})^i / i! \approx \frac{1}{n!}\left(\frac{t'_d}{t_{c_{\text{mix}}}}\right)^n \tag{30}$$

Tanks of different sizes may be analysed similarly, giving

$$\delta_a = 1 - \sum_{i=1}^{n} \frac{e^{-t'_d/t_{c\mathrm{mix}}(i)} t_{c\mathrm{mix}}^{n-1}(i)}{\prod_{j \neq i} (t_{c\mathrm{mix}}(i) - t_{c\mathrm{mix}}(j))} \approx \frac{1}{n!} \left(\frac{t'^n_d}{\prod_{i=1,n} t_{c\mathrm{mix}}(i)} \right) \tag{31}$$

If the reagent added to each tank is based on concentration measured at the tank exit without any measurement delay, then t'_d will simply be the minimum value of $t_{d\mathrm{mix}}$, where $t_{d\mathrm{mix}}$ is the delay between reagent addition and measurement response due to imperfect mixing.

To take account of the effect of uncompensated minor time lags on the performance of PI controllers, replace t'_d in the above formulas by the minimum value of $t_u/4$, where t_u is the natural period of a control loop around a single tank. Using this effective t'_d makes the calculated performance bound heuristic rather than rigorous. A good default is 10 seconds for t'_d when estimating the best practicable performance, while the heuristic method will give typical t'_d values of about 30 seconds.

Similar expressions can be derived for responses to ramp disturbances. If the disturbance is a decaying exponential, then this is equivalent to an extra lag on a step disturbance, and the response can be evaluated by using the formulas for non–equal–sized tanks. The improvement in achievable attenuation from replacing a step by an exponential disturbance of time constant τ' or a ramp reaching a new steady value after τ' is approximately $t'_d/(n + 1)\tau'$.

C. Optimization of Dynamic Systems

Dynamic optimization is an important subproblem underlying both dynamic worst-case design (Section II.A) and the general version of the disturbance rejection test (Section II.B.2). Here, we address methods for solving such problems and discuss the approach we have taken in our work.

The optimization problem of interest is

$$\begin{aligned} &\min_{u(t)} J(x(t_f), z(t_f), u(t)) \\ \text{s.t.}\quad & f(\dot{x}, x, z, u(t), t) = 0 \\ & g(x, z, u(t), t) = 0 \\ & q(x, z, u(t)) \leq 0 \qquad 0 \leq t \leq t_f \\ & L(x(\phi), z(\phi), u(t), \phi) \leq 0 \\ & M(x(\phi), z(\phi), u(t), \phi) = 0 \qquad \forall \phi \in \Phi \\ & \Phi \subset t \end{aligned} \tag{32}$$

where J is the objective of the optimization, $u(t)$ is a vector of optimization parameters that may vary over time, f and g define a model that will in general constitute a mixed system of differential and algebraic equations (a DAE system) with differential variables x and algebraic variables z, q defines path constraints that must be satisfied at all times, and L and M define interior point constraints that must be satisfied for a discrete set of times Φ.

It should be noted that $u(t)$ in formulation (32) can be replaced by time-invariant parameters θ by parameterizing the variation with time variation explicitly, e.g., by using a piecewise linear function of time. This allows standard finite-dimensional optimization methods to be applied and is the approach adopted in our work.

Dynamic optimization is a rapidly maturing field which we will not review in detail. Approaches can be distinguished by

- whether the dynamic equations are solved by conventional simulation techniques (feasible path) or as part of the optimization using approximation techniques to transform the problem to algebraic constraints (infeasible path)
- how constraints on values of variables over time (state or path constraints) are dealt with (see below)
- how gradients are calculated, e.g., adjoint equations or sensitivity equations
- the optimization methods applied, e.g., SQP, iterative dynamic programming.

A general review is given by Biegler (1990).

Our judgment is that feasible path methods in which the solution of the model equations over time is carried out by conventional integration software, which has been extensively developed and refined, are at present more reliable than infeasible path methods. Feasible path optimization methods are also easier to implement as the size of the optimization problem is much smaller. For these reasons, we have pursued feasible path methods despite evidence that infeasible path methods are more efficient on some problems.

We have used sensitivity equation methods (Leis and Kramer, 1985) for gradient evaluation as these are simple and efficient for problems with few parameters and constraints. In general, the balance in efficiency between sensitivity and adjoint methods depends on the type of problem being addressed. Adjoint methods are particularly advantageous for optimal control problems in which the inputs are represented as a large number of piecewise constant input values and few interior point constraints exist. Sensitivity methods are preferable for problems with few parameters and many constraints.

We have chosen to use an SQP method (Chen, 1988) as the basis for local solution of the dynamic optimization problem. This choice is based on the po-

tentially rapid convergence of this method and its computational efficiency on small problems.

We present below some easily implementable methods for improving the robustness and efficiency of feasible path dynamic optimization codes which have proved useful in our work. Here, we cover methods for preventing simulation error from disrupting optimization, representation of path constraints, and handling poor local approximations during the optimization.

1. Noise Control

All numerical integration methods solve problems only approximately and will usually provide some error control tolerances to adjust the accuracy of the solution. Typically, these include an absolute and relative tolerance on the local error during integration (*atol* and *rtol*), an event tolerance for the location of discontinuities (*evtol*), and a tolerance for the precision to which the equations must be satisfied on initialization (*sstol*). Generally, *rtol* and *atol* are combined to give an error control weighting, $w_i = rtol|y_i| + atol$, and are often given the same value. For feasible path dynamic optimization, these tolerances must all be chosen so that they meet the following criteria:

1. The numerical solution of the equations does not in itself introduce a significant modeling error.
2. Noise or precision error does not seriously disrupt progress of the optimization, thus causing slow convergence or failure.
3. The computation time used does not substantially exceed that required to satisfy these requirements.

The first requirement is generally easily met as the error in the equation solution typically becomes small compared to overall modeling error for moderate values of the error control tolerances. The tradeoff between the second and third requirements is more difficult and depends on the particular numerical characteristics of the system equations and the particular values of the optimization parameters. It is desirable to have some means of estimating and adjusting the precision error to optimize this tradeoff. This requires that the precision error be estimated, its effect on the optimization assessed, and the integrator tolerances adjusted appropriately.

a. Estimating Precision Error. When the sensitivity method for evaluating gradients is used (Dunker, 1984; Leis and Kramer, 1985), the system equations are reintegrated during the gradient evaluation following a successful line search by the optimizer. The variable trajectories $(x^s(t), z^s(t))$ generated during the gradient evaluation are not identical to the trajectories generated previously $(x(t), z(t))$. The difference is due to the jacobian of the system equations, which is used in

the iterative solution of the equations being updated more frequently when calculating gradients than when just integrating the model equations.

The precision error in the variables can be estimated by using the formula

$$\epsilon(y) = \frac{\max_i |\Delta^k y_i|}{\sqrt{\frac{(2k)!}{(k!)^2}}} \tag{33}$$

(Gill *et al.*, 1981) based on successive differencing of variable values generated using closely and uniformly spaced parameter values. Numerical trials indicate that

$$\delta(y) = |y^s - y| \tag{34}$$

is comparable to $\epsilon(y)$. The values often agree to within about 20%, with $\delta(y)$ usually being greater than $\epsilon(y)$. As $\epsilon(y)$ requires about six extra integrations to generate, there is a clear case for using $\delta(y)$, which is available without additional system integrations, as an estimate of the precision error.

This estimate strictly addresses only precision error in the variable trajectories, and not errors in the gradients. However, as noted by Dunker (1984), the gradients given by the sensitivity method are exact gradients of the computed solution trajectory regardless of truncation errors. This suggests that controlling the accuracy of the integration of the DAE system will generally be sufficient to give accurate gradients. Numerical experience indicates that $\epsilon\left(\frac{\partial y}{\partial \theta_i}\right) \Big/ \left|\frac{\partial y}{\partial \theta_i}\right|$ is indeed comparable to $\epsilon(y)/|y|$. This experience is all based on sensibly scaled problems with values of y and $\partial_y/\partial\theta_i$ within a few orders of magnitude of 1, but does support the suggestion that control of the variable accuracy is sufficient to control the gradient accuracy.

If the sensitivity method is not being used to generate the gradients, the precision can be computed using extra integrations to generate $\epsilon(y)$. This estimate could be mapped straightforwardly to precision error in gradients calculated by numerical differencing. Estimating errors in the gradients calculated using adjoints would require additional integrations of the adjoint equations to generate $\epsilon(\partial y/\partial\theta_i)$.

2. *Evaluating Effect on Optimizer*

The estimated error in objective and constraint values may be mapped through calculations in the optimizer to determine whether the noise is likely to impede progress. The appropriate measure will depend to some extent on the optimization code used. Noise in the objective and constraints may interfere

with the evaluation of the optimizer termination conditions or with comparisons between performance with different values of the optimization variables, e.g., during a line search. Noise in the gradients may result in selection of a search direction that is not a descent direction, particularly if the descent direction is chosen using a second-order algorithm with an ill-conditioned Hessian estimate. The Hessian estimate may itself be degraded by noise in the gradients.

A typical SQP termination condition for a constrained optimization problem

$$\begin{aligned} &\min_{\theta} J(\theta) \\ &\text{subject to} \quad c_i(\theta) = 0, \qquad i = 1, 2, \ldots, meq \\ &\qquad\qquad\; c_i(\theta) \le 0, \qquad i = meq + 1, \ldots, m \end{aligned} \tag{35}$$

is

$$\max\left[\sum_{i=1}^{meq} |c_i| + \sum_{i=meq+1}^{m} \max(0, c_i), \left(|\nabla_x J^T \delta_k| + \sum_{i=1}^{meq} |\lambda_i c_i| + \sum_{i=meq+1}^{m} \lambda_i \max(0, c_i)\right)/(1 + |J|)\right] \le optacc \tag{36}$$

where λ is the vector of estimated Lagrange multipliers, δ_k is the step estimated from the quadratic program on the kth iteration, and *optacc* is the specified optimization accuracy.

Comparisons between points are usually based on some form of merit function, P_m, e.g.,

$$P_m(\theta) = J(\theta) + \sum_{i=1}^{meq} \mu_i |c_i(\theta)| + \sum_{i=meq+1}^{m} \mu_i \max(0, c_i(\theta)) \tag{37}$$

where the penalty parameters μ_i are closely related to the $|\lambda_i|$.

Controlling the combined precision error

$$\begin{aligned} error = \frac{\delta(J)}{1 + |J|} &+ \sum_{i=1}^{meq} \max\left(1, \left|\frac{\lambda_i}{1 + |J|}\right|\right) \delta(|c_i|) \\ &+ \sum_{i=meq+1}^{m} \max\left(1, \left|\frac{\lambda_i}{1 + |J|}\right|\right) \delta(\max(0, c_i)) \end{aligned} \tag{38}$$

approximately controls the noise in the termination condition and merit function. If *error* is kept below *optacc*, noise effects on the termination condition and on the comparisons of values of the merit function are unlikely to disrupt convergence.

Evaluation of the effect of gradient errors would be much more difficult but, as discussed above, has not been found necessary.

3. *Adaptation of Integrator Error Tolerances*

The value of *error* can be used for feedback control of the integrator tolerances. This feedback is complicated by the fact that in the algorithm used for integration there are four tolerances that can be adjusted to control accuracy. Experimentation indicated that the computational cost of the integration showed little dependence on the initialization tolerance, *sstol*, or on the accuracy to which discontinuities were located in time, *evtol*. These tolerances could, however, have a substantial effect on noise levels. It was therefore decided to set these tolerances to very stringent levels, 10^{-10} being chosen as a default. As noted previously, the remaining tolerances *atol* and *rtol* that govern the acceptance of an integrator step are generally equated. Adopting this practice gives a single parameter *simacc* to adjust, with *rtol* = *atol* = *simacc*.

The mean value of *error* may be expected to increase monotonically as *simacc* increases while the computation time may be expected to decrease monotonically as *simacc* increases. The control objective is therefore to maintain a value of *error* that is *just small enough* to prevent serious noise problems. The precise relationship between *simacc* and *error* is unclear and will vary from system to system and integrator to integrator. As in any feedback scheme with a poorly defined system model, the feedback system must be suitably conservative to avoid instability. The scheme below has been found to be effective in extensive trials.

if (*error* > 2.*optacc/factor*) *simacc* = *simacc*/10

if (*error* < *optacc/factor*) *simacc* = *simacc*.min(1/.5, (*optacc/error.factor*)$^{1/5}$)

where *factor* defines the target margin of safety from levels of noise comparable to the termination threshold *optacc*. A value of 5 was found to be adequate as further increases in *factor* did not reduce the number of iterations required for convergence. The factor of 2 in the above algorithm provides a deadband so as to prevent the adaptation mechanism from responding to insignificant changes in the measured error. If the error seems too large, *simacc* is tightened vigorously to eliminate the high noise promptly. If the response to noise levels below the target is similarly vigorous, the error control is liable to oscillate continuously, giving repeated spikes of excessive noise. Hence a more cautious adaptation to low noise levels is used. The precise detail of the adaptation is not very important. Any scheme incorporating vigorous response to high noise and sufficiently cautious response to low noise so that oscillation does not occur would give essentially the same results. To prevent the error control correction from disrupting line search convergence, the function and constraint values should be recalculated whenever the integration accuracy is tightened.

This method has been found to eliminate the need for application-specific consideration of integrator error control while requiring no significant extra computation and obviating the use of very tight error control on the integrator. The number of iterations required for optimization was found not to differ significantly from that required with fixed small values of *simacc*, while the time for integration was reduced substantially. Failures attributable to noise were eliminated.

The main limitation of the method is that gradient errors are not directly estimated, analyzed, or controlled. If problems occur for this reason, the parameter *factor* could be increased, or fixed values of the integrator tolerances could be chosen by trial and error.

4. *Path Constraint Representation*

Problems may also arise regarding the representation of constraints of the form

$$q(x,\ z,\ \theta,\ t) \leq 0 \qquad \forall t \in [0,t_f] \tag{39}$$

known as path or state constraints. The most common approach to this problem is to use the method of Sargent and Sullivan (1979) in which the original infinite constraint is transformed to

$$\int_0^{tf} \max(0,\ q(x,\ z,\ \theta,\ t))^2 dt = 0 \tag{40}$$

This representation is single-valued, differentiable, and precisely equivalent to the original constraint; but it has two undesirable properties:

1. As the constraint violation approaches zero, its gradient approaches zero and the Lagrange multiplier associated with the constraint goes to infinity.
2. A path constraint that is inactive by even a small amount is invisible to the optimizer.

These limitations do not usually prevent convergence in practice, provided noise is controlled as discussed above. However, they may lead either to slow progress due to constraints that flip repeatedly from active to inactive or to the need for stringent noise control as the constraint and its gradients go to zero and the Lagrange multiplier grows.

A simple modification of the path constraint representation (40), is given by

$$\int_0^{tf} \max(0,\ q(x,\ z,\ \theta,\ t))^2 dt \leq c_{\text{sig}} \tag{41}$$

This representation makes slightly feasible constraint values visible to the optimizer, and reduces noise problems by avoiding the multiplication of a small constraint by large Lagrange multipliers or penalty factors within the optimizer calculations. The disadvantage is that the modified problem may deviate significantly from the actual problem if c_{sig} is too large. For abstract mathematical problems, this disadvantage is significant, but for engineering problems in which the constraint violation has a meaning it should always be possible to define c_{sig} to be the smallest violation judged to give a significant deviation from the original problem specification.

The scaling of the constraint is also important. The scaling below was found to work well. Within the model

$$\dot{x} = max(0, q(x, z, \theta, t))^2/c_{sig}, \qquad x(0) = 0 \tag{42}$$

was used, giving sensible values for the states associated with active constraints.

Many optimizers use the feasibility test $\sum |c| \leq optacc$ (cf. Eq. 36). The model generates a path constraint, $x(t_f)$, for which a value of 1 represents an insignificant constraint violation. Using

$$c = (x(t_f) - 1)optacc \leq 0 \tag{43}$$

is enough to make the significance levels consistent, but would allow numerically feasible points with path constraint violations greater than c_{sig}. The constraint

$$c = (2x(t_f) - 1)optacc \leq 0 \tag{44}$$

ensures that any numerically feasible point will have path constraint violations less than c_{sig}. The numerical solution may therefore have active constraints with path constraint violations between $c_{sig}/2$ and c_{sig}, which seems appropriate.

This modified path constraint representation was found to give a reduction of up to a factor of 2 in the number of iterations required for successful optimization.

5. *Poor Local Approximation*

Local approximations (linear or quadratic) are often particularly poor in dynamic optimization problems. For instance, this situation is found to occur when taking the full step predicted from the local approximation, δ_k, causes a path constraint to become active or the system to become unstable.

The step based on the local approximation may have to be reduced by several orders of magnitude to satisfy the conditions for a new value to be accepted. This may be accomplished using a line search method to select a scaling factor α for the step to satisfy some criteria. If the line search is to be accomplished

efficiently, large reductions in step size must be allowed ($\alpha_{j+1} \approx .1\alpha_j$, where j is the line search iteration). If the line search accepts the first point satisfying the search criteria, it may reduce the step magnitude much more than necessary and be unable to improve the local approximation. This can require many full iterations of the optimization to approach the optimum, each involving costly gradient evaluations.

This problem may be tackled by modifying the line search procedure. The line search should be allowed to reduce the step size until a point α_j is found satisfying the line search criteria. This point then defines a minimum step size, $\alpha_{\min}$, while the previous point, α_{j-1}, defines a maximum step size $\alpha_{\max}$. The line search then proceeds as follows:

```
if α_max/α_min > 2 i = 1
do until i = 0
    α_{j+1} = min(2α_j(α_j + α_max)/2)
    θ_{j+1} = θ_{k-1} + α_{j+1} δ_k
    Evaluate J,c for new θ
    If line search criteria satisfied and merit function reduced then
        if α_max/α_{j+1} > 2
            j = j + 1
        else
            θ_k = θ_{j+1}
            i = 0
        end if
    else
        θ_k = θ_j
        i = 0
    end if
end do
```

This algorithm essentially tries to locate the largest step satisfying the line search criteria, α^*, within a factor of 2, assuming that the criteria are met for all $\alpha \in [0,\alpha^*]$ and not satisfied elsewhere.

This method has been found to reduce the overall computational effort in approaching the optimum by up to 30% compared to that obtained by accepting the first point satisfying an Armijo cone condition on the merit function while allowing step reductions of up to a factor of 10 at each line search iteration and using a quadratic interpolation formula to estimate the step reduction.

A more precise, gradient-based line search such as that discussed by Gill *et al.* (1981) is unlikely to be beneficial as the increased line search cost would require an implausibly large reduction in the number of line searches to give a net benefit.

6. Discussion

We have developed a code incorporating these characteristics and applied it successfully to more than ten industrial design problems, some of which are discussed later in this chapter. Limited testing on standard problems also supports its effectiveness. This is problem 5.1 in Vasantharajan and Biegler (1990) illustrating that some difficulties with feasible path methods were solved without any difficulty using the code developed.

III. Neutralization of Waste Water

Neutralization of waste water has been chosen to illustrate the integrated design approach as it is a relatively simple system of considerable industrial importance and it is subject to large disturbances and a high degree of uncertainty. We start by explaining the general background to neutralization system design. Following this, a procedure applying the general integrated design approach to this application area is presented.

The design of waste water neutralization systems for an industrial plant or site generally involves an interaction between "end-of-pipe" treatment system design and decisions about the design and operation of other processes. The end-of-pipe neutralization system typically consists of a small number of CSTRs or in-line mixers arranged in series. Reagent is added to each mixer based on measurement of its exit pH. The design of the end-of-pipe treatment system is a compact design problem amenable to general treatment. The more general design decisions are application-dependent and open-ended and do not lend themselves to explicit incorporation within a design method for waste water neutralization systems. A useful design method for waste water neutralization should, however, facilitate interaction with this broader design context as decisions at this level can have a more substantial effect on total treatment cost than the end-of-pipe design itself.

Before presenting our procedure for waste water neutralization system design, we discuss general design methods for the selection of structural or discrete design variables.

A. Making Discrete Design Decisions

Real design problems involve discrete design decisions, such as layout of reactors and choice of control scheme, as well as selection of continuous design parameters, such as reactor sizes and controller tuning. The design tools pre-

sented have focused on techniques for analyzing performance of particular designs and for optimizing continuous design variables.

One approach to making discrete design decisions is to pose a mixed-integer nonlinear program (MINLP) and apply a suitable optimization method to the selection of discrete design parameters. We have not adopted this approach for the following reasons.

1. The MINLP approach requires all the objectives and design options (including models) to be fully defined before solution can commence. In typical waste water neutralization system design problems, models and costs are generated and refined for options as the results of the design process indicate a need to do so. The problem definition itself may be modified substantially as the design progresses owing to interaction with design decisions in the broader process context. These characteristics are not conducive to solution methods based on one pass through a single large optimization problem.
2. MINLP algorithms typically capture the results of NLP solutions for particular sets of discrete variables by using some form of linearization at the NLP solution. A design engineer can potentially extract much more information than this and bring judgment and experience to bear in deciding which option for the discrete design parameters to examine next. If the NLP problem can be solved in seconds, a MINLP algorithm will often have the edge over a designer as it can extract limited information about many design options in the time it will take the designer to extract a lot of information about a single option. If the NLP solution time is around an hour or more, as may be the case in dynamic optimization or worst-case design, the designer can be expected to gain the edge over the algorithm.

An alternative to using a MINLP algorithm is to use a heuristic-based, designer-driven search, coupled with a bounding strategy to efficiently eliminate structures. This approach has considerable potential when good rules exist for modifying designs and when bounds on achievable performance can be obtained for candidate structures without too much effort. Examples of this approach are given by Chan and Prince (1988) and Mizsey and Fonyo (1990). Chan and Prince present a heuristic strategy for modifying flowsheets to synthesize a flotation cell separation circuit and show that it gives reasonable results. Mizsey and Fonyo looked at the use of heuristics in design more generally, coining the term "predictor-based bounding strategy" to describe the use of bounds to eliminate structures as candidates selected by a search are evaluated rigorously.

Predictor-based bounding appears appropriate for the integrated design approach proposed. A range of tools can be used for generating bounds on performance:

1. nominal and worst-case ideal delay-limited control evaluation
2. nominal and worst-case steady-state evaluation
3. nominal and worst-case dynamic evaluation

By applying these tools appropriately and using sensible rules to guide the design choices it is possible to develop efficient integrated design procedures.

B. Design Procedure for Neutralization of Waste Water

1. Problem Definition

The basic design problem is defined by which streams are to be treated and the effluent characteristics to be achieved prior to discharge from the treatment system. This definition may be a design decision at a higher level. Therefore a range of problems may need to be considered to provide information on which higher-level design decisions can be taken. This places particular emphasis on efficient preliminary design methods that will allow higher-level options to be considered without costly detailed design analysis or experimental work.

A complete characterization of a neutralization problem for design purposes might require the following:

1. a set of titration curves (pH values tabulated as a function of reagent concentration) or a thermodynamic model covering the range of conditions the treatment system will encounter
2. a definition of the acceptable variation in output characteristics, usually in terms of pH
3. a definition of any additional process constraints (e.g., lower bound on pH within the treatment system to avoid gas formation)
4. a description of the variation over time of effluent flows and titration characteristics/compositions (disturbances)
5. properties of reagents that may be used for neutralization, including reagent particle size distribution and reaction kinetics if the reagent is used in solid form
6. a definition of any equipment outside the scope of the treatment system that provides smoothing of the composition or flow variations in the effluent streams
7. constraints on achievable mixing characteristics, particularly minimum mixing delay and minimum residence time
8. constraints on process equipment selection and sizing, e.g., whether in-line mixers can be used and maximum total volume
9. constraints on measurement type and positioning, e.g., whether flow measurements are available and whether in-line probes are acceptable

10. constraints on reagent addition system type and positioning, e.g., only enough acid reagent available for a final stage pH adjustment
11. maintainable performance levels of equipment, particularly measurements, e.g., maximum lag and maximum bias
12. constraints on controller type, e.g., only direct PI control of reagent addition to be used
13. costs for all design options that have a significant effect on the cost of the final system.

It is impractical to expect to know all these characteristics *exactly* for any problem; moreover, they may be highly variable over time. Thus the specification must be allowed to include sets of possible characteristics rather than just nominal values. Default values and bounds for many of the above characteristics are given in Section IValong with advice on experimental procedures to generate design data. Approximate costs for most options are given in Walsh (1993). It is not necessary to have all the information in order to begin the design.

In defining the disturbances, the need to start the plant up should be considered. A reasonable base-case set of disturbances is startup from zero flow to at least 25% of maximum flow in a single step and a 10% load change at maximum flow. A treatment system that cannot satisfy these requirements is unlikely to be adequately operable. These base-case disturbances may be used in preliminary analysis if actual data are not available.

For preliminary analysis of dynamic response, the information required is

1. the allowable pH range for the treated waste water
2. the minimum time between the disturbance reaching the outlet of the system and the feedback control response reaching the outlet of the system
3. the disturbance condition giving rise to the maximum pH change at the outlet within this time (this may be found by worst-case design if necessary)
4. the relationship between pH and reagent concentration within the allowable pH range under this disturbance condition, i.e., at least a partial titration curve.

This information provides an adequate summary of the overall characteristics for the purpose of evaluating whether an ideal (delay-limited) controller could achieve the required exit pH performance. The worst-case disturbance is often obvious and is therefore included as part of the specification information. It should be noted that if there is more than one tank between the initial disturbance effect and the treatment system exit, the worst-case condition for a given change in effluent load will involve maximum flow. The effect of flow in reducing the concentration change for a given load change (flow^{-1}) will be outweighed by degraded disturbance rejection (flow^{-n}, where n is the number of tanks). If the

worst-case condition is not obvious, the methods for worst-case design may be used to identify this condition given a model and uncertainty description.

Steady-state properties of the neutralization system are generally an issue only when neutralizing acid effluents with solid alkali reagents. For evaluating steady-state properties, the following information is required:

1. the allowable pH range for the discharged stream
2. the properties of the reagent to be considered
3. the combination of acid concentration and flow in terms of maximizing the pH change downstream of the treatment system exit (this may be found by worst-case design if necessary)
4. the relationship between pH and reagent concentration within the allowable pH range under this load condition.

The overall design procedure is summarized in Fig. 3. It should be noted that the design techniques are deployed in order of increasing computational load and design data requirements. This allows most design options to be rejected

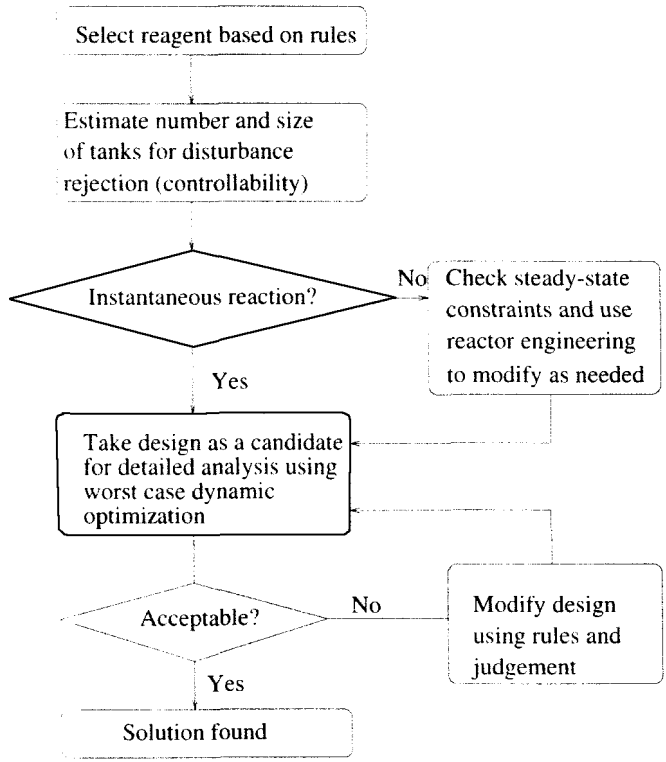


FIG. 3. Integrated design procedure for waste water neutralization.

with little effort. Only a few options will be analyzed using worst-case dynamic optimization.

2. Preliminary Design

The recommended procedure for preliminary design is

1: Select reagent

2: Evaluate design requirements for adequate transient control performance using ideal control analysis or optimal PI control performance estimates

3: If solid alkali reagents are being used then investigate steady-state conversion

If system chosen by control analysis has acceptable reagent conversion then take as a candidate design for more detailed analysis

Otherwise modify design based on reaction engineering

4: Take the resulting design as a candidate design for more detailed analysis

The first three steps are discussed further below.

a. Reagent Choice. Many rules exist for guiding reagent choice, scattered over the open literature. These rules range from clear constraints on when certain reagents may feasibly be used to fuzzy preferences and warnings. These rules are summarized in Walsh (1993), along with approximate costs for common reagents.

The following procedure for selecting a reagent is recommended:

1. Identify the set of feasible reagents.
2. Identify relevant design considerations for these reagents.
3. Decide whether any of these considerations override using the cheapest feasible reagent.

The key treatment system properties to consider in carrying out this analysis are (1) whether the effluent is acidic or alkaline, (2) the desired final pH, (3) the presence of ions forming poorly soluble salts with the reagent, and (4) the total acid load.

A typical feasibility criterion is "Carbonate-based reagents cannot be used to give a final pH above 5." A typical design consideration is "Caustic (NaOH) gives reduced sludge formation compared to calcium- or carbonate-based reagents."

This analysis will usually give a clear indication of the best reagent for a given application. The most common choice is calcium hydroxide.

b. Control Analysis. The data requirements for the general case are defined in the previous section. In the most difficult cases, preliminary control analysis requires the solution of a worst-case design problem (see Section II.B.2). In most cases, a much simpler approach can be used. The worst-case disturbance condition may often be evident, e.g., a wash operation on the plant or startup of the treatment system. The worst-case disturbance characteristics can usually be bounded by the maximum step change in effluent load, combined with the maximum pH sensitivity to reagent concentration and the maximum flow. If the worst-case condition is clear, it is possible to calculate the required disturbance attenuation δ_c (Section II.B.2). The acceptable δ_{cnet} can be computed from the titration curve (or a thermodyamic model) and the required output pH properties, with due allowance for the effect of pH measurement bias. The inlet concentration variation, Δ_{cnet}, can be computed from the flows and concentrations associated with the worst-case disturbance. The required overall disturbance attenuation is then given by $\delta_{cnet}/\Delta_{cnet}$.

The disturbance attenuation capability, δ_a, of a particular design can be calculated using Eqs. (31) or (52). Equation (31) will give either a rigorous or a heuristic bound, depending on whether the true or effective delay is used for t_d' (Section II.B.2). Equation (52) gives an *estimate* of the optimal performance *with PI control* (see Section V.A.3). The calculated disturbance attenuation can then be compared to the required disturbance attenuation. If the rigorous bound indicates that a design is inadequate, then it can be straightforwardly rejected. If the heuristic bound ($t_d' = t_u/4$) indicates that a design is inadequate, it should be rejected unless some means of compensating for the minor lags in the dynamic response is to be used. If the PI performance estimate indicates a design to be inadequate by a reasonable margin, say 30%, then the design can be rejected if PI controllers are to be used.

It is possible to determine the optimal volume of n CSTRs in series to achieve the required disturbance attenuation by equating the required and estimated attenuation (see Section V.A.4).

This analysis is a good starting point for preliminary design as it can give performance bounds and estimates for all neutralization systems regardless of reagent kinetics.

If on–off control is being considered, the achievable performance can be estimated by considering a limit-cycle at the natural period, t_u. The change in concentration from maximum to minimum reagent addition, Δ_{cnet}, must be attenuated sufficiently so that the outlet concentration variation is less than the acceptable variation, δ_{cnet}. Attributing the output variation to the first harmonic

of the square-wave input and assuming reagent flow is negligible compared to effluent flow, we have the following necessary condition for successful on–off control:

$$\delta_{cnet} \geq \frac{4\Delta_{cnet}}{\pi} \prod_{i=1,n} \frac{1}{(1 + (2\pi t_{c_{mix}}(i)/t_u)^2)^{1/2}} \tag{45}$$

If the worst-case disturbance is not evident, the ideal delay-limited control analysis is better approached by using the worst-case design tools on a dynamic model, as discussed in Section II.B.2. This optimization can be solved in a few minutes (on a SPARC 2) for typical problems.

c. Solid Reagent Conversion. Given a design that is potentially adequate in its dynamic response but uses slowly reacting solid alkali reagents, the next stage is to consider whether the reagent conversion is adequate. Even with adequate pH response at the exit of the treatment system, unreacted reagent at this point may cause downstream pH drift, giving rise to violation of the specified limits. If the design fails this test, it is necessary to modify the design to improve its conversion characteristics. This may be tackled as a reactor engineering problem, bearing in mind that reaction order is greater than 1 for these reagents (Section IV.B). Adding a plug-flow reactor stage early in the mixing scheme is likely to be beneficial, and moving to an increased number of backmixed reactors may be desirable. Alternatively, total volume may be increased.

There are a number of options for carrying out this analysis:

1. Carry out a full experimental investigation of reagent kinetics to generate a model and uncertainty description and apply worst-case optimization.
2. Use the default models from Section IV.B.
3. Apply a shortcut heuristic procedure.

A shortcut procedure should ideally allow simple computation and avoid optimistic results. The procedure outlined below meets these requirements.

The worst-case conditions for reagent carryover must be identified, and the associated reagent conversion requirements δ_l (permissible fractional carryover of reagent from the treatment system exit) computed. It should be noted that the worst-case load condition may be quite different from the worst-case disturbance condition for dynamic analysis. The worst-case load characteristics will normally be bounded by the maximum neutralization load at the maximum flow and the maximum pH sensitivity and may be identified more precisely by worst-case design if desired.

The apparent reaction time constant, τ_r, may be estimated from a batch neutralization experiment based on the pH response as it approaches a steady value in the target pH range. If δ_l is the allowable fractional carryover of reagent, the

time from $3\delta_l$ of the reagent remaining unreacted to the time for δ_l of the reagent remaining provides a reasonable estimate of the effective τ_r. Estimated in this way, τ_r may be used as a cautious estimate of an effective first-order kinetic time constant and the residual reagent, δ_{res}, may be estimated as

$$\delta_{res} = \frac{e^{-\tau_{PFR}/\tau_r}}{\prod_{i=1,n} (1 + t_{c_{mix}}(i)/\tau_r)} \tag{46}$$

where τ_{PFR} is the total plug flow reaction time available. δ_{res} should be less than δ_l to ensure that carryover of reagent is not excessive. This method has a sound basis in that it looks at the response associated with the portion of the particle size distribution which will dominate the residual carryover at the critical level, δ_l. It is pessimistic in that it then treats this response as characteristic of the entire particle-size distribution.

If a full steady-state model is used, the reagent conversion may be analyzed by steady-state simulation. All the reagent flows should be specified within the model in some way, e.g., by target pH values within the treatment system. Fixing the pH of each CSTR and analyzing the conversions of reagent added to each tank in turn may be useful to provide insight into the design problem (see Section V.D for an example).

d. Summary. At the conclusion of preliminary design analysis, there should be a clear picture of what the worst-case conditions are likely to be as well as a candidate design that is likely to meet both dynamic response requirements and reagent conversion requirements. The tools used are all quite efficient, so that a large number of options can be explored quickly. This allows the problem specification to be reviewed if necessary and provides a good starting point for more detailed design analysis.

3. Complete Design

Complete design utilizes a full model of the process, control system, and disturbances, including all *significant* uncertainty and variability, to determine whether performance requirements can be met. The design structure is initially based on the preliminary design work, which can also be used to define an initial set of models for worst-case design and initial values for some of the design variables. The controller and process parameters may be adjusted using worst-case optimization. If steady-state issues have not been fully explored in the preliminary design stage, it is sensible to run a steady-state worst-case design prior to the full design. This may identify the need for structural modifications or modifications to the problem formulation in a small fraction of the time it takes to solve a dynamic worst-case design.

Including all design issues within the mathematical formulation may complicate the model excessively, potentially slowing the solution down by an order of magnitude. Measurement noise, controller sampling effects, and detailed reagent delivery characteristics all fall into this category.

In a well-engineered system, measurement noise is unlikely to affect the performance of a PI controller significantly (Section V.A.2), but it may limit the use of derivative control action and will certainly limit the performance of advanced control schemes that attempt to approach ideal control (Section V.A.5). Our approach to PID control has been to optimize PI controllers and leave the possible benefit of derivative action to the commissioning engineers.

The effect of controller sampling time is not as bad as the effect of an increased delay in the system and can if necessary (sampling interval greater than about 2 s) be approximated by a delay equal to the sampling interval. This avoids the introduction of frequent discontinuities in the control trajectory.

Reagent delivery requires careful attention. However, it may be considered after the other design issues have been settled because it is generally amenable to satisfactory solution by detailed design without making a substantial difference to the overall design or to total cost (Walsh, 1993). In initial analysis, deliberately oversized idealized valves can be considered so as not to interfere with the control response. A conservative sizing of the actual valves can be based on evaluating the maximum reagent delivery capacity obtained assuming ideal valves. If this sizing indicates extreme rangeability requirements, then specialized equipment may be needed. If the rangeability requirement is moderately high, say 100:1, the effect of reducing the maximum reagent capacity should be checked, as valves saturating for short periods of time do not necessarily degrade performance significantly. If the rangeability requirements are less than about 50:1, use of standard equipment is acceptable. Equal-percentage valves, characterized to appear linear to the controller and fitted with positioners, are a good default reagent addition element. The effect of valve errors can be considered by including a detailed valve model in the simulation, although this is likely to slow down the integration dramatically. Section V.A.2 presents some evidence that reagent valve errors are not usually critical together with an expression for estimating the exit deviations caused by deadband errors (Eq. 51).

Reliability and maintenance issues may have a significant effect on the total cost, but are unlikely to justify reworking of the design provided some basic issues are accounted for. The main effect of reliability/maintenance analysis on existing plants is that most measurements and valves must be duplicated with appropriate isolating equipment to allow maintenance. Provided this is allowed for approximately in defining any costs considered in optimizing the design, the detail of the implementation of maintenance systems is not likely to affect the design choices. Operation in conditions likely to severely degrade reliability should be avoided by adding constraints or cost penalties to the design as nec-

essary. For example, steady pH values outside the range 2–12 are associated with increased measurement bias, and pH values above 7 encourage fouling of electrodes with carbonate. Therefore, pH operating points between 2 and 7 are preferred for good reliability.

Compliance with the specified pH limits is taken to be required 100% of the time. This is increasingly the accepted design basis and is actually easier to handle than requirements in which pH limits must be satisfied $x\%$ of the time. One method of handling "$x\%$" constraints would be by excluding worst-case combinations that are expected to occur less than $(100 - x)\%$ of the time. This introduces an extra level of iteration to the design problem. Specifications that the pH must be outside certain bounds for less than a certain time can be readily included in analyses based on the full dynamic model.

4. *Rules for Modifying Design*

The procedures presented so far largely leave the structural decisions to the designer. There are, however, a number of rules that may be useful, particularly if the designer does not have extensive experience with these problems.

If the worst-case disturbance is caused by a flow change as opposed to a concentration change, consider using ratio reagent addition with feedback, adjusting the ratio rather than the reagent itself. This requires gain-scheduling for control around a backmixed tank (dividing controller error by flow would be effective). If the worst-case disturbance is a concentration disturbance, use feedforward if the load can be estimated within about $\pm 20\%$.

If steady-state reagent conversion of solid alkali is the limiting constraint, consider the following options:

1. Use a plug-flow reactor or a small CSTR as the first stage, using underneutralization to enhance overall reagent conversion. Feedforward control of reagent addition to the PFR, if practicable, avoids the poor feedback control characteristics of the PFR (see Section V.E).
2. Provide acid reagent to downstream stages to compensate for carryover of reagent at intermediate stages in the treatment system when this is the limiting factor.
3. Increase the reactor volumes.

If control is adequate apart from large limit cycles following a major disturbance, apply input conditioning so as to eliminate the limit cycle. This is much preferable to eliminating the limit cycle by reducing controller gain as this would cause an overall degradation in performance. As input conditioning can be trivially implemented in modern control hardware, it is worth implementing in the initial design if there is a well-defined and significant nonlinearity.

On–off control should be considered if titration curve variability is high and disturbances are comparable to the total neutralization load.

If the appropriate control scheme is still not effective, increase the backmixed volume or add another controlled stage, possibly an in-line mixer.

If actuator saturation causes problems, consider increasing valve size, shifting operating points to achieve a better distribution of load between stages, or speeding up response of upstream control stages by reducing mixing delay and using fast response injector probe assemblies so as to reduce transient loading of downstream stages.

If actuator hysteresis effects cause excessive pH fluctuation, consider reducing valve size, using parallel valves, or using specialized reagent addition equipment. It is not likely that such effects will require an increase in backmixed volume as the 1% hysteresis associated with valves using positioners is unlikely to be the worst disturbance to exit pH with sensibly sized equipment (see Section V.A.2).

IV. Modeling of Waste Water Neutralization Systems

> We must remember that the most elegant and high-powered mathematical analysis based on a model which does not match reality is worthless for the engineer who must make design predictions (Levenspiel, 1972).

This section discusses the development of the models used in the case studies presented in Section V, providing an illustration of the challenges involved in developing models and uncertainty descriptions to support integrated design.

All design methods must include a model (at least implicitly) to be effective. As models are rarely precise, it is important to associate an uncertainty description with each model to allow the appropriate degree of robustness to be incorporated into designs utilizing that model. An inadequate model or uncertainty description may lead to an unworkable design or to unnecessary and expensive overdesign.

In choosing between several models giving equally good predictions, we prefer the model requiring least process-specific experimental input. For each model selected for use in design, the sources of uncertainty and variability are discussed, and guidelines for experimental work to reduce the uncertainty are given where appropriate.

The following modeling issues are considered:

Steady-state relationships between pH and reagent addition (Section IV.A)

Reaction kinetics of hydrated lime ($Ca(OH)_2$) (Section IV.B)

Dynamic response and bias errors of measurement (Section IV.C)

Effect of mixing on system characteristics (Section IV.D)

A. Steady-State pH Characteristics

The complex nonlinear relationship between pH and neutralizing reagent concentration, often expressed graphically as a titration curve, is a key characteristic of waste waster neutralization systems. It determines the sensitivity of pH to concentration disturbances and hence the required disturbance attenuation. Its nonlinearity may lead to large-amplitude limit cycles in response to disturbances if not adequately compensated for, and may force controller detuning due to the variability in the sensitivity of pH to reagent concentration. Steady-state pH characteristics are therefore an appropriate starting point in considering the modeling of waste water neutralization systems. Modeling the relationship between pH and reagent addition requires a steady-state mapping from compositions to the pH of the solution. The main approaches to constructing this mapping are the use of thermodynamic equilibrium relationships and the direct use of titration curves.

Titration curves have the virtue of simplicity, while thermodynamic models have, at least in principle, greater capacity to be adapted to changing process conditions. In practice, the amount of effort required to develop a thermodynamic model based on real chemical species and physical properties is usually prohibitive. Semiempirical models based on notional species and concentration equilibria can be developed quite readily (Gustafsson and Waller, 1983). However, these have extrapolation properties identical to the original titration curves. In both cases the pH of a mixture of components can be predicted accurately, subject to the assumptions that mixing two solutions of equal pH results in a solution with the same pH and the pH measurement is ideal (Gustafsson, 1982; Luyben, 1990). The validity of this assumption is discussed in Walsh (1993).

Titration curves are therefore used as the standard model representation in the case studies discussed in Section V. The titration curves are represented as tables of pH values and the reagent concentration changes required to move between consecutive pH values. This representation facilitates specifying uncertainty in the sensitivity of pH to reagent concentration within particular pH ranges. The concentration of reagent is given in molar units (M) with concentration taken by convention as zero at a reference pH and negative below that pH. The table function is interpolated using monotonic cubic splines.

In generating titration curves, care should be taken that any reactions are allowed to go to completion (check for pH stabilization over about ten minutes if slow reactions are observed). The material to be titrated should be at the same temperature as the effluent to be treated, and care should be taken to avoid the loss of volatile components from samples to the atmosphere. The reagent used should be the main reagent to be used for neutralization. If problems are experienced in metering small quantities of lime-based reagents, caustic (NaOH) can be used near neutral to get sufficient resolution of the titration curve—at least one point per unit pH change should be generated.

In characterizing uncertainty in titration curves (variability of slope), it is important to ensure that a representative set of samples are titrated. In particular, consideration should be given to titration characteristics during shutdowns, as the shutdown of a plant that provides a lot of weak acid or base wastes may lead to dramatically increased sensitivity of pH to disturbances and reagent adjustment. If backmixed buffering tanks are to be introduced as part of the design, it is useful to look at flow-averaged samples (samples are collected in proportion to total flow over a period of time and the aggregate sample is titrated) as well as grab samples (instantaneous samples). An averaging period comparable to or less than the residence time of the buffer tank will give a picture of the titration characteristics downstream of the buffer tank with less effort and error than combining results from many grab samples.

Titration curves may be misleading when the reagent and effluent reach an equilibrium pH on a time-scale comparable to or longer than the residence time in the treatment system. Unless the reactions involved are modeled, predictions based on the titration curve(s) may be quite inaccurate. The most common instance of this involves the use of calcium hydroxide reagents (slaked or hydrate lime), which is discussed below. If other slow reactions take place, considerable modeling effort may be needed, or very careful interpretation of results.

B. Characteristics of Calcium Hydroxide Reagent

The most common reagents for neutralizing acidic waste water are solid lime-based alkalis, particularly $Ca(OH)_2$. The kinetics of these reagents are not well defined in the literature on waste water treatment (Shinskey, 1973) and may be the dominant factor in equipment sizing in many applications. It is therefore important to develop and validate improved models for these reactions.

Experimental results by Haslam *et al.* (1926) and Yagi *et al.* (1984) support a mass transfer limited mechanism for the dissolution of CaO and $Ca(OH)_2$ in water. This mechanism was verified to be consistent with observed experimental data, using the correlation of Levins and Glastonbury (1972) to predict mass transfer coefficients for suspended particles in a tank. Available correlations for mass transfer were consistently accurate in predicting the observed rate of shrinkage of spherical $Ca(OH)_2$ particles with a radii of 100 microns (μ).

There is some controversy as to how the mass transfer rate varies with particle size, with the transition mechanism between boundary layer transfer and molecular diffusion appearing to depend on particle density. Levins and Glastonbury simply sum the molecular diffusion and boundary layer terms, giving a smooth transition between the two regimes. Brucato *et al.* (1990) present experimental results with dense particles ($\rho_p > 2$ kg/m^3), showing little effect of particle size down to a 15-μ particle radius (the Levins and Glastonbury correlation predicts an increase in dissolution rate of about 66% compared to large

particles). This observation is consistent with some early experimental work on lime reagents by Haslam *et al.* (1926), which showed less than 5% variation in mass transfer coefficient between a 45-μ and 500-μ particle radius. Most of Levins and Glastonbury's work was with low-density particles. However, they did have some data which indicated that high-density particles showed less variation of mass transfer coefficient with size than lower-density particles. Asai *et al.* (1988) indicate that the transition between turbulent boundary layer mass transfer and molecular diffusion mass transfer is sharper than indicated by Levins and Glastonbury and is better approximated by the maximum of the two contributions rather than by their sum.

This supports the use of a simplified correlation, for dense particles, in which the mass transfer is taken as the maximum of the mass transfer at 100-μ radius (determined experimentally or using Levins and Glastonbury's correlation) and the mass transfer rate due to molecular diffusion, which typically becomes dominant around 10-μ diameter in a well mixed solution (mixing power density $\approx$ .2 W/kg). Including terms for the diffusion of acids to the particle surface (Walsh, 1993) we obtain the model below.

$$\frac{\partial r_p}{\partial t} = -k_{\text{mix}}.\max\left(1, \frac{10}{r_p.k_{\text{mix}}}\right)(k_1 + {}_2[\text{H}^+] + k_3 cweak)\ \mu/\text{s} \tag{47}$$

where r_p is the particle radius in microns, $[\text{H}^+]$ is the concentration of free hydrogen ions, and *cweak* is the concentration of weak acids that will dissociate below a reference pH (pH_{ref}) of 7. k_{mix} is used to adjust the degree of turbulent mixing and should be about 1 for a well mixed vessel. k_1 was estimated to be about 0.1, k_2 about 7.5, and k_3 about 1.5. It should be noted that the concentration of $Ca(OH)_2$ in the bulk solution is assumed to be negligible.

The above model assumes complete suspension of the particles. Fluid velocities in pipes of 1 m/s are ample for suspension of typical solid alkali reagents. Typical agitation intensities of around .2 W/kg are adequate for suspension of typical calcium hydroxide preparations in tanks.

To complete the model of lime dissolution, a discretized particle size distribution is defined, with $[x_i]$ being the molar concentration of size fraction i with radius r_{p_i}. For an ideal continuous stirred tank reactor (CSTR) with residence time τ the equation for updating the concentrations of the solid lime size fractions is

$$\begin{aligned}\frac{\partial [x_i]}{\partial t} = {} & 3\frac{\partial r_{p_i}}{\partial t}[x_i]/r_{p_i} - \frac{\partial p_{p_{i-1}}}{\partial t}[x_{i+1}]/(r_{p_{i+1}} - r_{p_i}) \\ & + \frac{\partial r_{p_i}}{\partial t}[x_i]/(r_{p_i} - r_{p_{i-1}}) + ([xin_i] - [x_i])/\tau\end{aligned} \tag{48}$$

where $r_{p_{i+1}} > r_{p_i}$ and $[xin_i]$ is the inlet concentration.

This model was validated against industrial experimental data as discussed in Walsh (1993). Between 6 and 12 particle size fractions were found to be needed to characterize the particle size distributions, which resembled truncated log-normal distributions.

The main sources of uncertainty in this model are discussed below.

1. Errors due to variations in process chemistry: This will mainly affect k_3, which is strongly influenced by the diffusivity of the weak acid species and has about a 50% uncertainty band. Reaction inhibiting impurities may have a very major effect, which should be excluded by experimentation as it cannot readily be compensated for.
2. The variation of the reagent particle size distribution: This depends very much on the quality of the reagent source and should be tested by sampling of the supply.
3. The error due to variation in mixing conditions and uncertainty about their effect: This error is expected to be small for stirred tank reactors using typical power levels of around .2 W/kg ($k_{\text{mix}} \approx 1$). There is significant uncertainty in the effect of plug-flow mixing; but so long as the particles are properly suspended, $k_{\text{mix}} \in [.5, .9]$ can be expected.

To characterize the reagent properties for particular applications, batch (dynamic) titrations should be carried out with a range of final pH values across the operating region. The batch should be mixed using an agitator rather than a bead stirrer to get realistic mixing conditions. The suggested sampling rates for monitoring the pH response are every 10 seconds for the first minute after addition of lime, every 30 seconds for the next 4 minutes, and every 2 minutes thereafter. The particle size distribution should be measured either by *wet*-screening or by use of a laser counter such as a Coulter counter. If the time between preparing the suspension and completing the measurement is more than a few seconds, water should not be used in making up the suspension unless the fraction of $Ca(OH)_2$ required to saturate the water is small. The particle size distribution can then be used with the default model parameters given to check whether the standard model is appropriate. If deviations are substantial, the model parameters should be adjusted empirically. The most likely need for adjustment lies in the alkali region where reaction inhibition or back-reactions may occur.

C. Measurement Response

The response of electrodes used to measure pH is commonly modeled as a first-order lag ranging from one second to several minutes. The actual behavior of electrodes is known to be very complex and a number of researchers have

attempted to develop improved models (McAvoy, 1979; Hershkovitch *et al.*, 1978; Johansson and Norberg, 1968). These models are more complex than the first-order lag model and require much more data to define (compositions, diffusion coefficients, dissociation constants). The structural and parametric mismatch to the actual response has not been demonstrated to be reduced relative to the first-order lag approximation, so the use of these models does not seem justified. As a first approximation, the work on more complex models seems to support a first-order lag as being a good match to the form of the electrode response, although the appropriate lag coefficient is expected to vary with operating conditions.

pH probe dynamic response in CSTRs will be taken, based on industrial experience, as a first-order lag with a default maximum value of 30 seconds and minimum value of 5 seconds. A bias error of $\pm .25$ pH will also be taken as a default and it will be assumed that the pH setpoint lies between 2 and 12 to avoid the potentially large errors outside this region (Walsh, 1993). These defaults can, of course, be modified; e.g., for a well maintained multiple probe injector assembly, a faster response and smaller bias error would be appropriate (McMillan, 1984).

Attempts to reduce the uncertainty by experimentation should be approached cautiously as many design parameters affect the apparent response, including fluid velocity near the probe, operating pH, effluent composition, and probe aging and fouling over time.

D. Mixing

The most commonly used model of a mixed vessel is the "fractional tubularity" (delay-lag) model in which some part of the reactor is taken as exhibiting plug-flow conditions and contributing a delay ($t_{d_{\text{mix}}}$) and the rest of the reactor is taken as perfectly mixed (uniform concentrations) contributing a first-order lag ($t_{c_{\text{mix}}}$). The delay and lag in series are taken as describing the reactor residence time distribution (RTD). The delay-lag representation was validated using both CFD analysis and experimental residence time distributions (Walsh, 1993).

The delay for this model can be estimated for stirred tanks using the correlation of Hoyle (1976)

$$t_{d_{\text{mix}}} = kV_t^{0.85}/(Fagit) \tag{49}$$

For a tank with reagent added at the top of the tank and removed near the bottom at the opposite side $k = .9$, and $k = 1.8$ for a tank with the reverse flow pattern. V_t is the tank volume in cubic meters and *Fagit* is the agitator pumping rate in cubic meters per hour. It is observed that it is difficult to reduce $t_{d_{\text{mix}}}$ below about 9 seconds without causing excessive splashing and air entrainment.

This observation is supported by published correlations for pumping rate and air entrainment which indicate a minimum delay of about 7 seconds (see Walsh, 1993).

The dynamic response of in-line mixers generally approximates a pure delay (Chemineer, 1988). In using in-line mixers, it is important to ensure that injection velocities are high compared to stream velocities and to recognize that about 7 seconds elapses before reagent added to the line reaches the measuring device. These measures ensure smooth reagent addition and adequate micromixing, thereby avoiding excessive variation on the measured pH.

To avoid the need for special procedures and modification of the integration algorithm, delays may be modeled by using rational approximations, e.g., Padé functions or multiple first-order lags in series. Experimentation suggests that 10 series lags is adequate for most applications, so this is used as a default. The approximation should be checked by comparison with a more detailed model where it is believed to be particularly significant.

The above discussion considers mixing only in terms of time response associated with mixing and does not consider the interaction between mixing and reaction kinetics. For fast neutralization reactions and reactions with first-order kinetics, this emphasis is correct; but the time response or residence time distribution (RTD) is insufficient to characterize the effect of mixing on more general reactions. This is because the RTD does not uniquely determine the degree of micromixing of particles; e.g., a set of parallel plug-flow reactors can match the RTD of an ideal CSTR to any required precision, but the time at which particles of different ages first become mixed is very different from that of an ideal CSTR.

When RTDs are well represented by fractional tubularity models, the extremes of micromixing can be achieved by placing the CSTR before the plug-flow section (maximum mixing) or by placing the CSTR after the plug-flow section (maximum segregation). For predominantly second-order reactions, maximum segregation will improve conversion. In the absence of further information, placing the CSTR first will generally give the worst yield for a given RTD and provide a conservative approximation to the micromixing effects for premixed reactants. If the reactants are not premixed, it is possible that the reactants will not mix at all in the plug-flow stage. In this case, the conservative approach is to model the effect of delay on the control but to neglect its effect on reagent conversion. This may be implemented very efficiently by delaying only the reagent flow and not delaying any concentrations, a method that gives a general-purpose conservative model. This model is used in the case-studies on solid reagents (Section V).

The delay associated with mixing in a tank is assumed to be greater than 7 seconds and can be estimated from correlations. A sensible default value for the delay in a stirred tank is 10 seconds, as it will usually be desirable to approach

this lower bound. No uncertainty is associated with the delay, on the understanding that the value used in the model is the maximum expected value and that reduced delay will simply improve controller stability and/or performance. The effect of delay on reagent conversion may depend on the detailed mixing pattern, and conservative approximations are used.

One source of uncertainty that has not been tackled is the deviation of the age distribution of solid particles from the age distribution of the liquid. This is not very well understood and varies with particle size. This effect can be minimized by ensuring a good margin from the conditions under which particles are just suspended.

E. Summary

Models for the steady-state relationship between pH and reagent concentration have been explored. In general, titration curve models are more convenient and more readily determined, and these are used in the case-study work presented below. Guidelines for generating titration curves are presented.

A model for reaction of the solid alkali $Ca(OH)_2$ has been developed using general mass-transfer analysis principles. The model has been validated against experimental data and results in the literature. Guidelines for experimental validation and tuning of the model are given.

Measurement characteristics have been reviewed and a default model including an uncertainty description presented.

Mixing properties have been reviewed and a series delay-lag model chosen for general use. Methods for predicting the value of the delay and a lower bound on the achievable delay are discussed and a default value given.

Once the basic modeling foundations have been laid as above for a given application area, the modeling effort for any particular problem is greatly reduced.

Modeling issues in waste water neutralization are discussed in more detail in Walsh (1993), which also discusses redox and precipitation reactions and the reaction of carbonate reagents.

V. Design Examples

The design examples presented in this section cover a wide range of waste water neutralization problems. The examples studied confirm the power of the design methods developed. We consider that the success of the integrated design approach on this challenging and industrially important problem area establishes

its industrial relevance and motivates further work to explore the use of the method on more general process problems. The examples examined cover the following:

Exploration of a generic neutralization problem (Section V.A)

Preliminary design of a central effluent plant (Section V.B)

pH control of a strong-acid/strong-base system (Section V.C)

Neutralization and precipitation with $Ca(OH)_2$ in a central effluent plant with highly variable effluent characteristics (Section V.D)

Neutralization of a well-defined, highly concentrated acid stream with $Ca(OH)_2$ (Section V.E).

The last four examples are based on industrial case studies.

A. Exploration of a Generic Neutralization Control Problem

Before presenting the industrial case studies, it is helpful to use the tools developed to explore a generic problem, allowing some general results and insights to be developed.

1. Problem Definition

To qualify as a generic problem, the disturbance conditions, treatment system structure, and performance targets should be typical of real problems while avoiding unnecessary complexity. These aspects of the generic problem are defined and discussed below.

a. Disturbance Conditions. Disturbances involve both flow and concentration changes. Concentration changes are the most challenging, as it is always possible to minimize the direct effect of flow by use of flow measurements, e.g., by using pH feedback to the *ratio* between reagent and effluent flow. The disturbance was therefore taken as a disturbance in concentration only, for simplicity. The most difficult disturbances are those that occur as step changes, so the disturbance is taken as a pulse change in concentration at the inlet to the treatment system. The duration of the pulse is taken as 30 minutes, allowing recovery to steady-state between the rising and falling edge of the pulse. The precise duration of the pulse is not important.

b. Treatment System. The treatment system was assumed to be made up of continuous-flow stirred-tank reactors (CSTRs) in series as this is the standard industrial system. Most of the examples presented below use equal-sized tanks

with $t_{d_{\min}} \approx .05\ V_t/F_T$, as this is typical of industrial systems. Variations from this typical system were made to check the generality of conclusions, where appropriate. It should be noted that only the *ratio* of time constants and delays will affect performance, as an overall scaling of the model time constants simply scales the time axis of the response obtained. Measurement lags were varied to illustrate the effect of minor lags on performance. Actuator dynamics and reagent kinetics are neglected for simplicity, although the associated lags have an effect on control performance similar to measurement lags if they are all small compared to the residence time. Both acid and alkali reagent addition is assumed to be available. For most of the analysis, the reagent addition is assumed to have infinite rangeability and precision, as adequate rangeability and precision can normally be provided by appropriate detailed design. A deadband error effect is considered where an approximation to a practical "sticky" valve is required (Section V.A.2). Where controllers are implemented, they are assumed to be PI controllers from the exit pH of a CSTR to reagent addition at its inlet, as this is the standard industrial system. The final part of this section considers the changes required to move from typical PI performance to the ideal delay-limited control bound.

c. *Performance Requirements and Setpoints.* The target pH is assumed to lie in the range pH 5–9 (typical discharge consents). The titration characteristic is assumed to be strong-acid/strong-base

$$\begin{aligned} \mathrm{pH} &= -\log_{10}\left(\frac{-cnet + \sqrt{cnet^2 + 4.10^{-14}}}{2}\right), && cnet < 10^{-4} \\ \mathrm{pH} &= -\log_{10}\left(\frac{10^{-14}}{cnet}\right), && cnet \geq 10^{-4} \end{aligned} \tag{50}$$

as this gives the strongest steady-state nonlinearity in the target pH range. This means that the acceptable concentration range is $\pm 10^{-5}$ N. As the disturbance is approximated by two equal but opposite steps in concentration and hence produces equally severe acid and alkali deviations, the optimal operating point in relation to ideal delay-limited control or a linear feedback controller (such as a PI controller acting on concentration) is 7 pH (0.0 N concentration, midway between the upper and lower concentration bounds). As the disturbance is in concentration only and perfect actuators are assumed, setpoints other than that of the final stage do not affect the dynamic response for ideal or linear control. Setpoints therefore need to be optimized only when PI controllers based on uncompensated pH measurements are considered. A requirement for the control response to recover to near a steady-state is included by requiring that the concentration variations one hour after the disturbance be less than 10% of the

concentration variation that would cause violation of the 5–9 pH bounds. In terms of pH, this corresponds to bounds of 6–8.

The overall design problem is to select number of tanks, sizes of tanks, number of controllers and controller gains, and integral action times to achieve a required disturbance rejection at minimum cost. The generic problem is represented in Fig. 4.

2. *Why pH Control Is Difficult*

A useful starting point is to examine the characteristics of the pH control problem that make it difficult. The most fundamental reason for the difficulty is the stringency of the performance requirements, which may make it necessary to add reagent to within less than .1% of the ideal amount just to stay within legal limits on the pH of the treated stream.

Pure delays in dynamic response impose fundamental bounds on achievable disturbance rejection, as discussed in Section II.B.2. Uncertainty imposes additional limitations. For example, if the measurement lag were accurately known, it could be canceled and would not limit control performance. But the measurement response is more complex than this and it is variable over time; thus it cannot readily be canceled and combines with the delay to limit controller performance (see Sections V.A.3 and V.A.5). Limits on the reagent addition rate prevent the ideal delay-limited control bound from being achieved in some cases, even in the absence of uncertainty, as infinite actuator range is required (see Section V.A.5). However, these limits are not observed to have a major effect on PI control performance as the degree of transient overshoot is moderate compared to the steady-state control output change required.

Three effects that have a less clear-cut impact on achievable performance are the pH nonlinearity (if uncanceled), the precision limits on the reagent addition, and measurement noise. These are investigated below, in the context of the generic problem of Section V.A.1, to complete the picture of why pH control is difficult.

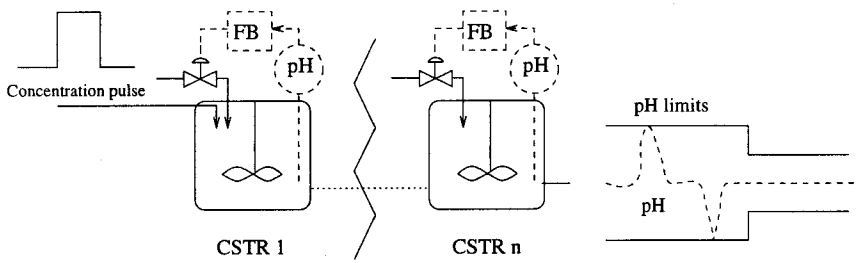


FIG. 4. Generic pH control problem.

a. Effect of Nonlinearity. Using a tank having a residence time (V_T/F) of 3 min, $t_{d_{mix}} = 9$ s, and $t_{c_{mix}} = 171$ s, the achievable disturbance rejection was found by optimizing a PI controller based on pH and a PI controller based on concentration. Secondary lags were neglected.

Delay-limited feedback analysis using Eq. (30)

$$\delta_a = 1 - e^{-t'_d/t_{c_{mix}}} \sum_{i=0}^{n-1} (t'_d/t_{c_{mix}})^i/i! \approx \frac{1}{n!}\left(\frac{t'_d}{t_{c_{mix}}}\right)^n$$

indicates that the maximum concentration pulse disturbance level is 1.95×10^{-4} N (calculated as $10^{-5}/\delta_a$ using $t'_d = t_{d_{mix}}$). A PI controller based on pH measurement was found by optimization to be able to handle 1×10^{-4} N disturbance level. A PI controller based on concentration measurement can handle a 1.6×10^{-4} N concentration pulse. The nonlinearity therefore has a significant quantitative effect on the achievable performance, reducing the maximum disturbance from 80% of the ideal control bound to 50% of the bound. At least for this problem, the dominant limitation is given by the process dynamics, as the PI controllers in both cases approach the theoretical bound on feedback controller performance imposed by the dynamics. This indicates that the effect of the nonlinearity on controller response does not necessarily have a strong influence on the design, even with the extreme strong-acid/strong-base nonlinearity.

Input conditioning can often be applied to minimize the effect of nonlinearity and chemical buffering usually reduces the nonlinearity compared to the strong-acid/strong-base case considered above. Given that the example above shows that even the extreme strong-acid/strong-base nonlinearity does not necessarily degrade performance qualitatively, the rest of the analysis is carried out assuming the titration nonlinearity to be sufficiently compensated as to be negligible.

b. Effect of Precision Error in Reagent Addition. McMillan (1984) notes that reagent control valve precision error is dominated by deadband error, where "[d]eadband error is the change in signal required to start the stroke from a stationary position or to change the stroke direction upon a change in signal direction." The deadband in valve stem position error is typically about 1% of range when a positioner is used. The discussion below assumes a linear valve characteristic (relation between flow and stem position for a fixed pressure drop) for simplicity.

This effect was explored by using a PI controller based on concentration around a tank with $t_{d_{mix}} = 9$ s and $t_{c_{mix}} = 171$ s as above. Again, there is no reason to expect the conclusions to be sensitive to the particular mixing conditions. The definition of deadband given leaves some ambiguity as to the valve behavior. Two models were considered for a unit deadband error:

If $u_d > u + 1$ then $u = u_d - 1$

If $u_d < u - 1$ then $u = u_d + 1$

and

If $abs(u_d - u) > 1$ then $u = u_d$

The first model corresponds to the valve position tracking the demanded output, with an offset of 1 unit (valve sticking and then gliding). The second model corresponds to the valve position not responding to discrepancies, with the demanded output smaller than 1 unit and moving promptly to the demanded value when the difference exceeds this (valve sticking and lurching). The "sticking-and-gliding" model is appropriate if the friction resisting movement is independent of the rate of change of valve position. The "sticking-and-lurching" model is appropriate if friction becomes negligible once motion commences. Sticking and gliding is probably closer to the real behavior, but sticking and lurching is also considered as it is not ruled out by the definition and is likely to give worse performance.

Numerical experiments (Walsh, 1993) indicate that the peak-to-peak deviation of the exit concentration can be bounded, for fairly tightly tuned controllers, by calculating the response of the exit concentration to a reagent valve exhibiting a square wave oscillation with peak-to-peak amplitude equal to the deadband error. The exit concentration variation can therefore be estimated as

$$\frac{4\Delta_{cnet}}{\pi} \prod_{i=1,n} \frac{1}{(1 + (\pi t_{c\text{mix}}(i)/t_u)^2)^{1/2}} \tag{51}$$

where Δ_{cnet} is the inlet concentration change resulting from a change in valve position equal to the deadband. This error would not be the limiting factor on transient performance unless the disturbances were much smaller than the total load or the deadband error was much higher than the value of 1% of range achievable using a valve positioner.

c. Effect of High-Frequency Measurement Noise. In a well-designed pH measurement system, a moderate amount of high-frequency noise is usually present (about $\pm .1$ pH). This noise has not to our knowledge been analyzed in the detail required to carry out a rigorous evaluation of its likely effect. In our experience, it does not present any significant difficulty when using PI control. It will combine with other forms of process uncertainty to limit the performance achievable with more advanced control strategies (see Section V.A.5).

d. Summary. pH nonlinearity, valve precision, and high-frequency measurement noise do not have a major impact on the design requirements when using

series CSTRs and PI control. The key factors limiting design performance are *delays and uncertainty.*

3. Examination of Bounds from Delay-Limited Control Analysis

The performance bound from delay-limited control analysis (Section II.B.2) may be computed for a number of series CSTR systems and compared to the performance obtained by optimizing PI controllers. For simplicity and generality, this section assumes the pH nonlinearity to be canceled at the controllers.

Extensive numerical optimization results (Walsh, 1993) allow a number of observations to be made.

1. The optimal PI performance was within 30% of the correlation

$$\delta_a = \frac{3t_d'^n}{2\prod_{i=1,n} t_{c\max}(i)} \tag{52}$$

   This represents a degradation of performance by a factor of $1.5n!$ compared to ideal delay-limited control.
2. If one tank in a multitank chain was uncontrolled, the disturbance attenuation was further increased by about 50%.

For preliminary design purposes, the estimate from Eq. (52) is an adequate predictor of achievable PI performance when all tanks are tightly controlled. If control on one tank of a multiple CSTR system is rendered ineffective—due to uncertainty, high delay compared to the minimum delay, or simply the absence of a controller—the predicted disturbance attenuation should be increased by 50%. An exception to this is the case of variation in the sensitivity of pH to concentration on the *final* CSTR. In this case, no degradation of performance from that obtained with the minimum buffering (maximum titration curve slope) will occur as the performance required in terms of concentration deviations relaxes along with the controller performance.

Section II.B.2 suggested that when minor lags are present in addition to pure delays, a heuristic "effective" delay could be computed based on the natural period of the control loop divided by 4. Numerical optimization suggested that Eq. (52) remained approximately valid, although the optimized performance was slightly improved compared to the pure delay case.

Summary. Estimating t_d' based on physical delays and using Eqs. (30) and (31) as appropriate gives a rigorous bound on the achievable performance with any controller. Estimating t_d' as $t_u/4$ and using these equations gives a sensible heuristic performance bound for PI feedback control (and probably for any controller that does not cancel the measurement dynamics and hence is subject to

the same bandwidth limitations). Estimating t'_d as $t_u/4$ and using Eq. (52) provides a realistic, albeit heuristic, estimate of achievable performance with PI controllers, neglecting uncertainty. Loss of effective control on a single CSTR in a multiple CSTR system does not have a dramatic effect on the achieved disturbance rejection with PI control.

4. *Optimal Sizing of Series CSTRs for pH Control*

A striking discrepancy between theory and practice in pH control systems using individually controlled series CSTR systems is the choice of the relative size of the tanks. There is a strong consensus in the literature that equal-sized CSTRs are undesirable for control purposes (Shinskey, 1973). However, equal-sized CSTRs are the industrial norm. Using the delay-limited control performance bound (Eq. 31)

$$\delta_a = 1 - \sum_{i=1}^{n} \frac{e^{-t'_d/t_{c'\text{mix}}(i)} t_{c\text{mix}}^{n-1}(i)}{\prod_{j\neq i} (t_{c'\text{mix}}(i) - t_{c\text{mix}}(j))} \approx \frac{1}{n!}\left(\frac{t_d'^n}{\prod_{i=1,n} t_{c\text{mix}}(i)}\right)$$

and a typical cost function of the form $k_1 + k_2 V_T^x$, it is possible to evaluate the desirability of using tanks of different sizes. The requirement to provide a disturbance attenuation, δ_c, can be approximated well by

$$\delta_c \geq \frac{1}{n!}\left(\frac{(F_T t'_d)^n}{\prod_{i=1,n} V_T(i)}\right) \tag{53}$$

where F_T is the total flow through the system and $t'_d \ll t_{c\text{mix}}$. The magnitude of the disturbance that can be rejected is approximately proportional to the product of the tank volumes, regardless of the number of tanks, if the minimum delay, t'_d, is assumed to be independent of the tank volumes. The economic selection of number and size of tanks to give a required disturbance attenuation, δ_c, at a flow F_T therefore takes the form

$$\min_{n,V_T(i),i=1,n} \left\{k1.n + k2 \sum_{i=1,n} V_T(i)^x\right\} \quad \text{s.t.} \quad \prod_{i=1,n} V_T(i) = (1/n!)(F_t t'_d)^n/\delta_c \tag{54}$$

where $x \neq 0$. For fixed n the problem becomes

$$\min_{V_T(i),i=1,n} \left\{\sum_{i=1,n} V_T(i)^x\right\} \quad \text{s.t.} \quad \prod_{i=1,n} V_T(i) = (1/n!)(F_t t'_d)^n/\delta_c \tag{55}$$

Cost for a given number of tanks and disturbance rejection capability is therefore always minimized by equal-sized tanks and the overall problem simplifies to

$$\min_{n} n(k1 + k2V_{T_{\text{opt}}}^{x}) \tag{56}$$

with the optimal volume, $V_{T_{\text{opt}}}$, given by $V_T t'_d (n! \delta_c)^{-1/n}$. A similar analysis can be carried out using the PI performance estimate Eq. (52), as the predicted performance is again approximately proportional to the product of the tank volumes for a given problem. The optimal volume for this case is given by $F_T t'_d (1.5/\delta_c)^{1/n}$. The difference between $V_{T_{\text{opt}}}$ with n tanks using the PI performance estimate and using the delay-bound equation with the same value of t'_d corresponds to a factor of $(1.5\ n!)^{1/n}$, that is, a factor of 1.5 for one tank rising to a factor of 2 for three tanks.

The use of this analysis is illustrated below. It might be required to estimate the optimal series CSTR system to achieve a disturbance attenuation of .001 with a flowrate of .03 m^3/s and $t'_d = 10$ s, using PI control. Using the cost function $20{,}000 + 2000V_t^{0.7}$ (based on Cushnie, 1984) and Eq. (52) gives a minimum cost CSTR system with two tanks having a volume of 11.6 m^3 each and a cost of £62,000. For comparison, a single-tank system would cost £164,000 and a three-tank system would cost £75,000. Increasing t'_d or F_T proportionately increases the required volume for a given n. An increase of a factor of 4 in F_T would shift the economic optimum to three tanks with a cost of £100,000 and a tank volume of 13.6 m^3. It should be noted that the optimal volume can be expected to lie below $(k1/k2)^{1/x}$ (27 m^3 in this example) as the optimization of n will tend to increase the number of tanks until $k1$ becomes the dominant part of the tank cost. Optimal volumes can be expected to lie between 10 and 30 m^3 if the cost expression used is appropriate.

For typical values of F_T of 30–300 m^3/hr, the predicted optimal volumes of 10–30 m^3 correspond to residence times of between 2 and 60 min. It should be economically attractive to achieve mixing near the achievable bound (see Section IV.D) with tanks of this size, so that $t_{d_{\text{mix}}} \approx 10$ s independent of tank volume. The main secondary lag in typical applications is the measurement lag. The value of this lag varies with installation conditions but is not dependent on tank volume and may be generally taken as less than 30 s. This gives an effective $t'_d \approx 25$–30 s and a rigorous $t'_d \approx 10$ s, with both values independent of tank volume. This shows that the two main assumptions of the above analysis ($t_{d_{\text{mix}}} \ll t_{c_{\text{mix}}}$ and $t'_d \neq f(V_T(i),\ i = 1,\ n)$) are roughly satisfied at the optimum for typical problems. This confirms the validity of the analysis for typical treatment systems. If flowrates were much above 300 m^3/h or the CSTR cost function was quite different, these assumptions might not hold.

The above analysis does not consider other known economic benefits that arise from choosing equal-sized tanks—buying in bulk, simplified construction, and reduction in number of spares required. These factors all reinforce the economic desirability of equal-sized tanks.

Summary. It is possible to rapidly estimate the economically optimum number and size of CSTRs in series required to achieve a given disturbance attenuation. The design obtained from this analysis will always have equal-sized tanks due to the dependence of the constraints on the product of the tank residence times. The tanks will typically be between 10 and 30 m^3 in size with residence times between 2 minutes and 1 hour. This is consistent with industrial practice and contradicts the recommendation in the literature, based on frequency response arguments, that tank sizes should be split in a ratio of about 1:4 or greater (Shinskey, 1973; Moore, 1978; McMillan, 1984). This provides a simple illustration of the ability of the integrated approach to balance operability and economic considerations in a way that qualitative argument cannot.

5. *Toward Ideal Control*

In the light of the analysis in the rest of this section, let us review what the cost of using PI control is in comparison to achieving the ideal delay-limited control performance bound, and let us examine some measures that may reduce this gap.

In Section V.A.4, we showed that optimal volume is proportional to t'_d and that there is a factor of about $(1.5n!)^{1/n}$ between the optimal volume with PI control (without uncertainty) and the volume required, assuming the performance bound is reached. The PI performance is governed by the effective delay ($t'_d = t_u/4$) rather than the actual pure delay. The maximum effective delay, assuming a mixing delay $t_{d_{\text{mix}}}$ of 10 s, a probe lag of up to 30 s, and a CSTR with $t_{c_{\text{mix}}} \gg t_{d_{\text{mix}}}$, is 28.7 s. The measurement lag may therefore imply almost a threefold increase in the required volume, compared to that required with an instantaneous measurement response.

The volume factor between PI and ideal control for a given t'_d, $(1.5n!)^{1/n}$, may be ascribed to the fact that the PI control response differs from the ideal control response. It is necessary to characterize the ideal control response in order to identify the characteristics preventing its attainment. The initial control action taken by the ideal controller must be sufficient to reverse the direction of change of concentration at the exit of the nth tank of a n tank treatment system. For one tank, this is achieved by any control action that turns a reagent excess into a deficit or vice versa. For more than one tank, infinite control action at the treatment system inlet is required to reverse the direction of change because the initial process gain to a pulse input is zero. Finite reagent delivery capacity therefore prevents the ideal control bound from being reached for more than one tank with control on the first tank only. In order not to make the disturbance worse, the ideal delay-limited control must not take a control action more than twice that required to cancel the disturbance—at least if the con-

straints are symmetric about the initial value. This means that the ideal controller must identify the disturbance magnitude and type quite accurately from the initial rate of change of pH observed. This will be obstructed by noise and errors in the model relating pH changes to concentration changes. PI control forms an estimate of the disturbance magnitude without using derivative information and is therefore slower and more cautious than an ideal controller.

The most significant factor that would allow closer approach to the ideal bound is reducing or canceling the pH measurement lag. Cancelation of the lag by lead-lag filtering would be a very effective method of improving control performance in the absence of uncertainty. As discussed in Section IV.C, the probe response is very complex (although the effect on the control performance can be approximated by an uncertain first-order lag). A greatly improved understanding of probe behavior would be needed to allow effective cancelation of the probe dynamics within a control system. High-frequency measurement noise will also limit the degree of lead-lag filtering that is practicable.

Probe lag is a function of installation conditions, in particular decreasing with increasing fluid velocity past the probe. This velocity is relatively low in a stirred tank compared to a properly designed sampling system or injector probe assembly (McMillan, 1984); thus the use of alternatives to mounting the probe directly in the CSTR should be given careful consideration. The other key factor in probe response time is fouling and probe maintenance. The long-term solution to slow response due to fouling may come from advances in measurement technology; but until such time, the potential benefits of careful maintenance and operating in conditions that minimize fouling (underneutralization rather than overneutralization) should be noted.

At the design stage the choice of the maximum probe lag to be considered has a major impact on the final design if transient performance is the limiting factor. In most of the case studies presented in this section, 30 seconds has been used based on discussion with experienced engineers familiar with systems using probes mounted in stirred tanks. This figure can probably be reduced for other types of installations, as in the case study in Section V.C. Improved understanding of measurement response would be helpful in choosing the appropriate value.

If it is desired to evaluate advanced control schemes, the models used in this work should be regarded as generating an upper bound on practically achievable performance; further work is required to generate models suitable for this purpose. At a minimum, the performance of advanced control algorithms should be checked in the presence of correlated noise and actuator deadband error, and an improved model of the probe response is likely to be needed.

Summary. The greatest scope for moving performance toward the ideal control bound lies in reducing measurement lag and improving understanding of meas-

urement response. Constraints on reagent addition rate prevent ideal control from being achieved for more than one tank with control on the first tank only. Finally, model errors and measurement noise obstruct the achievement of ideal control.

6. *Summary*

The exploration of the generic problem has allowed a number of issues to be clarified. The key conclusions are summarized below.

1. pH control is difficult because of delays and uncertainty. The pH nonlinearity, reagent addition precision errors, and measurement noise play a relatively minor role.
2. The bounds on disturbance rejection based on the ideal delay-limited control analysis are reasonably tight. The achievable performance with PI control (in the absence of uncertainty) may be estimated quite accurately using Eq. (52).
3. The optimal series CSTR configuration to minimize cost for a given disturbance attenuation requirement and instantly reacting reagent comprises equal-sized tanks for typical treatment systems.
4. The greatest potential for moving control performance toward the ideal delay-limited control bound lies in minimizing or compensating for pH measurement lags.

B. Preliminary Design of a Central Effluent Plant

In this example, we illustrate the use of preliminary design tools in exploring design options efficiently. The design problem is represented schematically in Fig. 5.

The aim of the analysis was to determine the likely benefit from local containment of the main pulse flow disturbance in terms of requirements for the treatment system. The treatment system was being required to move from 95% compliance with discharge consents of 6–9.5 pH to 100% compliance. Two CSTRs with 10 minutes residence time at normal flows were already available.

Reagent selection was not an issue as NaOH/HCl were already being used. The reaction was assumed to be virtually instantaneous, eliminating steady-state considerations. A complete design analysis was not considered justified until the decision on local containment had been made. The preliminary control analysis (Section III.B.2) was therefore used as the sole design tool. As two CSTRs with 10 minutes residence time at normal flow already existed, the analysis was carried out based on n CSTRs of this size. Application of Hoyle's correlation (Eq.

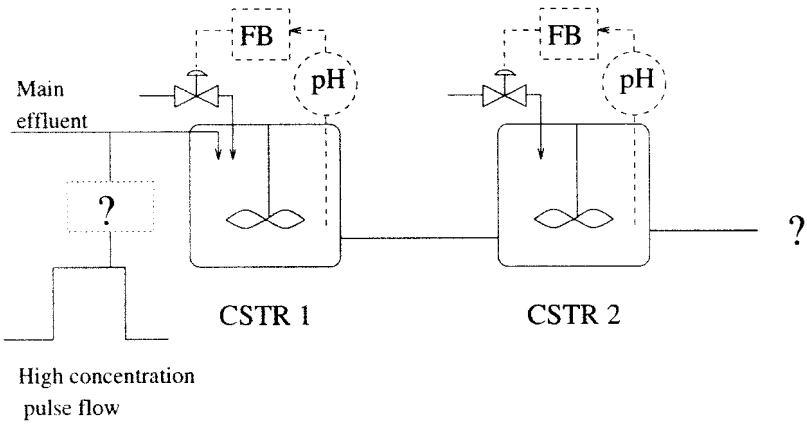


FIG. 5. Design problem for preliminary design.

49) to the existing tank design indicated a mixing delay of 10 seconds. Assuming typical industrial probes situated within the CSTRs, the measurement could contribute a maximum lag of about 30 seconds. Each CSTR therefore has a natural period, t_u, of about 2 minutes and a heuristic effective delay ($t_u/4$) of 30 seconds (see Section II.B.2).

Inlet effluent concentration was normally about $\pm .01$ N, but there was a regular pulse flow discharge of 30 minutes duration, at comparable flowrate to the main effluent streams and between -1 and -3.5 N concentration. There was no available means of measuring the effect of this disturbance on the inlet concentration, so feedforward control was not considered as an option. This discharge gives the worst inlet concentration variation and the maximum flow and was therefore taken as the worst case. This disturbance gives a residence time per CSTR of about 5 minutes.

The first scenario explored was the plant requirement to handle all the streams with no local containment of the flow pulse. Extensive titration data were provided which indicated a minimum concentration change across the consent range of .0005 N. This compares to the strong-acid/strong-base concentration change of about .00003 N, showing that significant buffering is present even in the worst case. The slope of the titration curve ($\partial \mathrm{pH}/\partial_{cnet}$) was observed to vary by more than 10:1 between experimental samples. The inlet concentration change, Δ_{cnet}, in the worst-case disturbance is 1.75 N. The rising edge of the pulse is much more severe because of the higher associated flow; thus the setpoint can be biased toward the upper consent level, giving an allowable concentration deviation, δ_{cnet}, of approximately .0005 N. The required disturbance attenuation, δ_c, is therefore about .0003 (.0005/1.75). The attenuation available from n tanks whose design is the same as those already available can be examined by using the bounds on the performance computed using the pure delay and the effective

delay (Eq. 30) and by using the heuristic estimate of attenuation with PI control (Eq. 52). The results of this analysis are tabulated below:

n	1	2	3	4
Pure delay bound	.03	.0005	6×10^{-6}	1.2×10^{-7}
Effective delay bound	.1	.004	.00015	3×10^{-6}
PI estimate	.15	.015	.0015	.00015

The rigorous bound implies that at least three tanks would be required, and the heuristic estimate indicates that a four-tank system would be required when using PI control. Even with a four-tank system, it can be expected that meeting the performance requirements will be difficult, considering the variation in titration curve slope—and hence process gain—noted.

If local containment is applied to the worst-case disturbance, the new worst-case disturbance can be assumed to be bounded by startup to maximum flow, giving $\Delta_{cnet} = .01\ N$. This disturbance may be in either the acid or the alkali direction, so δ_{cnet} is reduced to .00025 N. This corresponds to a required disturbance attenuation of .025. The residence time has been doubled and the achievable attenuation decreases by a factor of 2^n compared to the table above. The rigorous bound suggests that one tank might be just adequate (.017 versus .025), while the heuristic estimate indicates that two tanks would be required. The two-tank system has an estimated attenuation of .004 compared to a requirement of .025 and should be able to deliver the required performance despite the titration curve variability.

The rules presented in Section II.B.4 indicate that on–off control should be considered (maximum disturbance comparable to maximum load and titration curve slope varying by more than 5:1). The corresponding inlet concentration change, Δ_{cnet}, is at least .02 N (to allow effluent concentration variations to be accommodated). The residence time is 10 minutes, and the natural period is about 2 minutes. Applying Eq. (45) gives a predicted output concentration variation of $2.5 \times 10^{-5}\ N$, which is much less than the minimum concentration variation spanning the consent range ($\delta_{cnet} \approx .0005\ N$), confirming that on–off control could be a viable option.

In summary, the analysis indicates that, to achieve 100% compliance with the discharge pH consents, either (1) two CSTRs with PI or on–off control and local containment of the worst-case disturbance or (2) four CSTRs with careful controller design are required. The preliminary design analysis gives a useful picture of the tradeoff between the higher-level design decision as to whether to contain the main disturbance locally and the "end-of-pipe" treatment cost without requiring substantial design effort.

This example follows the procedures in Section III.B.2 quite closely, with some judgment being exercized in the selection of the appropriate worst-case scenarios. This application is fairly typical of preliminary designs in that good

approximations to the worst-case disturbance can be identified *a priori*, allowing computation of the required disturbance attenuation for comparison to the predicted performance of series CSTR systems. The main experimental requirement was obtaining sufficient titration curves to provide a good estimate of the minimum buffering. Once the key high-level design decision has been made, these curves would allow a good uncertainty description for the pH characteristics to be developed for the worst-case design.

C. pH Control of Several Strong Acid Streams with NaOH/HCl

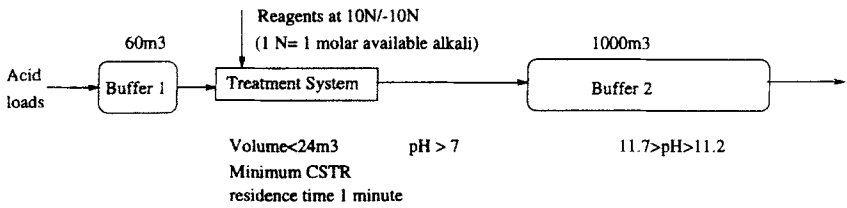
1. Problem Definition

The process requirements and constraints are summarized in Fig. 6. The distinctive aspects of this problem definition are the large delay in measurement response due to the sampling arrangement employed, the tight constraint on total volume of the treatment system, and the presence of tanks upstream and downstream which assist in meeting the specification.

The pH relationship was taken as strong-acid/strong-base, as this was found to be a good approximation in the target pH range.

Mixing in treatment system CSTRs was modeled by a mixing delay, $t_{d_{mix}}$, of 7 seconds in series with a mixing lag $t_{c_{mix}}$. The 7-second delay was based on avoiding both air entrainment and flow short-circuiting with small residence times. Careful engineering is required to achieve this mixing delay. The first buffer tank had a mixing delay of 30 seconds. The second buffer tank was treated as well-mixed.

All delays were represented by 20 identical first-order lags in series. This representation is more precise than the default 10-lag model as the time delays



Acid loads;
200-600m3/hr at -.2N, varying slowly
Up to 80m3/hr pulse at -2N
Up to 40m3/hr pulse at -6.N
Pulses last 2 hours

Measurements available within treatment system :
pH +/-.1pH bias, 1-5 seconds lag, 20 seconds delay
flow +/- 1% of value (bias), 5 seconds lag (filter)
External measurements;
None

FIG. 6. Problem definition.

are unusually large compared to the mixing lags, which are less than 2 minutes at maximum flow.

Measurement biases were assumed constant for each simulation of the system. It should be noted that the measurement performance assumed corresponds to a "clean," well-maintained system in which the pH probes are placed in an in-line sampling system rather than in the CSTR.

The slowly varying flow was treated as constant but uncertain.

As in the previous example, reagent selection had already been made and this part of the design procedure was omitted. Reaction was effectively instantaneous.

2. *Design*

a. Preliminary Design. Due to the virtually instantaneous neutralization reaction, any nontrivial screening test for this problem must be based on dynamics (Section III.B.2). The fraction of the worst-case disturbance controllable with ideal delay-limited feedback was calculated to provide a controllability measure (Section II.B.2). Disturbance concentrations were contracted toward zero by the "disturbance fraction." The disturbances have separate effects due to flow and concentration changes. A change in flow at the inlet to the first buffer tank is assumed to propagate immediately to the treatment system, changing the net concentration following a reagent addition point and creating a disturbance propagating from this point. A change in concentration at the inlet of the buffer tank must propagate through the buffer tank mixing delay before beginning to affect the treatment system. The flow and concentration effects are therefore separated in time due to the mixing delay in the buffer tank and may be analyzed separately for this system, even though they are not independent.

The first control structure considered was pH measurement and reagent addition at the inlet to the treatment system. This gives a minimum delay of 20 seconds between the disturbance reaching the exit of the treatment system and the control response reaching the exit. All feedback control structures have at least this delay, owing to the delay in measurement response. Delay between reagent addition and the pH sampling point was assumed to be negligible, the required micromixing to attenuate noise being provided by turbulent diffusion in the sampling line. Optimization of a dynamic model was used to identify the worst-case disturbances and measurement bias, calculate the disturbance fraction, and select the controller setpoint. Solution of each problem took only a few minutes on a SPARC 2. The worst-case disturbances were found to involve both acidic flow pulses starting simultaneously for the concentration effect, and the lower-concentration, higher-flow pulse starting with the other pulse disturbance already active for the flow effect. The worst-case "slowly-varying" flow

varied with the number of tanks in the disturbance path, being at its minimum value with no tanks in the path, and at its maximum value with one or more tanks in the path. The optimized controller setpoint was 11.6 to satisfy the steady-state pH constraint while giving maximum margin from neutral. The initial pH during the disturbances was 11.5 due to the measurement bias. If no backmixing is provided in the treatment system, only 1% of the flow effect or 8% of the concentration effect can be tolerated. Putting the maximum volume into a single tank allows 30% of the flow effect or 93.7% of the concentration effect to be tolerated. Splitting the volume evenly, giving two equal-sized tanks, allows 81.9% of the flow effect or 388% of the concentration effect to be handled. Uneven division of volume between two tanks or the use of more than two tanks is prevented by the requirement of at least one minute of residence time per CSTR.

The dominance of the flow effect suggests that flow information should be used directly to cancel this effect. This can be achieved by adding a flow measurement and controlling the *ratio* of the effluent and reagent flows based on the pH measurement (Section III.B.4). Despite the measurement lag and minor biases in flow measurement, this makes the concentration effect dominant. As only the two-tank configuration allows the concentration effect to be tolerated, we have two 12-m^3 tanks with ratio control of reagent flow at the inlet to the treatment system as the starting point for the search for an effective plant.

The results of the disturbance fraction analysis are summarized in Fig. 7. The worst-case disturbance (for the concentration effect) is indicated to be the maximum acid pulses occurring simultaneously, so this becomes the starting point for the worst-case design analysis.

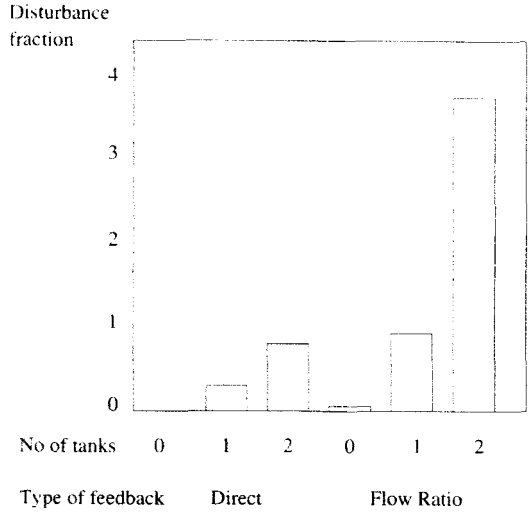


FIG. 7. Relative performance of alternative process/control configurations.

Identifying the worst-case condition was less clear-cut in this example than in the example presented in Section V.B. The worst-case disturbance combination varied with the process design and disturbance type considered. Using the optimization formulation from Section II.B.2 to identify the worst-case was therefore very useful. The preliminary control analysis is sufficient to eliminate most design options, indicating the need to use two tanks despite the fact that this implies unusually low minimum residence times (1 minute compared to the norm of at least 3 minutes) and indicating the need for ratio control of reagent addition.

b. Complete Design. PI controllers were used to implement feedback control. All controllers in the analysis below attempt cancelation of the strong-acid/strong-base characteristic, by passing the measured pH through the inverse of the nominal strong-acid/strong-base characteristic. As the screening analysis coupled with equipment constraints has defined the process parameters for the example considered, our optimization objective was to minimize control system cost plus the cost of excess reagent compared to the ideal delay-limited control case. The design parameters were the controller tuning and setpoints. The uncertain parameters were the disturbance characteristics and the varying measurement characteristics identified in Fig. 6. The disturbances were characterized by steps with switching times between 1 second and 10 minutes and variable step levels between zero and the maximum flow before and after the switch. The steady flow was allowed to vary between its minimum and maximum value. This gave a total of 11 uncertain parameters. In addition to the constraints specified in Fig. 6, it was required that the concentration recover to within $\pm 3 \times 10^{-4}$ N of the setpoint within 30 minutes to ensure well damped control.

The basic control scheme used for the perfect control analysis, a single in-line feedback loop, gives 10 times the allowable concentration variation at the exit of the second tank for the predicted worst case. Including feedforward reagent addition would be insufficient by itself to give the tenfold improvement required, due to a $\pm 20\%$ error in the estimated load (inferred from pH), so an additional in-line ratio feedback controller was added between the two tanks. As an additional actuator was therefore available at no extra cost, lead-lag feedforward from the load error at the first controller to the second controller actuator was added. To reduce feedforward dynamic mismatch, the lag was set to approximately the residence time of the first tank. The lead constant was added to the design parameters.

This scheme (shown in Fig. 8) was successful in meeting the performance requirements with the tuning parameters given below:

setpoint1	gain1	iat1	setpoint2	gain2	iat2	FF gain	lead constant
1.8	.061	23.3	11.31	.0082	2.	1.12	20.

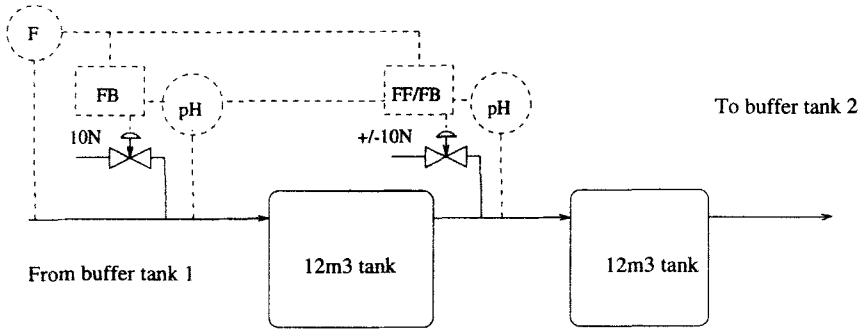


FIG. 8. Final process and control scheme.

Feedforward (FF) gain is given as a multiple of the gain based on the nominal steady-state model. All times are in seconds. Feedback controller gains are given in $m^3\ m^{-3}\ N^{-1}$.

The outer-approximation algorithm (Section II.A) took six iterations to identify this solution, with a projection factor, ϵ_p, of .05 on the disturbance amplitude. Both vertex and nonvertex constraint maximizers were identified, confirming the need to consider nonvertex maximizers. The variables that contributed nonvertex maximizers were the step switching times (several times) and the measurement lags (once). Robustness was verified with respect to all vertex combinations of uncertain values and a random selection of interior points (*ivert* = 1, $nrand_{max}$ = 1000).

The ideal delay-limited control optimum gives a reagent cost of about £3,500,000 p.a. based on a unit reagent cost of £50/m^3 and a typical operating condition of 400 m^3/h acid flow at $-.2$ *N*. The scheme in Fig. 8 with the parameters above gave an additional reagent cost of £950 p.a. compared to the ideal control case. The robust design obtained seems likely to be close to the economic optimum.

Increasing the projection factor to .1 reduced the number of iterations to generate this solution to three, with a 20% increase in the excess reagent cost. This shows the effectiveness of an increased projection factor in generating an approximate solution more rapidly. The progress of the algorithm on this problem is discussed in more detail in Appendix B, which illustrates the operation of the algorithm more clearly.

3. *Conclusions and Review*

A robust design was developed in the face of strong requirements on disturbance rejection, tight constraints on the equipment, and unusually large measurement delay. The preliminary control analysis was effective in narrowing down the possible designs for consideration. The worst-case design algorithm

allowed the development of a feasible design which was shown to be close to optimal. This example shows that the worst-case design algorithm developed is capable of solving realistic problems, confirms the existence of nonvertex constraint maximizers and indicates the use of a projection factor to trade off accuracy of solution and computation time to be effective.

D. $Ca(OH)_2$ Neutralization with Highly Variable Titration Characteristic (Central Effluent Plant)

1. Problem Definition

The objective of this case-study was to *evaluate* a proposed design for a treatment system for a central effluent plant dealing with waste streams from many sources.

The proposed design is a two-CSTR system with minimum residence time per tank of 24 minutes. Each tank is supplied with 5% w/w (1.45 *N*) hydrated lime ($Ca(OH)_2$) reagent controlled based on pH measured at the exit of the tank. Detailed tank design was such that a mixing delay of less than 10 seconds was expected. The neutralization tanks are followed by a flocculation tank with a minimum residence time of 10 minutes. There is an agitated level-controlled buffer tank upstream of the neutralization tanks, giving about 12 hours of concentration and flow equalization at maximum flow. The scheme to be evaluated is shown in Fig. 9.

The aim is to be able to control with setpoints between 7.5 and 8.5 pH $\pm$ .5 pH. The $\pm$.5 pH bounds are applied *at the flocculation tank.* This has the effect of reducing variability in the concentration (as compared to the second tank) and taking account of the discrepancy between average pH in the second controlled tank and in the flocculation tank due to carryover of reactive lime.

The presence of the buffer tank upstream means that process disturbances are greatly attenuated. Nevertheless it was desired to be able to make step flow changes of at least 50% of design flow, introducing significant disturbances.

The pH probe response is described by a lag of 5–30 seconds. Biases of up to $\pm$.25 pH are assumed to be possible on pH measurement.

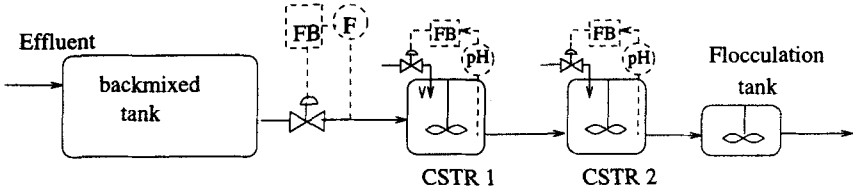


FIG. 9. Treatment system to be evaluated.

Twenty-four-hour flow-averaged titration curves collected over one month were taken as representative of the titration curve variability to be expected at the exit of the buffer tank. The curves are shown in Fig. 10 with pH plotted against net concentration in normals, pH 5 being taken as zero net concentration.

The key characteristics of titration data for design purposes are the slope of the titration curve (∂pH/$\partial cnet$) and the variability of this slope. The titration data are shown in Fig. 11 in the form of (∂pH/$\partial cnet$) as a function of pH. The titration curve slopes show a variability between 4:1 to 10:1 above pH 3; e.g., the slope between 7 and 9 varies 5:1 with values between 1000 pH/*N* and 200 pH/*N*. The maximum slope of 2000 pH/*N* was 1000 times less than the maximum slope for the strong-acid/strong-base titration characteristic, showing substantial buffering even in the worst case. The peak sensitivity in the target region (7–9 pH) was 1000 pH/*N*. Inlet concentration ranged from .08 *M* to .035 *M* acid (−.08 to −.035 *N*). For the analysis carried out, it is assumed that any titration

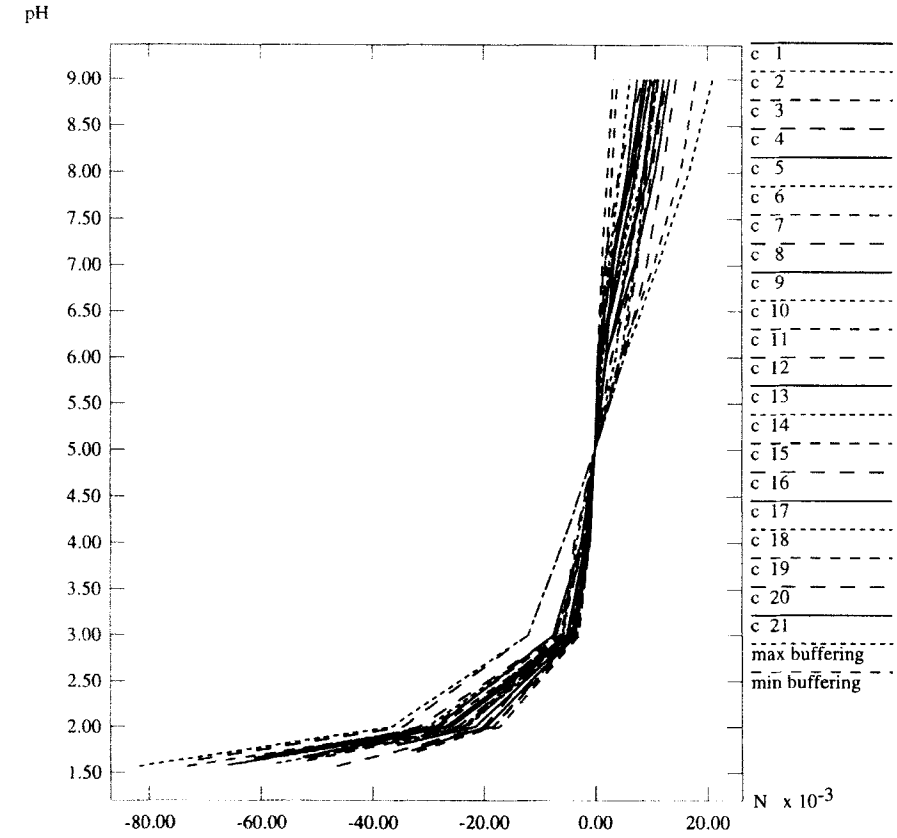


FIG. 10. Titration curves.

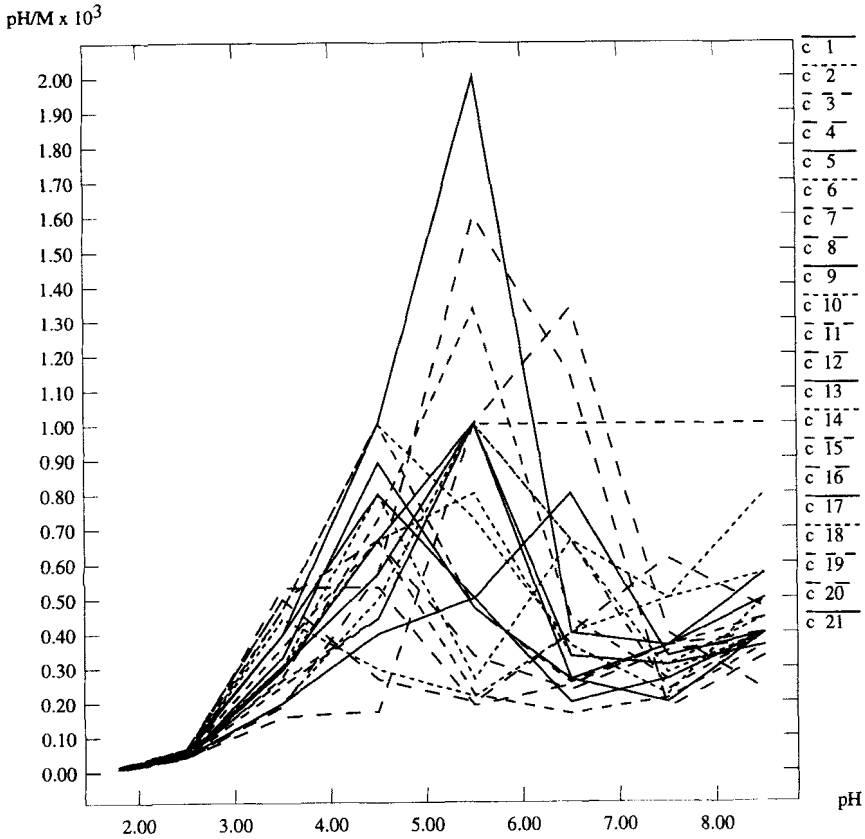


FIG. 11. Variation of titration curve slope with pH.

curve lying within the range of sensitivities observed may occur; i.e., there is no correlation between buffering at different pH levels. The "max buffering" and "min buffering" curves in Fig. 10 represent the extremes under this assumption and do not depart substantially from the observed curves.

The titration curves are given below in the form of ranges of concentration change spanning a given pH range. The reagent concentration required to reach pH 7 (*cnet* = 0) varies from about .035 *N* to .08 *N*, due to the inlet concentration variation noted above.

PH range	2–3	3–4	4–5	5–6
Concentration (*N*)	.015–.03	.0015–.007	.001–.005	.0005–.005

pH range	6–7	7–9
Concentration (*N*)	.0007–.005	.002–.01

Reagent dynamics were modeled using the standard model (Eq. 47):

$$\frac{\partial r_p}{\partial t} = -k_{\text{mix}}.\max\left(1, \frac{10}{r.k_{\text{mix}}}\right)(k_1 + k_2[\text{H}^+] + k_3 cweak) \ \mu/\text{s}$$

As the acid concentrations are moderate, k_2 and k_3 may be set to zero without introducing significant errors. k_1 was taken as .08–.12 to allow for modeling error and k_{mix} was taken as 1 as there was no reason to expect atypical mixing.

The particle size distribution was assumed to be given by

$$\phi(\ln(r_p)) = \frac{1}{1.15\sqrt{2\pi}} \exp(-((\ln(r_p) - 1.39)/1.15)^2/2), \qquad r_p < 250\ \mu \quad (57)$$

2. *Analysis of Design*

a. Preliminary Control Analysis. The maximum sensitivity of pH to concentration in the target pH range is 1000 pH/*N*. Allowing for the measurement bias of ±.25 pH the acceptable transient deviation is reduced from .5 to .25 pH corresponding to δ_{cnet} = .00025 *N*. The worst-case disturbance will be a step from 50 to 100% of flow corresponding to a maximum inlet concentration change, Δ_{cnet}, of up to .04 *N*. This disturbance is worse than a step from 0 to 50% of flow because although it has Δ_{cnet} = .08 *N*, this step has half the associated flowrate, which with 3 tanks in the disturbance propagation path implies 8 times the disturbance attenuation available in the worst case. The required disturbance attenuation at maximum flow is about .006 ($\delta_{cnet}/\Delta_{cnet}$).

The mixing delay is 10 seconds. The maximum measurement lag is 30 seconds. The reagent dynamics are complex, but the effective lag associated with the initial response is no more than 10 seconds (based on examination of time for 50% conversion of reagent using the model given above). The effective delay, calculated as $t_u/4$ (Section II.B.2), is therefore about 41 seconds. The estimated disturbance attenuation with PI control is 8×10^{-5} (Eq. 52). Transient control performance is not expected to be a problem for this system and a basic control scheme should be adequate.

b. Steady-State Analysis. To meet the performance requirements at steady-state, the reaction of $Ca(OH)_2$ downstream of the controlled tanks must not raise the pH by more than .5 pH in the flocculation tank. This assumes that the setpoint in the second tank is chosen .25 pH below the target value so that maximum margin for carryover is provided, given the measurement errors. This is equivalent to at least .625% of the reagent reacting downstream. Also, the reagent carrying over from the first tank to the second tank must not cause the

pH in the second tank to rise above its setpoint, preventing effective control. For a standard control scheme with fixed setpoints, it is necessary to establish a choice of setpoints for the first and second tanks that meets these requirements.

Shortcut analysis can be used to indicate whether a problem is likely. The time from 97 to 99% reagent conversion is about 1 minute based on the dynamic titration curves shown in Fig. 12 and the associated titration curve. Using this value of 1 minute as the equivalent reaction lag, τ_r, the carryover of reagent may be estimated by using Eq. (46).

The following table indicates carryover of reagent added to one tank from the tank exits downstream.

	tank 1	tank 2	tank 3
tank 1	.04	.0016	.00015
tank 2		.04	.0036

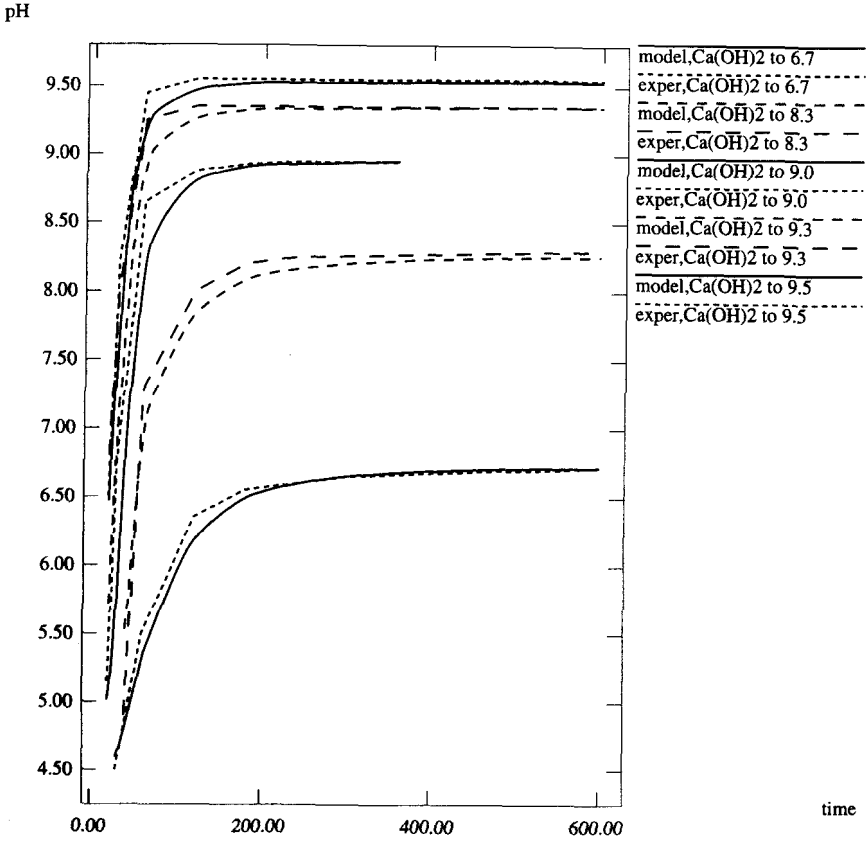


FIG. 12. Dynamic titration experiments with $Ca(OH)_2$ (1).

The shortcut calculation results indicate that there should be no problem with excessive carryover provided less than 11.6% [$.0016(1 - x) + .04x \leq .00625$] of the reagent is added to the second tank. The second tank should not be swamped by reagent from the first tank, provided at least 4% of the total reagent is required to make the shift between the first and second tank pH setpoints. This margin between setpoints allows all the reagent to be added to the first tank without the 4% carryover to the second tank causing the pH in the second tank to increase above the second tank setpoint. Choosing setpoints to keep the reagent addition to the second tank between 4 and 11.6% of the total reagent addition is not possible as the titration curve slope shows variability of at least 5:1 when most of the reagent has been added. As the shortcut calculation should be pessimistic (at least for carryovers of the order of 1%), it is likely that the system will work. Analysis of the full reagent model is required to confirm this.

Using the full model gives the carryover table below as $k1$ varies from .12 to .08.

	tank 1	tank 2	tank 3
tank 1	.0084–.0124	.00027–.00057	.000035–.0001
tank 2		.0084–.0124	.0006–.0012

It should be noted that, as k_2 and k_3 are assumed to give no contribution to the rate, the calculated carryovers are independent of the particular pH setpoints in the tanks. If the pH setpoint in the first CSTR is below about 3, carryover of reagent added to this tank should be slightly lower owing to acid acceleration of the reaction.

The analysis indicates that up to about 50% of the reagent can be added to the second tank without excessive downstream pH drift in the worst case. Provided at least 1.2% of the reagent is required to make the shift between the first and the second tank pH setpoints, the second tank should always be adding some reagent. Despite the titration curve variability of up to 10:1, there should be no difficulty in accommodating these requirements. Steady-state conversion is not likely to be problematic.

The more precise analysis shows the shortcut analysis to have been pessimistic—as expected—although the results tally within an order of magnitude. If the effective time constant had been based on the initial part of the batch titration responses (τ_r less than 10 seconds), the shortcut analysis would have been optimistic compared to the full analysis.

3. Conclusions and Review

The proposed design was shown to be capable of meeting the performance requirements quite comfortably without having to carry out a complete design.

The preliminary analysis (Section III.B.2) indicated sufficient performance margin to allow confidence that the design was adequate.

The plant described has now been commissioned and is working satisfactorily as predicted. pH differences between the second tank and the flocculation tank have been observed to be small (less than .3 pH) and may be dominated by measurement errors. These observations are consistent with the analysis above.

E. $Ca(OH)_2$ Neutralization of a Single Effluent Stream at High Intensity

1. Problem Definition

This section discusses work carried out on the design of an effluent treatment system for neutralization of a highly concentrated (-11.8 *N*) acid stream, predominantly composed of HCl.

The design problem is to produce a neutralization system that can maintain a pH of between 6.5 and 7.5 after all reagent has reacted. Effluent load is high and low suspended solids are required. A good-quality high-calcium hydrated lime was used as it was the cheapest reagent meeting these requirements and reacting reasonably rapidly.

The reagent was to be delivered as a 20% w/w slurry (standard practice), giving a concentration of about 5.9 *N*. The concentration of acid in the neutralized stream is therefore about -4 *N* due to dilution effects (reagent flow is twice the effluent flow near neutral).

It is desired to vary the flow of effluent readily. This requirement was captured by requiring that the system be able to start up with a step to 50% flow, followed by a step to 100% of flow one hour later and a step to 25% of flow after a further hour.

Cost analysis indicated that the total treatment system residence time should be kept below about 30 minutes (qualitative reduction in civil costs) while minimizing the number of CSTRs (low marginal cost with volume and substantial cost per unit). Additional cost reductions could be obtained by reducing the total residence time further.

Imperfect CSTR mixing is estimated to introduce 10 seconds dead time between reagent addition and measurement response.

pH measurement properties are characterized by a ± 0.25 pH bias and a first-order lag between 5 and 30 seconds. To avoid degraded measurement performance, it is required that the pH controller setpoints lie between 2 and 12 pH.

The titration data are tabulated in terms of the alkali concentration change required for a given pH change.

PH range	1–2	2–2.5	2.5–3	3–5	5–6	6–8
Concentration change (*N*)	.01	.006	.003	.001	.001	.008–.02

The only significant variability in the buffering was between 6 and 8 pH. The effluent to be treated has a slowly varying acid concentration around -11.8 N. Zero net acid/alkali concentration was taken as giving a pH of 7.

The dynamic titration data for this problem (Fig. 13) were consistent with the model discussed earlier, so the following model is taken to be appropriate (Eq. 47):

$$\frac{\partial r_p}{\partial t} = -k_{\text{mix}}.\max\left(1, \frac{10}{r.k_{\text{mix}}}\right)(k_1 + k_2[\text{H}^+] + k_3 cweak)\ \mu/\text{s}$$

There is uncertainty in the kinetics due to model mismatch, errors in fitting the model, and variations in mixing conditions, temperature, and particle size distribution. For this application, an overall uncertainty in the rate, $\partial r_p/\partial t$, of $\pm 40\%$ was included to allow for these factors, with $k_1 = 0.1$, $k_2 = 7.5$, and $k_3 = 1.5$.

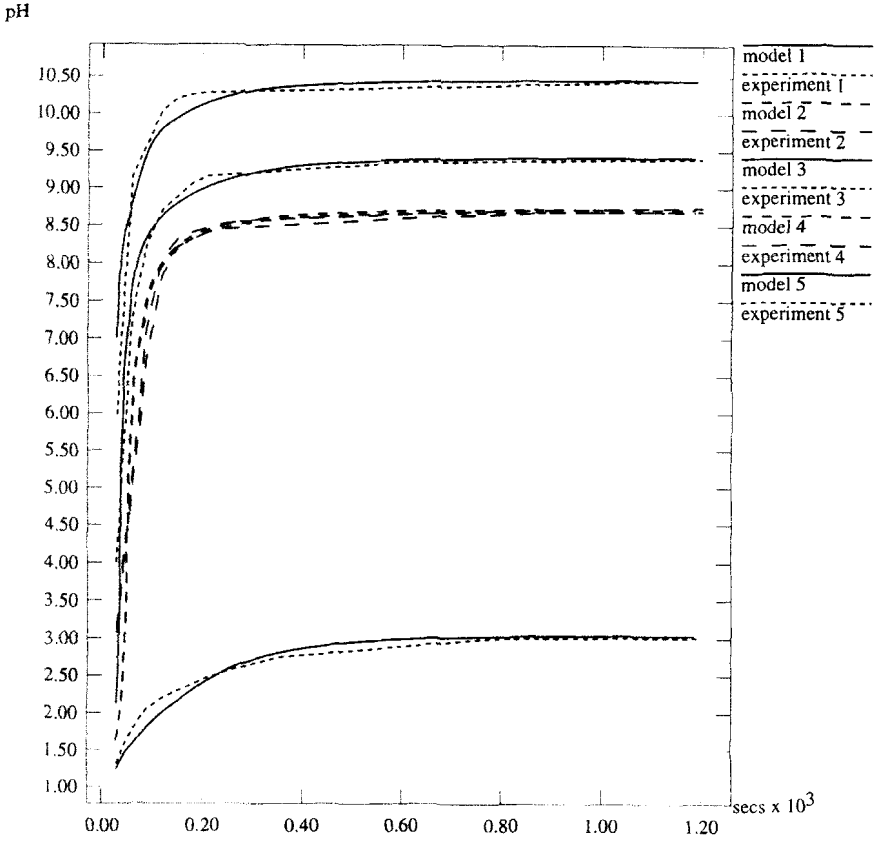


FIG. 13. Dynamic titration experiments with $Ca(OH)_2$ (2).

For CSTRs, k_{mix} was assumed to be 1, while for in-line mixers k_{mix} was assumed to be between 0.5 and 0.9. A surface reaction rate of 1.5 μ/s was assumed to allow for a possible change in reaction mechanism at high acid concentrations not ruled out by the experimental data. The particle size distribution was assumed to be given by Eq. (57):

$$\phi(\ln(r_p)) = \frac{1}{1.15\sqrt{2\pi}} \exp(-((\ln(r_p) - 1.39)/1.15)^2/2), \qquad r_p < 250\ \mu$$

(same reagent source as previous example).

2. Design

a. Preliminary Control Analysis. The maximum sensitivity of pH to concentration is 250 pH/*N*. Allowing for measurement bias, the acceptable transient deviation is about .25 pH, giving $\delta_{cnet} = .001\ N$. The worst-case disturbance will be the step from 50 to 100% of flow corresponding to a concentration change at the treatment system inlet, Δ_{cnet}, of about 2 *N*. The required disturbance attenuation, δ_c, at maximum flow is therefore about .0005.

Ideal delay-limited feedback analysis with n equal-sized tanks (Eq. 30) may be used to give a rigorous bound on achievable performance. This indicates that with simple control from CSTR pH to reagent addition, the minimum residence time required for one CSTR is 5 hours while the total residence time required for two CSTRs in series is 10 minutes. Combined feedforward/feedback control was also considered. The feedforward reagent addition was estimated to have a relative error made up of $\pm 2\%$ varying linearly with flow and $\pm 3.5\%$ varying slowly with time in the flow range of interest. This implies a potential benefit from feedforward of about a factor of 10 in disturbance rejection, requiring the feedback system to achieve a disturbance attenuation of 0.005. The minimum residence times for the combined control scheme are then 35 minutes for a single CSTR and 3 minutes for a two-CSTR system. A single CSTR system using combined feedforward/feedback control as above can therefore be eliminated on control grounds without the need for controller design and simulation.

Using Eq. (52) and an effective delay of about 41 seconds (as in Section V.D) to estimate the likely PI performance, indicates a total residence time of about 25 minutes for a two-CSTR system with combined feedforward and feedback and a total residence time of 75 minutes without feedforward. As the cost of the feedforward controller is small compared to the cost of an extra tank or the extra civil work to accommodate a residence time above 30 minutes, the scheme with feedforward control and two CSTRs seems likely to be the preferred option to achieve good control.

b. Steady-State Analysis. A second fundamental limit on system design is the steady-state reagent conversion. The pH downstream of the treatment system must lie between 6.5 and 7.5 for all values of the relevant uncertain parameters—measurement biases, rate of reaction, and the relationship between pH and concentration. If sufficient unreacted reagent carries over from the final stage, it will cause the pH to drift upward excessively downstream. Considering two CSTRs in series with the second tank using caustic reagent (NaOH) due to potential metering problems at low flows, the minimum total residence time is found to be 85 minutes. This value was generated using a steady-state worst-case optimization in which the downstream pH following complete conversion of reagent was required to be between 6.5 and 7.5 and the second tank was required to be adding reagent continuously at steady-state. The pH setpoint for the first tank was set to 2, the allowable value most favorable to reagent conversion. The second tank setpoint was set to 6.75 as this is the lowest value consistent with meeting the specification, given measurement errors. Flow was set to 100% of design rate. The CSTR residence times were optimized. The measurement biases, the uncertainties in the rate expression, and the uncertainty in the titration curve buffering were the relevant uncertain parameters. Ten CSTRs in series were used to model the plug-flow reactor.

The 85-minute residence time required for reagent conversion in a two-CSTR system greatly exceeds the desired maximum residence time of 30 minutes. It is therefore necessary to modify the treatment scheme so as to improve reagent conversion while retaining adequate control response. Previous design practice indicates adding a third CSTR, but this would result in an undesirable increase in cost. The approach adopted was to use a simple low-cost plug-flow reactor (PFR) before the first CSTR to achieve improved reagent conversion. A PFR has very poor feedback control characteristics as it increases the delay in the controller response considerably. It was therefore decided to use feedforward control to add most of the reagent to the PFR, while allowing feedback control from the first CSTR pH to add reagent directly to the CSTR. An additional feedback loop (feedforward trim) was included to monitor the ratio of reagents added to the PFR and CSTR and adjust the feedforward ratio so that the actual ratio was at its desired value. This minimizes the effect of feedforward calculation errors, allowing maximum benefit to be obtained. The maximum PFR residence time was estimated to be 2 minutes. The idea of using a PFR to assist in meeting steady-state requirements, which emerged in this case-study, has been incorporated into the design procedure (Section III.B.4). This scheme is shown in Fig. 14.

Repeating the steady-state worst-case design with the new scheme gave a minimum total residence time of 20.3 minutes, which is satisfactory and comparable to the requirement predicted from the control analysis. The PFR residence time went to its maximum value of 2 minutes. The fraction of total

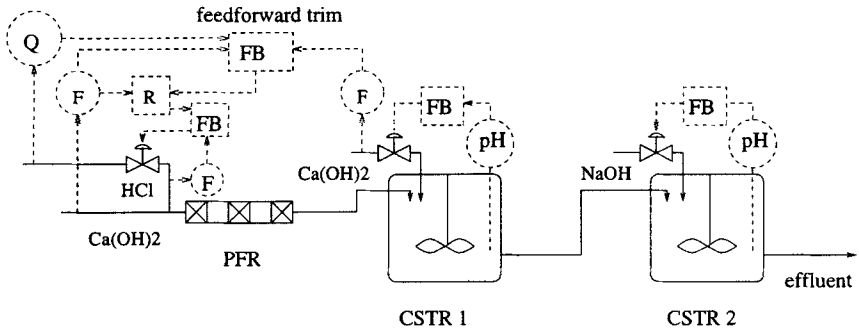


FIG. 14. Final process and control scheme.

$Ca(OH)_2$ added to the PFR was optimized along with the reactor volumes and found to have an optimum value of about 94.5%. Requiring the two tanks to be of equal size did not significantly alter the minimum residence time. Equal-sized tanks were therefore assumed for further analysis, as this has inherent benefits in terms of cost and should be optimal for control.

At this point in the design process, a structure that meets the steady-state requirements and is likely to meet the control requirements has been obtained.

Each steady-state worst-case optimization took about 15 minutes to run (on a SPARC 2). The worst-case combination was minimum rate of reaction with the measurement biases pushing the effective setpoints upward by .25 pH and the minimum buffering. This was a relatively simple worst-case optimization and no projection factor was used. This vertex worst case was in fact correctly chosen *a priori* based on physical considerations, and was promptly reidentified in the local phase of the constraint maximization. Most of the optimization time was spent verifying this choice of the worst case by vertex enumeration and random search.

c. Complete Design. At this point it becomes appropriate to carry out a full dynamic worst-case design, as a design structure has been identified that is likely to satisfy the design requirements and be close to optimal. The uncertainty in measurement dynamics and in the feedforward calculation was added to the sources of uncertainty considered when looking at the steady-state behavior. PI feedback controllers were assumed initially. The feedforward controller was started with a fixed ratio of 92% for the first increase in flow (feedback adjustment must be disabled when there is no flow), after which the feedback adjustment of the ratio was enabled to bring the fraction of $Ca(OH)_2$ added to the PFR to 95%. The first tank setpoint was set to 2 pH as before. The feedforward trim adjustment is predominantly integral, so the gain was set to a small value (.01 with a process gain of about 1) and only the integral action time was

adjusted. The tuning of the three feedback controllers and the setpoint for the second tank were optimized along with the CSTR volume and the PFR volume.

The input conditioning function tabulated below was used on the first CSTR pH controller, as the steep titration curve slope between 3 and 6 pH poses a risk of limit-cycling (see Section III.B.4):

pH in	1.	2.5	3.	6.	8.
"pH" out	1.	2.5	2.75	3.	5.

The main effect of this conditioning function is to prevent the rapid pH shift between 3 and 6 pH, giving rise to a sharp kick in the control response which could otherwise cause oscillations. The input conditioning is assumed to be implemented as a piecewise linear function, as this is the method generally available in control software.

The optimization attempted to minimize total reactor volume, subject to maintaining the downstream pH between 6.5 and 7.5 and to the second CSTR pH recovering to within .005 pH of the setpoint 30 minutes after a load change. The latter requirement was imposed to ensure reasonable damping.

The optimization generated a minimum residence time of 27.4 minutes with 2 minutes residence time in the PFR and 12.7 minutes in each CSTR. The controller tuning obtained was

	K_c	iat (s)	setpoint
CSTR 1	.13	1000	2 pH*
CSTR 2	.08	300	6.88 pH
FF trim	.01*	8.4	95%*

The variables indicated by an asterisk (*) were assigned *a priori*. The feedforward (FF) trim controller gain is dimensionless as both the input and the output are ratios of flows. The CSTR controller gains are in m^3/h reagent per m^3/h acid at maximum flow per pH. Some of the tuning parameters were on bounds, but the associated Lagrange multipliers did not indicate a significant incentive to rerun the optimization.

This design is likely to be close to the optimum design for this problem as control scheme modifications have limited potential to reduce the equipment size and there is no apparent means to eliminate any of the reactors without violating the constraint on residence time.

The worst-case dynamic optimization took 4 days on a SPARC 2, highlighting the importance of using efficient screening tools rather than plunging into the full design analysis for each alternative design considered. There were 10 uncertain parameters altogether; feedforward error was characterized by a fixed bias ($\pm 2\%$ of value) and biases at minimum and maximum flow ($\pm 3.5\%$ of value) with linear interpolation at intermediate flow, two pH measurement lags,

two pH measurement biases, the titration curve slope between 6 and 8 pH, the overall uncertainty in rate of reaction, and the uncertainty in the rate of reaction in the PFR. Vertex enumeration and up to 200 random points were allowed for the global phase of the search and $\rho_{min}^{max} = 1$ was used for the local search. As the model was slow to run (about 2 minutes for a function evaluation and 10 minutes for a gradient evaluation), careful *a priori* selection of worst-case combinations was used. Three worst cases were set *a priori* and an additional worst case was identified by vertex enumeration, although this did not greatly alter the solution. All the worst cases were vertex points. A projection factor of .05 on the maximum flow was used and only two iterations were needed. The Lagrange multipliers indicated two of the *a priori* worst cases to be dominant. In both of these the measurement lags were at their maximum value of 30 seconds. In one worst case the setpoints were biased high with low reaction rates (the steady-state worst case) and the feedforward error as a function of flow was set so that the effective feedforward ratio would increase with flow, tending to push the transient toward the high pH limit. In the other case the setpoints were biased low with high reaction rates, and the feedforward error as a function of flow was set so that the effective feedforward ratio would decrease with flow, tending to push the transient toward the low pH limit. The pH responses associated with these two worst cases are shown in Fig. 15, illustrating the diversity of response the treatment system is designed to accommodate.

The maximum transient deviation is actually somewhat less than expected from the earlier analysis (0.12 versus 0.25 pH). This largely reflects the fact that the feedforward trim allows a consistent benefit from feedforward of almost a factor of 20 compared to the factor of 10 assumed in the earlier analysis.

This example is at the limit of what could be handled on a SPARC 2, despite all the measures discussed in Section II.A to improve efficiency.

VI. Conclusions

The proposed approach to integrated design has been applied successfully to chemical wastewater neutralization systems. This approach has been shown to be effective in allowing the efficient development of robust low-cost designs. The results achieved show that an integrated attack on design problems using computer-aided process engineering tools is feasible.

A number of techniques have been presented that can be used in many other contexts. The delay-limited control analysis and the worst-case design method may be directly applied to a wide range of problems. Development of a suitable design procedure, models, and additional tools for any new problem area would require considerable effort. We consider that appropriate design procedures for

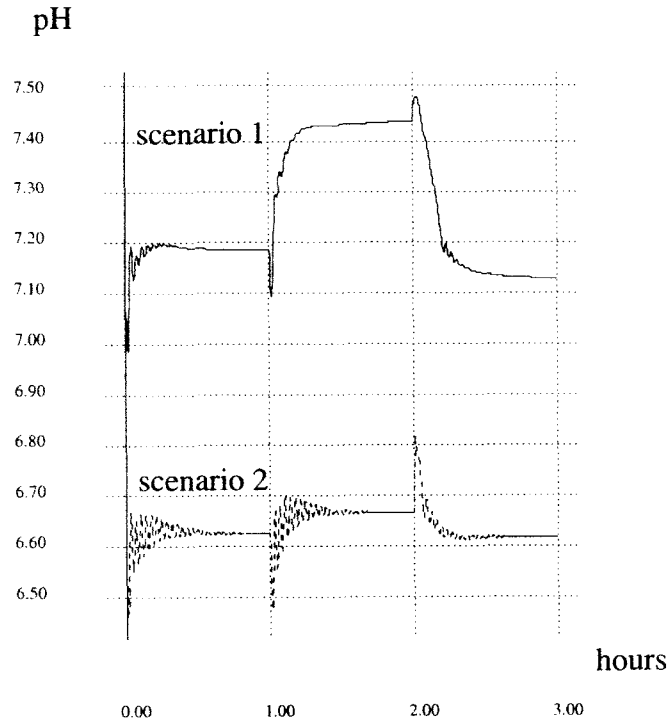


FIG. 15. Worst-case downstream pH responses.

many problems would retain the basic structure of preliminary control analysis, steady-state analysis, and full design with dynamic models and uncertainty. There are many opportunities for extension of the work described.

Dynamic optimization of realistic problems remains time-consuming and model reliability can be a major obstacle on large problems. Many industrial models require considerable hand-holding to get them to work at all. Improvements in computing power, solution techniques, and model-building tools are clearly needed for the widespread utilization of this technology. It is likely that no single approach to dynamic optimization will prove universally appropriate; thus, the software packages that intelligently select appropriate strategies are needed.

The algorithm presented for design with uncertainty seems quite effective, but a number of useful extensions can be identified.

1. The use of a single polyhedron for the set of uncertain parameters is restrictive and should be generalized to allow more complex restrictions on the allowable parameter set. While this is straightforward conceptually, its implementation needs careful attention to maintain the efficiency of the algorithm.

2. Logical constraints could be introduced to the vertex search to identify and eliminate redundant vertices and make the search more efficient.

The use of the design techniques in other problem areas should be explored. We are currently exploring the design of batch distillation columns and the design of distributed chemical plant.

Notation

Symbols

atol	absolute tolerance on integrator error control
c	vector of constraints
c_i	the *i*th constraint
c_x	concentration of acid/base x in normals
c_{sig}	the smallest significant constraint violation
cnet	the net acid/base concentration in normals (alkali = positive); taken as zero at the reference pH (pH_{ref})
cweak	the concentration of weak acids that dissociate below the reference pH (pH_{ref})
$\tilde{c}$	"majorized" constraints
d	disturbances
evtol	tolerance on location of time of discontinuity
$f(\dot{x}, x, z, \theta, t)$	residual function of DAE equations
$f(x)$	a general function of variable x
F	flexibility index
F_T	total flow
Fagit	recirculating flow generated by agitator
Fr	reagent flow
Fr_h	reagent flow corresponding to upper pH limit
Fr_l	reagent flow corresponding to lower pH limit
$g(x, z, \theta, t)$	residual function of algebraic equations
$g(\theta)$	gradient vector of objective, J, w.r.t. θ
H_t	height of tank
iat	controller integral action time
ivert	flag for whether vertex enumeration is used in constraint maximization
I_j^c	the set of i corresponding to inputs, u_i, connected by a controller to the jth output, y_j
I_j^s	the set of i corresponding to inputs, u_i, having a strong effect on the jth output, y_j
J	objective function
J_d	the set of j corresponding to outputs, y_j, from which the effect of the disturbance, d, can be compensated
k_1, k_2, k_3, k_{mix}	constants in solid alkali reaction equation
$L(x(\phi), z(\phi), \theta, \phi)$	equality interior point constraints
m	number of elements of constraint vector c
meq	number of equality constraint elements of constraint vector c
$M(x(\phi), z(\phi), \theta, \phi)$	inequality interior point constraints

neq	number of equations in the model
*nloc*1	number of local searches during local constraint maximization
*nloc*2	number of local searches during global constraint maximization
nmods	number of uncertain parameter vectors in V_0
nrand	number of random searches
nset	number of elements of V_0 that have been preset
np	number of parameters for which gradients are required
N_i	number of parameter sets in V_i
o	vector of n_o operating variables
optacc	specified optimization accuracy
O	set of operating variables
p	vector of n_p design variables
P	set of design variables
pH	the potential indicated by a pH probe
$\mathrm{pH}_{\mathrm{ref}}$	the reference pH used in defining *cnet* or *cweak*
$q(x, z, \theta, t)$	dynamic path constraints
r_p	particle radius
rtol	relative tolerance on integrator error control
simacc	specified simulation accuracy
sstol	tolerance on residual norm for initialization
t	time (seconds, unless otherwise specified)
t_d	delay between disturbance occurring and controller response reaching output
t_d'	delay between disturbance reaching output and controller response reaching output
$t_{d_{dv}}$	delay between disturbance and output
$t_{d_{uv}}$	delay between input and output
$t_{c_{\max}}$	the lag time constant associated with mixing
$t_{d_{\min}}$	the minimum deadtime associated with mixing
$t_{d_{\max}}$	the deadtime associated with mixing
t_f	final time of integration
t_u	the ultimate period of a control loop
v	vector of n_v varying or uncertain variables
$\bar{v}$	the normalized deviations in v
v_0	the nominal value of v
v_i	the ith element of the vector v
v^i	vector of uncertain parameters lying in the subset V^i
v^s	a vertex element of V
V	set of varying or uncertain variables
V_0	outer approximation to V at first iteration of worst-case design
V_i	outer approximation to V at ith iteration of worst-case design
V^i	ith subset of the uncertain parameter space
V_x	the volume of solution x
V_t	the tank volume
W_u	weighting matrix on u
W_y	weighting matrix on y
x	state variables
x^s	state variables evaluated while evaluating sensitivities
u	process inputs/manipulated variables/controller outputs

y	process outputs/measured variables
y_s	target values or setpoints for measured variables
z	algebraic variables
z^s	algebraic variables evaluated while evaluating sensitivities
α	scaling factor on full optimization step, δ_k
δ	flexibility index for a single value of ν
δ_a	estimated achievable disturbance attenuation
δ_c	required disturbance attenuation ($\delta_{cnet}/\Delta_{cnet}$)
δ_{cnet}	acceptable variation in *cnet* at treatment system exit
δ_f	the fraction of a disturbance that can be rejected ($\delta_f = \partial_c/\partial_a$)
δ_l	permissible fractional carryover of reagent
δ_k	full optimization step at kth iteration
$\delta(y)$	change in variable value when sensitivities are evaluated
Δ_{cnet}	variation of *cnet* at treatment system inlet
ϵ	a small number (defined more precisely where required)
ϵ_p	a projection factor to accelerate design with uncertainty
$\epsilon(y)$	precision error in variable y
$\rho^{\max}$	threshold for testing for potential change in active set
θ	optimization variables
Θ	set of optimization variables
λ	Lagrange multiplier
Λ	Relative gain array matrix
μ	penalty factor in SQP
τ	time constants (various)
τ_{PFR}	time in plug-flow reactor
τ_r	first-order reaction time constant
ϕ	times in Φ
Φ	discrete subset of time
Operators	
E	expected value of
$\| \|_n$	the nth norm
Δ	change in value
$[x]$	molar (g-moles/liter) concentration of component x
Superscripts	
*	optimum value
Subscripts	
max	maximum value
min	minimum value
Acronyms	
CSTR	continuous (flow) stirred-tank reactor
PFR	plug-flow reactor
PID	proportional-integral-derivative control
psd	particle-size distribution (fraction by mass)
Units	
M	concentration in moles per liter
N	concentration in normals [available OH^- (M); acids have negative normality]
w/w	concentration measured as weight of one component over total weight
μ	length in microns [micrometers (μm)]

APPENDIX A: WORST-CASE DESIGN ALGORITHM

1: Initialization
Define the sets P and V.

$\rho^{\max} = 1$, $nloc1 = 1$, $nloc2 = 1$, $nrand = 1$
Set $\rho_{\min}^{\max}$, $nloc1_{\max}$, $nloc2_{\max}$, $ivert$, $nrand_{\max}$
to control the maximum and minimum effort in the constraint maximization algorithm.

For each v_i define $pvert_i$

Set ϵ_p

Set the initial number of models, $nmods$, the number of models defined by the user, $nset$, and the values of v_j, $j = 1 \ldots nset$

If $nmods > nset$
  Maximize each constraint separately in a local vertex search
    Generate a random initial point
    Move to vertex maximizing the constraint based on a local linearization
    Repeat until search fails to increase constraint
Eliminate duplicate vertex maximizers
If the projected model gives an increased violation replace the original model with the projected model.
Select the best $nadd$ vertex maximizers.
($nadd$ = max($nmods - nset$, no of distinct vertex maximizers))
Reset $nmods$ to $nset + nadd$.

Proceed to multimodel design (2:)

2: Multimodel design problem

A: Identify the predicted active set of models to be used in the design.
   All models which have shown an active constraint in any of the three previous design iterations are predicted as active.
   Initial models are held for at least 3 iterations.
   Models which have been incorrectly dropped from the predicted active set are not dropped again.

B. Carry out design optimization subject to constraints from active models
   If the optimization fails to find a feasible solution STOP
   (Worst-case design may be infeasible)

C: Check violations for models not in the predicted active set.
If necessary update the predicted active set and return to B:

3: Constraint maximization

LOCAL SEARCH:
Local vertex search
lv1: Choose a random initial value of v
lv2: Update v using

$$v_i = v_i^l \text{ if } \frac{\partial c_{\max}}{\partial v_i} \leq 0$$

$$v_i = v_i^u \text{ if } \frac{\partial c_{\max}}{\partial v_i} > 0$$

until either $c_{\max}$ fails to increase as predicted or the search predicts a vertex already examined
lv3: For each $c_k < c_{\max}$ compute the maximum increase Δc_k in the constraint, based on the constraint gradients, and identify the corresponding vertex.
Compute the projected fractional change $\rho = \Delta c_k/(c_{\max} - c_k)$
If the maximum value of ρ is less than $\rho^{\max}$
or no new vertices are identified than local vertex search is complete
Else select the new vertex value of v giving the maximum value of ρ and go back to lv2:
If no constraint violations have been found set $\rho^{\max}$ to $\rho_{\min}^{\max}$
to increase the depth of search and go to back to lv3:

Carry out local searches into the interior of V from up to $nloc1$ vertices identified during the local vertex search as having ascent directions into V which did not lead to an increased value of $c_{\max}$ on the corresponding vertex.
If no new constraint violation is obtained then increment $nloc1$ until either $nloc1 = nloc1_{\max}$ or no more suitable vertices are left

If a constraint violation has not been found or
a global search has been used previously then

GLOBAL SEARCH:
Vertex search
If $ivert = 1$ test all the vertices not already examined
Carry out a local search from the largest new maximizer with an ascent direction into V

Multistart random search
Carry out random trials until the number of trials exceeds *nrand*
  Generate trial points using

$$v_i = v_i^l + (v_i^u - v_i^l) \min\left(1, \max\left(0, \frac{r_i}{1 - pvert_i} + .5\right)\right)$$

  where r_i is a uniform random variable $\in[-0.5, 0.5]$.
  and v_i^u and v_i^l are the upper and lower limits on the ith parameter in v
  If a random point is found increasing the maximum constraint violation then
    If no new constraint violation has been found and *nloc*2 local searches have been used in the global search increment *nloc*2 up to a maximum of $nloc2_{\max}$
    Carry out local search unless the number of local searches in the global search phase exceeds *nloc*2
  If no new constraint violation has been found and *nrand* random trials have been used increment *nrand* up to a maximum of $nrand_{\max}$

OUTER APPROXIMATION UPDATE:
If a constraint violation has been found then
  Update models for design, V_i
    If projected model gives increased constraint violation then add projected model else add the unprojected model to active set
    Proceed to multimodel design (2:)
else
  Optimum of worst-case design has been found, STOP

APPENDIX B: EXAMPLE OF PROGRESS OF ALGORITHM

The discussion below details the operation of the worst-case design algorithm for the example discussed in Section V.C.2.

The tuning parameters for the optimization were set as follows: $\rho_{\min}^{\max} = 1$, $nloc1^{\max} = 10$, $nloc2_{\max} = 10$, *ivert* = 1, $nrand_{\max} = 1000$. *pvert* was set to .99 on the initial and final levels of the two step disturbances and to .8 on all other variables. $\epsilon_p = 0.1$ is used in the example below, corresponding to increasing the concentration of the disturbance streams by 10%.

The nominal case was taken as a base flow of 400 m^3/h with no disturbances and no measurement bias. As the disturbances were infrequent, the economic objective for a given control structure was taken to be minimize the excess reagent used at steady state compared to the reagent required with ideal delay-

limited control. The ideal delay-limited controller had been able to operate with an exit pH setpoint of 11.3 (actual pH can drop to the limit of 11.2 due to measurement bias). Increasing this setpoint to accommodate disturbances or overneutralizing in the first tank results in increased reagent cost.

As the preliminary analysis suggested that the ratio control should make the concentration disturbance dominant over the flow disturbance, the initial scenarios were all based on the disturbances occurring simultaneously to give the maximum feed concentration variation. It was clearly necessary to consider both pulses starting at once and both pulses ending at once. Normally, the maximum measurement lag will be the worst case, so the measurement lags for the initial scenarios were set to 5 for the pH measurement on both tanks. The key pH measurement was that on the first tank, which was used for both feedback and feedforward control, so a bias of both .1 and −.1 pH units was considered. Finally, a bias of .1 was assumed on the last pH measurement as this reduces the effective pH at the treatment system exit, pushing the final setpoint up and increasing requirements for alkali reagent. This gave five scenarios in all (nominal case plus combinations of pulses starting/finishing bias positive/negative). This approach of choosing the initial scenarios based on a mixture of preliminary design analysis and qualitative reasoning is quite typical.

The progress of the design optimizations is shown below. The objective function increases monotonically as the optimization progresses and the outer approximation is refined by adding additional scenarios. All the additional scenarios added use the projection factor of 0.1.

iteration	setpoint1	gain1	iat1	setpoint2	gain2	iat2
0	11.32	.05	10	11.34	.05	10
1	1.89	.051	11.8	11.308	.0083	2
2	1.75	.073	41	11.31	.01	2
3	1.7	.07	47.4	11.312	.082	2

iteration	FF gain	lead constant	excess reagent cost
0	.8	30.	not feasible
1	1.16	19.4.	£628
2	1.19	18.4	£833
3	1.18	18	£1100

The third design passed the test for feasibility and was accepted as the optimum.

The first scenario added involved the low-concentration, 80 m^3/h, disturbance being applied for the first 10 minutes and then removed. The high-concentration, 40 m^3/h, disturbance was initially absent, but was applied after 344 seconds. The first pH probe was biased so that the actual pH was .1 higher than the measured pH. The second probe was biased in the opposite way. The first probe was at maximum lag, while the second probe had a lag of 1.6 seconds.

This scenario was identified as follows. The local vertex search failed to identify a constraint violation. A local search from a vertex identified in the local vertex search as not being a local maximum converged to a solution that differed from one of the initial scenarios only in that the second measurement lag was reduced to 1 second. This gave a constraint violation of 8×10^{-6}, which was just less than the optimization tolerance of 10^{-5}. No other suitable vertices from which to carry out a local search had been found. The global phase of the maximization was therefore activated. Vertex enumeration identified a constraint violation of 40×10^{-5}. This corresponded to the low-concentration disturbance being removed and the high-concentration disturbance being applied simultaneously, with the true pH being biased high on the first stage and low on the second stage and the first probe having maximum lag and the second probe having minimum lag. This vertex was not a local maximizer, so a local search was initiated using it as a starting point. This local search found the scenario described above, giving a constraint violation of 96×10^{-5}. As the number of random trials in the global phase, *nrand,* is initialized to 1 and a significant constraint violation had been found, the search stopped after just one random trial which failed to give an increased constraint violation. Setting ϵ_p to 0.1 for the scenario identified increased the integrated squared constraint violation so the projected scenario was added to the scenarios for the next design iteration.

The second constraint maximization did not find a violation until the vertex enumeration phase and gave a violation of 92×10^{-5}. The maximizing vertex occurred with minimum base flow, the low-concentration disturbance being applied after 10 minutes and the high-concentration disturbance being applied after 1 second, with both pH values biased low and probe lags at their maximum. Again this vertex was not a local maximizer. A local search increased the constraint violation to 97×10^{-5} by delaying the onset of the high-concentration disturbance to 51 seconds into the simulation. Again, only one unsuccessful random trial was made. Increasing ϵ_p to 0.1 again increased the violation, so the projected constraint was added to the scenarios for design.

On the third constraint maximization the vertex enumeration and subsequent local search failed to identify a significant violation. The random search therefore was allowed to run until either a violation was found or $nrand_{\max}$ trials were made. The allotted 1000 random trials failed to identify an increased constraint violation (from which a further local search would have been initiated), so the optimization terminated with the solution noted above.

In relation to the noise-control algorithm, it is of interest to note that the models corresponding to the scenarios for design were integrated with values of *simacc* ranging from 10^{-3} to 4×10^{-7}, illustrating the difficulty of optimizing *simacc* without an adaptation mechanism.

References

Asai, S., Konishi, Y., and Sasaki, Y. "Mass Transfer Between Fine Particles and Liquids in Agitated Vessels," *J. Chem. Eng. Jpn.* **21,** 107–112 (1988).

Biegler, L. T. "Strategies for Simultaneous Solution and Optimization of Differential-Algebraic Systems,"*in* "Proceedings of the Third International Conference on Foundations of Computer-Aided Process Design, Snowmass, CO, 1989," p. 155–179 (1990).

Brucato, A., Brucato, V., Rizzuti, L. and Sanfilippo, M. "Particle Dissolution Kinetics with Batch Stirred Vessels, *Int. Chem. Eng. Symp. Ser.* **121,** 327–341 (1990).

Chan, W. K., and Prince, R. G. H. "Heuristic Evolutionary Synthesis with Non-Sharp Separators," *in* "Third International Symposium on Process Systems Engineering, Sydney, 1988," p. 300–312 (1988).

Chemineer. "Kenics Static Mixers: KTEK Series," Tech. Rep. K-TEK-4. Chemineer, 1988.

Chen, C. L. "A Class of Successive Quadratic Programming Methods for Flowsheet Optimisation," Ph.D. Thesis, University of London (1988).

Cushnie, G. C. "Removal of Metals from Wastewater: Neutralisation and Precipitation." Noyes Publications, Park Ridge, NJ, 1984.

Dunker, A. M. "The Decoupled Direct Method for Calculating Sensitivity Coefficients in Chemical Kinetics," *J. Chem. Phys.* **81,** 2385–2393 (1984).

Gill, P. E., Murray, W., and Wright, M. H. "Practical Optimization." Academic Press, New York, 1981.

Grossmann, I. E., and Floudas, C. A. "Active Constraint Strategy for Flexibility Analysis in Chemical Processes," *Comput. Chem. Eng.* **6,** 675–693 (1987).

Grossmann, I. E., Haleman, K. P., and Swaney, R. E. "Optimization Strategies for Flexible Chemical Processes," *Comput. Chem. Eng.* **7,** 439–462 (1983).

Gustafsson, T. K. "Calculation of the pH Value of a Mixture of Solutions," *Chem. Eng. Sci.* **37,** 1419–1421 (1982).

Gustafsson, T. K., and Waller, K. V. "Dynamic Modelling and Reaction Invariant Control of pH," *Chem. Eng. Sci.* **38,** 389–398 (1983).

Haslam, R. T., Adams, F. W., and Kean, R. H. "The Rate of Solution and Availability of Commerical Limes," *Ind. Eng. Chem.* **18,** 19–23 (1926).

Hershkovitch, H. Z., McAvoy, T. J., and Liapis, A. L. "Dynamic Modelling of pH Electrodes," *Can. J. Chem. Eng.* **56,** 346–353 (1978).

Holt, B. R., and Morari, M. "Design of Resilient Processing Plants. V. The Effect of Deadtime on Dynamic Resilience," *Chem. Eng. Sci.* **40,** 1229–1237 (1985).

Hoyle, D. "Designing for pH Control," *Chem. Eng.* **83,** 121–126 (1976).

Johansson, G., and Norberg, K. "Dynamic Response of the Glass Electrode," *J. Electronal. Chem.* **18,** 239–250 (1968).

Leis, J. R., and Kramer, M. A. "Sensitivity Analysis of Systems of Differential and Algebraic Equations," *Comput. Chem. Eng.* **9,** 93–96 (1985).

Levenspiel, O. "Chemical Reaction Engineering." Wiley, New York, 1972.

Levins, D. M., and Glastonbury, J. R. "Particle-Liquid Hydrodynamics and Mass Transfer in a Stirred Vessel. II. Mass Transfer," *Trans. Inst. Chem. Eng.* **50,** 132–146 (1972).

Luyben, W. L. "Process Modelling, Simulation and Control for Chemical Engineers." McGraw-Hill, New York, 1990.

Mayne, D. Q., Michalska, H., and Polak, E. "An Efficient Outer Approximations Algorithm for Solving Infinite Sets of Inequalities," *in* "Proceedings of the 29th Conference of Decision and Control, 1990," p. 960–955 (1990).

McAvoy, T. J. "Dynamic Modelling of pH in Aqueous Systems," *in* "Proceedings of AIChE Workshop in Industrial Process Control, 1979," p. 35–39 (1979).

McMillan, G. K. "Tuning and Control Loop Performance." Instrument Society of American Publications, 1983.

McMillan, G. K. "pH Control." Instrument Society of America Publications, 1984.

Mizsey, P., and Fonyó, Z. "Toward a More Realistic Overall Process Synthesis—The Combined Approach," *Comput. Chem. Eng.* **14,** 1213–1236 (1990).

Moore, R. L. "Neutralisation of Waste Water by PH Control." Instrument Society of America Publications, 1978.

Morari, M. "Effect of Design on the Controllability of Chemical Plants," *in* "Interactions between Process Design and Process Control (J. D. Perkins, ed.), p. 3–16. Pergamon, Oxford, 1992.

Perkins, J. D., and Wong, M. P. F. "Assessing Controllability of Chemical Plants," *Chem. Eng. Res. Des.* **63,** 358–362 (1985).

Polak, E. "An Implementable Algorithm for the Optimal Design Centering, Tolerancing, and Tuning Problem," *J. Optim. Theory Appl.* **37,** 45–67 (1982).

Polak, E., and Stimler, D. M. "Majorisation: A Computational Complexity Reduction Technique in Control System Design," *IEEE Trans. Autom. Control* **AC-31,** 1010–1021 (1988).

Prett, D. M., and Garcia, C. E. "Fundamental Process Control." Butterworth, London, 1988.

Prett, D. M., Garcia, C. E., and Ramaker, B. L. "The Second Shell Process Control Workshop." Butterworth, London, 1990.

Sargent, R. W. H., and Sullivan, G. R. "Development of Feed Changeover Policies for Refinery Distillation Columns," *Ind. Eng. Process Dis. Dev.* **18,** 113 (1979).

Shinskey, F. G. "pH and pIon Control," Wiley Environ. Sci. Technol. Ser. Wiley, New York, 1973.

Straub, D. A., and Grossmann, I. E. "Integrated Stochastic Metric of Flexibility for Systems with Discrete State and Continuous Parameter Uncertainties," *Comput. Chem. Eng.* **14,** 967–985 (1990).

Swaney, R. E., and Grossmann, I. E. "An Index for Operational Flexibility in Chemical Process Design. Part I. Formulation and Theory," *AIChE J.* **31,** 621–630 (1985a).

Swaney, R. E., and Grossmann, I. E. "An Index for Operational Flexibility in Chemical Process Design. Part II. Computatonal Algorithms," *AIChE J.* **31,** 631–641 (1985b).

Tits, A. L. "On the Optimal Design Centering, Tolerancing, and Tuning Problem," *J. Optim. Theory Appl.* **45,** 487–494 (1985).

Vasantharajan, S., and Biegler, L. T. "Simultaneous Strategies for Optimisation of Differential-Algebraic Systems with Enforcement of Error Criteria," *Comput. Chem. Eng.* **14,** 1083–1100 (1990).

Velguth, F. W., and Anderson, R. C. "Determination of Minimum Capacities for Control Applications." *ISA J.* **1,** 33–38 (1954).

Walsh, S. P. "Integrated Design of Chemical Waste Water Treatment Systems," PhD Thesis, Imperial College, London (1993).

Yagi, H., Motouchi, T., and Hikita, H., "Mass Transfer from Fine Particles in a Stirred Vessel," *Ind. Eng. Chem. Process Des. Dev.* **23,** 145–150 (1984).