INTEGRATED DESIGN OF CHEMICAL WASTE WATER TREATMENT SYSTEMS

by

Stephen Walsh, MEng
May, 1993

A thesis submitted for the degree of Doctor of Philosophy of the University of London
and for the Diploma of Membership of the Imperial College

Department of Chemical Engineering and Chemical Technology
Imperial College of Science, Technology and Medicine
London, SW7

Abstract

Current practice in chemical waste water treatment is reviewed. Strengths and weaknesses are highlighted and some useful cost data extracted.

Tools, models and procedures for the integrated design of systems for the chemical treatment of waste water are developed and tested.

The key design tools developed are

1. a robust dynamic optimisation package;
2. a screening method for evaluating the effect of delays on achievable performance;
3. an optimisation method for dealing with parametric uncertainty.

Validated models are developed for the most common solid reagents, $Ca(OH)_2$ and $CaCO_3$. Methods of representing the steady state pH characteristics of aqueous solutions are examined and their advantages and limitations identified. A model for the oxidation of cyanide with hypochlorite is developed. Modelling of CSTR mixing is examined using computational fluid dynamics and results in the literature. Models of pH probe characteristics are critically reviewed. The difficulties in modelling precipitation behaviour are reviewed. Default model parameters and uncertainty descriptions are provided.

A design procedure is presented which sets out the data requirements and appropriate use of the tools and models for various levels of design, ranging from generating lower bounds on cost to detailed system specification. Rules are provided to assist development of a design.

The design framework developed is validated by the application of the techniques to realistic industrial design problems tackled in collaboration with ICI. The methods used are compared to previously published design guidelines.

The results of the work are reviewed and some needs for further work identified.

Acknowledgements

I wish to thank my supervisor, Prof. John Perkins, for his valuable guidance and support throughout this project. I would also like to thank my colleagues for their support and advice over the course of this work. My thanks to the many engineers in ICI who have helped me during the project. Finally, my thanks to my family and friends for being there when I needed them.

Contents

List of symbols and special terms **7**

List of Figures **13**

1 Introduction **14**
1.1 Objectives of project 14
1.2 Design requirements 15
1.2.1 System characteristics 15
1.2.2 Project context 16
1.3 Overview of thesis 18

2 Review of current practice in chemical waste water treatment system design **20**
2.1 Process characteristics 20
2.2 Design options 22
2.2.1 Process options 22
2.2.2 Control options 25
2.3 Published design guides 27
2.3.1 Process options 28
2.3.2 Control options 38
2.4 Trends in current practice 40
2.4.1 Adaptive control 40
2.4.2 Nonlinear control 41
2.5 Conclusions 42

3 Design tools **43**
3.1 Optimisation with continuous variables 43
3.1.1 Local solution 44
3.1.2 Global solution 49
3.2 Optimisation of dynamic systems 50
3.2.1 Choice of method 51
3.2.2 Robust and efficient feasible path dynamic optimisation 54
3.3 Design with uncertainty 65
3.3.1 Review 65
3.3.2 New algorithm 76
3.4 Screening tools 82
3.4.1 Review 82
3.4.2 A screening test for disturbance rejection in nonlinear processes subject to dead time 86
3.5 Summary of design software 95

4 Modelling **96**
4.1 Steady-state pH characteristics 97
4.1.1 Thermodynamic equilibrium models 97
4.1.2 Titration curve models 102
4.1.3 Examples 105

4.2 Characteristics of solid alkali reagents 111
  4.2.1 Mass transfer 112
  4.2.2 Reaction of $Ca(OH)_2$ 119
  4.2.3 Reaction of $CaCO_3$ 125
  4.2.4 Other solid alkalis 129
4.3 Redox reactions 129
4.4 Measurement response 132
4.5 Mixing 134
4.6 Precipitation reactions 139
4.7 Previous pH modelling environments 140
4.8 Summary 140

5 Design procedures 142
5.1 Making discrete design decisions 142
5.2 Exploration of a generic neutralisation control problem 143
  5.2.1 Problem definition 144
  5.2.2 Why pH control is difficult 145
  5.2.3 Examination of bounds from delay-limited control analysis 149
  5.2.4 Optimal sizing of series CSTRs for pH control 152
  5.2.5 Towards ideal control 154
  5.2.6 Summary 156
5.3 Neutralisation of waste water 156
  5.3.1 Problem definition 156
  5.3.2 Preliminary design 159
  5.3.3 Complete design 163
  5.3.4 Rules for modifying design 164
5.4 Redox reactions in waste water 166

6 Design examples 167
6.1 Preliminary design of a central effluent plant 167
6.2 pH control of several strong acid streams with NaOH/HCl 170
  6.2.1 Problem definition 170
  6.2.2 Design 171
  6.2.3 Conclusions and review 175
6.3 $Ca(OH)_2$ neutralisation with highly variable titration characteristic (Central Effluent Plant) 175
  6.3.1 Problem definition 175
  6.3.2 Analysis of design 178
  6.3.3 Conclusions and review 180
6.4 $Ca(OH)_2$ neutralisation of a single effluent stream at high intensity 181
  6.4.1 Problem definition 181
  6.4.2 Design 182
  6.4.3 Conclusions and review 188
6.5 $CaCO_3$ neutralisation for a small number of effluent streams 188
  6.5.1 Problem definition 188
  6.5.2 Design 191
  6.5.3 Conclusions and review 194
6.6 Oxidation of cyanide with hypochlorite 194

6.7 Conclusions and review 195

7 Conclusions and future research **196**

7.1 Results and conclusions 196

7.2 Suggestions for future work 197

References **198**

A Solution of DAEs to generate gradients using the sensitivity approach **208**

B Worst-case design algorithm **211**

List of symbols and special terms

Symbols

a	effective ionic radius
a_x	the activity of a component x ($a_x = \gamma_x m_x$)
A	Debye-Hückel constant
$atol$	absolute tolerance on integrator error control
$\hat{A}(\theta)$	matrix with the transposed gradient vectors of the active constraints as its rows.
B	a positive definite approximation to the Hessian matrix, H
c	vector of constraints
c_i	the *ith* 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 which will dissociate below the reference pH (pH_{ref})
$\tilde{c}$	"majorised" constraints
C	coefficient of ionic strength effect in Hückel equation
d	disturbances
D_{85}	diameter below which 85% of particles by mass lie
D_a	diameter of agitator
D_p	diameter of particle
D_{pipe}	diameter of pipe
D_t	diameter of tank
$\mathcal{D}$	diffusion coefficient
e_i	unit basis vector with the *ith* element equal to 1
$evtol$	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
$G(s)$	process transfer function
$g(\theta)$	gradient vector of objective, J, w.r.t. θ
H_t	height of tank
$H(\theta, \lambda)$	Hessian of the Lagrangian
iat	controller integral action time
ie	integrated error in PI controller
$ivert$	flag for whether vertex enumeration is used in constraint maximisation
I	the ionic strength ($I = \frac{1}{2}\sum_i z_i^2 m_i$, where z_i is the charge on the ith ion and m_i is its molal concentration)
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_*	equilibrium constant
k_1, k_2, k_3, k_{mix}	constants in solid alkali reaction equation
k_m	mass transfer coefficient
k_w	dissociation constant of water
K_c	controller gain
$L(\theta, \lambda)$	$J(\theta) + \lambda^T c(\theta)$, the Lagrangian
$L(x(\phi), z(\phi), \theta, \phi)$	equality interior point constraints
m	number of elements of constraint vector c
m_x	the molal (gmole/kg) concentration of component x
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
$nloc1$	number of local searches during local constraint maximisation
$nloc2$	number of local searches during global constraint maximisation
$nmods$	number of uncertain parameter vectors in V_0

$nrand$	number of random searches
$nset$	number of elements of V_0 which have been preset
np	number of parameters for which gradients are required
N_i	number of parameter sets in V_i
N_{js}	agitator speed at which particles are just suspended in a tank
o	vector of n_o operating variables
$optacc$	specified optimisation accuracy
O	set of operating variables
p	vector of n_p design variables
P	set of design variables
P_a	pressure in atmospheres
pH	the potential indicated by a pH probe — by convention $-log_{10}(a_{H^+})$
pH$_{ref}$	the reference pH used in defining *cnet* or *cweak*
pk_*	$-log_{10}(k_*)$
$q(x, z, \theta, t)$	dynamic path constraints
q_x	the quantity of component x in normals×litres
r_p	particle radius
r	distance from particle centre
r'	transformed distance from particle surface $(1\text{-}r_p/r)$
$rtol$	relative tolerance on integrator error control
R	selection matrix for initialisation
Re	Reynolds number
$\mathcal{R}$	the rate of reaction $\frac{\partial[x]}{\partial t} = \mathcal{R}$
s	Laplace operator
$simacc$	specified simulation accuracy
$sstol$	tolerance on residual norm for initialisation
S_x	the matrix $\frac{\partial x}{\partial \theta}$
Sc	Schmidt number
Sh	Sherwood number
t	time (seconds unless otherwise specified)
t_{99}	time to 99% homogeneity
t_d	delay between disturbance occurring and controller response reaching output
t'_d	delay between disturbance reaching output and controller response reaching output

$t_{d_{eff}}$	effective delay used by McMillan
$t_{d_{d,y}}$	delay between disturbance and output
$t_{d_{u,y}}$	delay between input and output
$t_{c_{mix}}$	the lag time constant associated with mixing
$t_{d_{min}}$	the minimum deadtime associated with mixing
$t_{d_{mix}}$	the deadtime associated with mixing
t_f	final time of integration
t_u	the ultimate period of a control loop
$tc(.)$	a titration curve function
T	temperature in Kelvin
v	vector of n_v varying or uncertain variables
$\tilde{v}$	the normalised deviations in v
$\bar{v}$	the mean velocity for Levins and Glastonbury's correlation
v_0	the nominal value of v
v_e	the effective velocity due to turbulence
v_i	the *ith* element of the vector v
v^i	vector of uncertain parameters lying in the subset V^i
v_{js}	the fluid velocity at which particles are just suspended in a pipe
v_j	the *jth* vector of uncertain variables in V_i
v_s	the fluctuating slip velocity
v_T	the terminal velocity
v^v	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	a reaction invariant
W_i	*ith* component of the reaction invariant
W_u	weighting matrix on u
W_y	weighting matrix on y
x	state variables
x^s	state variables evaluated while evaluating sensitivities
x'	delayed state variables

X_{pH}	$10^{pH-pk_w} - 10^{-pH}$
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
z_i	charge on *ith* ion
$Z(\theta)$	matrix whose columns form a basis set for the vector space defined by $\hat{A}p = 0$
Z_t	clearance of agitator from tank base
α	scaling factor on full optimisation step, δ_k
β	a Hückel equation constant
γ_x	the activity coefficient of x ($a_x = \gamma_x m_x$)
δ	flexibility index for a single value of v
δ_a	estimated achievable disturbance attenuation
δ_c	control precision ($\delta_{cnet}/\Delta_{cnet}$)
δ_{cnet}	acceptable variation in *cnet* at treatment system exit
δ_f	the fraction of a disturbance which can be rejected ($\delta_f = \frac{\delta_c}{\delta_a}$)
δ_l	load precision
δ_k	full optimisation step at *kth* iteration
$\delta(y)$	change in variable value when sensitivities are evaluated
Δ_{cnet}	variation of *cnet* at treatment system inlet
ΔpH	required change in pH moving towards neutral
ϵ_m	the power dissipated per unit mass
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
ρ_l	liquid density
ρ_m	solid-liquid mixture (slurry) density
ρ^{max}	threshold for testing for potential change in active set
ρ_p	particle density
θ	optimisation variables
Θ	set of optimisation variables
λ	Lagrange multiplier
Λ	Relative gain array matrix

μ	penalty factor in SQP
μ_l	liquid viscosity (dynamic)
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
∇	gradient operator
Δ	change in value
$[x]$	molar (gmoles/litre) concentration of component x

Superscripts

T	transpose
$*$	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
p.s.d.	particle size distribution (fraction by mass)
STR	continuous (flow) stirred tank reactor

Units

M	concentration in molar
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

List of Figures

2.1 Strong-acid/ strong-base titration curve 21

3.1 Ideal and PI control 88

4.1 Room temperature titration data for concentrated acid 106
4.2 75 °C titration data for concentrated acid 108
4.3 Predicted and measured titration curves for mixtures 110
4.4 Comparison of mass–transfer correlations 115
4.5 Batch experiments with $Ca(OH)_2$ (1) 122
4.6 Batch experiments with $Ca(OH)_2$ (2) 123
4.7 Batch experiments with $Ca(OH)_2$ (3) 124
4.8 Batch experiments with $CaCO_3$ 128
4.9 Predicted time response using a CFD model 135
4.10 Predicted and measured residence time distributions 136

5.1 Generic pH control problem 145

6.1 Design problem for preliminary design 167
6.2 Problem definition 170
6.3 Relative performance of alternative process/control configurations 173
6.4 Final process and control scheme 174
6.5 Treatment system to be evaluated 176
6.6 Titration curves 177
6.7 Variation of titration curve slope with pH 177
6.8 Final process and control scheme 184
6.9 Worst-case downstream pH responses 187
6.10 Typical response of cascade control scheme to acid disturbance 193
6.11 Worst response of cascade control scheme to acid disturbance 194

Chapter 1

Introduction

This chapter sets out the overall objectives of the work presented in this thesis. The design requirements for chemical waste water treatment systems are presented. These requirements are examined in terms of the characteristics of the systems of interest and in terms of the project context of the design. Finally, an overview of the thesis is presented.

1.1 Objectives of project

The scope of this project is to develop, implement and test an integrated design framework for the design of systems for the chemical treatment of waste water.

Waste water treatment encompasses a very broad range of technologies including neutralisation (pH adjustment), reduction/oxidation, precipitation, solvent extraction, clarification, sludge dewatering and disposal, and biological treatment. This project focuses on the chemical treatment of wastewater which encompasses neutralisation, reduction/oxidation and precipitation. All of these are normally accomplished by mixing chemical reagents with the waste water so as to obtain the required chemical compositions to allow the water to be discharged or reused. The selection of problems to be considered is based on selecting a broad enough area to stimulate general results on design techniques and have strong industrial relevance while not creating several virtually independent projects by including problems involving fundamentally different technologies.

The treatment systems selected are primarily of interest in an industrial, as opposed to municipal, context. The industrial context is therefore used in considering the overall design process, though much of the analysis would be relevant to municipal systems as well.

Waste water treatment systems (plant and controllers) are required to achieve extremely precise regulation of exit chemical compositions, often against large, rapid variations in flows or compositions of effluents to be treated. Meeting this requirement generates frequent and recurring problems in design, control and operation of the treatment system which are likely to intensify in the face of current trends in public awareness and legislation.

There is a recognised need to develop a more systematic design approach in this area to assist in

1. routing effluents to optimise plant/site performance;
2. choice of reagent(s);
3. sizing of units and configuration of plant;
4. selection of control structure;
5. control system tuning and choice of operating points.

Choices in each of these areas interact strongly. In particular inappropriate process design can make good control of exit compositions impossible regardless of control system design. Variability in process characteristics and uncertainty in the models used to predict performance have a major impact on the evaluation of possible designs. The design issues to be addressed therefore require an integrated design approach.

The integrated design objective is to obtain minimum cost designs while ensuring that safety and performance constraints are satisfied for all possible plant parameters and disturbances. Research in integrated design aims to provide tools dealing with all aspects of the above objective in a well-coordinated way, so as to generate good designs efficiently.

1.2 Design requirements

1.2.1 System characteristics

Chemical wastewater treatment systems are made up of a reagent dosing system, a mixing system to achieve the required blending and reaction, a measurement system to monitor performance and a control system to adjust the reagent addition to meet performance objectives. The major characteristics affecting performance of chemical wastewater treatment systems are given below.

1. Highly dynamic operation with disturbances only partially measurable.
2. Variability and nonlinearity in the steady state relationship between manipulated variables and measured variables (typically pH or redox potential).
3. Complex (nonlinear) reaction kinetics.
4. Measurements subject to variable dynamic response and to bias errors.
5. Delays (dead time) in the dynamics of the mixing operations.

6. Legally enforced constraints on the exit measurement value or chemical composition which are increasingly being treated as hard constraints which must be satisfied 100% of the time. There may be additional hard constraints to be satisfied for safety reasons, e.g. to avoid generating toxic gases during treatment.
7. Operating demands which vary over time due to production changes in the plants producing the effluent requiring treatment.

These characteristics of chemical waste water treatment systems have some clear implications for the design requirements and appropriate design techniques.

A set of design tools for this problem must be able to handle all the characteristics identified above. That is, they must be able to deal with nonlinearity, constraints on dynamic response, delays and uncertainty. The tools, together with appropriate models, must allow the performance achieved as a result of the design choices to be predicted. Given tools to predict performance, optimisation methods *may* be able to determine good values for the variables more efficiently than a competent design engineer. Tools should be provided to automate the choice of those design variables for which suitable optimisation methods can be provided.

To evaluate performance accurately nonlinear dynamic models must be used with due consideration being given to the uncertainty of this model and the variability of the process. Key aspects of this model are the relations defining the nonlinear steady state characteristics, the mixing dynamics, the reaction kinetics and the measurement dynamics. The models must predict performance with acceptable accuracy while making realistic demands on the availability of process data.

Developing the required tools and models is the central focus of the work presented in this thesis.

1.2.2 Project context

The design of waste water treatment 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 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 treatment systems. A useful design method for waste water treatment 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.

Decisions in the broader process affect the end-of-pipe design in several ways.

1. Total flow of waste-water: The volume of the treatment system required to provide the residence time needed for reagent conversion or for attenuation of disturbances in effluent concentration is proportional to total flow. Additionally, reducing the total load (molar flow) of a waste component is often much easier when the component is at high concentration, e.g. in precipitation of trace impurities. Reduction in total flow may have a major impact on cost of the end-of-pipe treatment and may be achieved by segregating uncontaminated water streams, minimising use of water in washing equipment and by process modification.

2. Production of waste components: Process design may eliminate certain wastes completely either by using a different process route or physically separating and recycling some components.

3. Segregation of streams: Decisions as to which streams are treated together may have a crucial impact on treatment feasibility and cost. Certain components if allowed to combine become qualitatively more difficult to treat. In particular metals should where possible be separated from complexing agents. Some combinations of streams may generate hazards, e.g. acid conditions may lead to release of ammonia or chlorine gases. Other combinations may be beneficial, e.g. merging acid and alkali streams from different plants on a site may reduce the reagent cost substantially. Combination of weak acid/base streams with strong acid/base streams may make control significantly easier.

4. Disturbance reduction: The worst disturbances to the effluent treatment system usually do not arise from the normal operation of continuous plant. Batch operations, shutdowns and equipment washes often generate much more severe disturbances to the effluent treatment system. These disturbances can often be reduced more economically by changing operating procedures or providing containment facilities local to the plant.

In some cases process measures simplifying the end-of-pipe treatment may be justified purely in terms of their effect on the main processes. In other cases the best decision can only be reached by consideration of the tradeoff between the cost of modifications in the process design and the cost of the end-of-pipe treatment. An end-of-pipe design procedure can assist the optimisation of this tradeoff by

1. providing approximate cost estimates for alternative problem specifications without requiring large amounts of effort on design or data acquisition;
2. identifying characteristics of the problem specification which are critical to the end-of-pipe treatment costs.

The design tools and procedure presented in this thesis directly address these two needs.

1.3 Overview of thesis

Chapter 2 reviews current practice in the area of chemical effluent treatment system design. The design options available for chemical waste water treatment are reviewed with regard to potential use in this work. Previous design guides are critically reviewed to identify useful principles and highlight limitations which need to be addressed. Published cost data are summarised.

Due to the range of problems and methods considered in this work reviews relating to the design tools and to modelling are integrated with the relevant sections of the thesis.

Chapter 3 discusses the tools developed to solve the design problems of interest. Continuous optimisation techniques are examined and a suitable technique chosen. A robust and efficient dynamic optimisation technique is developed. A new algorithm for design with uncertainty is presented. Approaches to screening designs are reviewed and a new method for assessing the effect of dead time on achievable disturbance rejection is presented. Finally, the software implementation of the tools is briefly reviewed.

The key design tools developed are

1. a robust dynamic optimisation package;
2. an optimisation method for dealing with parametric uncertainty;
3. a screening method for evaluating the effect of delays on achievable performance.

These tools are not restricted to chemical wastewater treatment. They have potential application to other design problems, particularly those involving significant constraints on transient response, delays in the process dynamics and uncertainty or variability in the process characteristics.

Chapter 4 presents the models developed to allow the application of the design tools to the problems of interest. The modelling work has centred on the neutralisation characteristics of solid reagents. Validated models are developed for the two most common solid reagents, $Ca(OH)_2$ and $CaCO_3$. Methods of representing the steady state pH characteristics of aqueous solutions have been examined and their advantages and limitations identified. Modelling of mixing in stirred vessels is reviewed using computational fluid dynamics and results in the literature. Models of pH probe characteristics are critically reviewed. A model for the oxidation of cyanide with hypochlorite is presented. The difficulties in modelling precipitation behaviour are discussed. Default parameters and uncertainty descriptions are presented for the key models and advice is provided on data acquisition and experiment design.

Chapter 5 presents the general framework for carrying out designs. A generic problem is used to develop some useful results and insights. The design procedure developed sets out the data requirements and appropriate use of the tools and models for various levels of design, ranging from generating lower bounds on cost to detailed system specification.

Chapter 6 consists of design examples. Most of the examples are closely based on actual industrial problems. The examples include systems with rapid neutralisation reactions and systems using solid alkali reagents. The problems examined include the effects of highly variable titration characteristics and of high intensity neutralisation. The methods used are compared to previously published design guidelines.

Chapter 7 reviews the design framework developed and draws some general conclusions. The areas of original work are emphasised and some possible directions for future work are presented.

Chapter 2

Review of current practice in chemical waste water treatment system design

Before developing the tools, models and procedures for use in design of chemical waste-water treatment systems, it is important to consider how this problem is tackled in current practice. Firstly, some of the underlying process characteristics which should be considered in looking at current practice are introduced (2.1). Then the key design options for end-of-pipe treatment systems are discussed (2.2). Current design practice is reviewed primarily by examining several published design guides which are used industrially (2.3). To complement the review of established guides some research work which may point to trends in current practice is examined (2.4). In conclusion, the strengths and weaknesses of current practice are summarised. This discussion clarifies the background for the rest of the thesis which aims to build on the strengths of current practice and contribute to the elimination of the weaknesses identified.

2.1 Process characteristics

The most distinctive property of chemical waste-water treatment is the complex, nonlinear and variable relationship between reagent addition and measured output properties, even at steady state. This relationship is normally discussed in terms of a titration curve. The titration curve is given as a plot of the output property, usually pH or redox potential, against the concentration/volume of reagent added. For pH titrations one extreme is given by the strong-acid/ strong-base titration curve (figure 2.1).

This curve is extremely steep near neutral pH values (around 7 pH). Reagent therefore needs to be added very precisely to adjust the pH towards neutral for this case. In most cases, the actual curve will deviate from this extreme showing lower slope of pH with concentration, particularly near neutral. This effect is known as buffering and can make achieving a given exit pH specification qualitatively easier. If the titration characteristic is constant and known, then its contribution to the process nonlinearity can be cancelled by conditioning of the measurement signal. Unfortunately, buffering is rarely constant. Variability of buffering makes all aspects of treatment system design difficult and processes with highly variable buffering are generally accepted to present the most

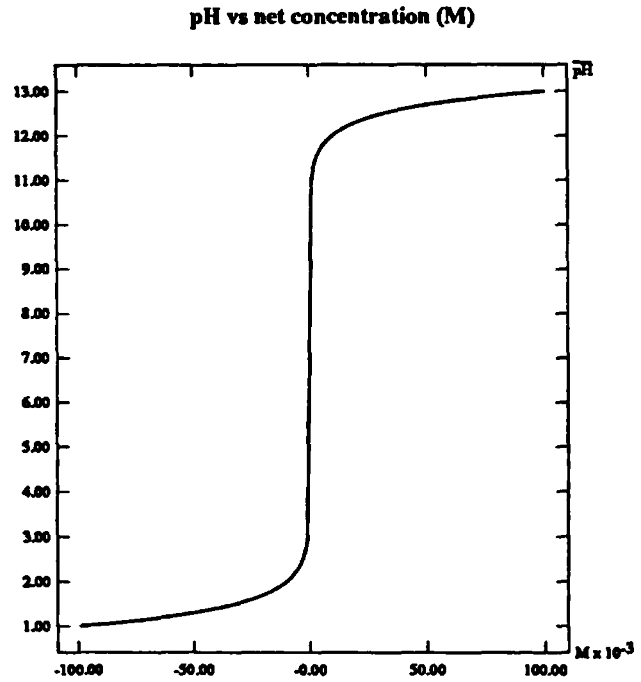


Figure 2.1: Strong-acid/ strong-base titration curve

challenging and difficult design and control problems (Trevanthan, 1979; Jacobs *et al.*, 1987).

Many reagents for neutralisation react practically instantly. The cheaper alkali reagents are generally in solid form and exhibit slow heterogeneous reactions with acid wastes. This requires careful design to ensure adequate conversion of reagent. Redox reactions used for eliminating wastes such as cyanide are typically slow and pH dependent. The slow reactions introduce additional nonlinearity, uncertainty and variability to the process.

The primary measurement is usually based on an ion selective electrode. All these measurements are vulnerable to bias errors. Many exhibit complex dynamic response. The measurements introduce substantial uncertainty to the observed behaviour.

All these characteristics are discussed extensively in chapter 4.

2.2 Design options

As noted in the introduction process and control system design interact strongly. Process design options have the most direct influence on cost. Inappropriate process design may prevent control system design from achieving the required performance. Integrated design of process and control system may allow significant reductions in total cost. The discussion below therefore considers both process and control options.

2.2.1 Process options

2.2.1.1 Batch or Continuous treatment

Existing effluent treatment practice is dominated by continuous treatment systems which have been found to have lower costs in dealing with typical effluent loads. Historically, batch treatment has been used for treatment of effluents which either had particularly small continuous flows or which were generated from a small number of batch processes giving very variable load. At very low continuous loads it could be difficult to meter the continuous reagent flows required. It may also be significant that at low flows equipment costs will be dominated by fabrication costs so that cost is not highly sensitive to required material holdup. Batch systems have inherent advantages for very variable loads which a continuous treatment system would find particularly difficult to control.

In short, batch treatment is favoured by low and highly fluctuating effluent load and continuous treatment by high and steady load. The choice may also be influenced by whether local expertise is strong or weak in batch system design and operation. This is clearly an important choice which may have a major impact on system cost, and should be addressed in a general design procedure.

2.2.1.2 Choice of reagent

Reagent choice influences plant and controller design very strongly and has a major impact on both capital and operating cost. Reagent costs for neutralisation may range from about £0.50/kgmole neutralised (for waste-product solid alkalis) to about £12/kgmole (for high quality nitric acid). A "typical" effluent treatment system taking waste water from several plants (about 200m^3/hr at about .1M concentration) would neutralise about 175,000 kgmoles of effluent in a year giving an annual reagent cost of between £90,000 and £2,100,000 depending on reagent choice. The capital cost of a "typical" plant (reagent delivery system, control system and 3 continuous reactors) is about £500,000. Lower reagent cost usually results in increased capital cost. Chemical properties of the effluent may make the use of certain reagents infeasible or undesirable. Selection of reagent will depend on reagent cost and chemical properties and the impact of reagent choice on equipment size and design. This is the key design option determining cost of end-of-pipe treatment. Any design procedure should identify constraints on reagent selection and provide a means of estimating relative costs to enable initial choice of reagent. Finally, it should assist development of a design to satisfy performance requirements using the chosen reagent.

2.2.1.3 Mixers

There are four standard mixing units with application to effluent treatment each with distinctive control and reaction properties.

Stirred Tank Reactors (STRs) These provide the most common mixing method for effluent treatment. They can be used effectively in feedback mode since they provide good attenuation of disturbances. They are however expensive. Costs depend strongly on materials of construction. A "typical" STR would cost about £30,000–60,000. STRs with continuous flow are known as CSTRs.

Plug flow reactors Plug flow reactors (PFRs), such as static mixers and pipes with turbulent flow, are favoured by their low cost and space requirements in many applications. They are also more efficient in converting reagent than CSTRs, in terms of conversion/unit volume. As a control element they suffer from substantial noise and poor attenuation of disturbances. PFR cost may vary immensely depending on design. In their simplest form, they are an injection point and a pipe with turbulent flow. In this case they add a negligible amount to the cost of the reagent delivery system. In the more complex form, they may use a large number of in-line mixing elements or small internal channels and the equipment costs may be more than £10,000.

Jet-mixed tanks These are in many ways a variant of the STR, but often do not achieve as good mixing as an STR. Controller tuning is highly sensitive to throughput if the effluent flow is used to provide the mixing as the deadtime due to imperfect mixing varies at least in inverse proportion with the effluent flow. If a pumped recycle is used for mixing, rather than using the process flow, then this sensitivity is suppressed and deadtimes comparable to a well mixed STR can be achieved. This latter configuration is however unlikely to have a cost advantage over an STR. The former configuration may well have a considerable advantage in mixing cost for effluents which have a high static head available as the energy for mixing can then be obtained at minimal cost. When working with clean fluids jet mixers may be easier to maintain than agitators. This may be a significant issue if maintenance access is difficult. Jet mixed tanks have similar costs to STRs except that the agitator is replaced by an injection nozzle and possibly a pump.

Unstirred tanks/ lagoons Unstirred tanks are sometimes used as buffer capacities to damp out disturbances at the inlet to the treatment plant or oscillations at the outlet of the treatment plant. They may also be used as containment facilities for abnormal events. Unstirred tanks between controlled stages are not usual in industrial practice, but may be effective in some cases, especially if some of the controlled stages use plug flow reactors. If a level controlled tank is used, it may be used to smooth flow variations as well as concentration variations. Process units provided for other purposes may sometimes be used as unstirred tanks with minimal cost. An unstirred tank will have a typical dead-time of about 30% of residence time (volume divided by flow), while a large lagoon may have dead-time as high as 99% of residence time. This means that they are not very useful in themselves for feedback control, though they may be used to smooth out disturbances to assist feedback control around other mixing elements. Lagoon costs are governed by land costs and civil costs. Costs have increased recently due to the move to above ground construction to facilitate inspection and leak detection. As an indication, a 50,000 m^3 lagoon would cost about £150,000.

Summary The selection, linking and sizing of mixing units is central to obtaining an effluent treatment system which satisfies its performance requirements. Choices in this respect have a substantial impact on cost and should be considered carefully.

2.2.1.4 Reagent addition systems

Reagent addition systems generally include storage facilities, pumps, valves and pipework. This in itself can be a substantial part of treatment plant costs (typically £200,000 for caustic). Solid reagents (all the low cost alkalis) are normally delivered as a slurry and require solid handling facilities and a slurry tank which must be vigorously mixed. With the most common alkali reagent, $Ca(OH)_2$, it may be economically preferable to purchase the reagent as quicklime, CaO, and react it with water on-site in an additional reactor known as a "slaker". The costs for reagent handling equipment are considerably increased by moving from caustic to solid alkalis (typically £350,000).

If high reagent addition rangeability is required (greater than 50:1) then it may be necessary to use some form of parallel reagent addition system typically a small and large valve in parallel. This is a detail in terms of overall cost and treatment system behaviour, but should be noted as neglecting it can cause the treatment system to fail.

The reagent addition system is often a packaged unit which can be designed very effectively using general tools and expertise. Its main impact on the design of the overall system is in its effect on variation of cost with reagent type.

2.2.2 Control options

2.2.2.1 Measurement systems

Measurement system characteristics may affect the control system response significantly and must be considered in the design process. pH/ redox measurement maintenance may be a significant operating cost as the measurement electrode may need frequent cleaning or replacement. Failure of the measurement system to perform ideally is a common reason for poor performance of effluent treatment systems.

2.2.2.2 Controller options

The options for continuous control are outlined below.

Standard PID feedback control Due to noise effects and nonlinearities this is often reduced to PI control. At the design stage, when the noise effects cannot be readily predicted, assuming a PI control will give a conservative result leaving the benefit of derivative action, if any, as a bonus for the commissioning team. In batch systems PD control is normally the preferred option as overshoot is normally highly undesirable and PI controllers tend to overshoot or oscillate.

Feedforward disturbance compensation It may be possible to get substantial disturbance reduction by feedforward. It may be possible to reject up to about 95% of the disturbance by this means which would make a qualitative difference to control performance. However, in cases where the effluent/reagent properties vary substantially feedforward may actually make control worse. The cost of adding feedforward is generally the cost of one or two new measurements which would be preferable to adding an additional controlled CSTR.

Cascade/ratio control At its simplest, and least useful, cascade control may simply involve an inner flow control loop to linearise the flow response to the composition controller output. Cascade control of one pH controller from another is unusual but may be used to accommodate changes in process operating conditions. The main use of ratio control is to provide rapid response of reagent addition to flow disturbances. The output of the main controller adjusts the ratio of reagent flow to effluent flow rather than adjusting reagent flow directly, so that changes in effluent *flow* are compensated for. This scheme would be most effective in situations where concentrations vary slowly, but flows vary rapidly. It simplifies pH feedback control around a PFR by removing the variation in the optimal controller gain with flow. With a CSTR it complicates feedback control by introducing a variation in this gain which may require compensation. Ratio control is also sometimes used to tie two feed streams together, e.g. chlorine gas and sodium hydroxide feeds which are sometimes used to generate hypochlorite for oxidation reactions.

Input conditioning and compensation Input conditioning applies a static nonlinear function to the measured variable to create an input with better control properties — typically a more linear relationship between the controller input and reagent addition. Input conditioning may be used to reduce any degradation of achievable control due to measurement nonlinearity or to take account of asymmetry in the performance requirements. Achievable improvement may be limited by variation in the nonlinearity and by measurement bias. It may also be useful to explicitly compensate one measurement using another, e.g. compensating redox potentials for the effect of pH to reduce interaction. These options entail minimal extra cost with modern control hardware.

Gain scheduling The PID controller parameters may be switched according to measured values or supervisory inputs. A simple example of this is altering the parameters to compensate for the effect of variable deadtime in a jet-mixed tank based on measurement of flow. The effect of gain scheduling is to reduce the range of process variation which must be accommodated by each set of control parameters. This should allow tighter control to be achieved.

Adaptive control Adaptive control involves the use of a search procedure to monitor control performance and adjust the controller parameters. Adaptive control represents a significant increase in controller complexity, but has the potential to improve performance significantly in appropriate applications.

Nonlinear control Nonlinear control utilises a nonlinear process model either to compute optimal control response or to linearise the system response. This is an area which has generated a lot of theoretical interest but few applications to date. There appear to be unresolved issues of robustness to unmeasured disturbances and manipulated variable constraints. Input conditioning and gain scheduling may be regarded as special cases of nonlinear control which are relatively well understood and appropriate to this application area.

Multivariable control General multivariable control involves the simultaneous use of more than one measurement to determine the appropriate value of one or more inputs. Cascade/ ratio controllers are special cases of multivariable control. The opportunities to gain benefit from more general multivariable control appear limited as measured and manipulated variables generally have an obvious and effective pairing based on control around a single mixing unit. General multivariable control is not considered further in this thesis.

All the controller options considered in this thesis are readily implemented on modern control hardware, usually in stand-alone controllers (typical cost £1500). The main cost of using options other than standard PID is the design and commissioning effort required to ensure they are effective.

Continuous control may need to be supplemented by supervisory logic to deal with multiple operating modes or startup considerations.

Design procedures and tools should guide the selection of controller structure, type and tuning and allow the combined process/control system performance to be evaluated and, if appropriate, optimised.

2.3 Published design guides

A number of design guides have been published with useful information for effluent treatment design, and in particular pH control system design, over the past two decades. Many of the guides (Shinskey, 1973; Hoyle, 1976; Moore, 1978; McMillan, 1984) have a control focus and centre on a heuristic-based approach to pH and pIon control system design which is associated with the Instrument Society of America. This work is widely referenced and used and represents the major approach to pH control system design available in the public domain. Other guides (National Lime Association, 1983; Cushnie, 1984; Eilbeck and Mattock, 1987) have a process engineering emphasis. All these guides focus on rules and procedures for designing systems based on stirred tanks in series without recycles as these are the most common industrial systems.

The above guides have complementary strengths (and weaknesses). Shinskey is good on the basic pH chemistry and gives some discussion on redox treatment systems. Per[1] presents a number of concrete suggestions for system design though the supporting arguments are sometimes weak and the suggestions sometimes debatable. Hoyle presents a correlation for predicting tank deadtime which is widely cited and used and some heuristics on the use of lime-based reagents. Moore gives an interesting empirical discussion of the properties of lime-based reagents. McMillan gives the best discussion of the analysis of control dynamics and the implications for CSTR design. There are some errors in pers analysis, but it does draw together all the important elements for rapidly reacting reagents. Per gives a strong presentation on design and control of in-line mixer systems. Per also gives detailed and interesting discussions on reagent control valves and pH measurement systems. Pers discussion of mixing and of the effect of valve precision is however quite weak, and the treatment of reagent conversion is misleading. The report by the National Lime Association deals mainly with the neutralisation properties of limestone-based reagents and indicates some limitations on the use of these reagents. Cushnie gives a systematic approach to analysing reagent conversion and makes a useful attempt at defining procedures flowing from experimental analysis to process design, including cost estimation. However, per is not very helpful on control. Eilbeck and Mattock present an extensive qualitative discussion of redox measurement and chemistry which should be read by anyone intending to model these systems. None of the above guides have much to say on computer-based modelling and simulation, though some (Moore, 1978; Cushnie, 1984) note its critical importance as a final performance check.

In the discussion below, I critically review the information and guidelines presented by these guides. The design choices discussed in section 2.2 are used to focus and structure the review. I do not attempt to summarise all the guides' advice on these options but instead present the advice I judge to be best. In some cases, I note guidelines which are either ill-founded or potentially misleading and which highlight the need for improvements.

2.3.1 Process options

2.3.1.1 Batch or continuous

Shinskey indicates that batch systems are typically only used below 100gpm (20 m^3/hr) continuous flow. This appears consistent with the general conjecture that batch treatment may be favoured when reactor costs are fabrication dominated. Data presented by Cushnie indicate that CSTR cost increases only about 50% with size up to 10m^3 volume. With the maximum flow indicated by Shinskey, this would allow 30 minutes residence time to be achieved with little extra cost compared to the 3 minutes minimum residence time recommended by Hoyle for a CSTR using rapidly reacting reagent.

---

[1] I use per as a genderless third personal singular and pers as the corresponding possessive, as English inconveniently fails to provide appropriate terms

McMillan suggests that batch systems have a longer natural period than continuous systems and are therefore harder to control. This actually equates continuous control with control around a PFR (plug flow reactor). The natural period of a CSTR (continuous stirred tank reactor) is normally not significantly different from a batch STR.

Overall, this issue is given little consideration. As most effluent treatment plants deal with flows around 100–200m^3/hr (Proudfoot, 1983) it is unlikely that batch operation will be used often.

2.3.1.2 Choice of reagent

This choice is subject to many simple constraints which can be used together with rough estimates of costs to guide reagent selection. Some constraints will only emerge from application/specific experimental analysis, but a number of constraints can be extracted / of from published guides which may be used to determine which reagents are feasible for particular applications.

1. $CaCO_3$ and $MgCO_3.CaCO_3$ cannot be used for neutralisation above about pH 5 due to saturation of the solution with $CaCO_3$ (Cushnie, 1984; National Lime Association, 1983). This limitation also applies to many waste alkalis.

2. $CaCO_3$ and $MgCO_3.CaCO_3$ cannot be used for neutralisation of solutions containing more than about .3% by weight of H_2SO_4 due to inactivation of the reagent by gypsum precipitation (National Lime Association, 1983; Jacobs, 1951).

3. Use of calcium-based reagents with effluents containing compounds forming poorly soluble calcium salts (e.g., fluorides, phosphates, sulphates, tartaric and oxalic acid) must be approached with caution if concentrations are such that the solubility product of the salt is exceeded. The consequences of precipitate formation (sludge formation, suspended solids and probe coating) must be considered and the reactivity of the reagent may be seriously inhibited by formation of a layer of insoluble salt on the reagent particles (National Lime Association, 1983; Haslam *et al.*, 1926). Application specific experimentation is needed to check whether problems with reaction rate are likely to occur.

4. Similarly, sulphuric acid may be unsuitable for use with solutions containing high concentrations of calcium ions. $CaSO_4$ precipitates are particularly problematic as supersaturated solutions are readily formed and precipitates may be deposited downstream in inconvenient locations forming a hard coating on equipment.

5. Magnesium-based reagents are not effective for over neutralisation (pH much above 9) as $Mg(OH)_2$ will precipitate.

6. Soda ash ($NaHCO_3$) should not be used for overneutralisation as a substantial amount of the reagent will remain as HCO_3, in equilibrium with atmospheric CO_2.

The choice between feasible reagents will be strongly influenced by costs but a number of other design considerations can be identified which may result in a more expensive reagent being chosen.

1. If it is required to precipitate fluorides, phosphates or sulphates in addition to neutralising an acid then a calcium-based alkali would normally be chosen.

2. $Ca(OH)_2$ is known to be more effective in precipitating trace metals than caustic (NaOH) and is therefore preferable when this is important.

3. If it is desired to avoid sludge formation NaOH is preferable to alkalis containing carbonate or calcium.

4. Magnesium hydroxide may have better sludge formation properties than calcium hydroxide and does not readily precipitate sulphates.

5. Carbonate reagents may introduce a useful buffering effect around pH 6 which can make control easier (Mattock, 1964).

6. Dolomitic hydrated lime ($Ca(OH)_2.Mg(OH)_2$) is regarded as reacting more slowly than high calcium hydrated lime (Hoyle, 1976) and having more variable properties (Cushnie, 1984).

7. $CaCO_3$, $MgCO_3.CaCO_3$ and some waste alkalis may contain a significant proportion of large particles (about 1mm diameter) which it will be impractical to react completely (National Lime Association, 1983). This precludes their use when there is a tight specification on suspended solids, unless provision is made to separate the excess reagent from the treated waste.

A minor consideration with solid alkalis is that they are handled as slurries or solids for which control of reagent addition rate is more difficult than for reagents delivered as a solution. This is particularly important at low reagent flowrates (less than about 1 litre/minute) so it may be necessary to use another reagent for final stage neutralisation, even when a solid alkali is used as the main reagent. The alternative reagent is required in very small quantities. Therefore it should not require expensive reagent handling equipment and not have a significant impact on operating cost.

Aside from these considerations and any problems identified in experimental testing, the decision as to which reagent to use for neutralisation will be based on cost. Estimated reagent costs are tabulated below, mainly based on Cushnie (1984) and Chemical Marketing Reporter. It should be noted that these costs are quite approximate and are likely to vary significantly depending on location and grade of material. Acid costs in particular may be much lower than those given if a source of low grade acid is available. Costs are given in terms of pounds per kgmole effluent neutralised to allow a direct comparison of cost per unit neutralising capability.

Reagent costs

reagent name	formula	cost £/kgmole neutralised
hydrochloric acid	HCl	7.00
nitric acid	HNO_3	12.00
sulphuric acid	H_2SO_4	5.00
caustic	$NaOH$	8.00
limestone (high calcium)	$CaCO_3$	0.75
quicklime (high calcium)	CaO	1.70
hydrated/slaked lime (")	$Ca(OH)_2$	2.20
dolomitic limestone	$MgCO_3.CaCO_3$	0.50
dolomitic quicklime	$MgO.CaO$	1.20
dolomitic hydrated lime	$Mg(OH)_2.Ca(OH)_2$	1.50
magnesia	$Mg(OH)_2$	6.00
soda ash	$NaHCO_3$	6.00

Of the guides reviewed, only Cushnie (1984) provides data on the costs associated with the treatment system and reagent selection. These costs are given as graphs in 1978 US dollars. I have approximated the graphs by correlations and made a rough conversion to 1993 pounds using the Chemical Engineering plant cost index and the exchange rate. The key data are given below

Equipment costs

cost item	capital cost	operating cost/ annum
Stirred tank reactor	$20,000+2,000V_t^{0.7}$	
NaOH feed system	$30,000+1300q^{0.75}$	$600q^{0.4}$
$Ca(OH)_2$ feed system	$40,000+370q$	$40q$
pH control system	40,000	40,000

q is the reagent feed capacity in kg/hr of the pure reagent and V_t is the tank volume in m^3. These correlations are crude and many companies will have better data internally, but they seem to provide the best approximation available in the published guides. The control system cost is particularly crude as it is not related to number of measurements or valves. They are *broadly* consistent with costs I have encountered in my case-study work. The correlation for $Ca(OH)_2$ does not include a lime slaker which would typically add about £50,000 to the cost. In terms of doing a cost analysis for reagent selection, the main thing missing is a means of estimating size and number of reactors (see section 2.3.1.3).

For most systems the choice of reagent will be straightforward as the above constraints, preferences and costs will result in a single clearly preferable candidate (usually HCl/H_2SO_4 for alkali effluent and $Ca(OH)_2$ for acid effluent). In some cases, approximate design of the mixing system will be required to clarify the economics. Detailed design may be necessary to establish whether the reagent reaction properties allow the performance requirements to be met (particularly with alkalis under £1/kgmole neutralised).

The discussion above has centred on neutralisation reagents as these are the most extensively used and studied. A similar picture of a highly constrained cost-based decision appears in other effluent treatment applications. For example, for cyanide oxidation hypochlorite, ozone, peroxide or permanganate may be used. Hypochlorite is the most common reagent for cyanide oxidation due to cost, applicability to typical concentrations of cyanide in industrial effluent and the existence of well established guidelines for "best current practice" (Cushnie, 1985). For precipitation, it may be necessary to add seed crystals or flocculants to achieve the required behaviour. This is usually tackled by using previous experience to guide selection of options for experimental investigation as the predictive models of behaviour are inadequate (Cushnie, 1984).

2.3.1.3 Mixers

The most widely cited rule (Shinskey, 1973; Hoyle, 1976; McMillan, 1984; Cushnie, 1984) for determining mixing configuration is a heuristic rule for selecting the number and type of reactors in series based on the required change in pH towards pH 7 (neutral). The most recent version is given by McMillan and summarised in the table below.

ΔpH	no. of "well-mixed" CSTRs
0–2	1
2–4	2
worse	3

ΔpH is the pH difference between the untreated effluent and the constraint on the discharged effluent pH. Well-mixed is taken to mean that the response approximates to a deadtime ($t_{d_{mix}}$) plus first order lag ($t_{c_{mix}}$) with the deadtime less than 5% of the lag. McMillan notes that one tank can be deleted if the titration curve is fixed and pH and flow feedforward information are available. Per also suggests that a single in-line mixer can be used if the open-loop gain is less than 10 (presumably %in to %out) and ΔpH is less than 2. Per describes the overall rule as conservative. As it does not consider the actual disturbances or titration curve buffering it *should* be conservative.

McMillan gives a "more rigorous" rule based on estimating the disturbance attenuation (δ_a) of a PID controlled mixer as the steady state disturbance effect with the proportional control component only ($1/(1+K_p.K_c)$). K_p is the steady-state process gain and K_c is the controller proportional gain. This is used as an estimate of the peak deviation in response to a step disturbance. δ_a is the ratio of the peak concentration deviation with control to the peak open loop deviation. For each controlled well-mixed reactor this can be approximated as $\frac{t_{d_{eff}}}{t_{c_{mix}}}$, where $t_{d_{eff}}$ is the effective deadtime. The effective deadtime is taken as the deadtime in the model $\frac{k.e^{-t_{d_{eff}}s}}{1+\tau s}$ with the parameters chosen as for Ziegler-Nichols tuning. For a CSTR, $t_{d_{eff}}$ can be approximated as the sum of all deadtimes and minor lags from reagent addition to pH measurement. For n tanks McMillan's approach gives

$$\delta_a \approx \left(\frac{t_{d_{eff}}}{t_{c_{mix}}}\right)^n \tag{2.1}$$

If the disturbance takes the form of a first order lag step response, $1 - e^{-t/\tau}$, then δ_a is reduced by a factor $e^{-4\tau/t_u}$ where t_u is the ultimate period of the loop.

The predicted disturbance attenuation, δ_a, can be compared to the required control precision, δ_c. δ_c is defined, by McMillan, as the ratio of the change in reagent flow required to cancel the step disturbance, ΔFr, to half the change in reagent flow which would take the pH from the high to the low pH constraint, $Fr_h - Fr_l$.

$$\delta_c = \frac{Fr_h - Fr_l}{2\Delta Fr} \tag{2.2}$$

This is the fractional change in reagent addition which would take the pH from midway between the constraints to one of the constraints. The number of tanks is increased until $\delta_a < \delta_c$.

This analysis represents a major advance on the "ΔpH" heuristics such as that discussed above. Incorporating consideration of the actual disturbance rejection requirements and the effect of process dynamics on achievable disturbance rejection allows a sensible design analysis. The more rigorous approach still has several weaknesses as a design tool for selecting number of tanks.

1. The disturbance attenuation estimate, δ_a, is derived based on the choice of a PID controller and does not constitute a bound on the performance achievable. It does not therefore allow the designer to say *unambiguously* that a design will not work.
2. It does not explicitly address the use of uncontrolled tanks or the use of tanks with different sizes/ mixing intensity.
3. The factor used in calculating the disturbance attenuation for a first order lag disturbance is incorrect. A controlled tank contributes an attenuation $t_{d_{eff}}/t_{c_{mix}}$, typically about .05. The factor used implies that an uncontrolled tank used to convert a step disturbance to a first order lag would contribute an attenuation of roughly $e^{-t_{c_{mix}}/t_{d_{eff}}}$, typically less than 2.10^{-9}!

McMillan's "more rigorous" analysis indicates the "conservative" approach not to be generally conservative. Assuming strong-acid/ strong-base titration characteristics, typical pH constraints of 5–9pH and a step disturbance from zero to full flow, then the required control precision is $10^{-\Delta pH}$. The achievable attenuation for n well-mixed tanks ($t_{d_{mix}} \leq .05 t_{c_{mix}}$) and instant measurement and valve response is about $.05^n$. The "conservative" rule may therefore recommend a two tank system for a problem with a required precision of .0001 and a predicted attenuation of .0025! On the other extreme, a well buffered system may have a "ΔpH" of 4 and a required control precision of .1, so that a single "well-mixed" CSTR may well be adequate. These examples highlight the weakness of the ΔpH rule which should no longer be used.

McMillan notes that an in-line mixer can be used in conjunction with a large moderately backmixed volume to give very small deadtime to backmixed residence time ratios and correspondingly good disturbance rejection without high energy requirements for mixing.

With fast neutralisation reactions the optimum CSTR mixing is generally held to be such that the fractional deadtime ($\frac{t_{d_{mix}}}{t_{c_{mix}}}$) is about 0.05 (Shinskey, 1973; Hoyle, 1976). It is implicitly assumed that further reduction in fractional deadtime offers little benefit as loop response is dominated by other minor lags or deadtimes or that mixing costs increase excessively. Hoyle (1976) notes that this fractional deadtime is difficult to achieve for tanks with residence times below about 3 minutes without excessive splashing and air entrainment. The difficulty in reducing $t_{d_{mix}}$ below about 9 seconds observed by Hoyle is borne out by more general work on mixing discussed in chapter 4 and appears to represent a real design constraint in mixed tanks. 3 minutes is often treated as the minimum residence time for good control performance, based on Hoyle's observations.

Examination of the effect of mixing from first principles shows that neither an "optimal" fractional deadtime of .05 nor a minimum residence time of 3 minutes are appropriate general rules. Mixing cost is primarily governed by tank-size and recirculating flow rates in the tank from either agitation or jet mixing. Neither of these bears any necessary relationship to residence time which is inversely proportional to the flow through the tank. If flow rates are low then the cost of using high residence times and low fractional deadtime is also low, as the marginal costs associated with increased volume or recirculating flow are small. Reducing the fractional deadtime may allow much improved disturbance rejection and avoid extra cost associated with a higher number of reactor stages. In cases where volume is highly constrained it may be appropriate to use residence times less than 3 minutes and accept a fractional deadtime greater than .05 to give the optimal disturbance rejection for a given total volume. Both these heuristics are therefore inappropriate. Given this it is evident that the decision on residence time and mixing intensity should be based on achieving the required disturbance rejection while minimising cost (at least for rapid reactions). This requires a method of predicting $t_{d_{mix}}$.

Several correlations are given for predicting $t_{d_{mix}}$. Shinskey estimates the delay as $V_t/(2(Fagit+F_T))$ where V_t is the tank volume, $Fagit$ is the agitator pumping rate and F_T is the total flow through the tank. This correlation is based on a simple approximation of flow patterns in a stirred tank. Hoyle gives graphical correlations, based on experimental studies, which correspond to a delay of $0.9V_t^{0.85}/(Fagit)$, with all volumes in m^3. This applies to tanks with reagent being added at the top of the liquid on the opposite side of the tank to an exit near the tank bottom. If flow is reversed the deadtime was noted to approximately double ($1.8V_t^{0.85}/(Fagit)$). McMillan uses $V_t/Fagit$ which is unnecessarily conservative. Moore cites both Shinskey's and Hoyle's correlations and indeed presents a tank sizing procedure which is effectively to equate the two correlations! Cushnie accepts Hoyle's correlation. Shinskey's correlation is derived from simple flow visualisation and McMillan's is not explained. As Hoyle's correlation is based on a specific empirical study, I believe it has most credibility. Hoyle's correlation strictly applies only to baffled tanks with: comparable diameter and liquid height; downward pumping axial impellers; inlet and outlet positioned at opposing sides and levels in the tank; agitator diameter about 1/3 the tank diameter. This fortunately is the most common configuration as it is regarded as a good compromise for many applications. $Fagit$ can be computed from general correlations or from agitator manufacturers' data.

??

The effect of $t_{d_{mix}}$ on control performance can be evaluated and the appropriate value chosen based on the trade-offs between the cost of mixing and the cost of other methods of achieving the required control performance.

For slow reactions the minimum residence time requirements may be determined by steady-state reaction engineering considerations. This is addressed by most of the guides in a purely heuristic way. For example, for solid alkali reagents, there is a clear need to allow sufficient residence time for most of the reagent to dissolve. Hoyle estimates at least 5 minutes for high calcium hydrated limes and about 20 minutes for dolomitic hydrated limes. These heuristics have similar problems to the "Δ pH" heuristics discussed above, in that they simply do not include enough information to assess the requirements of a particular design. That is, they do not consider what levels of residual reagent are acceptable (required conversion precision) or how variability in reagent/effluent properties affects the requirements. This requires a more fundamental modelling approach. Cushnie presents a more classical reaction engineering approach. Per discusses methods of fitting models to batch titration experiments (all reagent added in one dose and the response over time recorded) and graphical methods of analysing the conversion of reagent in back-mixed reactors for general homogeneous reactions. The main limitation I can see to pers approach is that per does not address the fact that many of the key reactions are heterogeneous (solid/ liquid) so that procedures for homogeneous reactions will not work properly. Per also uses rather arbitrary functions to fit the responses rather than attempting to develop a model based on first principles which would (hopefully) have better predictive properties. McMillan gives a graph of conversion versus residence time which implicitly assumes a homogeneous first order reaction. This could be very misleading as the reactions are not well described by this model (see chapter 4).

For redox reactions, the minimum residence time is often substantial being a function of the particular reaction(s) taking place. It can be estimated using steady state calculations. There is often an interesting coupling between operating pH and required residence time (Shinskey, 1973). This introduces a trade-off between reagent usage and capital cost which seems best examined using optimisation-based techniques and detailed models or extensive experience, as the relationships are complex and problem dependent.

With multiple reactors the sizing decision in principle becomes much more complex. There is very broad support in the guides reviewed for the principle that if two controlled tanks are used in series the second tank should be about 4 times the size of the first tank (Shinskey, 1973; Moore, 1978; McMillan, 1984). Shinskey motivates this by arguing that for identical controller tuning strategies and constant $\frac{t_{d_{mix}}}{t_{c_{mix}}}$ on both stages the peak in the closed loop frequency response will be reduced by using tanks of different volumes. The argument is developed using concentration feedback (perfect linearisation) so Shinskey's inference that the second tank as opposed to the first should be larger is invalid. The argument does not address whether this effect is really significant enough to justify a major process decision such as this. On simple loops such as these the controller will typically reduce the open loop disturbance attenuation by at most 30% around the natural frequency. Even this effect can be alleviated simply by tuning the controllers on successive tanks differently. Industrial practice conflicts with the suggested need to split tank sizes — equally sized tanks appear to be the norm (Proudfoot, 1983). This discrepancy is explored in section 5.2.4.

?
achew

McMillan (1984) gives a useful discussion of conditions for use of an in-line mixer. These conditions are summarised below.

1. The control system does not see variations in pH giving rise to large gain changes.
2. A backmixed volume exists to attenuate rapid disturbances.
3. Rapid reagent injection / multiple injectors are used to give even reagent flow and some premixing.
4. An in-line (injector) probe assembly is used minimising probe lag.

Based on experience and recent discussions with a Kenics engineer I would add that it may be necessary to allow a significant delay (about 7 seconds) between reagent injection and pH measurement to ensure reasonably low noise levels by allowing time for adequate micro-mixing of reagent and effluent.

2.3.1.4 Reagent delivery systems

McMillan (1984) gives a useful discussion of the factors limiting valve precision. The main limitation is hysteresis in the response of valve position to controller signal. This is noted to be about 5 to 10% of range without a positioner and about 1% of range with a positioner. This imposes limitations on the control performance. McMillan incorrectly states that the valve precision must be less than the control precision, δ_c. This is only true when using in-line mixers which have no inherent attenuation of high frequency variations. For a tightly controlled CSTR valve precision limitations can be expected to give a limit cycle on the reagent addition at around the natural frequency of the process and with an amplitude corresponding to the precision error (see section 5.2.2). The effect of this limit cycle on pH will be attenuated by the back-mixing in the CSTR(s) allowing the valve precision to be much larger (worse) than the control precision.

For very small reagent loadings, e.g. final stage neutralisation, a metering pump may be preferable to a control valve. In particular, caustic above 30%w/w is liable to go solid in small valves (Moore, 1978). Cushnie suggests that reagent feeds below 2 litres/second should use metering pumps. This limit seems *much* too high.

It is generally recommended (Shinskey, 1973; McMillan, 1984) that reagent addition elements should be characterised to appear linear to the controller. Even if the nonlinearity of the valve appears to balance a nonlinearity in the pH characteristic this procedure should be followed as the use of integral action in the control means that the measurement nonlinearity cannot be properly compensated for by another nonlinearity at the controller output (McMillan, 1984).

2.3.2 Control options

2.3.2.1 Measurement systems

The detailed design of measurement systems is addressed very well by McMillan (1984). The measuring probe can be placed directly in the tank (through a dip-pipe), in the exit line from the mixer (injector assembly), or in a sampling system drawing a continuous flow from the exit stream. The in-tank arrangement gives minimum pure delay in response but becomes unwieldy if the outlet is submerged by more than about 3 metres (30m^3 tanks or larger) and is vulnerable to fouling or damage from solids in suspension. It may also result in increased probe lag if the circulating fluid has a velocity of less than about .3m/sec. The in-line arrangement can create an additional risk of a leak, but generally gives excellent speed of response to concentration changes. The sampling arrangement is highly reliable and easily maintained but adds a few seconds to the effective delay in the control loop. The choice between these options will vary according to company guidelines, size of tank and allowable materials of construction. McMillan favours injector assemblies as giving the most rapid response (provided flow rate is not very variable).

Overall this is a detailed design issue whose main impact on the overall design lies in its effect on measurement dynamics and errors. Dynamic response may in practice be severely degraded by probe coating and fouling or abrasion. Probes tend to drift as the glass electrode ages and often show clear long-term measurement biases, particularly at the extremes of pH range (below 2 or above 12). Both pH and redox electrodes exhibit variable dynamic response depending on the size and direction of the measured value change and the chemical composition of the solution (McMillan, 1984; Eilbeck and Mattock, 1987).

These effects should be allowed for in evaluating performance. Modelling of probe response is discussed further in section 4.4. Section 5.2.5 discusses the effect of probe response on treatment system design.

2.3.2.2 Controller

Only PID and on-off feedback strategies are considered in the guides reviewed. The use of on-off control may be appropriate if disturbances are of a similar magnitude to the total load and the titration nonlinearity is very variable. Both these factors reduce the performance gain which would normally arise from replacing on-off with PID control. No clear PID tuning guidelines are provided except for the strongly buffered case for which standard tuning rules are appropriate. In general controller tuning can be expected to require optimisation for the particular problem tackled as it is a function of both the disturbances and the nonlinearity.

McMillan (1984) recommends the use of input conditioning to eliminate as much of the nonlinear measurement characteristic as possible. This seems to be a useful measure if the curve slope varies substantially over the range of deviations encountered during disturbances and the curve is reasonably consistent.

For processes with variable titration curves Myron and Shinskey (1979) recommend a special adaptive controller, which increases the deadband if limit cycling is detected (high frequency bandpass filter output) and decreases the deadband for drifting measurement values (low frequency bandpass filter output). The high frequency bandpass must be carefully tuned to avoid noise sensitivity, and the low frequency adaptation should be set to adapt more slowly than the high frequency adaptation to avoid the controller inducing cyclic bursts of oscillations. Trevanthan (1979) refers to this approach and notes that neither it nor other special adaptive procedures such as using a fast response secondary treatment system to characterise the process have had universal success. The special adaptive procedure above has been replaced within Foxboro (for whom Greg Shinskey was a senior consultant until his recent retirement) by the EXACT controller which is based on the Ziegler-Nichols procedure and hence is only suited to fairly well buffered (moderately nonlinear) systems. Adaptive control is discussed further in section 2.4.

In many pH or pIon problems the use of feedforward is of little benefit or even counterproductive. This is due to feed concentration variability and the difficulty in measuring it. If the feedforward estimate is not within ±100% of the actual required input then it should not be used as it is liable to amplify disturbances. Feedforward should be tuned over the full range of load variation (Shinskey, 1973). McMillan's observation that combined pH/flow feedforward allows the elimination of one well mixed tank if the titration curve is fixed is over-optimistic for most applications as inferring concentration from pH can be very inaccurate. Despite these qualifications feedforward control should always be considered as a design option since in a suitable case it may give up to about a tenfold attenuation of disturbances. The achievable performance may be evaluated by analysing the sources of error in the feedforward calculation. If feedforward can be expected to give an improvement then it should be considered in more detail to establish whether its use allows an overall reduction in cost.

2.4 Trends in current practice

pH control has been the subject of a considerable amount of academic research due to its non-linearity and variability/ uncertainty. This has made it an interesting problem for researchers in adaptive and nonlinear control. Major benefits are claimed for both these control strategies so it is of interest to evaluate whether and how they are likely to influence industrial practice.

2.4.1 Adaptive control

Adaptive control using linear models has been extensively tested in industrial applications. If the process characteristic is highly nonlinear then these schemes generally have difficulty with large disturbances. Jacobs *et al.* (1987) note that the applicability of these schemes appears to be limited to moderately nonlinear systems. That is, linear adaptive control can be expected to be beneficial where the titration relationship is fairly linear but varies significantly over time.

Other work on adaptive control has tried to estimate the nonlinear relationship between pH and concentration and to cancel this relationship through input conditioning. Work at Abo Akademi by Gustafsson and Waller generated some initial results with a version of this approach tested on a bench-scale rig (Gustafsson, 1985). Variations in the nonlinearity were quite slow and the variation was in the quantity rather than the nature of the buffering. Even so, it was found necessary to identify all parameters (nonlinear output relationship and linear concentration dynamics) except a single gain off-line. Only this gain was adapted to the changed conditions. This approach is essentially equivalent to using a fixed input conditioning scheme to reduce the nonlinearity to a level at which a very simple linear adaptive controller can be applied.

More recent work (Williams *et al.*, 1990) has tried to overcome problems in acquiring enough information to track the variations in the nonlinearity rapidly by using two reagent addition points and three measurements in an in-line mixing system. A simplified model is used to parameterise the titration curve so that the steady-state reagent inputs and pH measurements are adequate to fit the model parameters. The method is assessed in simulation. The required control precision is about .1 and the neutralised effluent would require about 100 seconds downstream smoothing to stay within the limits 5–9pH. This method is interesting and *might* extend the range of application of adaptive control to effluent treatment systems in which the titration variations are simple, but rapid.

Overall, adaptive control appears appropriate for a subset of effluent treatment problems satisfying the conditions below.

1. The nonlinearity is moderate so that linear adaptive control can be applied.
2. The nonlinearity varies substantially so that adaptation can be beneficial.
3. The nonlinearity varies more slowly than load disturbances so that its variation can be tracked before the peak deviations have already occurred.
4. The improvement in control performance results in a cost saving which can justify the additional controller complexity and potential failure modes.

Most linear adaptive strategies should be effective in this context, and their use can be expected to become more common, within this niche.

2.4.2 Nonlinear control

To justify considering general nonlinear control it must be demonstrated that this control offers significant benefits over conventional "LFPID" (Linearised Feedforward PID) control. Several papers have made this claim (Jayadeva *et al.*, 1990; Parrish and Brosilow, 1988). Both apply nonlinear control techniques to the problem of pH control in a single tank with no delays or minor lags included in the model. Their performance results are meaningless due to the unrealistic assumption that mixing and measurement response are ideal. Many other applications of nonlinear controllers to pH control have appeared in the literature over the past few years. All such applications that I am aware of share the same basic flaw, though at least some of them achieve perfect control as permitted by the model used.

Results presented in section 5.2.2 show that LFPID control achieves performance levels very close to a theoretical bound for control of a single tank with a delay. The main potential benefit from nonlinear control lies in allowing compensation for the complex measurement dynamics. Much more accurate models than those reported would be needed to evaluate its performance in this respect (see section 5.2.5).

2.5 Conclusions

Having critically reviewed current practice, some design choices emerge as being quite well supported (so long as all the relevant information is pulled together). Reagent choice can be largely decoupled from more detailed design and tackled as a simple constrained optimisation. Detailed evaluation is needed mainly to establish a feasible design given the reagent choice and will normally change the reagent choice only if a feasible design cannot be found. Reagent delivery system design and measurement system design are well established technologies. The consequences of these design choices need to be assessed in terms of cost/performance, but the detailed implementation need not be considered further. The practical control options are fairly well defined, but the choice of controller type and parameters is largely left to the designer's judgement.

Some design choices are weakly supported. The choice between batch and continuous treatment systems does not have clear guidelines but the only heuristic available suggests that a batch system would not often be considered. Mixing scheme design is governed by heuristics which are at best incomplete and at worst inappropriate. Properties of solid reagents are poorly defined.

In general the use of computer modelling is not well supported by the established design guides. No tools are provided which allow designs to be rejected rigorously without detailed analysis. Variability and uncertainty are acknowledged as major problems, but systematic design methods for addressing uncertainty are not given in any of the established guides.

The review of current practice confirms the need for

1. models suitable for computer simulation, particularly models of solid alkali reagents;
2. rigorous, generally applicable screening tools to evaluate *achievable* control performance for different mixing schemes;
3. systematic methods for dealing with uncertainty.

These needs are addressed in the rest of the thesis.

Chapter 3

Design tools

This chapter covers all aspects of the design tools developed and used in this research project. Tools are developed to address the needs for systematic design with uncertain nonlinear dynamic models and rigorous and efficient screening of design options which were discussed in the previous chapters.

The tools presented are primarily based on optimisation of continuous design variables. Continuous optimisation methods are examined and appropriate optimisation methods identified (3.1). The specific problems associated with dynamic optimisation are then discussed and a robust and efficient dynamic optimisation method developed in this project is presented (3.2). Design with uncertainty poses special problems. A new algorithm for solving optimisation problems with uncertainty is presented in section (3.3). Methods for screening designs rigorously and efficiently are then discussed (3.4). In particular, a new method for analysing the effect of dead time on achievable disturbance rejection is presented (3.4.2). Finally, section 3.5 gives a brief outline of the software developed to implement these tools.

3.1 Optimisation with continuous variables

Optimisation-based design techniques are centred on the solution of mathematically posed optimisation problems the results of which provide the design engineer with information on which to base design decisions. The basic optimisation problem of interest is the minimisation of an objective subject to both equality and inequality constraints, where the objective and constraints may be nonlinear. This is known as a nonlinear program (NLP).

The general form for a NLP may be written (Gill *et al.*, 1981) as

$$\begin{aligned} \min_{\theta \in \Re^n} \quad & J(\theta) \\ \textit{subject to} \quad & c_i(\theta) = 0, \quad i = 1, 2, \ldots, meq \\ & c_i(\theta) \leq 0, \quad i = meq + 1, \ldots, m \end{aligned} \tag{3.1}$$

where θ is a vector of optimisation variables, J is the objective function to be minimised, and c is a vector of m constraints, of which the first meq constraints are equality constraints.

Techniques for solving this problem may be divided into local methods which aim to a find a single local minimum and global methods which aim to find the best local minimum.

3.1.1 Local solution

In discussing solution techniques it is useful to start with some definitions and assumptions.

The following terminology will be used.

1. *Active constraints* are equality constraints and inequality constraints equal to zero at the optimum.
2. $\hat{A}(\theta)$ denotes the matrix whose rows are the transposed gradient vectors of the active constraints.
3. $Z(\theta)$ denotes a matrix whose columns form a basis set for the vector space defined by $\hat{A}p = 0$.
4. The *Lagrangian* function is $L(\theta, \lambda) = J(\theta) + \lambda^T c(\theta)$, where λ is a vector of Lagrange multipliers.
5. The Hessian of the Lagrangian is denoted by $H(\theta, \lambda)$.

Unless otherwise stated it is assumed that

1. the objective function and constraints are C^2 continuous, i.e. derivatives up to and including the second exist and are continuous;
2. the gradients of active constraints with respect to the optimisation variables are linearly independent (constraint qualification).

A local solution of a NLP is a feasible point from which no small variation of the optimisation variables can give another feasible point with an improved objective. Provided the assumptions above are satisfied, necessary conditions for a local minimum (θ^*) are given by the Karush–Kuhn–Tucker (KKT) conditions

$$
\begin{aligned}
&\exists\lambda^{*} \textit{such that} \\
\nabla_{\theta}L(\theta^{*},\lambda^{*}) &= 0 \\
c_i(\theta^{*}) &= 0, \quad i = 1,\ldots,meq \\
c_i(\theta^{*}) &\leq 0, \quad i = meq+1,\ldots,m \\
\lambda_i^{*}c_i(\theta^{*}) &= 0, \quad i = meq+1,\ldots,m \\
\lambda_i^{*} &\geq 0, \quad i = meq+1,\ldots,m \\
p^{T}Z(\theta^{*})^{T}H(\theta^{*},\lambda^{*})Z(\theta^{*})p &\geq 0 \quad \forall p
\end{aligned}
\tag{3.2}
$$

If $Z(\theta^{*})H(\theta^{*},\lambda^{*})Z(\theta^{*})$ is positive definite, i.e. the last equation is a strict inequality, then these conditions are sufficient for a local minimum. All local optimisation algorithms attempt to generate a point satisfying the KKT conditions, though they do not usually guarantee that such a point has been obtained.

A very wide range of techniques is available for the local solution of nonlinear programs. Extensive reviews can be found in many textbooks with a particularly useful practical review in Practical Optimisation (Gill *et al.*, 1981). The key consideration in the choice of a method for optimisation with dynamic models is the number of function and gradient evaluations required. The most efficient methods in this respect are those which make use of second order information on the curvature of the performance indices.

Sequential quadratic programming methods have come to dominate solution methods for moderate sized (less than 100 variables) optimisation problems for which function evaluations are expensive/ time consuming. A detailed discussion of SQP methods and factors effecting their performance is given by Chen (1988). SQP methods centre on the solution of a quadratic program

$$
\begin{aligned}
\min_{\delta_k} \quad & \nabla_{\theta}J(\theta_{k-1})^{T}\delta_k + \tfrac{1}{2}\delta_k^{T}B_k\delta_k \\
\textit{subject to} \quad & c_i(\theta_{k-1}) + \nabla_{\theta}c_i(\theta_{k-1})^{T}\delta_k = 0 \quad i = 1,\ldots,meq \\
& c_i(\theta_{k-1}) + \nabla_{\theta}c_i(\theta_{k-1})^{T}\delta_k \leq 0 \quad i = meq+1,\ldots,m
\end{aligned}
\tag{3.3}
$$

at each iteration to generate a search step, δ_k, which may move the optimisation variables closer to a KKT point. This method may be interpreted as taking a Newton step with respect to the system of nonlinear equations defining the KKT optimality conditions, provided the QP active set is correct and B_k equals the Hessian of the Lagrangian.

As it is generally computationally expensive to calculate the Hessian directly a quasi-Newton formula is used to update a positive definite approximation of the Hessian as the search progresses, given an initial estimate. The local convergence properties of this algorithm with suitable choice of B_k (e.g. a quasi-Newton update based on changes in the estimated Lagrangian) have been shown to be very good. Some form of line search or trust region method must be applied to ensure that the optimisation progresses even when the full step, δ_k, given by the solution of the quadratic program does not give an improved point.

The specific code used in this project is the SRQPD code written by C. L. Chen (1988). Based on testing reported by Chen this code is at least comparable to and possibly more robust than routines such as VMCWD (Powell) and NLPQL (Schittowski). It is also accessible and portable. The code comes in two versions, using sparse and dense jacobian representations. The dense code is considered adequate for this project as the problems tackled do not have more than about 20 optimisation variables. A number of aspects of the code are discussed below. It should be noted that the discussion is based on the latest version of the code which has been updated since Chen's thesis (Chen, 1988).

If infeasible QP subproblems arise due to no feasible point within the linearised constraints, the code provides three relaxed QP formulations which may allow the search to proceed. These are documented in Chen (1988). This facility is vital for a robust code and was not altered in this project.

Chen's code provides a wide range of merit functions for implementing the necessary tradeoff between constraint satisfaction and objective minimisation during the line search or trust region search phases of the optimisation. The performance of the optimiser does not appear to be sensitive to these options and an exact penalty function line search is used as a default.

$$P_e(\theta) = J(\theta) + \sum_{i=1}^{meq} \mu_i |c_i(\theta)| + \sum_{i=meq+1}^{m} \mu_i \max(0, c_i(\theta)) \tag{3.4}$$

The code uses a different penalty parameter update to that given by Chen (1988) in that Powell's update formula

$$\mu_{i_k} = \max(|\lambda_{i_k}|, \frac{1}{2}(|\lambda_{i_k}| + \mu_{i_{k-1}})) \tag{3.5}$$

is replaced by

$$\begin{array}{ll} \mu_{i_k} = \max(0, \lambda_{i_k} c_{i_k}) & i = 1, \ldots, meq \\ \mu_{i_k} = \max(\max_i |\lambda_{i_k}|, \mu_{i_{k-1}}) & i = meq + 1, m \end{array} . \tag{3.6}$$

If this does not give a descent direction then

$$\mu_{i_k} = |\lambda_{i_k}| \tag{3.7}$$

is used.

This modification has several apparent disadvantages.

1. All constraints are weighted equally regardless of the estimated sensitivity of the objective function to their variations. This could make line searches less efficient on difficult problems.

2. The use of large penalty factors on constraints with small Lagrange multipliers amplifies noise on those constraints unnecessarily.

Discussion with Chen has not yielded clear justification for this change for *general* problems, though it appears to have some benefit on the flowsheeting problems per considered. The version of the code used in this project uses Powell's update (equation 3.5) as in Chen (1988).

The line search method in the code reduces the QP step, δ_k, by a decreasing factor, α, until any of the following conditions is met.

1. The merit function (by default P_e), decreases sufficiently as judged by a cone condition ($\Delta P_e \leq .1\alpha\nabla_\theta P_e^T\delta_k$).

2. The estimated Lagrangian decreases.

3. A merit function is obtained which is no worse than that given by the previous best J, c pair *and* a reduced objective or reduced constraint violation norm (compared to the values at the previous iteration) is obtained.

The second and third conditions are used to encourage full steps, as this is required for good local convergence properties. The line search method is discussed further in section 3.2.2.3.

Hessian initialisation was found to be a key factor in performance of the examples considered by Chen (1988). The best strategy was found to be

$$(B_0)_{ii} = \max\left(\epsilon, \left| \frac{\frac{\partial J(\theta_0)}{\partial \theta_i}}{\theta_i^+} \right|\right) \tag{3.8}$$

where θ_i^+ is $|\theta_i|$ if $|\theta_i|$ is greater than ϵ and one otherwise. ϵ is the square root of the machine precision. Chen did not use this strategy as a default as per believed it vulnerable to gradient inaccuracies. Using the gradient evaluation method discussed in section 3.2 it is unlikely that noise problems would arise and this strategy is therefore used. Qualitative consideration of the strategy shows that it gives an initial Hessian allowing steps of length about θ_i^+ to be taken in each variable in the initial QP subject to the linearised constraints. If the variables are also scaled to lie between .5 and 1.5 (see below) then an initial step of the order of the variable range may be taken. This means that the first step in the optimisation is primarily determined by the objective and constraint gradients and constraint values, rather than by the initial Hessian. This appears appropriate. It should be noted that initialising to the identity matrix will discourage changes in variables with small gradients.

The termination condition of the optimisation is

$$\begin{aligned}\max[\quad & \sum_{i=1}^{meq} |c_i| + \sum_{i=meq+1}^{m} \max(0, c_i), \\ & (|\nabla_x J^T \delta_k| + \textstyle\sum_{i=1}^{meq} |\lambda_i c_i| + \sum_{i=meq+1}^{m} \lambda_i \max(0, c_i))/(1 + |J|)] \leq \mathit{optacc}\end{aligned} \tag{3.9}$$

This condition ensures that constraint violations are below a level regarded as significant in itself and that the predicted change in the objective based on the QP step or reducing constraint violations to zero is small. No explicit gradient test is carried out and the step size is not itself required to be small. A value of *optacc* of 10^{-5} has been found to force convergence to optimal solutions for problems with known solutions and to give sensible answers for other problems. Values of *optacc* larger than 10^{-3} often allow termination of the search far from the optimum. Values below 10^{-5} appear unnecessary for well-scaled problems. 10^{-5} has therefore been adopted as a default value for *optacc*. Values of $|J|$ much less than one at the optimum make the test on the predicted change in the objective an absolute rather than a relative test and may lead to premature termination. This potential problem is eliminated by appropriate scaling.

Scaling of the problem is observed to have a considerable effect (Chen, 1988; Gill *et al.*, 1981) on performance of NLP solution algorithms. An engineering problem posed in its natural units may be poorly scaled. Good scaling is more or less ensured (Gill *et al.*, 1981) by

1. scaling the objective so that its value near the optimum is about 1;
2. scaling the design variables so that their value near the optimum is about 1;
3. scaling the constraints so that violations have approximately equal weight and that deviations within the solution accuracy are insignificant.

It was found necessary in the case studies to follow these guidelines, at least approximately, to obtain good convergence properties. In this work, all design variables are scaled so that their bounds correspond to 0.5 to 1.5 as seen by the optimiser, though the models may still use conventional engineering units. This scaling is similar to that tested by Biegler and Cuthrell (1985) and found to be useful on the steady-state optimisation problems considered. Scaling issues specific to dynamic models are discussed in section 3.2.2.2.

Multiple local optima may exist, unless certain *convexity* properties are satisfied (Bazaara and Shetty, 1979). The problems of interest are expected to be non-convex.

Multiple local optima have been identified in some case studies. This motivates the consideration of global solution techniques.

3.1.2 Global solution

Finding the global optimum of a NLP is a much more difficult problem than finding a local optimum, if the problem is not convex. A very useful recent review of unconstrained global optimisation for general functions is given by Törn and Žilinskas (1989) and provides the basis for the observations below. The difficulty of global optimisation centres on the need to find the best local minimum while having no general criterion for recognising when this has been obtained. A large number of methods is available but there is no consensus on the best general purpose method. Passive enumeration of grid or random search points is reliable but extremely inefficient. These search techniques may be augmented by local searches in several ways.

1. A local search is carried out to improve the best random or grid point (single-start).
2. A local search is carried out from each random or grid point (passive multi-start).
3. A local search is carried out when random search or grid enumeration identifies a better point than the best local minimum so far (active multi-start).

Passive multi-start will tend to repeatedly converge to the same local minima which wastes computational effort spent in redundant local searches. Active multi-start guarantees an improved local minimum for each local search but makes limited use of information gathered in the global/random phase of the search. A group of methods making more use of the available information is *clustering* methods. These work by identifying clusters of points associated with distinct local minima and initiating local searches in the distinct regions identified. Clustering is carried out by retaining points with relatively good function values or by taking a few steps of a local algorithm from each point.

Other methods are discussed by Törn and Žilinskas but do not appear to have the general applicability of the above group of techniques. The review notes that single-start and multi-start modifications of random search techniques are often used to solve practical problems.

A limitation of these approaches is that they do not in general provide guarantees of solution accuracy. This is unfortunate but as methods which provide such guarantees are much more inefficient there is no real choice.

> From a practical point of view the problem may be stated in a different way: There exists a goal (e.g. to find as small a value of f(.) as possible), there exist

> resources (e.g. some number of trials), and the problem is how to use these resources in an optimal way (Törn and Žilinskas, 1989).

Heuristics may be usefully employed to this end.

> [One] would expect a global optimisation algorithm to be neither purely heuristic nor purely mathematical because of the complexity of the problem to be solved, but to be a proper composition of both (Törn and Žilinskas, 1989).

Extension to constrained problems may be made by using a penalty function on the global phase of the search and using a standard local search method.

Visweswaran *et. al.* (1990) has developed global optimisation strategies which can exploit problem structure. Unfortunately the required structural information is not available for the problems of interest here as the objective or constraints are normally generated by integration of a dynamic model.

Some form of active multi-start method with heuristics applied to improve efficiency appears to be most appropriate for those problems in this work where it is desired to approximate global optima.

This is most important in the procedure for design with uncertainty as it is desired that the design should satisfy the requirements for all allowable values of the uncertain parameters. That is, the *global* maximum of the inequality constraints with respect to the uncertain parameters should be less than zero. Typical problems do exhibit multiple local maxima. This is discussed further in section 3.3.

For design optimisations it is of course *desirable* that the global optimum be obtained. Failure to find the global optimum for the design variables however is not as critical as failing to find the global maximum of the constraints, as the former implies the cost may be greater than necessary while the latter may imply the process will not meet its requirements at all. For design variables, sensible initialisation and bounding of the design variables will enhance the chances of a global optimum. If qualitative examination of the design indicates surprising characteristics or if adequate computer resources are available then a multi-start strategy may be useful. Physical insight could be used to guide selection of starting points. *Most* design problems examined do not seem to have multiple local optima.

3.2 Optimisation of dynamic systems

This section addresses the special problems arising from optimisation using dynamic models. The optimisation problem of interest is

$$
\begin{aligned}
\min_{\theta} J(x(t_f), z(t_f), \theta) \quad & s.t. \\
f(\dot{x}, x, z, \theta, t) &= 0 \\
g(x, z, \theta, t) &= 0 \\
q(x, z, \theta, t) &\leq 0 \\
L(x(\phi), z(\phi), \theta, \phi) &= 0 \ \forall \phi \in \Phi \\
M(x(\phi), z(\phi), \theta, \phi) &\leq 0 \ \forall \phi \in \Phi \\
0 \leq t \leq t_f & \\
\Phi \subset t &
\end{aligned}
\tag{3.10}
$$

where, J is the objective of the optimisation, θ is a vector of optimisation parameters, f and g define a DAE model with states 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 subset of time, Φ.

3.2.1 Choice of method

Two approaches can be taken to formulating NLP problems based on dynamic models.

1. Feasible path approaches

   In this approach the differential equations are solved to generate the performance indices for each choice of design variables. The results, along with gradients if required, are then passed to the optimiser to update the design variables. This approach has been applied to chemical engineering problems within the Chemical Engineering Department at Imperial College (Morison, 1984; Howell, 1984; Gritsis, 1990; Vassiliadis, 1993) and elsewhere (McAuley and MacGregor, 1992) with considerable success.

2. Infeasible path approaches

   In this approach the dynamic equations are discretised to give a set of algebraic equations, e.g. using orthogonal collocation on finite elements, which are then included as equality constraints in a NLP problem. This problem has many added variables but the same number of degrees of freedom as the original problem and is therefore potentially soluble with techniques such as projection methods or sequential reduced quadratic programming (Chen, 1988). An advanced implementation of this approach with a focus on chemical engineering problems is that of Biegler (Biegler, 1990; Vasantharajan and Biegler, 1990; Logsdon and Biegler, 1993).

In choosing between these approaches a number of factors were considered.

1. Ease of solution

   The feasible path approach splits the problem into two subproblems

   (a) Integration of DAEs
   (b) Solution of small to moderate (less than 100 variables) size NLPs

   Each of these subproblems has been extensively researched and can be solved using standard, well-established software.

   To solve problems using an infeasible path strategy it is necessary to use more specialised optimisation codes for large-scale optimisation and to develop new error control strategies for the DAEs (Vasantharajan and Biegler, 1990; Logsdon and Biegler, 1993).

   This factor strongly favours feasible path approaches.

2. Efficiency of solution

   Comparisons of feasible path and infeasible path methods by Vasantharajan and Biegler (1990) indicate that the two methods have comparable efficiency on problems for which both were found to work.

   This factor does not allow any conclusion to be drawn.

3. Generality of method

   Leaving aside ease of solution, it may be argued (Vasantharajan and Biegler, 1990) that infeasible path techniques can solve problems with constraints on the value of the model variables over all time (path constraints) more effectively than feasible path methods. This argument is based on the capability of infeasible path methods to translate the path constraints to point constraints at each of the discretised points on the state trajectory so that the extent to which the constraint is not violated is visible. In the feasible path approach the path constraint is normally converted to a terminal constraint requiring that the integrated constraint violations be zero which provides no information about proximity to a constraint violation. An example which appears to support this is given by Vasantharajan and Biegler (1990). As discussed in 3.2.2 this example is solved successfully by a feasible path method developed in this project. On the conceptual level it should be noted that transformation of a path constraint to a set of interior point constraints may be carried out by sampling a feasible path solution so that there is no fundamental divide on this issue. Path constraints are discussed further in section 3.2.2.2.

This factor does not allow any conclusion to be drawn.

4. Usefulness of intermediate results

   If the problem fails to converge to the optimum the intermediate solution for the infeasible path method may not satisfy the DAEs and may therefore be meaningless. Failure to converge with the feasible path method should usually give a meaningful improved design.

   This factor favours feasible path methods.

Feasible path methods therefore appear preferable at present, though it is possible that as infeasible path methods mature they will become more competitive. *The feasible path approach is therefore used in this project.*

Having decided on a feasible path method, it is necessary to consider methods for generating performance index gradients. The three major approaches to generating the required gradients are

1. finite differences;
2. backward integration of an adjoint system for each state dependent constraint or objective function (Morison, 1984; Gritsis, 1990);
3. evaluation of the sensitivity equations in parallel with the forward integration using efficient methods which avoid the need to explicitly integrate the sensitivity equations (Leis and Kramer, 1985; Caracotsios and Stewart, 1985).

The finite differences approach can give very inaccurate gradients without careful, problem dependent, choice of perturbation sizes and is relatively inefficient for more than about 4 parameters. It can therefore be disregarded if either of the other methods can be applied. The adjoint approach has a major disadvantage if sensitivity information for a number of state dependent functions is needed in that it requires a separate backward integration for each function. Additionally, the adjoint equations may be more difficult to integrate than the original model equations (Rosen and Luus, 1991). For the sensitivity equation approach a single forward integration generates all the required gradients and the integration difficulty is not normally affected. The adjoint approach has some special advantages for piecewise constant optimal control problems as the number of equations to be integrated to evaluate the gradients is particularly small (Rosen and Luus, 1991) but otherwise the sensitivity-based approach seems marginally preferable. *The sensitivity equation approach is used in this work* and is discussed in more detail in appendix A.

Dealing with noise in the solution of the dynamic model is an important issue which does not seem to be systematically dealt with in the literature. This issue is addressed in section 3.2.2.

An alternative to applying a conventional gradient-based NLP solution method to the performance indices generated as above is to pose and solve a variational problem based on Pontryagin's maximum principle. This was used by Nishida *et al.* (1976) to solve a robust design problem subject to simple bounds and equality constraints on the design variables. This approach is shown to be efficient, at least for simple problems, and a method for assessing the optimality of the solution with respect to extra design variables which were originally fixed is presented. The requirements given for the underlying model are restrictive as the technique presented requires ordinary differential equations with twice differentiable right hand side. This type of approach is observed by Biegler (1990) to be limited to the simplest optimal control problems (path or final time constraints require an additional outer loop making the approach inefficient) and will not be considered further.

3.2.2 Robust and efficient feasible path dynamic optimisation

This section presents some methods developed in this project for improving the robustness and efficiency of feasible path dynamic optimisation codes.

3.2.2.1 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 tolerances 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 steady-state tolerance for the precision to which the equations must be satisfied on initialisation of the equations (*sstol*). *rtol* and *atol* are generally combined to give an error control weighting, $w_i = rtol|y_i| + atol$ and are usually given the same value. For feasible path dynamic optimisation these tolerances must all be chosen so that

1. the numerical solution of the equations does not in itself introduce a significant modelling error;

2. noise or precision error does not seriously disrupt progress of the optimisation, 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 modelling 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 values of the optimisation parameters.

Morison (1984) gives a heuristic that the local error tolerance on the integration should be an order of magnitude below the optimiser accuracy. This is implausibly simple as the relationship between local error control and model variable noise is not straightforward. In case studies, it was found that this heuristic does not give reliable performance and that the relationship between local error tolerances and solution noise may vary substantially with the model parameters. In one example noise led to an optimisation run failing to progress after the search led to a region in which limit cycle behaviour occurred causing increased levels of noise. Tightening the local error tolerances for the integration resolved this problem. The local integration tolerance required may vary by at least several orders of magnitude over the course of the optimisation as well as varying substantially from problem to problem.

It is therefore desirable to have some systematic means of estimating and adjusting the precision error to optimise the tradeoff between the requirements for low noise and efficient integration. This requires that the precision error be estimated, its effect on the optimisation assessed and the integrator tolerances adjusted appropriately.

Estimating precision error When using the sensitivity method (Appendix A) the system equations are reintegrated during the gradient evaluation following a successful line search by the optimiser. 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 was estimated using the formula

$$\epsilon(y) = \frac{\max_i |\Delta^k y_i|}{\sqrt{\frac{(2k)!}{(k!)^2}}} \tag{3.11}$$

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

$$\delta(y) = |y^* - y| \tag{3.12}$$

was comparable to $\epsilon(y)$. The values often agreed to within about 20%, with $\delta(y)$ usually being greater than $\epsilon(y)$. As $\epsilon(y)$ requires about 6 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 only addresses 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 $\epsilon(\frac{\partial y}{\partial \theta_i})/|\frac{\partial y}{\partial \theta_i}|$ is indeed comparable to $\epsilon(y)/|y|$. This experience is all based on sensibly scaled problems with values of y and $\frac{\partial y}{\partial \theta_i}$ within a few orders of magnitude of one, 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 then 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 $\epsilon(\frac{\partial y}{\partial \theta_i})$.

Evaluating effect on optimiser The estimated error in objective and constraint values may be mapped through calculations in the optimiser to determine whether the noise is likely to impede progress. The appropriate measure will depend to some extent on the optimisation code used. Noise on the objective and constraints may interfere with the evaluation of the optimiser termination conditions or with comparisons between performance with different values of the optimisation variables, e.g. during a line search. Noise on 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.

The SQP termination condition used for a constrained optimisation problem is (equation 3.9)

$$\max[\quad \sum_{i=1}^{meq} |c_i| + \sum_{i=meq+1}^{m} \max(0, c_i),$$
$$(|\nabla_x J^T \delta_k| + \textstyle\sum_{i=1}^{meq} |\lambda_i c_i| + \sum_{i=meq+1}^{m} \lambda_i \max(0, c_i))/(1 + |J|)] \quad \leq optacc$$

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 optimisation accuracy.

Comparisons between points are based on the exact penalty merit function (equation 3.4).

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

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

Controlling the combined precision error

$$error = \tfrac{\delta(J)}{1+|J|} + \textstyle\sum_{i=1}^{meq} \max(1, |\tfrac{\lambda_i}{1+|J|}|)\delta(|c_i|) + \sum_{i=meq+1}^{m} max(1, |\tfrac{\lambda_i}{1+|J|}|)\delta(max(0, c_i)) \tag{3.13}$$

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 below, has not been found necessary.

Adaptation of integrator error tolerances The value of *error* can be used for feedback control of the integrator tolerances. As noted above, there are often four tolerances to consider. The general practice of equating the relative and absolute tolerances, *atol* and *rtol*, was adopted as there was no clear general criterion for using different values. In dynamic problems, the steady-state and event tolerances, *sstol* and *evtol* make a fairly minor contribution to total computational cost, and so can be set to stringent *a priori* values, say 10^{-10}. This approach leaves a single parameter *simacc* to be chosen to determine *rtol* and *atol*.

The mean value of *error* may be expected to increase monotonically as *simacc* increases and 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 cautious to avoid instability. The scheme below has been found to be effective in extensive trials.

if $(error > 2.optacc/factor)$ $simacc = simacc/10$
if $(error < optacc/factor)$ $simacc = simacc.\mathrm{min}(1.5, (\frac{optacc}{error.factor})^{\frac{1}{5}})$

The use of a deadband in the adaptation and the use of rapid decrease and slow increase of *simacc* gives a good compromise between stability of adaptation and speed of response. *factor* allows the tradeoff between robustness and efficiency to be tuned. A value of 5 has been found to be adequate. To avoid the error control correction 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 not requiring significant extra computation to implement and not requiring the use of very tight error control on the integrator. The number of iterations required for optimisation was found not to differ significantly from those 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, analysed or controlled. If problems did occur for this reason, the parameter *factor* could be increased, or fixed values of the integrator tolerances could be chosen by trial and error.

It should be noted that the method should not be used for purely algebraic problems as the gradient evaluation is completely decoupled from the equation solution in this case.

3.2.2.2 Path constraint representation

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

$$q(x, z, \theta, t) \leq 0 \; \forall t \in [0, t_f] \tag{3.14}$$

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

$$\int_0^{t_f} max(0, q(x, z, \theta, t))^2 dt = 0 \tag{3.15}$$

This representation is single-valued, differentiable and precisely equivalent to the original constraint, but 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 optimiser.

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

A simple modification of the path constraint representation above, is given by

$$\int_0^{t_f} max(0, q(x, z, \theta, t))^2 dt \leq c_{sig} \tag{3.16}$$

This representation makes slightly feasible constraint values visible to the optimiser, and reduces noise problems by avoiding a small constraint being multiplied by large Lagrange multipliers or penalty factors within the optimiser 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 choice of c_{sig} is application dependent, but the examples below illustrate some methods of choosing c_{sig}. A typical path constraint q is that a particular variable, say pH, should be less than some limit, say 9, for all time. A typical measurement range for pH is 0–14 so that the smallest change resolvable on a 10-bit A/D signal is .014pH. Changes below this level can therefore be considered insignificant. For integrated violation representations it is necessary to consider the time for which such a violation is insignificant. In general one second — the fastest sampling rate on most computer control systems — may be assumed to be a small time. By this reasoning a c_{sig} of .0002 $(1 * .014^2)$ would be appropriate for the integral squared pH trajectory. This approach corresponds to choosing c_{sig} as a violation so small that the control/monitoring systems will never notice it. This gives a sensible value of c_{sig} for constraints concerned with measured conformance to specification. An alternative approach is to consider how large a violation could be induced by trivial changes in the model, e.g. a 1% variation in a time constant, and ensure that c_{sig} is less than this induced violation. This approach corresponds to ensuring that the effect of the choice of c_{sig} is small compared to modelling approximations. This approach is more general than that based on measured conformance.

Both approaches aim to give a solution which *in relation to the real-world problem* does not differ significantly from a solution based on exact satisfaction of constraints.

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

$$\dot{x} = max(0, q(x, z, \theta, t))^2 / c_{sig} \quad , x(0) = 0 \tag{3.17}$$

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

The optimiser uses the feasibility test $\sum |c| \leq optacc$ (equation 3.9). The model generates a path constraint, $x(t_f)$, for which a value of 1 represents an insignificant constraint violation. Using

$$c = (x(t_f) - 1) optacc \leq 0 \tag{3.18}$$

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

$$c = (2x(t_f) - 1) optacc \leq 0 \tag{3.19}$$

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 optimisation.

Gritsis (1990) suggests another method for feasible path optimisation with path constraints in which the maximum value of the constrained variable is required to be less than zero:

$$\max_{t \in [0, t_f]} \{q(y, t)\} \leq 0 \tag{3.20}$$

This method overcomes the main weaknesses of the integral violation approach but is dubious, as a general method, unless non-differentiable optimisation techniques are used. Algorithms assuming differentiability may encounter difficulties as the first derivatives of the constraint are undefined when there are two or more equal local maxima of the constraint. If discontinuities are infrequent and do not cluster near the optimum then conventional local search methods which assume differentiability will usually succeed. When optimising subject to path constraints reducing one local maximum often implies increasing another which tends to give closely spaced maxima near the optimum. This procedure is therefore not generally robust and cannot replace the integrated violation method for the general case. However, in the particular case where the constraint is being maximised in the overall optimisation so that there is reason to expect just one local maximum in the vicinity of the optimum then this approach should be effective. It allows a local constraint maximisation to choose a sensible search direction from an initial point in which no path constraints are actually violated. It has the additional advantage of removing the integral path constraint violation equation from the model which may make the model integration quicker.

3.2.2.3 Poor local approximation

Local approximations (linear or quadratic) are often particularly poor in dynamic optimisation problems. For instance, this situation is found to occur when taking the full step predicted from the local approximation, δ_k, results in either a path constraint becoming active or the system becoming 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 then it may reduce the step magnitude much more than necessary and be unable to improve the local approximation. This can result in many full iterations of the optimisation being required 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, α_{min}, while the previous point, α_{j-1}, defines a maximum step size α_{max}. The line search then proceeds as below.

```
if $\alpha_{max}/\alpha_{min} > 2$ ifor=1
do until ifor=0
    $\alpha_{j+1} = \min(2\alpha_j, (\alpha_j + \alpha_{max})/2)$
    $\theta_{j+1} = \theta_{k-1} + \alpha_{j+1}\delta_k$
    Evaluate $J, c$ for new $\theta$
    If line search criteria satisfied and merit function reduced then
        if $\alpha_{max}/\alpha_{j+1} > 2$
            j=j+1
        else
            $\theta_k = \theta_{j+1}$
            ifor=0
        end if
    else
        $\theta_k = \theta_j$
        ifor=0
    end if
end do
```

This algorithm essentially tries to locate the largest step acceptable to the line search criteria, α^*, within a factor of two, assuming that the criteria are satisfied 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, Murray and Wright (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.

3.2.2.4 Failure tolerance

During an optimisation it is quite possible that the integrator will take an abnormal exit from an integration run. This *may* indicate a fundamental flaw in the model but is more often due to design variables entering a region giving rise to instability or initialisation difficulties. A heuristic response to this is to assume that regions for which the integrator cannot solve the equations are not desirable regions to operate and to try to force the optimiser to retreat from such regions and continue the search. This can be achieved by ensuring that, if an abnormal termination occurs *during a line search* for which this heuristic seems reasonable, large positive values are placed in the constraints and objective so that the line search will retreat from this point. This has proved effective in practice in allowing a local minimum to be found. If this strategy fails then the model formulation and optimisation variable bounds require careful review.

The assumption behind these measures is that whenever possible the optimiser should continue despite numerical problems in the integrator.

3.2.2.5 Avoiding discontinuities

> [P]resumably no one would deliberately include significant discontinuities in the modelling function or its derivatives (Gill *et al.*, 1981).

Discontinuities are unfortunately part of the real world and pose particular problems for numerical integration and optimisation software. Optimisation methods assuming differentiability are generally much more efficient than algorithms for non-differentiable optimisation, but are less robust to discontinuities in the derivatives. The standard predictor-corrector methods for integration such as BDF and Adams-Moulton work best on smooth variable trajectories and are forced to reinitialise and reduce order to traverse discontinuities. Sensitivity evaluation may fail if the values of the derivatives of the states are discontinuous and the discontinuity time is a function of the parameters, unless special measures are taken (Appendix A). C^2 continuity is generally sufficient to avoid difficulties with the above methods and C^1 continuity will avoid many problems.

As with path constraint formulation it is desirable to exploit the fact that a perfect match to the real-world is neither practical nor necessary to improve the mathematical properties of the model. A common example of a real discontinuity is integral desaturation in a controller. A typical integral desaturation strategy is to freeze the integrated error when the scaled controller output, u, reaches the limits 0 or 1 (integral freezing). This gives a discontinuity in the derivative of the integral error which would disrupt gradient evaluation (Appendix A) and a discontinuity in the controlled variable trajectory which would slow down the integration. It can also result in continuous cycling across the discontinuity if freezing the integral leads to the output coming off the limit and unfreezing the integral pushes the output back to the limit. This has been found to occasionally effectively halt the integration as the integrator repeatedly locates the discontinuity and reinitialises. A simple method for dealing with this, which should not differ from integral freezing excessively, is to replace the integral freezing equations

$$\begin{aligned} ie &= \int K.error \\ K &= 1 \quad \forall u \in [0,1] \\ K &= 0 \quad \forall u \notin [0,1] \end{aligned} \tag{3.21}$$

by

$$\begin{array}{ll} ie = \int K.error & \\ K = 1 & \forall u \in [\epsilon, 1-\epsilon] \\ K = \frac{3}{\epsilon^2}u^2 - \frac{2}{\epsilon^3}u^3 & \forall u \in [0, \epsilon] \\ K = \frac{3}{\epsilon^2}(1-u)^2 - \frac{2}{\epsilon^3}(1-u)^3 & \forall u \in [1-\epsilon, 1] \\ K = 0 & \forall u \notin [0, 1] \end{array} \tag{3.22}$$

where ie is the integrated error and ϵ is a small number typically between .01 and .05. The above approximation gives a quite respectable desaturation strategy while being C^1 continuous and thereby avoiding the problems with full integral freezing.

Another major source of discontinuities lies in model approximations. The use of piecewise linear approximations to, for example, a titration curve, may force frequent reinitialisations of the model, which slow down the integration. This can be avoided by using an appropriately smooth interpolation method to interpolate values. Cubic splines may be used in some cases, but may introduce undesirable non-monotonic characteristics even given a monotonic set of breakpoints. This disadvantage is overcome by Fritsch and Butland (1984) who give a method for constructing local monotone piecewise cubic interpolants with C^1 continuity. Their method was implemented and used for titration curve representation in this project.

3.2.2.6 Dynamic response requirements

It is invariably a design requirement for control systems that they should be stable at least in the sense that bounded inputs do not give rise to unbounded outputs. For example, for a pH control system it may be required that the response in the absence of disturbances should not exhibit limit cycles over 1 pH in amplitude.

It may also be required that the system should return to a steady state or some approximation of this after disturbances have passed. This is desirable in terms of meaningful results, if nothing else, in that if the system does not return to roughly the condition it started from then the results become questionable.

There are a number of possibilities for implementing these requirements.

1. Enforce a stability margin based on a linearised model. This can be useful, but has the disadvantage that for nonlinear systems it may be too strict or too lax. Firstly, local stability may be a tighter constraint than requiring limit cycles to be small. Secondly, as is often observed in pH systems, local stability does not preclude large amplitude limit cycles.

2. Require that the integral absolute/squared error over the last part of the trajectory (after disturbances have passed) be small.
3. Require that the trajectories satisfy tapered rather than constant path constraints.

All the above methods have been tried, and there appears to be little to choose between the last two, except that the shaped path constraint method is the more flexible method in that any trajectory envelope can be imposed precisely.

In order to start the process from its true resting state rather than a possibly artificial exact steady state, the process may be subjected to a small disturbance to excite any natural limit cycle modes before being subjected to the main disturbances of interest. This is not necessary if the response is forced to return to close to a steady state in some way.

3.2.2.7 Testing

The main test of the algorithm proposed was in its success in handling the problems discussed in chapter 6. The examples discussed in that chapter include a wide range of problems, most involving active path constraints. It is interesting to note that problem 5.1 presented by Vasantharajan and Biegler (1990) illustrating some difficulties with feasible path methods was solved without any difficulty using a code based on the above principles. The code has also been used in controller structure selection (Narraway and Perkins, 1993; Narraway, 1992) with very high reliability.

3.3 Design with uncertainty

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 the relevant techniques from previous work and attempts to exploit the characteristics of the design problems.

3.3.1 Review

The design problems of interest in this project involve substantial uncertainty in the model on which the design is based. This has two main aspects. Firstly, there is uncertainty and variability in the parameters of the treatment system including measurement characteristics, reaction kinetics and titration curves. Secondly, the particular set of disturbances entering the plant at a given time may not be known and performance for all possible disturbance conditions should be considered. This gives rise to the need to develop designs which can meet the performance requirements robustly.

This review only considers methods potentially applicable to the design of wastewater treatment systems, i.e. optimisation-based methods applicable to general nonlinear systems. Algorithms designed for linear or bilinear systems are not reviewed. The uncertainty is assumed to be continuous, i.e. discrete events such as equipment failures are not included in the analysis.

3.3.1.1 Selection of approach

As noted by Grossmann *et al.* (1983) the problem of design with uncertainty is not well-defined and many different approaches exist. Optimisation-based approaches for design with uncertainty can be divided according to the method of handling constraint violations; penalise violations in objective, trade off probability of violation against cost explicitly, design to prevent violations.

Pai and Hughes (1987) apply a smooth penalty function to constraint violations and include this penalty in the objective to be minimised. A probability distribution is assumed for the uncertain variables. The tradeoff between risk and cost is then managed implicitly by minimising the expected cost. Construction of a suitable penalty function requires that violations of constraints should map smoothly to an economic penalty. It is not clear how to do this for the problems of interest, particularly for legal and safety constraints. Therefore, this approach will not be pursued in this project.

Another approach (Grossmann and Straub, 1991) is to construct an explicit tradeoff curve between a "flexibility index", such as the probability of feasible operation (stochastic flexibility), and the design cost. This curve may then be used to assist the designer in making the tradeoff between risk and cost. The curve has the virtue of simplicity, but does not capture the full complexity of the tradeoffs to be made. Some of the relevant issues are noted below.

1. Violations caused by fixed but uncertain parameters represent a risk that the process may never operate. Violations caused by variable parameters indicate that the plant will not operate as required some of the time. These are quite different risks.

2. Constraint violations which cause temporary out-of-spec operation and constraint violations which imply temporary or permanent shutdown have quite different implications for risk.

It is not clear how the designer can make a good tradeoff between cost and risk based simply on the curve provided. Therefore this approach will not be pursued in this project.

Design to prevent constraint violation is subject to the criticism (Pai and Hughes, 1987) that the cost of constraint violation is not infinite and that this design approach may, at least in principle, accept an arbitrary increase in cost in order to avoid a constraint violation. This problem is not unique to design with uncertainty, but occurs in any optimisation-based design which uses constraints. Constraints should be used where they are a better approximation to the design problem of interest than a term in the objective function. In some cases the constraint formulation is natural and straightforward, e.g. where any violation of a constraint may trigger loss of business or prosecution with costs well above the cost of avoiding the violation. In other cases the constraint formulation is tentative and subject to review, e.g. where the constraint represents an initial judgement of a desirable property based on information which is not readily incorporated in the design objective function. This is part of the normal iterative process in which a design specification may be modified in the light of design analysis.

Design to prevent constraint violation requires that the uncertainty should be considered to be bounded. It is impossible, in general, to avoid constraint violations if the possible parameter values are unbounded, e.g. subject to a normal distribution. In many cases, this requirement is not restrictive. Some examples of uncertainty sources for which a bounded uncertainty description is appropriate are given below.

1. The design specification may include explicit bounds on flows and compositions to be handled by the process.
2. Disturbance parameters may be bounded by the design and operation of the processes generating the disturbances.
3. Feed variability will often be bounded by quality control procedures which will specify bounds on feed properties.
4. Measurement error in normal operation may be bounded based on manufacturer's specifications and testing and potentially maintained within this bound by fault detection and maintenance procedures.

In other cases the definition of bounds does not come from the nature of the uncertainty but from the exercise of judgement, e.g. bounds on physical properties, model parameters and market forecasts. In these cases bound definition is a key part of risk management. There will usually be insufficient data for the bounds to be chosen based on a statistical confidence limit even if it was clear what this limit should be. As with constraints some of the bounds will need to be reviewed in the light of design results. In my experience this is a realistic approach which facilitates the exercise of design judgement and is closer to addressing the complex multi-dimensional tradeoff between risk and cost than other methods. Design to prevent constraint violations, known as worst-case design, is therefore adopted for this project.

3.3.1.2 Worst-case design

Two approaches to worst-case design which have been applied successfully to engineering design problems with uncertainty are those of Grossmann *et al.* (Grossmann *et al.*, 1983; Swaney and Grossmann, 1985a; Swaney and Grossmann, 1985b; Grossmann and Floudas, 1987) and Polak et al (Polak, 1982; Tits, 1985; Polak and Stimler, 1988; Mayne *et al.*, 1990). The problem which both approaches tackle is how to deal with an effectively infinite number of constraints (constraints must be satisfied for the infinite number of parameter realisations which must be considered) which makes the design optimisations *semi-infinite*. 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 (is an outer approximation to) the feasible region associated with $v \in V$.

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, $c^*$, 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, go to Step 2
```

The methods differ mainly in step 3.

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

$$\begin{aligned} &\min_{p\in P} \mathop{E}_{v\in V}\{\min_{o\in O} J(p,v,o) \mid c(p,v,o) \le 0\} \\ &s.t. \forall v \in V\{\exists o \in O(\forall k \in K[c_k(p,v,o) \le 0])\} \end{aligned} \tag{3.23}$$

where 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 which 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

$$\max_{v\in V} \min_{o\in O} \max_{k\in K}[c_k(p,v,o)] \le 0 \tag{3.24}$$

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

$$\begin{aligned} &\min_{p\in P} \mathop{E}_{v\in V}\{J(p,v)\} \\ &s.t.\ \forall v \in V\{\forall k \in K[c_k(p,v) \le 0]\} \end{aligned} \tag{3.25}$$

The formulation without operating variables is qualitatively easier to solve as one level of optimisation is eliminated. More specifically, finding v^* for the optimisation problem defined by equation 3.24 is a non-differentiable global optimisation problem which 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 there are variables which may be adjusted during plant commissioning or operation to give improved performance in the light of the actual plant behaviour. Requiring such variables to be chosen so as to accommodate all possible uncertain variables, i.e. as design variables, introduces an element of conservatism to 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 non-ideal way. A design generated using optimisation 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 should not therefore be used in determining design parameters, though they may be useful in certain screening tests if the resulting problem can be solved efficiently (see section 3.4). 6

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 this "would make the problem virtually unmanageable" at the design stage. In this project, 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 this project is that without operating variables. Solving general problems with operating variables is not necessary.

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 which are useful for this project.

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 maximise that particular constraint at its upper
limit, if the constraint were monotonic.

2: Solve for new design parameters, p^*, which minimise the
expected cost subject to the constraints associated with each $v_j \in V_i$
$\min_{p,o_j,j=1..N_i} \sum_{j=1}^{N_i} w_j J(p, o_j, v_j)$

s.t. $c(p, o_j, v_j) \leq 0, j = 1..N_i$
(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 minimise the maximum
constraint violation for the new design parameters.
$\min_{o \in O}\{c_{max} \mid c_{max} \geq c_k(p, v_j, o), \forall k\}$
(It is not necessary to carry out the minimisation exhaustively as
the minimisation 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, go to 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 which will not be satisfied in all the problems of interest in this project. 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. 1D- quasi-convexity implies that

$$\max(f(x_1), f(x_2)) \geq f(\alpha x_1 + (1-\alpha)x_2) \quad \forall \alpha \in [0,1], \forall \alpha_2 \in \mathcal{R}, \forall x_1, x_2 = x_1 + \alpha_2 e_i \tag{3.26}$$

where e_i is a vector with *ith* element 1 and all other elements 0. A non-vertex 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. Non-vertex 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 maximisers, 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; Swaney and Grossmann, 1985b). These are discussed below.

The flexibility index, F is defined by

$$\begin{aligned} F &= \max \delta \\ s.t.\ & \max_{v \in V_\delta} \min_{o \in O} \max_{k \in K} c_k(p, v, o) \leq 0 \\ V_\delta &= \{v \mid (v_0 - \delta \Delta v^-) \leq v \leq (v_0 + \delta \Delta v^+)\} \end{aligned} \tag{3.27}$$

where v_0 is the nominal value of v. F is therefore 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

$$\begin{aligned} F &= \min_{\tilde{v} \in \tilde{V}} \max_{\bar{\delta}, o} \bar{\delta} \\ s.t.\ & c(p, v, o) \leq 0 \ \forall \delta \in [0, \bar{\delta}] \\ v &= v_0 + \delta \tilde{v} \\ \tilde{V} &= \{\tilde{v} \mid -\Delta v^- \leq \tilde{v} \leq \Delta v^+\} \end{aligned} \tag{3.28}$$

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

$$\begin{aligned} F &= \min_{\tilde{v} \in \tilde{V}} \max_{\delta, o} \delta \\ s.t.\ & c(p, v, o) \leq 0 \\ v &= v_0 + \delta \tilde{v} \\ \tilde{V} &= \{\tilde{v} \mid -\Delta v^- \leq \tilde{v} \leq \Delta v^+\} \end{aligned} \tag{3.29}$$

This index could be used for trading off risk and cost but as argued above such indices oversimplify a complex tradeoff. In addition to the problems discussed previously 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 optimisation procedure to maximise constraint violation instead of minimising 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 in this project 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 $sign(\tilde{v}_i) = -sign(\frac{\partial \delta}{\partial v_i})$
until either $\delta$ fails to decrease as predicted or
the method predicts a vertex already examined

3: For each $c_k < 0$ compute the maximum increase $\Delta c_k$ 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 $\rho$ over all the constraints is less than $\rho^{max}$      ←   ??
or no new vertices are identified then STOP
Else select the new vertex value of $v$ giving the maximum value of $\rho$
and goto 2:
```

Swaney and Grossmann (1985b) state that setting $\rho^{max} < 0$ implies enumeration of the vertices. Enumeration would actually only be *guaranteed* if the search for a new constraint violation included all vertices rather than the predicted maximisers only. For example, the search above can terminate on a vertex which was a local maximiser of all the constraints as no new vertices would be identified for exploration. ρ^{max} does, however, provide a useful means to adjust the amount of effort applied to the local search.

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 non-vertex worst point.

Kabatek and Swaney (1992) present another method which attempts to identify variables, v, which are "compensatable" by o and to avoid exploring these variables.

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 which 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 MILP. For nonlinear constraints a modified approach is presented to decompose the problem into a series of smaller 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 are 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 which 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
>
> $$\min_{p}\{J(p)|c(p) \leq 0; \max_{v^i \in V^i} c^i(p, v^i) \leq 0, i \in \{1, \ldots, I\}\}$$

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 optimisation methods such as grid enumeration for solving the constraint maximisation problem, without operating variables, o.

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

$$c(p,v) \leq 0 \quad \forall v \in V \tag{3.30}$$

by

$$\tilde{c}(p,v) \leq 0 \quad \forall v \in V \tag{3.31}$$

where

$$c(p,v) \leq \tilde{c}(p,v) \quad \forall v \in V \tag{3.32}$$

This operation is referred to as majorisation 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 optimisation, 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 3.3.2.

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

None of the methods reviewed provide a complete solution to the type of problems tackled in this project. They suffer from either making assumptions which are not likely to be satisfied or relying on a degree of brute force which is computationally unrealistic. The review above has however highlighted a number of methods which could usefully be exploited in an algorithm for worst-case design.

1. Exploiting the heuristic that many local minimisers 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 maximisation as the optimisation progresses.

These ideas are used below in developing a new algorithm.

3.3.2 New algorithm

This section presents the new algorithm for design with uncertainty used in this project. The discussion below is supplemented by a pseudo-code listing in Appendix B. The discussion in this section should be read before the appendix.

3.3.2.1 General observations

No operating variables are included in the optimisation 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 optimisation problem. This gives the optimisation formulation

$$\min_{p} \{J(p,V) | \max_{v \in V} \max_{k \in K} [c_k(p,v)] \leq 0\} \tag{3.33}$$

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

$$J(p,V) = \epsilon | J(p,v) \leq \epsilon \ \forall v \in V \tag{3.34}$$

or a weighted cost

$$J(p,V) = \sum w_j J(p,v_j) \ \ , v_j \in V_i \tag{3.35}$$

have been used in this project. 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 that

1. The local search procedures are more likely to suffer problems due to discontinuities from either changes in the identified local minima with respect to o (discontinuity in value or first derivative) or 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 optimisation replaces an integration.
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 ($v \in V$). This is the simplest case and was found to be adequate for the case studies examined. The advantage of this description is that the boundaries of the uncertain parameter space, V, are easily identified and many constraint maximisers 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, though 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 in this project.

Vertex constraint maximisers are not *assumed*, though 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 with 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 should not take too many iterations to identify an outer approximation which is 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 maximisers 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 where many maximisers 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 heuristic used is to build a projection factor, ϵ_p, into the model such that increasing ϵ_p makes the design more conservative, expanding the subset of V for which the design is feasible.

The use of $\epsilon_p > 0$ should accelerate the achievement of a feasible solution at the expense of giving a design which has a cost between the optimum cost for the actual problem and the optimum cost for the "projected" problem. E.g., the design might be able to handle a throughput or disturbance amplitude of between the nominal value and $1+\epsilon_p$ times the nominal value. This introduces a sensibly bounded degree of conservatism to the design in order to obtain a more rapid solution. The projection factor ϵ_p may be large on initial evaluation of designs and be reduced for a more precise optimisation of the final design.

The key conditions for this heuristic to be appropriate are

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 does not give a net benefit. The heuristic was found to work effectively in the design examples (chapter 6) 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; initialisation, multi-model design and constraint maximisation. Constraint maximisation is presented first to assist the flow of information.

3.3.2.2 Constraint maximisation

The constraint maximisation algorithm makes use of the heuristic that the maximising values of the uncertain variables will often lie on the simple bounds. It also uses the idea of progressively increasing the amount of effort applied to the constraint maximisation up to a maximum level.

1. In the early iterations of the design local vertex searches as used by Grossmann and Swaney provide an efficient way to generate new maximisers and push the design towards the required robustness. Local searches from a subset of the vertices may be used to try to identify non-vertex maximisers. ρ_{min}^{max} is set to control the depth of the heuristic vertex search. $nlocl_{max}$ is set to limit the number of local searches used to supplement the local vertex search (see Appendix B for details).

2. If the local search methods are unsuccessful, then global search procedures are activated. In *global* constraint maximisation vertex enumeration may be used ($ivert$ = 1). A multi-start random search (3.1) may be used in which the search is biased towards the vertices by specifying an a priori probability for each variable lying on a bound at the worst case ($pvert$). 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 $\frac{\partial c_{max}}{\partial v}$ indicates an ascent direction within V. The use of *pvert* was found to be essential for an effective random search as non-vertex maximisers 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 when either the computational effort employed exceeds the maximum used in previous iterations and a constraint violation has been identified or 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 is consistent with Törn and Žilinskas' (1989) description of practical global optimisation as being concerned with the efficient application of finite resources to find the best solution. It 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 maximisers to push the design towards its final form before applying the more expensive procedures.

In carrying out local searches for

$$\max_{v \in V} \max_{k \in K} c_k(p, v) \tag{3.36}$$

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

$$c_{max} = \epsilon | \log(\sum_{k=1}^{N} exp(\frac{c_k(p, v) - \epsilon}{c_{sig_k}})) = 0 \tag{3.37}$$

can be used. This approach is preferred because

$$\max_{v \in V} c_{max} \tag{3.38}$$

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

$$\max_{k \in K} \max_{v \in V} c_k(v, x) \tag{3.39}$$

When $\epsilon_p > 0$ 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 and otherwise the unprojected model is added.

In deciding whether a new constraint violation has been obtained the violation is required to be greater than the multi-model design optimisation tolerance. This requirement combines with error control as discussed in section 3.2.2.1 to prevent noise causing reidentification of a previous maximum.

3.3.2.3 Initialisation

Initialisation includes defining the sets V and P. 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 design constraints are simply added to the outer approximation constraints for each multimodel design.

The initialisation must also define the parameters controlling the maximum effort to be applied to the optimisation. 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 which 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.

Initialisation 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 judgement or analysis (first *nset* elements of V_0) or based on a simplified constraint maximisation 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 maximising the individual constraints. An initial application of a local vertex search algorithm (see Appendix B) to each constraint seems more appropriate. Only distinct vertex maximisers 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 maximisation 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.

3.3.2.4 Multi-model design

In the multi-model design stage it is desirable to carry out the design with the minimal set of models which define the optimum while not dropping and reidentifying models which do affect the solution. The approach adopted is to predict the "active set" of models, design with this active set, check the solution against all models identified and if any additional model indicates constraint violations add it to the predicted active set and rerun the optimisation. The active set is predicted as all models which 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 optimisation. This strategy is quite heuristic but appears reasonable.

The use of the "ϵ_p" heuristic can potentially slow convergence if the unprojected constraint maximisers included in the multi-model design become active again as the design progresses. This behaviour would not be expected in typical problems and has not actually been observed, but 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.

3.3.2.5 Some practical considerations

It is important to emphasise that despite the measures introduced to make the worst-case design algorithm more efficient for the problems of interest, it is not practical to deploy the global optimisation 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 judgement as to the key parameters to be given the most rigorous treatment. Engineering judgement may identify some variables which are expected to make all performance measures worst at one extreme of their range or groups of parameters which 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 is 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 optimisation level.

The performance of the algorithm is discussed in chapter 6 in the context of the design examples tackled.

3.4 Screening tools

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 modelling and analysis is often 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 analysing the effect of delays on achievable disturbance rejection.

Within an integrated design procedure (see chapter 5) steady-state modelling and optimisation may be used as a screening tool within the overall process of developing a design which 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. The methods required to carry out such analyses for nonlinear systems are covered in the previous sections of this chapter.

This section focuses on methods which attempt to extend steady-state analysis to capture the effect of the process dynamics on the design.

3.4.1 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 which are likely to be difficult to control, but have an inherent ambiguity when placed 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 then they are particularly useful as they are unambiguous.

3.4.1.1 Controllability indicators

With the usual waste-treatment scheme — several reactors in series, with reagent added to each reactor and composition measured at the exit of each reactor — the process transfer function G(s) is triangular. This means that measures of interaction such as the RGA and minimised condition number will indicate no significant interaction problems

(Narraway, 1992). Controlling each series reactor with a feedback loop around the reactor gives a simple system in which the disturbance seen by each controller is affected by the controllers on preceding reactors, but there is no two-way interaction. Non-minimum phase zeros have not been noted within the feedback loops around individual reactors. For these reasons, many of the general purpose controllability indicators are not appropriate for this problem and are not considered further.

One group of indicators which is relevant is measures of the effect of process delays on process control. These are discussed together with the presentation of a new method for analysis of the effect of delays in section 3.4.2.

3.4.1.2 Design validation

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

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. This approach is of general interest as it represents an attempt in the control area to deal with uncertainty and constraints systematically. It is emphasised 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 to determine whether performance requirements can be satisfied irrespective of controller structure and type and 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. This 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 of only vertices of V

to be applied (see section 3.3.1). The mathematical formulation is

$$\min_{\Delta u(k)} \max_{v \in V} \{ \| y(k+1) - y_s(k+1) \|^2_{W_y^T W_y} + \| \Delta u(k) \|^2_{W_u^T W_u} \} \tag{3.40}$$

subject to bounds on the outputs, y, the control moves, $\Delta u(k)$, and the control values, $u(k)$. y_s is a vector of target values for the outputs. 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.

To illustrate this consider the following problem. Suppose a typical stable system $y(s) = \frac{k1e^{-k2s}}{1+k3s}u(s)$ is known exactly except for an uncertainty in $k1$ and it is required that the control bring the system from a given initial condition to steady state at a target value within an infinite time 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 as would have been determined very rapidly by designing a PI controller.

This problem of posing 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 optimisation 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 is itself 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 in this project (see section 3.3).

A possible method of salvaging the validation problem is to solve *both* a minmax (equation 3.40) and a maxmin optimal control problem.

$$\max_{v \in V} \min_{\Delta u(k)} \{ \| y(k+1) - y_s(k+1) \|^2_{W_y^T W_y} + \| \Delta u(k) \|^2_{W_u^T W_u} \} \tag{3.41}$$

If the maxmin analysis indicates the problem to be infeasible then 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 then 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 control having perfect knowledge of the plant parameters and disturbances, including *future* disturbances. This future knowledge can eliminate the restrictions imposed by pure time delays completely. 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 *realisable* control system exists which 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 which can satisfy the constraints) and the minmax is infeasible (there is no single input sequence which can satisfy the constraints for all v). This is unfortunate as this outcome is the least informative as to whether a realisable controller exists (sufficient condition not satisfied and necessary condition satisfied).

For example, optimising 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 cancelled at the system inlet by adjusting flows to maintain the correct blend (ensuring that for each disturbance a control sequence exists which can cancel that disturbance perfectly) so that 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 optimisation by implementing a control system in the model used to evaluate performance. This unfortunately 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 (sections 3.3) 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.

3.4.2 A screening test for disturbance rejection in nonlinear processes subject to dead time

3.4.2.1 Introduction

A screening test for analysing achievable disturbance rejection in dynamic systems subject to dead time 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 which can meet the specified disturbance rejection requirements. A variation of this test can be used to calculate the fraction of the disturbances which can be rejected. This provides a measure of controllability which 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.

Dead time or pure delays frequently arise in process control. They are usually associated with the physical transport of mass or energy, e.g. in a pipe, or with the processing of analytical measurements. They are widely recognised as being a key source of control difficulties. Previous process controllability work on analysing the effect of delays on achievable control performance has centred 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.

3.4.2.2 Review of previous work

Previous work on controllability analysis of systems with delays has centred 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. The paper notes 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. Later work has concentrated on the interactions of delays with non-minimum phase zeros which is not of particular relevance.

These measures of delay all provide useful indicators of the effective delay but do not in themselves indicate whether the effect of the delay prevents a disturbance 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 must be included in order to assess disturbance rejection.

It should also be noted that the analyses of delay times discussed above do not differentiate 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 distinct 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 then 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 PID controller and a number of variants including computing the response time using the Ziegler-Nichols rules for computing effective delay have appeared, e.g. the work of McMillan discussed in section 2.3.1.3. 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 meeting constraints.

3.4.2.3 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 figure 3.1).

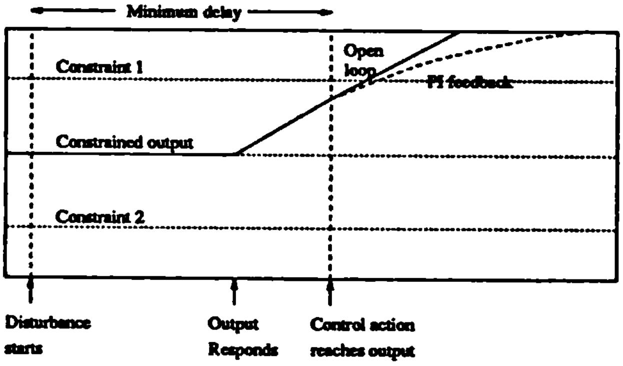


Figure 3.1: Ideal and PI control

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 as 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 disturbance.

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.

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

$$t_d = t_{d_{d,y}} + t_{d_{u,y}} \tag{3.42}$$

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

$$t_{d_{\bar{j}}} = \min_j\{t_{d_{d,y_j}}\} + \min_i\{t_{d_{u_i,y_{\bar{j}}}}\} \tag{3.43}$$

This expression for the minimum delay from control action to the constrained output y_j, $\min_i\{t_{d_{u_i,y_j}}\}$ 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 then the upper bound delays could be substituted for those used. The delay measure of Perkins and Wong could also be used but would be more conservative in this context as it does not differentiate between individual outputs. 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 where paths with minimal delay are technically present but are not realistically useful. E.g., 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, the disturbance can only usefully be detected using a subset of the measured variables, $y_j | j \in J_d$. Similarly, 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. These restrictions combine to eliminate weak connections which would not form 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

$$t_{d_{\bar{j}}} = \min_{j \in J_d} \{ t_{d_{d,y_j}} + \min_{i \in I_j^s \cap I_j^c} \{ t_{d_{u_i,y_{\bar{j}}}} \} \} \tag{3.44}$$

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.

Non-independent entry points for disturbances may occur if a disturbance gives rise to changes in several process characteristics simultaneously, e.g. flow and concentration. Each entry point may have different delay characteristics. There are a number of options for considering this class of problem, none of which are generally appropriate.

1. The minimum delay for rejection of any of the disturbance components may be treated as the minimum delay for rejection of the entire disturbance. This is often an unreasonably optimistic assumption, but will always give a valid bound.
2. Each component may be treated as an independent disturbance. If any of the disturbance components is likely to partially counteract the effect of other components this will not necessarily give a valid bound on disturbance rejection.
3. Certain components may be judged negligible and their associated delays ignored.
4. The delays associated with each component may justify separating the effects of two components if one component could be rejected before the other component takes effect. For example, a disturbance in flow may reach a process well before any accompanying concentration disturbance, so that the process could be treated as having recovered from the flow disturbance before the concentration disturbance is encountered.

The best formulation for a particular problem of this type requires judgement to be exercised. This is discussed further in section 6.2.

3.4.2.5 Feasibility tests

For a nominal process model the test for feasibility of disturbance rejection is

$$\exists p \in P \; s.t. \quad \max_{k \in K} \max_{t \in [0, t_{d_k}]} \{c_k(p, t)\} \leq 0 \tag{3.45}$$

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 parameterised 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 but is a fair approximation to industrial practice, in that it is often desired that the need for operator intervention and control system adjustment in normal operation should be minimised. The feasibility test then becomes

$$\begin{aligned} \exists p \in P \; s.t. \quad & \max_{v \in V} \max_{k \in K} \max_{t \in [0, t_{d_k}]} \{c_k(p, v, t)\} \leq 0 \\ & V = \{v \mid -\Delta v^- \leq v \leq \Delta v^+\} \end{aligned} \tag{3.46}$$

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 3.3 this formulation is more difficult to analyse and will often give inappropriately optimistic results. The formulation with operating variables is, however, given below for completeness.

$$\begin{aligned} \exists p \in P \; s.t. \quad & \max_{v \in V} \min_{o \in O} \max_{k \in K} \max_{t \in [0, t_{d_k}]} \{c_k(p, v, o, t)\} \leq 0 \\ & V = \{v \mid -\Delta v^- \leq v \leq \Delta v^+\} \end{aligned} \tag{3.47}$$

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 annualised cost. The optimised cost would then give a rough estimate of the effect of process disturbances and dynamics (and uncertainty) on the optimal process cost.

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

The problems have the additional advantage that they provide a result which 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 shown in the case studies (chapter 6) to provide a useful bound on the performance achieved with implementable controllers.

Solution of validation problems involving uncertain parameters can be tackled using the techniques discussed in section 3.3.

In certain problems the delays will themselves be functions of the parameters p and v. This introduces a potential problem for the optimisation 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 non-smooth optimisation methods. As noted by Gill, Murray and Wright (1981) smooth algorithms may often be successful on problems such as those with occasional discontinuities and should be tried first.

The optimisation problems above may be used to provide a rigorous bound on the achievable performance if delays are calculated as discussed in section 3.4.2.4. If it is desired to obtain a heuristic estimate of performance rather than a rigorous bound then the control response delays, $t_{d_{u_i,y_j}}$, within control loops could be replaced by heuristic values such as $\frac{\pi}{2\omega_n}$, 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 5.2.3 confirms this heuristic to be effective in considering PI control of pH in a stirred tank.

3.4.2.6 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 feedback control would fail, but feedforward control might succeed, then 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 modelling errors is less than the calculated δ_f for feedback control. That is, if δ_f without feedforward control is $x\%$ then 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. As it can be computed more readily than a full design analysis can be carried out it can play a useful role in design.

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

$$\begin{aligned} \min_{p\in P} -\delta_f \; s.t. \quad & \max_{v\in V} \max_{k\in K} \max_{t\in[0,t_{d_k}]} \{c_k(p,v,t)\} \leq 0 \\ & V = \{v \mid -\Delta v^- \leq v \leq \Delta v^+\} \end{aligned} \tag{3.48}$$

δ_f should be included in the design variables, p.

3.4.2.7 Simplified analysis for series CSTRs

While general problems require optimisation of a nonlinear dynamic model as discussed above, the analysis can be greatly simplified for some special cases. The case of particular interest in this project is that of continuous flow stirred tank reactors (CSTRs) in series, where 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 control precision, δ_c, where

$$\delta_c = \frac{\delta_{cnet}}{\Delta_{cnet}} \tag{3.49}$$

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 pH constraints) then equation 2.2

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

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 pH limits.

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

$$y(t) = 1 - e^{-t/t_{c_{mix}}} \sum_{i=0}^{n-1} (t/t_{c_{mix}})^i / i! \tag{3.50}$$

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 pH measurement), the best possible disturbance attenuation, δ_a, is given by

$$\delta_a = 1 - e^{-t'_d/t_{c_{mix}}} \sum_{i=0}^{n-1} (t'_d/t_{c_{mix}})^i / i! \approx \frac{1}{n!}(\frac{t'_d}{t_{c_{mix}}})^n \tag{3.51}$$

Tanks of different sizes may be analysed similarly giving

$$\delta_a = 1 - \sum_{i=1}^{n} \frac{e^{-t'_d/t_{c_{mix}}(i)} t_{c_{mix}}^{n-1}(i)}{\prod_{j \neq i}(t_{c_{mix}}(i) - t_{c_{mix}}(j))} \approx \frac{1}{n!}(\frac{t'^n_d}{\prod_{i=1,n} t_{c_{mix}}(i)}) \tag{3.52}$$

If each tank has reagent added based on pH measured at the tank exit without any measurement delay then t'_d will simply be the minimum value of $t_{d_{mix}}$.

To take account of the effect of uncompensated minor time-lags on the performance of PI controllers replace t'_d in the above formulae 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. 10 seconds is a good default for t'_d when estimating the best practicable performance while the heuristic method will give typical values for t'_d 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 using the formulae for non-equally 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 $\frac{t'_d}{(n+1)\tau'}$.

3.5 Summary of design software

Tools have been provided to assist in efficient screening of design options and in the optimisation of the continuous parameters of dynamic systems, including systems subject to bounded uncertain parameters.

These tools have been implemented in FORTRAN. The optimisation software allows the solution of a wide range of optimisation problems

1. conventional NLPs;
2. flexibility index evaluation;
3. multi-model optimisation (parameterised realisations of a single model);
4. worst-case design.

Given a suitable model these tools can be invoked from a menu as required. The menu also allows for direct use of simulation. Checking routines are provided to verify

1. the consistency of model inputs with the residual equations;
2. accuracy of the gradient evaluation;
3. the precision of the simulation results (states and gradients).

Model implementation is based directly on the DASOLV code (Jarvis and Pantelides, 1992). Residual equations are entered in a subroutine and structural and jacobian information must also be provided directly.

J

Chapter 4

Modelling

> 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).

All design methods, like all control 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 utilising that model. An inadequate model or uncertainty description may lead to an unworkable design or to unnecessary and expensive overdesign.

Models reported in the literature for chemical waste water treatment systems are diverse; often qualitative rather than quantitative; and in some cases contradictory. The first source of models lies in the literature on effluent treatment and pH control. In cases for which this source seems inadequate, I have attempted to resolve discrepancies or develop improved models drawing on the general chemistry and chemical engineering literature.

Where possible, I have tested and refined the models using data provided by ICI. If more than one model appears adequate for design, I have favoured 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 where appropriate guidelines for experimental work to reduce the uncertainty are given.

The preliminary review of design requirements (section 1.2) identified the major aspects of system behaviour affecting pH and pIon effluent treatment systems. Many of these raise important modelling issues:

1. steady-state relationships between manipulated and measured variables, usually governed by ionic equilibria (4.1, 4.3, 4.6);
2. complex (nonlinear) reaction kinetics (4.2,4.3);
3. dynamic response and bias errors of measurement (4.4);
4. effect of mixing on system characteristics (4.5).

The analysis of these specific modelling problems is followed by a review of some existing design packages for pH control systems (4.7) and a summary of the work presented in this chapter (4.8).

4.1 Steady-state pH characteristics

As discussed in section 2.1 the complex nonlinear relationship between pH and neutralising reagent concentration, often expressed graphically as a titration curve, is a key characteristic of waste waster treatment systems. It determines the precision to which reagent must be added to meet specifications on exit pH, may introduce large 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 modelling of chemical waste water treatment systems. Modelling 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.

4.1.1 Thermodynamic equilibrium models

A first principles approach to this mapping is to model the thermodynamic equilibrium relations between the actual effluent components using a suitable activity coefficient convention. This approach is necessarily approximate, even with good estimates of the model parameters and effluent compositions as there is, as yet, no thermodynamically rigorous method of predicting the effects involved.

The conventions and terminology used in this work are summarised below for clarity.

1. Equilibria are expressed in terms of the activities, a_x, e.g. $a_x a_y = k_{xy} a_{xy}$ is used as the equilibrium relationship for the reaction $x + y \rightleftharpoons xy$.

2. Activities are expressed as $a_x = \gamma_x m_x$ where m_x is the molal (gmoles/kg) concentration of x, γ is the activity coefficient, and $\gamma_x \rightarrow 1$ as $m_x \rightarrow 0$.

3. Concentrations are also expressed in molar/M (gmoles/litre), denoted by $[x]$ and normals/N (available OH^- (M)) denoted by c_x. It should be noted that a 1M acid solution is described as a -1N solution.

A key problem in the thermodynamic approach is the calculation of activity coefficients (γ) which are used to correct the equilibrium relationships which hold at infinite dilution . Recent work on estimating physical properties of electrolytes, particularly activity coefficients, up to ionic strengths of 6 molal does not amount to a thermodynamically rigorous model but is more accurately described as "semi-empirical".

> For each mixture it is necessary to construct an appropriate mathematical model for representing the properties of that mixture. Whenever possible, such a model should be based on physical concepts, but since our fundamental understanding of fluids is severely limited, any useful model is inevitably influenced by empiricism (Reid *et al.*, 1988).

Many of the models require test data for ternary solutions and involve many parameters, a combination which makes their validity questionable.

> Seek simplicity; beware of models with many adjustable parameters. When such models are extrapolated even mildly into regions other than those for which the constants were determined, highly erroneous results may be obtained. (Reid *et al.*, 1988)

> Face the facts: you cannot get something from nothing. Do not expect magic from thermodynamics. If you want reliable results you need some reliable data. (Reid *et al.*, 1988)

Databases are available for the main correlation based approaches but in many cases recourse to experimental data or actual experimentation would be needed to supplement these. Data for molecule-ion interactions is particularly sparse. There will therefore only be a certain subset of problems for which a thermodynamic model will be feasible.

An extensive review and summary of aqueous electrolyte thermodynamics is given in Zemaitis *et al.* (1986). For solutions with less than 1 molal ionic strength a range of modified Debye-Hückel equations is discussed. The most accurate of these is probably the Hückel equation

$$\log_{10}(\gamma_i) = \frac{-Az_i^2\sqrt{I}}{1+\beta a\sqrt{I}} + CI \tag{4.1}$$

which has two adjustable parameters, a and C. A and β are standard constants. The ionic strength, I, is defined as $I = \frac{1}{2}\sum z_i^2 m_i$ where z_i is the charge on the *ith* ion and m_i is the concentration of the *ith* ion in molal (gmoles/kg). This equation has the advantage of the existence of extensive compilations of data, but the disadvantage that it becomes increasingly inadequate as the ionic strength increases above .1M and alternative approaches are regarded as necessary above 1M. The Hückel equation will therefore be used for thermodynamic modelling of solutions with moderate ionic strength.

The most extensively used approach to thermodynamic modelling of concentrated electrolytes is noted by Renon (1986) to be that of Pitzer which is shown in a comparison to have better accuracy than NRTL or MSA models developed later. This method has been extensively documented in Pitzer (1973) and subsequent articles. Modifications of Pitzer's approach have also been applied to weak volatile electrolytes, e.g. Edwards *et al.* (1975; 1978). Other existing models such as NRTL and extended-UNIQUAC have no clear advantages over Pitzer's model and less published data. Zemaitis (1986) discusses Pitzer's and other approaches and summarises available data for the main models, but does not clearly favour any particular method. Many chemists present their data using Pitzer's approach or a variant thereof. These approaches have been used for thermodynamic modelling of *concentrated* electrolytes in this project.

Limited thermodynamic data is available at temperatures other than 25 °C. The first requirement for equilibrium calculations at other temperatures is to have good correlations for the variation of equilibrium coefficients and activity coefficients with temperature. The variations of the Debye-Hückel parameters A and β (equation 4.1) with temperature are known, but there is little data on the variation of other parameters in activity coefficient models with temperature.

Apart from the difficulties in calculating equilibria accurately there is an additional fundamental problem that many aqueous reactions have very slow kinetics so that equilibrium is not established on the time-scale of interest. Metastable solutions (apparently stable non-equilibrium solutions) and practical irreversibility of reactions are often observed, particularly in solid-liquid equilibria and metal hydration.

The thermodynamic modelling approach also has several drawbacks from a practical viewpoint over more empirical methods.

1. It requires knowledge of the compositions of all ions in solution at significant concentrations which may be difficult and expensive to obtain.

2. It gives rise to models which are much more difficult to solve than those based on the common approach of using experimentally determined titration curves to provide the mapping from concentration to pH.

In addition to the difficulties of thermodynamic modelling, problems in predicting experimental titration characteristics arise from systematic deviations between pH measured by an electrode and the theoretical relationship

$$pH = -log_{10}(a_{H^+}) \tag{4.2}$$

Deviations from the ideal cell voltage are generally present due to electrode junction potentials in the cell which are difficult to correct for or due to interference effects from other ions. For the glass electrode, which is generally used industrially, the best known interference is from sodium ions which enter the glass lattice and cause the apparent pH to fall, with the error developing over a period of several minutes. This effect is particularly strong at high pH and high temperatures. Similar problems occur in acid conditions below pH 2. Special electrodes can reduce these effects but not eliminate them. Junction potentials vary with the activity of the ions in solution becoming more pronounced at high activities. These effects can generally only be partially modelled (Bates, 1964).

An example of the thermodynamic modelling of concentrated electrolytes which illustrates the issues discussed above is given in section 4.1.3.

Much previous work on modelling the relationship between pH and reagent concentration has used simplified thermodynamic models in which activity coefficients and cell non-ideality are completely neglected so that the model reduces to concentration equilibrium relationships. $[x]$ (molar) and m_x (molal) are treated as equivalent in the discussion below as the solutions are assumed to be sufficiently dilute that the distinction is unimportant.

For strong-acid/ strong-base systems this simplified approach gives the equilibrium equations

$$\begin{aligned} [OH^-]-[H^+] &= cnet \\ [OH^-][H^+] &= k_w \approx 10^{-14} \\ pH &= -log_{10}([H^+]) \end{aligned} \tag{4.3}$$

where *cnet* is the net concentration of acids and bases (total concentration in normals) and k_w is the dissociation constant of water. $[x]$ can be replaced in general thermodynamic models by the activity a_x where $a_x = \gamma_x m_x$. It should be noted that "strong" implies complete dissociation of the H^+/OH^- groups from the other components of the acid/base. A "weak" acid or base is one in which a significant proportion of the active groups remain associated with the other ions in the acid or base in the pH range of interest, say 2–12, e.g.

$$[H^+][A^-] = k_a[HA] \tag{4.4}$$

where $pk_a = -log_{10}(k_a)$ and $pk_a \in [2, 12]$.

Equations 4.3 may be solved explicitly as

$$\begin{aligned} pH &= -log_{10}(\tfrac{-cnet+\sqrt{cnet^2+4k_w}}{2}), & cnet < 10^{-4} \\ pH &= -log_{10}(\tfrac{k_w}{cnet}), & cnet \geq 10^{-4} \end{aligned} \tag{4.5}$$

where the conditional is used to give a numerically robust expression without a significant discontinuity. The titration curve generated using equation 4.5 exhibits the maximum sensitivity of pH to net concentration possible in ideal solutions due to the complete absence of partially dissociated (weak) acids or bases which "buffer" (resist) the change in pH resulting from a change in *cnet*.

Concentration equilibria were used by Gustafsson and Waller (1983) to develop a general representation for acid-base reactions using reaction invariants. Reaction invariants are solution components or combinations of components which do not vary as the equilibrium reactions shift and may therefore be used directly in mass-balances. They apply reaction invariant analysis to an aqueous solution of strong acids and bases, and weak acids in equilibrium with up to three hydrogen ions.

$$\begin{aligned} H_2O &\rightleftharpoons H^+ + OH^- \\ A &\rightleftharpoons H^+ + B \\ B &\rightleftharpoons H^+ + C \\ C &\rightleftharpoons H^+ + D \end{aligned} \tag{4.6}$$

with the equilibrium relations

$$\begin{aligned} [H^+][OH^-] &= K_w \\ [H^+][B] &= k_A[A] \\ [H^+][C] &= k_B[B] \\ [H^+][D] &= k_C[C] \end{aligned} \tag{4.7}$$

where k denotes an equilibrium constant. They use the total amount of each acid species ($[A]+[B]+[C]+[D]$) and the net charge of all groups participating in the acid-base reactions as reaction invariants and obtain a single implicit equation to compute pH given the concentration of the reaction invariants. For strong reagents this equation can be expressed in terms of the net concentration of reagents and the strong components of the effluents. The expression they obtain is linear in the reaction invariants, given the dissociation coefficients and the observed pH, and may be used to compute the net concentration and the total concentrations of the weak acids for use in control. This is essentially a linearising transform of the pH to improve control characteristics. The main purpose of their representation is to give a low order description of the titration characteristic which can be updated on-line for use in adaptive control as discussed in section 2.4.

In using thermodynamic models there will be uncertainty and variability in many constants and concentrations. The task of formulating and using an uncertainty description for each parameter is not practical. Judgement and sensitivity analysis should be used to identify the key sources of uncertainty which can then be treated as dominating the other components. The key sources of uncertainty are those which affect the titration characteristics of the process most substantially. The need to consider uncertainty therefore provides an additional incentive to consider the direct use of titration characteristics which may provide a simpler representation of the process input-output behaviour than a thermodynamic model.

4.1.2 Titration curve models

The main alternative to modelling using equilibrium equations is to characterise the effluent directly using titration curves. The pH is then calculated using the mapping between reagent concentration and pH defined by the titration curve. This approach is exact (neglecting variable measurement errors) for the particular effluent and reagent pair considered. However, the extrapolation of titration curve models to other effluent/ reagent pairs, e.g. to predict the curve for a more dilute effluent, is not "transparent" and has generated considerable confusion in the literature.

The titration curve is usually given as a table of pH as a function of the volume of reagent per unit volume of effluent. However, a titration curve for a single effluent and reagent may be expressed as a function of any variable which is a monotonic function of the volume fraction, such as net concentration, *cnet*, or concentration ratio. By convention the net concentration, *cnet*, is taken as zero at a reference pH (pH_{ref}), usually 7. This allows the effluent to be assigned a notional concentration in normals, c_e, based on the volume of reagent added to get to pH_{ref} (V_r), the reagent concentration in normals, c_r, and the effluent volume, V_e, used in the titration.

$$c_e = -\frac{c_r V_r}{V_e} \tag{4.8}$$

It should be recalled that the concentration in normals is negative for acids and positive for alkalis. Two common approaches to estimating the effect of a new effluent concentration c_e^{new} are

1. tabulate the titration curve against the net concentration ($\frac{c_e V_e + c_r V_r}{V_r + V_e}$) and interpolate for pH using the net concentration based on c_e^{new} (Badran, 1984);

2. tabulate the titration curve against the ratio of the reagent and effluent concentrations (q_r/q_e, where $q_x = V_x c_x$) and interpolate for pH using the concentration ratio based on c_e^{new} (Moore, 1978).

The first method corresponds to *shifting* the titration curve along a reagent concentration axis so that only the total reagent concentration required to reach a given pH is changed but the sensitivity of pH to reagent concentration remains constant. The concentration ratio approach is equivalent to *stretching* the titration curve so that sensitivity of pH to reagent concentration varies at every pH value in inverse proportion to the effluent concentration.

The limitations of these two approaches can be shown by considering the sensitivity of pH to concentration, $\frac{\partial pH}{\partial cnet}$, for the case of a single weak acid component (neglecting activity coefficient effects),

$$\frac{\partial pH}{\partial cnet} = \frac{1}{ln(10)(10^{pH-pk_w} + 10^{-pH} + \frac{[acid]10^{pk_{acid}-pH}}{(1+10^{pk_{acid}-pH})^2})} \tag{4.9}$$

where $pk = -log_{10}(k)$. From this expression it is clear that for pH values sufficiently close to pk_{acid}, $\frac{\partial pH}{\partial cnet}$ becomes almost inversely proportional to the concentration of the weak acid. For pH values far from pk_{acid} the buffering becomes independent of the weak acid concentration. That is, the concentration ratio approach is a good approximation for parts of the titration curve which are highly buffered by weak acids or alkalis while the net concentration approach is a good approximation for parts of the titration curve which are only slightly buffered. The net concentration approach would also be appropriate if only the strong components of the effluent varied in concentration.

Luyben (1990) has proposed a general method for predicting the titration curve of a mixture of effluents from the titration curves of the component solutions. This method relies on the premise that *if two solutions originally at the same pH are mixed the pH is unchanged.* The amount of reagent required to give a particular pH value is then the sum of the reagent quantities required to bring each of the original solutions to this pH. This relationship may be used to calculate pH given total reagent or total reagent given pH.

$$\begin{gathered} V_r = \sum_{i=1}^{N} V_{r_i} \\ V_{r_i} = tc_i^{-1}(pH)V_{e_i}, \quad i = 1,\ldots,N \end{gathered} \tag{4.10}$$

where V_{r_i} is the volume of reagent required to bring a volume of the *ith* effluent V_{e_i} to a given pH value and $pH = tc_i(V_{r_i}/V_{e_i})$ is the titration characteristic of the *ith* effluent.

Luyben does not present a justification of the basic premise that solutions of equal pH give the same pH on mixing. It can be justified as follows neglecting activity coefficient effects, i.e. assuming dilute solutions or assuming activity coefficients to be unaffected by mixing. All acid–base equilibria take the form

$$\frac{[H^+][A^{n-}]}{[HA^{(n-1)-}]} = k_a \tag{4.11}$$

or

$$\frac{[OH^-][M^{n+}]}{[MOH^{(n-1)+}]} = k_b \tag{4.12}$$

On mixing two solutions of equal pH, electroneutrality and the concentration of H^+ and OH^- are instantaneously preserved. Acids or bases in solutions at the same pH are present in identical ratios of compositions so mixing does not instantaneously alter these ratios. The acid-base equilibria are therefore unperturbed unless other reactions take place to absorb or generate components involved in the acid-base equilibria. In particular, precipitate or gas formation could completely invalidate the assumption that pH remains unchanged. Reactions independent of the acid-base reactions can affect the pH indirectly through factors such as temperature or activity coefficients.

Luyben's approach provides a sounder general method for extrapolating the effect of effluent concentration variations than either the net concentration or concentration ratio method. It is equivalent to generating models based purely on ideal acid-base equilibria which fit each curve exactly and using the combined model to generate the new curve.

This approach may be used to estimate the effect of concentration changes in the effluent and reagent. Given a titration curve based on a strong reagent

$$\frac{V_r}{V_e} = tc_1^{-1}(pH) \tag{4.13}$$

variations in the effluent and reagent concentrations can be accounted for by

$$\frac{V_r}{V_e} = tc_2^{-1}(pH) = \frac{(1-k_e)X_{pH} + k_e(c_r - X_{pH})tc_1^{-1}(pH)}{k_r c_r - X_{pH}} \tag{4.14}$$

where

$$X_{pH} = 10^{pH - pk_w} - 10^{-pH}, c_r^{new} = k_r c_r, c_e^{new} = k_e c_e$$

This relationship may be derived by using the premise that mixing two solutions with the same pH results in a mixture with unchanged pH. The relationship above can be decomposed into components representing the reagent required by the new system with all dilution due to concentration changes neglected and the amount of reagent required to bring this excess water to the desired pH

$$tc_2^{-1}(pH) = \frac{k_e}{k_r} tc_1^{-1}(pH) + \frac{X_{pH}((1-k_e) + (1-k_r)\frac{k_e}{k_r} tc_1^{-1}(pH))}{k_r c_r - X_{pH}} \tag{4.15}$$

Considering only the first term gives the concentration ratio method for dealing with concentration variations

$$\frac{q_r}{q_e} = tc_1^{-1}(pH)\frac{c_r}{c_e} = ktc_1^{-1}(pH) \tag{4.16}$$

showing this relationship to be a sensible approximation of the general titration relationship which does not involve solution of an implicit equation in pH. It should be noted that the above expression relies on the reagent being strong. If the reagent is weak then additional information to the single titration curve is required to construct the general relationship incorporating concentration variations. This case does not seem to be relevant for practical problems and will not be considered further.

In using titration curve models there will normally be uncertainty in the buffering effects at each pH value due to lack of repeatability in experiments, variation in the effluent or reagent properties or the presence of unmodelled dynamic effects. The first two sources of uncertainty are, at least in principle, readily estimated by repeated experiments on a single sample and repeated samples. In practice some degree of judgement needs to be exercised as the full range of effluent characteristics and experimental variability can never be realised in a finite (let alone practical!) number of trials. The experimental investigation should be guided by consideration of process characteristics such as semi-batch operations and the effect of shutdowns. A mixture of grab sampling and flow-averaged sampling is often required to get a clear picture. Care should be taken that results are not distorted by effects such as cooling of hot samples, settling of suspended solids and evaporation. As noted by Cushnie (1984) underestimating the variability of the effluent characteristics can often be the cause of design failures. In generating titration curves, care should be taken that at least one point is generated for each unit pH change. Little error is incurred by using a more dilute reagent near neutral pH if this is necessary to achieve this.

A convenient way of representing the titration characteristics and the associated uncertainty is as a range of concentration variations associated with particular pH ranges. The range of concentrations of reagent required to reach a reference pH should also be specified to complete the titration curve definition. This information can be used to generate a smooth interpolation of the titration characteristic suitable for numerical solution (see section 3.2.2.5).

The validity of Luyben's approach to predicting the titration curve of mixtures is examined below using an example.

4.1.3 Examples

The examples below show some results from applying each of the two main modelling approaches — thermodynamic modelling and titration curve manipulation — to some industrial data provided by ICI.

4.1.3.1 Thermodynamic modelling

The titration characteristics of a fairly pure concentrated strong acid (11.8M HCl) with a 5.9M $Ca(OH)_2$ reagent with typical impurities were provided. Data were provided at room temperature and 75 °C.

The low temperature titration characteristic was determined with good repeatability, including when reverse titration with acid was carried out. Titration data are plotted in figure 4.1 together with the results of a thermodynamic model (see below).

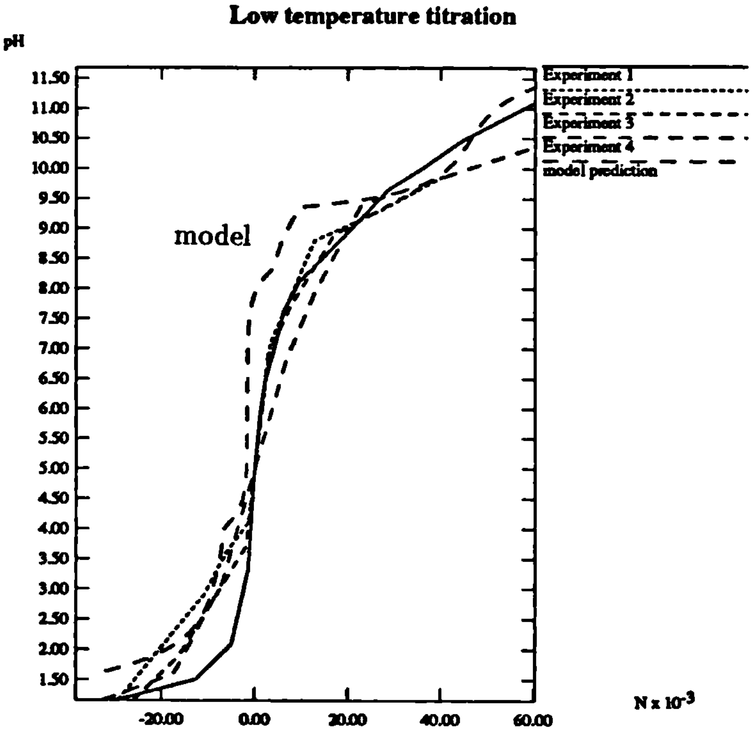


Figure 4.1: Room temperature titration data for concentrated acid

The data show high buffering in the key region of pH 7 to 9. Examination of typical impurities indicates the presence of significant Mg(.007), CO_3 (.01), SiO_2 (.007) and Al_2O_3 (.002). Figures in brackets are moles impurity/mole $Ca(OH)_2$. It seemed likely that these impurities were generating the buffering effects.

A thermodynamic model was written for this system based on data from Baes and Mesmer (1976) on activity coefficient effects at 25 °C (in similar form to Pitzer) and from Nordstrom *et al.* (1990) on temperature dependence of equilibrium constants. A key uncertainty in this type of model is which equilibria are effective on the time-scale of the experiment. Based on the match to the data and observations from Sposido (1989) on aluminium, Volosov *et al.* (1972) on silica and Mattock (1964) on carbonate the following assumptions were made.

1. Aluminium hydroxy complexes are not in equilibrium with their solid phase.
2. Silica species are not in equilibrium with their solid phase.
3. Carbonate is eliminated from the system as CO_2 gas.
4. Magnesium and calcium are in equilibrium with their solid phases.

The resulting model gave the results shown in figure 4.1 showing a fair fit to the data, particularly the level of buffering between about 8 and 9.5pH. The main discrepancies were that the buffering effect appeared at slightly higher pH values than in the experimental data and the model showed a flatter pH region associated with magnesium precipitation around pH 9.5 than was actually observed.

Applying the OLI thermodynamic modelling package to the above problem with the same assumptions on effective equilibria gave an inferior match to the data with a very pronounced plateau around 6 pH and little buffering between 7 and 8.5pH. This seems to be accounted for by the omission of three aluminium hydroxy species from the OLI public database. These species spread out the aluminium buffering effect and shift it to a higher pH.

It is clear from the match achieved between model and experiment that the buffering effect can be explained by the impurities associated with the lime reagent. It is also clear that it would not have been possible to use the model to predict the titration curve as results are sensitive to modelling assumptions as to which equilibria are closely approached during the experiment and give only a *fair* fit to the data.

The high temperature titration curves (figure 4.2) are dominated by two regions of slowly changing pH, one around 6–7pH and one around 8–9pH. In both regions an initial rapid rise in pH was observed on addition of reagent, followed by a slow downward drift over about 20 minutes. The downward drift is observed to be about half the initial response so long as the drift from the previous addition has already stabilised. The first "flat" region stretches over about .03N concentration change which corresponds to the typical level of magnesium impurity. The second region stretches over a much larger range and must therefore be associated with calcium. The successive precipitation of Mg and Ca seems very likely to be the explanation for this behaviour. The buffering associated with SiO_2 and the metal impurities at low temperatures is not as clearly in evidence, though it may be obscured by the precipitation reactions. Some buffering was observed between 5–6pH.

??

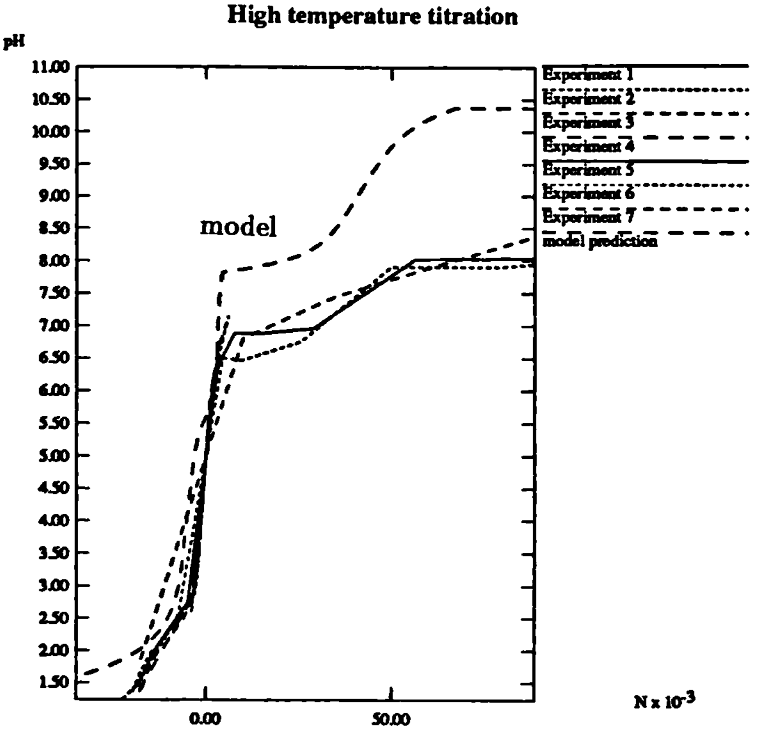


Figure 4.2: 75 °C titration data for concentrated acid

The thermodynamic model used above was applied to the high temperature conditions, but was less successful than in the low temperature case. The two flat pH regions are reproduced over the correct concentration ranges but the pH values predicted are substantially too high. The lowering of the precipitation pH values may reflect coprecipitation effects or simply the inadequacy of the thermodynamic data. The buffering effect between 4–5pH predicted due to aluminium supersaturation was not observed experimentally, probably indicating that the aluminium hydroxide approached equilibrium with its solid phase more rapidly with increasing temperature, as has been observed by researchers investigating aluminium (Sposido, 1989). The model correctly predicts the disappearance of buffering between 2–4 pH. The match between model and experiment is adequate to reinforce the conclusions above as to the nature of the main effects, though the model is inadequate for predicting the observed curves.

Additional titrations with NaOH, at 75 °C, showed substantial apparent buffering, with apparent saturation around 11.9pH due to the error induced in the glass electrode by the presence of high concentrations of Na^+. As the buffering is an illusion due to probe error and the system does not show significant buffering at ambient temperatures, the final pH on cooling could not be controlled to a good precision based on the high temperature pH measurement.

The variation of pH with temperature was examined for neutralisation with $Ca(OH)_2$ at the higher temperature followed by cooling to room temperature. Two scenarios can be hypothesised.

1. The base case for analysing the effect of temperature on pH is the shift in pH due to the change in the dissociation constant k_w of water. In the absence of other chemical effects temperature would have no effect in acid conditions, would shift the pH about -.0135pH/K at neutral and shift the pH about -.027pH/K in alkali conditions (in the range 0-100 °C).
2. If all chemical processes are completely reversible then the pH should vary between the values indicated by the low and high temperature titration curves.

Observed changes will be compared to these scenarios to clarify the nature of the chemical changes with temperature.

When the neutralised solutions were cooled from 75–85 °C immediately following turning off the agitator, the pH changes fell into two patterns.

1. When a large excess of reagent had been added leading to $Ca(OH)_2$ precipitation the change in pH is accounted for by the k_w effect.
2. When reagent addition was terminated towards the end of the $Mg(OH)_2$ precipitation region the shift in pH (about 2.7pH) was about twice the k_w effect and corresponded fairly well to a shift from the high to the low temperature characteristic.

This suggests that in the second case the precipitate is redissolving giving a greater pH rise than would be expected from a pure k_w effect (about 1.3pH).

This hypothesis was investigated by turning off the agitator, maintaining the solution as close to the original temperature as possible, allowing suspended solids to settle, removing some of the clarified liquid without disturbing the settled solids and allowing the clear liquid to cool. pH and temperature were monitored during settling and cooling. This showed an overall pH rise of between .9 and 1.15pH between 75 °C and 20–25 °C which is fairly consistent with the predicted k_w effect of up to 1.3pH rise (less when very close to neutral).

This example demonstrates the use of the thermodynamic approach in developing insights into observed pH variation with concentration and temperature.

1. The analysis indicated that the observed buffering is related to reagent impurities rather than the pure reagent. The existence of this buffering effect is in itself an interesting result which is not widely known.
2. The analysis allowed variations of pH with temperature to be understood.

It also illustrates the limitations of this approach as a predictive tool.

1. Assumptions on the species formed and the effective equilibria have a major effect on the predictions and are not straightforward.
2. The quantitative predictive performance achieved ranges from fair to poor.

Thermodynamic models are therefore not used directly in any of the industrial case studies discussed in chapter 6, and titration characteristics are used directly instead.

4.1.3.2 Predicting titration curves of mixtures

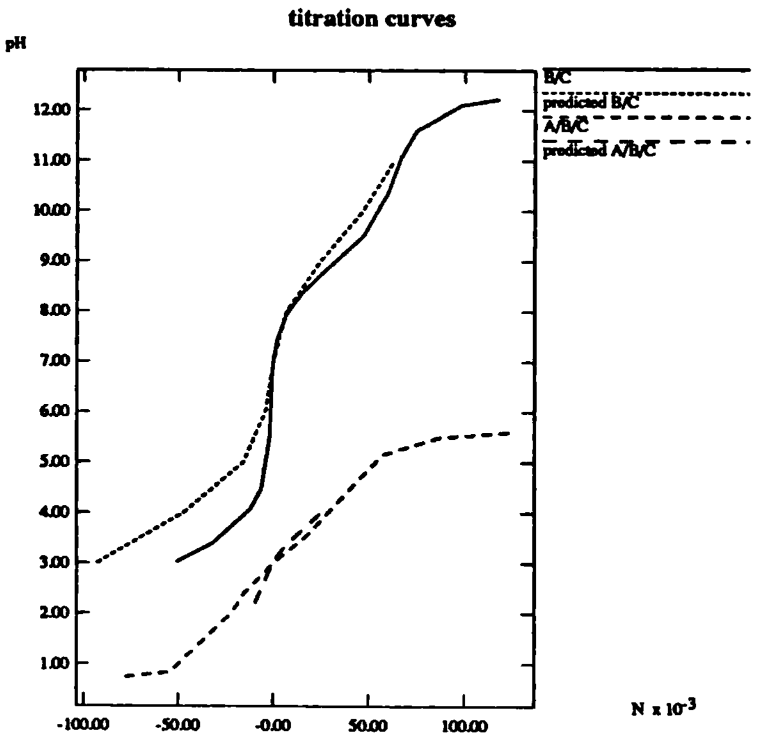


Figure 4.3: Predicted and measured titration curves for mixtures

ICI provided some data on the titration characteristics of three component solutions from different processes (A, B, C) and mixtures thereof (A/B/C, B/C). The solutions contained a complex mixture of weak acids and bases. This example is discussed further in section 6.5. These data allowed some experimental validation of the method proposed by Luyben (1990).

To check the ability to predict the titration curves of mixtures from the titration curves of components the curves for B/C and A/B/C were estimated from the titration curves for the individual streams. The comparison between the measured and estimated curves is shown in figure 4.3. The curve for A/B/C is only compared in the range pH 2–4 to avoid deviations due to incomplete reaction above pH 4 (titration with carbonate) and due to 2 being the minimum pH in the B titration.

The estimate for B/C is a good match to the observed curve except between pH 5–6. In this region the buffering observed is reduced from a predicted 10mM/pH to 2mM/pH. This reduction suggests that an irreversible reaction occurs between some components of B and C. The estimate for A/B/C is a fair match. Overall, the limited experimental corroboration available indicates that the method of constructing titration curves of mixtures from the titration curves of components is a useful approximation, but far from exact in general.

In conclusion Luyben's approach provides a useful approximation. However, its validity needs to be checked for particular applications due to the possibility of effectively irreversible reactions invalidating the basic assumption that mixtures of two components at the same pH remain at the original pH.

4.2 Characteristics of solid alkali reagents

The most common reagents for neutralising acidic waste water are solid lime-based alkalis, particularly $Ca(OH)_2$. As noted in chapter 2 the kinetics of these reagents are not well defined in the literature on waste water treatment 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.

Solid-liquid reactions are generally analysed in terms of a rate of reaction for a unit surface area. This rate may be governed by surface reaction, mass transfer from the surface of the particle or by mass transfer through an "ash" coating if the reaction leaves a residue. The "ash" mechanism is not relevant to solid alkalis as they dissolve in the acid solution without significant residue. Analysis of the surface reaction mechanism is similar to analysis of heterogeneous reactions except that rate of reaction of a particle is proportional to the effective surface area. Reactions governed by mass transfer from the particle surface into the bulk solution pose a complex problem in calculating the effective rate of mass transfer which is known to vary with particle size and mixing conditions.

Mass transfer control will be shown in sections 4.2.2 and 4.2.3 to be the rate controlling step for the key reagents $Ca(OH)_2$ and $CaCO_3$, under the conditions of interest. Mass transfer control also provides an upper bound on the rate of reaction of other solid alkali reagents. This section therefore starts by reviewing some general results on mass transfer from suspended particles before considering the specific reactions of interest.

4.2.1 Mass transfer

The analysis of mass transfer from suspended solid particles can be divided into two components:

1. predicting mass transfer in the absence of chemical reaction (hydrodynamic effects);
2. predicting mass transfer with chemical reaction.

4.2.1.1 Predicting mass transfer coefficients without chemical reaction

This problem is of general relevance and a large number of transfer coefficient estimation methods have been proposed. The methods of Harriott (1962a; 1962b), Brian and Hales (1969) and Levins and Glastonbury (1972b; 1972c) appear to be the most widely used. The correlations are given in terms of the mass transfer coefficient, k_m. The molar flux from the particle surface is given by k_m times the difference in concentration between the particle surface and the the bulk solution.

Harriott's (1962a) method for agitated tanks uses the terminal velocity of a suspended particle in a stagnant fluid, v_T, to estimate the particle Reynolds number, Re, by

$$Re = \frac{D_p v_T \rho_l}{\mu_l} \tag{4.17}$$

where D_p is the particle diameter, ρ_l is the liquid density and μ_l is the liquid viscosity (in consistent units).

This Reynolds number is then used to estimate a lower bound on the Sherwood number ($Sh = \frac{k_m D_p}{\mathcal{D}}$ where $\mathcal{D}$ is the diffusion coefficient).

$$Sh = 2 + 0.6 Re^{0.5} Sc^{1/3} \tag{4.18}$$

and hence a lower bound on the mass transfer coefficient, k_m. Sc is the Schmidt number ($\frac{\mu_l}{\rho_l \mathcal{D}}$). The bound represents the sum of molecular diffusion effects in a stagnant fluid (which give a Sherwood number of 2 for roughly spherical particles) and a boundary layer mass transfer term. This lower bound is then corrected using graphical correction factors for diffusivity and power per unit volume which typically increase k_m by about a factor of 2. A rough correction for the ratio of agitator and tank diameter, D_a/D_t, is also indicated. A criticism of this method is that large particles with very small density differences would be assigned inappropriately small mass transfer coefficients ($v_T \approx 0$), so that a minimum density difference of .3g/cm^3 must be assumed to obtain sensible results.

Brian and Hales' (1969) method uses a graphical correlation of Sh/Sc against $\frac{\epsilon_m^{1/3} D_p^{4/3} \rho}{\mu_l}$, where ϵ_m is the power per unit mass. This correlation was motivated by Kolmogoroff's theory of turbulence, though it can also be motivated by dimensional analysis. It is successful in fitting the data roughly over a wide range of the dimensionless term used. Nienow (1975) notes that Harriott's lower bound estimate gives points which fit this curve quite well, but that Brian and Hales' method fails to predict the variation of the coefficients with agitation power for particular fluids and tank geometries. Nienow (1975) and Levins and Glastonbury (1972a) both note that the mass transfer coefficient may vary quite significantly at a given power if agitation geometry is varied, limiting the usefulness of this method.

Levins and Glastonbury (1972c) propose a correlation which incorporates key features of both the above methods and seems to answer the main criticisms of the previous methods. The velocity, $\bar{v}$, used in the particle Reynolds number is treated as the r.m.s. sum of three components; the stagnant fluid terminal velocity (v_T), a fluctuating slip velocity estimated from Levins *et al.* (1972b) (v_s), and an effective velocity due to turbulence, v_e.

$$\bar{v} = (v_e^2 + v_T^2 + v_s^2)^{0.5} \tag{4.19}$$

v_s is noted to be less than a third of v_T and will therefore never make more than about a 3% contribution to the estimated mass transfer coefficient and is probably negligible considering that v_e is generally dominant. v_e is estimated from a correlation based on Levin and Glastonbury's (1972c) experimental work with neutral buoyancy particles to be

$$v_e = 0.93\frac{\mu_l}{D_p\rho}\left(\frac{D_p^{4/3}\epsilon_m^{1/3}\rho}{\mu_l}\right)^{1.23}\left(\frac{D_t}{D_a}\right)^{0.35} \tag{4.20}$$

The dimensionless variable used by Brian and Hales (1969) is introduced through this term. The velocity is then used in a similar equation to Harriott's except that the Schmidt number exponent was regressed as having a different value to the 1/3 associated with convective or boundary layer mass transfer

$$Sh = 2 + 0.44Re^{0.5}Sc^{0.38} \tag{4.21}$$

This correlation matches Harriott's method quite closely, with Harriott's approach giving somewhat higher transfer coefficients for particle diameters below about 100μ.

The dependence on agitator geometry is explained (Levins and Glastonbury, 1972c) by noting that more uniform power dissipation could be expected to give better mass transfer due to the dependence of transfer coefficient on $\epsilon_m^{1/6}$ at high enough power values. Little effect was observed due to agitator clearance "provided the particles are adequately suspended" and this was standardised at $Z_t/D_t = 0.3$. Nienow (1975) notes an effect of clearance on the power to just suspend the particles and hence achieve the lower bound associated with the terminal velocity. This should be accommodated by use of a suitable correlation for predicting when the particles become suspended and the v_T term in the correlation.

Levin and Glastonbury's method is supported by the reviews of Nienow (1975) and Treybal (1980).

Yagi *et al.* (1984) give a comparison between alternative correlations on 145μ ion exchange particles and 25.5μ $Ca(OH)_2$ particles. The methods of Brian *et al.* (1969) and Levins *et al.* (1972a) give excellent matches for the ion exchange particles with Harriott's method slightly underestimating the coefficients at all stirring speeds. The Levins and Glastonbury correlation overestimates the transfer coefficient for the $Ca(OH)_2$ particles by a factor of about 1.6 making it apparently worse than the correlation of Brian *et al.* This may be due in part to the use of a diffusion constant of $1.83.10^{-5}$ for $Ca(OH)_2$. Experimental data (Boynton, 1980) indicate a diffusion constant ranging from $1.93.10^{-5}$ at infinite dilution to $1.38.10^{-5}$ at saturation concentration. This suggests that the diffusion constant used could be a factor of 1.3 too high. With this correction the correlation of Levins and Glastonbury (1972c) becomes the best correlation for this case also with Brian *et al.* and Harriott following closely. Other correlations reviewed by Yagi *et al.* were inferior to these three.

All the work above refers to suspension in tanks. Harriott (1962b) reports experimental work on particles suspended in a pipeline. Per reports the mass transfer coefficients to be 10–30% less than in agitated tanks with the same power per unit volume and D/T of 0.5. The stagnant terminal velocity was observed to give a valid lower bound. The Levins and Glastonbury correlation with a reduction of 30% in the predicted coefficient subject to the terminal velocity lower bound should therefore provide a reasonable estimate for this case. An alternative approach is provided by Kikuchi *et al.* (1988) who sum the turbulent power dissipation with the power dissipated by the particle falling at its terminal velocity.

Figure 4.4 shows a comparison between the methods of Kikuchi *et al.* and Levins and Glastonbury applied to Ca(OH)$_2$ suspended in a pipe with a power dissipation of .001W/kg and suspended in a stirred tank with a power dissipation of .2W/kg.

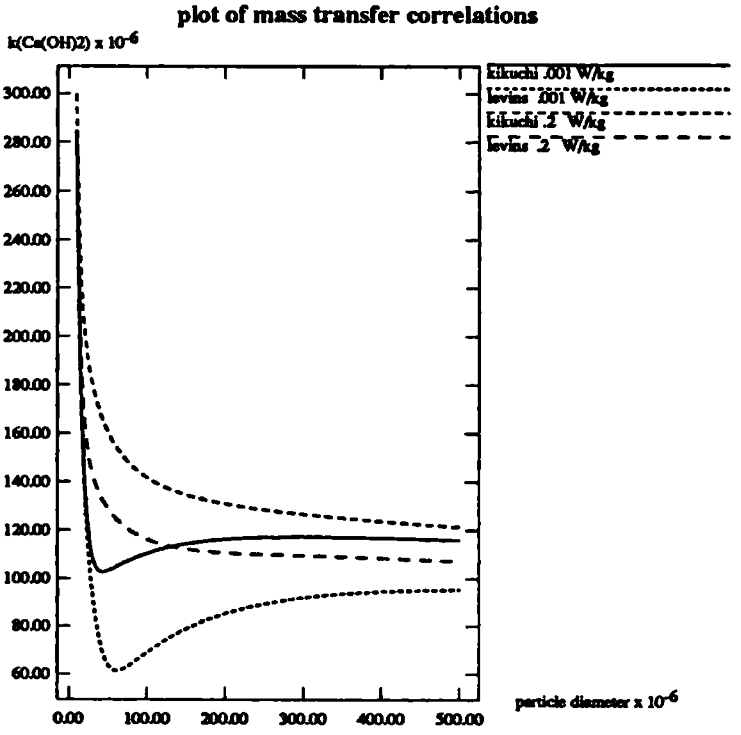


Figure 4.4: Comparison of mass–transfer correlations

The Levins and Glastonbury method is always more conservative and is therefore preferred. The two methods are fairly consistent, particularly for larger particles, and are in good agreement with experimental data for particles suspended in stirred tanks which indicates a transfer coefficient of about 130 × 10^{-6} m/sec (see section 4.2.2).

On the basis of the observations above, the correlation of Levins and Glastonbury (1972c) appears the best of the standard correlations. However, more recent work has indicated some weaknesses of this correlation. Brucato *et al.* (1990) note that their experimental results with dense particles ($\rho_p > 2kg/m^3$) show little effect of particle size down to 30μ particle diameter. This 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 90μ and 1000μ particle diameter. The effect of ϵ_m was however consistent with Levins and Glastonbury's correlation. 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. This suggests that for dense particles the correlation should be applied at a typical particle size and that this rate should be assumed constant for moderately sized particles. 200μ diameter is a suitable size as most of the correlated data is around this size and the available correlations are in good agreement.

Two recent papers (Asai *et al.*, 1988; Armenante and Kirwan, 1989) have focussed on micro-particles around the size at which molecular diffusion is expected to be the dominant mass transfer mechanism (Sh=2). This work has confirmed that the exponent relating k_m and $\mathcal{D}$ varies from $\frac{2}{3}$ to 1 as particle size decreases and has confirmed the limiting Sh of 2 for roughly spherical particles (molecular diffusion). Asai *et al.* 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, as in Levins and Glastonbury's correlation.

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 200μ diameter (determined experimentally or using Levins and Glastonbury's correlation) and the mass transfer rate due to molecular diffusion which typically becomes dominant around 20μ diameter in a well mixed solution ($\epsilon_m \approx .2$ W/kg).

The work of Yagi *et al.* (1984) indicates that mass transfer may fall substantially below the molecular diffusion limit for high particle concentrations (particles representing several percent by volume of the solution). They observed that as the particle concentration increases the mass transfer rate drops substantially and stirring dependence reduces. They explain this observation by several particles smaller in size than micro-eddies becoming trapped together in a single eddy and impeding diffusion from each other and propose a correlation based on surface area per unit volume, a. The explanation seems somewhat contrived to me as it is well known that small particles tend to agglomerate giving a reduction in effective surface area. A very detailed analysis of their experimental setup also seems appropriate to rule out experimental artefacts as their method relied on being able to measure very small deviations from the saturation concentration of the solution to determine the mass transfer rate. Asai *et al.* (1988) carried out similar investigations and did not note a fall in mass transfer rate with particle concentration indicating that Yagi *et al.*'s claim of a general observation is not valid. The possibility of particle agglomeration should always be considered in assessing experimental data.

All the correlations discussed require that the particles should be fully suspended before the correlations apply. The most generally supported correlation (Uhl and Gray,

1986) for predicting when this will occur in agitated tanks is that of Zwietering (1958).

$$N_{js} = \frac{s(\frac{\mu_l}{\rho_l})^{0.1} D_p^{0.2} \frac{g(\rho_p-\rho_l)}{\rho_l} X^{0.13}}{D_a^{0.85}} \tag{4.22}$$

where N_{js} is the agitation speed (r.p.s.) achieving complete suspension, s is a tabulated correction term depending on agitator type and geometry and X is the mass of solid particles as a percentage of the mass of liquid. All variables are in standard SI units. X should be fixed at 1 if the concentration drops below 1% (Nienow, 1975).

Similar correlations are available for particles in pipes of which the most appropriate for fine particles (less than about 400 μ diameter) seems to be that of Spells (1955). This correlation is based on the particle size, D_{85}, below which 85% of the particles (by mass) lie and may be rearranged to give

$$v_{js} = (.0251 g D_{85} \frac{\rho_p - \rho_l}{\rho_l} (\frac{D_{pipe}\rho_m}{\mu_l})^{.775})^{.816} \tag{4.23}$$

where v_{js} is the fluid speed at which particles are just suspended, ρ_m is the density of the suspension and μ_l is the viscosity of the liquid. Standard SI units are used. General purpose correlations such as that of Oroskar *et al.* (1980) are more conservative, probably unnecessarily so. Fluid velocities of 1m/sec are ample for suspension of typical solid alkali reagents. If trying to use high slurry concentrations, above about 8% by volume, experimentation may be needed to confirm that the slurry can be pumped reliably as the viscosity may increase dramatically.

4.2.1.2 Predicting mass transfer rates with chemical reaction

A lot of work in this area has focussed on finding analytical solutions for special cases. This work is not directly relevant as it is insufficiently general so it will not be reviewed here. The most relevant work, in terms of problem formulation, is that of Olander (1960). This deals with the solution of molecular diffusion problems with chemical reaction.

The general method for instantaneous reactions (following Olander (1960)) is given below.

1. Define conditions in the bulk fluid.

2. Define solid/liquid boundary condition(s) — fluid phase equilibria, solid/liquid equilibria. If a component, or combination of components, cannot penetrate the particle surface and is not generated by surface reaction then set its derivative with respect to distance from the surface to zero, at the solid/liquid boundary.

3. Define mass transfer equations under molecular diffusion using reaction invariants. For example, for $A + B \rightleftharpoons C$ define the mole balance for $W = A + B + 2C$ as the concentration of W does not vary with extent of reaction.

4. Set the rate of change of concentration with time to zero, allowing analytical integration of the mass balance equations and solve the resulting set of algebraic equations.

For spherical particles the mass balance equations take the form

$$\frac{\partial [W]}{\partial t} = \frac{1}{r^2} \sum \mathcal{D}_i a_i \frac{\partial r^2 \frac{\partial [W_i]}{\partial r}}{\partial r} \tag{4.24}$$

where r is the distance from the particle centre and W is a reaction invariant, $W = \sum a_i W_i$. $[W]$ denotes the molar concentration of W. Applying the transformation $r' = (1 - r_p/r)$, where r_p is the particle radius, gives

$$\frac{\partial [W]}{\partial t} = \frac{(1-r')^4}{r_p^2} \sum \mathcal{D}_i a_i \frac{\partial^2 [W_i]}{\partial r'^2} \tag{4.25}$$

The transformed variable r' is 0 at the particle surface and goes to 1 as r goes to infinity. The steady-state solution of this equation ($\frac{\partial [W]}{\partial t} = 0$) takes the form

$$\alpha + \beta r' = \sum \mathcal{D}_i a_i [W_i] \tag{4.26}$$

These coefficients may be solved for algebraically given the bulk solution compositions and particle surface boundary conditions. With concentrations in molar and diffusion coefficients in cm^2/sec the rate of dissolution of W becomes $10^{-3}\beta / r_p$ (gmoles/cm^2 sec) where r_p is the radius of the particle in cm. The rate of change of particle radius (cm/sec) is given by

$$\frac{\partial r_p}{\partial t} = -\frac{10^{-3} \beta M}{r_p \rho} \tag{4.27}$$

where M is the molecular weight and ρ is the density in g/cm^3.

If finite rate reactions are present then an additional term appears in the mass-balance equations to account for the rate of change of $[x]$ due to reaction rate, $\mathcal{R}$

$$\frac{\partial [x]}{\partial t} = \frac{1}{r^2} \mathcal{D}_x \frac{\partial r^2 \frac{\partial [x]}{\partial r}}{\partial r} + \mathcal{R} \tag{4.28}$$

which gives the transformed steady-state relationship

$$0 = \frac{(1-r')^4}{r_p^2} \mathcal{D}_x \frac{\partial^2 [x]}{\partial r'^2} + \mathcal{R} \tag{4.29}$$

The solution of the steady-state mass transfer equations becomes a dynamic boundary value problem.

It is possible to refine the above model to include effects such as activity coefficient variation with concentration altering the effective diffusion coefficients and electric fields influencing the mass transfer of ions. This was not found to be necessary for this work and will not be reviewed.

The analysis of molecular diffusion with chemical reaction can be combined with the general correlations for mass transfer without chemical reaction by using the correlations to estimate an effective distance from the particle surface at which the bulk fluid concentrations apply (film thickness). The reaction equations may then be solved with the bulk boundary conditions moved from $r' = 1$ to nearer the particle surface.

4.2.2 Reaction of $Ca(OH)_2$

Previous work on $Ca(OH)_2$ reaction kinetics presents a confusing picture. Two German papers (Becker and von Zander, 1977; Becker, 1986) assert that the dissolution of $Ca(OH)_2$ is *surface reaction controlled* with a rate proportional to effective surface area. Shinskey (1973) reviews the work of Docherty (1972) concluding that Docherty's observed responses "do not conform to any of the conventional reaction rate equations", but noting that "the *slower reaction below pH 8* can be avoided by conducting a first stage reaction at pH 9.5" [my italics]. This recommendation has had a strong influence on industrial practice, with overneutralisation in the first tank and acid addition to later tanks being common practice. Haslam *et al.* (1926) give a wide-ranging analysis of the dissolution and reaction behaviour of CaO, considering the dissolution to be *mass-transfer controlled by the diffusion of* $Ca(OH)_2$. The dissolution rates observed by Haslam are about 3.5 times those given by Becker (assuming the gross surface area to be the effective surface area). Haslam demonstrates that reaction rate is proportional to gross surface area for particles in the range 90–1000μ. Yagi *et al.* (1984) treat the dissolution of $Ca(OH)_2$ as mass transfer controlled, showing that the rate and its variation with agitation correspond well to mass transfer control. Both Becker and Haslam note that the presence of *acid accelerates the dissolution of* $CaO/Ca(OH)_2$ in contrast to Shinskey.

The work of Docherty involves a number of errors — incorrect particle sizing and poor experimental design — so that it does not justify detailed review.

The work of Haslam and Yagi is consistent and gives the most plausible model for the dissolution characteristics of $Ca(OH)_2$ — mass transfer control.

Haslam *et al.* (1926) give experimental results from which the reaction equation

$$\frac{\partial r_p}{\partial t} = -(.09 + 6.25[H^+])\ \mu/sec \tag{4.30}$$

may be inferred (for $[H^+] < .025M$ and $r_p \in [45, 500\mu]$), where $[H^+]$ is the concentration of hydrogen ion in solution and r_p is the radius of the particle in microns. Yagi *et al.*'s (1984) results indicate that the mass transfer coefficient (for dissolution in water) of 12.5 μ radius $Ca(OH)_2$ particles is about 40% smaller than that predicted by the standard mass transfer correlation of Levins and Glastonbury (see section 4.2.1.1). Haslam's data are broadly consistent with this correlation except that the correlation predicts a 30% variation in rate over the particle size range examined while no significant variation was observed. Exact comparison cannot be made as Haslam et al's mixing conditions are not fully specified, but an exact match is obtained for particles with a radius of 100μ assuming a power input of .2W/kg which is consistent with the complete suspension of 1mm diameter particles noted by Haslam *et al.* The acceleration of dissolution in acid conditions closely matches that predicted based on mass transfer control coupled with acid-base equilibrium reaction in solution (the reaction rate of .09 in the absence of acid is consistent with an acid acceleration coefficient of 7, compared to the observed 6.25, assuming a saturation concentration of $Ca(OH)_2$ of .022M, $\mathcal{D}_{Ca(OH)_2} = 1.4.10^{-9} m^2/sec$ and $\mathcal{D}_{H^+} = 9.10^{-9} m^2/sec$). It should be noted that the mass transfer analysis found to match the experiments was based on the *diffusion of $Ca(OH)_2$ for both CaO and $Ca(OH)_2$ solid phase* indicating that this is the rate limiting step in both cases.

The model of Becker is demonstrably incorrect in that the work discussed above shows that surface reaction is not the controlling mechanism for $Ca(OH)_2$ dissolving in water. The observations of Becker and von Zander (1976) can be reconciled with the mass transfer model, by noting that their rate equation was derived from a rotating disk experiment and that their analysis of particle dissolution relied on an effective surface area substantially greater than the gross surface area. Becker's theoretical argument for reaction control (Becker, 1986) is over simplistic.

Recognising that the dissolution reaction is mass transfer controlled, the general understanding of solid-liquid mass transfer can be drawn upon. From this it can be expected that rate of reaction will vary depending on mixing conditions, though variation will usually only be about 2:1 once full suspension of the particles is achieved. This means that consideration should be given as to whether the mixing conditions in experiments are consistent with the mixing conditions expected in the process. It can also be expected that the rate of reaction will increase as the particle size decreases below the value at which the rate due to molecular diffusion exceeds the rate observed for particles around 100μ radius (see section 4.2.1.1).

I have analysed data from three ICI applications using $Ca(OH)_2$ reagent. In each case the reagent particle size distribution was estimated experimentally and approximated by a discrete particle size distribution to allow a lumped parameter model to be used. The method chosen was to consider about 10 particle sizes to approximate the particle size distribution. The concentrations of particles, $[x_i]$, at each radius chosen, r_i, were then updated by considering fluxes due to solid dissolution, particle shrinkage and flow through a perfectly mixed volume with residence time τ. This gives the equation

$$\frac{\partial[x_i]}{\partial t} = 3\frac{\partial r_{p_i}}{\partial t}[x_i]/r_{p_i} - \frac{\partial r_{p_{i+1}}}{\partial t}[x_{i+1}]/(r_{p_{i+1}} - r_{p_i}) + \frac{\partial r_{p_i}}{\partial t}[x_i]/(r_{p_i} - r_{p_{i-1}}) + ([xin_i] - [x_i])/\tau \tag{4.31}$$

where $r_{p_{i+1}} > r_{p_i}$. For simulation of batch experiments τ is set to a very large value and the initial concentrations set to correspond to the reagent particle size distribution.

In the first two cases the mass transfer based model gave a very good match to the data, with coefficients similar to those given by Haslam and an additional effect included for weak acids based on mass transfer analysis

$$\frac{\partial r_p}{\partial t} = -k_{mix}.max(1, \frac{10}{r_p.k_{mix}})(k_1 + k_2[H^+] + k_3 cweak) \ \mu/second \tag{4.32}$$

where r_p is the particle radius in microns, $[H^+]$ is the concentration of free hydrogen ions and *cweak* is the concentration of weak acids which 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 one 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 results for these two cases are shown in figures 4.5 and 4.6. In both cases the particle size distribution was estimated as log normal, truncated at 250 μ.

$$\phi(\ln(r_p)) = \frac{1}{1.15\sqrt{2\pi}} \exp\left(-((\ln(r_p) - 1.39)/1.15)^2/2\right), \ r_p < 250\mu \tag{4.33}$$

A 12 point discretisation was found to give an adequate match to a 200 point discretisation used as a reference case. The initial quantity of reagent was treated as an adjustable parameter, in fitting the experimental data, to avoid large errors near neutral due to small errors in the quantity added. The first set of experimental data was based on the neutralisation of fairly pure 4M HCl while the second set of experimental data was based on the neutralisation of .1M sulphuric acid along with significant weak acid and heavy metal components.

In the second set of experiments, it was necessary to adjust the model to give a good fit to the response above pH 8. The deviation observed from the standard model was consistent with the larger particles not dissolving at the higher pH and was modelled empirically by multiplying the rate of reaction of the largest 7% of particles by a factor

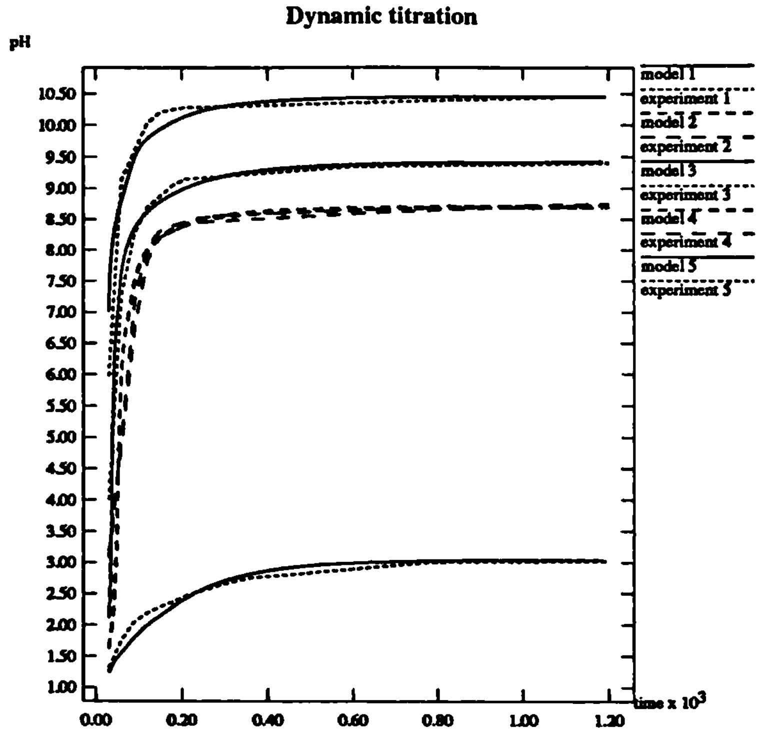


Figure 4.5: Batch experiments with $Ca(OH)_2$ (1)

of 10^{8-pH} for pH values above 8. This model must be regarded as very tentative as the data are limited and the apparent reduction in rate of reaction observed may also be due to the reaction being masked by backward reactions such as $CO_2 \rightarrow HCO_3^-$ or $Ca^{2+}(aq)+CO_3^{2-}(aq) \rightarrow CaCO_3(s)$. A possible explanation for slower reaction of the large particles at pH values above 8 lies in the observation by Boynton (1980) that carbonate impurities are concentrated in the larger particles. These may obstruct dissolution of the $Ca(OH)_2$. The "alkali inhibition" effect was not observed in the first set of experiments which had a much higher initial concentration of acid (4M versus .1M). The effluent examined was a complex mixture of weak acids and metal impurities.

This inhibition effect is likely to *increase* the *apparent* rate of reaction if this is calculated assuming that the final pH always corresponds to 100% conversion. This may account for Shinskey's (1973) suggestion that reaction is faster at high pH which is used to motivate pers recommendation to carry out a first stage reaction around 9.5 pH. The analysis above indicates that the apparent acceleration is either an illusion, which will not benefit control or reagent conversion, or a product of an inhibition of reaction which would actually increase the carryover of reagent from the first stage which may subsequently react in later stages, particularly if acid is added. Use of overneutralisation also implies a cost penalty due to the extra reagent required and creates conditions in which precipitation and probe fouling are more likely. The only clear benefit of the scheme proposed by Shinskey is that the provision of acid reagent to the second stage allows excessive carryover of reagent to be compensated for by acid addition provided the reagent carried over reacts with the added acid within the treatment system. This benefit may be achieved without overneutralisation by providing both acid and alkali addition to the second stage reactor. The general recommendation to use lime overneutralisation does not appear to be justified. Chapter 6 illustrates this further through two design examples with $Ca(OH)_2$ which do not use overneutralisation (sections 6.3 and 6.4).

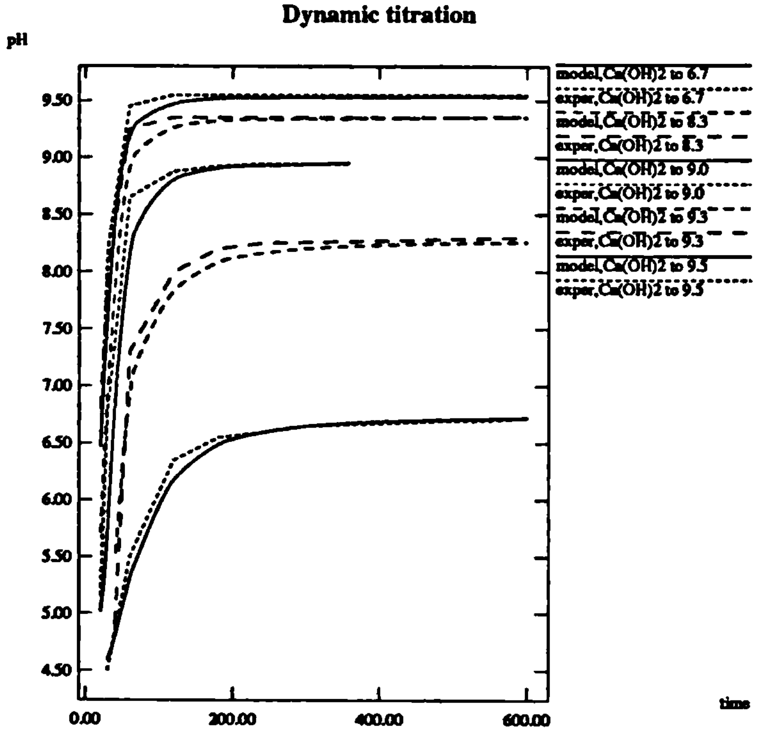


Figure 4.6: Batch experiments with $Ca(OH)_2$ (2)

For the third set of data the model fit was not as good, with the best fit being obtained for $k_{mix} \approx .3$ and the other parameters as before. In this case the experiments were carried out using a magnetic bead stirrer rather than an agitator, and it is believed that this resulted in markedly worse mass transfer (possibly involving incomplete suspension), accounting for the discrepancy between the data sets. The fit achieved is shown in figure 4.7. The particle size distribution showed a higher median size, but a shorter tail, so that the tail of the particle size distribution was fairly similar to the previous log normal distribution. An 8 point discretisation was used. The effluent examined was a complex mixture of weak acids.

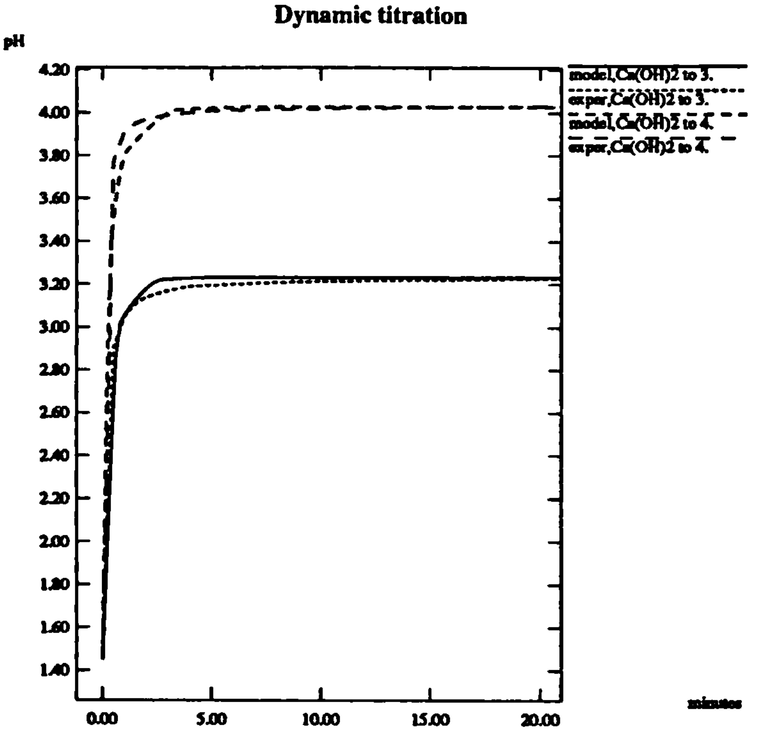


Figure 4.7: Batch experiments with $Ca(OH)_2$ (3)

None of the reactions above showed evidence of chemical inhibition in acid or neutral conditions. This can occur if any compound forming an insoluble calcium salt is present at concentrations above its solubility product. Reaction inhibition has been observed with sulphates, fluorides, oxalates and tartrates. This possibility is best tested for by batch titration experiments. If the results differ substantially from the model developed in this thesis and mixing can be shown to be good then an inhibiting reaction is very likely and possible causes and remedies should be investigated.

None of the examples above showed a significant effect due to the concentration of $Ca(OH)_2$ in solution approaching the saturation concentration, $[Ca(OH)_2]_{sat}$. The effect of the concentration of $Ca(OH)_2$ in the bulk solution is readily accommodated, based on standard mass transfer analysis, by replacing k_1 in equation 4.32 by $k_1([Ca(OH)_2]_{sat} - [Ca(OH)_2])/[Ca(OH)_2]_{sat}$.

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

1. The error in fitting the model to the data. This is a relatively small contribution to the total uncertainty.

2. 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.

3. The variation of the reagent particle size distribution. This depends very much on the quality of the reagent source, and can be quite high, possibly dominating the overall uncertainty.
4. 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 $.2W/kg$ ($k_{mix} \approx 1$). There is significant uncertainty in the effect of plug-flow mixing, though so long as the particles are properly suspended $k_{mix} \in [.5, .9]$ can be expected.

To characterise the reagent properties for particular applications batch (dynamic) titrations should be carried out with a range of final pH values as in the examples above. 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, 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 then 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. The most likely need for adjustment lies in the alkali region where reaction inhibition or back reactions may occur.

4.2.3 Reaction of $CaCO_3$

Work on $CaCO_3$ has been carried out by both geochemists and researchers on flue gas desulphurisation. There is a consensus that below pH 4 the reaction in strong acids is governed by mass transfer of the H^+ ion. Chou (1989) follows Plummer (1978) in attributing an observed acceleration of reaction by CO_2 in dilute acid solutions to a surface reaction mechanism. Chan and Rochelle (1982) present a wide-ranging investigation in which the effect of pH, CO_2 and weak acids are all modelled using a mass transfer model. The model used is quite successful in explaining the experimental data, including the effect of CO_2. Bjerle and Rochelle (1984) use a variation of this model to investigate dissolution from a plane surface with mixed success. Later work by Wallin and Bjerle (1989) directly compared the surface reaction and mass transfer based models and found that the predictive capability of the mass transfer based model was usually better than the Plummer surface reaction model, even with the coefficients of the Plummer model recomputed for the new data.

The mass transfer based model seems strongly supported by results in the literature with the exception of the work reported by Bjerle and Rochelle (1984). Its main limitation for the purposes of effluent treatment modelling is that it is quite complex and requires solution of a dynamic boundary value problem. I therefore decided to implement a mass transfer based model of carbonate dissolution, compare it to the reported results and attempt to derive a simplified model suitable for use in treatment system design.

In implementing the model, it was noted that the published models neglected variation of ionic strength and hence activity coefficients moving away from the material surface. Activity correlations presented by Truesdell and Jones (1974) using the Hückel equation (4.1) were used in this work, incorporating the effect of ionic strength variation. Equilibrium constant data from Nordstrom *et al.* (1990) were used. In solving the dynamic boundary value problem, the DAE equations were integrated from the particle surface ($r' = 0$) to $r' \leq .995$ (see section 4.2.1.2). In Chan and Rochelle (1982) the equations were integrated from $r' = 0$ to $r' = .85$ and cubic extrapolation used. This probably does not make a major difference. The species $CaCO_3^0$ and $CaHCO_3^+$ used by Chan and Rochelle were removed as their existence is disputed and the solid boundary condition was expressed in terms of the solubility product of $CaCO_3$ rather than as a concentration of $CaCO_3^0$. Rather than using an overall correction factor to scale rates calculated assuming molecular diffusion into an infinite medium (Chan and Rochelle, 1982) I used a film model and adjusted the distance from the particle surface at which the bulk solution concentrations were assumed to apply. The model was otherwise consistent with the papers above. Results using this model were consistent with Chan and Rochelle (1982) and Wallin and Bjerle (1989). I could not reproduce the results of Bjerle and Rochelle (1984) and in fact obtained a strikingly better fit to the data they presented, suggesting that there was some error in their version of the model. One of the equations in the paper by Bjerle and Rochelle gives an incorrect sign in the dynamic equation for the finite rate reaction $CO_2 \rightleftharpoons HCO_3$. The model developed was also quite consistent with the rate model developed by Plummer (1978). Based on these comparisons, the model was taken as validated.

Much of the complexity of the model comes from consideration of the reaction between CO_2 with HCO_3 as a finite rate reaction, as this requires the solution of a dynamic boundary value problem to model mass transfer. Two alternative models were considered. One model neglected this reaction and the other assumed the reaction was instantaneous. Both of these algebraic models showed discrepancies of up to a factor of 2 from the full dynamic model, even when the bulk concentration of CO_2 was negligible, and were not pursued further. Both simplified models failed to give predictions of the effect of CO_2 of the correct order of magnitude.

I applied the full model to a system representative of the condition for which some industrial data were available, a highly buffered system in which the buffering was provided by a mixture of weak acids with a range of *pk* values (the *pk* is the pH at which about half the acid has dissociated). In this system the accelerating effect of CO_2 was found to be insignificant. The effect of CO_2 in reducing the "saturation pH" at which $CaCO_3$ commenced precipitating and mass transfer was reversed was very significant. Weak acids with *pk* values below the saturation pH were found to be effective in accelerating the reaction. This acceleration effect was found to be approximately proportional to the available hydrogen ions released by the weak acid as pH varied between the bulk pH and the saturation pH at which carbonate would start to precipitate. The strength of the weak acid effect was proportional to the assumed acid diffusion coefficient. The effect of each weak acid varied only slightly with its *pk* value and the bulk solution pH, provided these were at least .5 pH below the pH at which carbonate would precipitate. The effect of $[H^+]$ concentration was found to vary markedly with bulk pH. The effect on rate of dissolution above about pH 4 was about twice the effect below about pH 3. This is due to the finite rate reaction between CO_2 and HCO_3 changing the effective stoichiometry of the reaction between H^+ and CO_3^{2-} within the boundary layer. The precise transition in effectiveness of H^+ with pH was a function of the boundary layer thickness. With the weak acids, the apparent variation in effectiveness with pH appeared less (about 30%). In practice the effectiveness of the weak acids will also be substantially influenced by the diffusion coefficients of the particular weak acids.

The above analysis suggests that calcium carbonate dissolution can be modelled similarly to calcium hydroxide, with the main changes being that the coefficients for the effect of acids on reaction rate can be expected to be somewhat higher due to the higher molar volume of calcium carbonate (30% greater) and to exhibit up to a further factor of 2 increase as neutrality is approached. The accelerating effect of CO_2 is not large enough to be significant in applications in which $CaCO_3$ is used as a reagent to be added under control as in this case.

The mass transfer analysis can be used to predict approximate values for the model coefficients (equation 4.32).

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

For $CaCO_3$ k_1 is about .0001 (insignificant) and k_2 and k_3 vary with pH, ranging from about 20 and 2–4 respectively at high pH (above about 4 pH) to about 10 and 1–2 respectively at low pH.

Unfortunately, the only industrial data available for this case were from the same experimental set-up suspected of giving bad mixing with $Ca(OH)_2$. The best fit was obtained with $k_{mix} \approx .3$ (as before) and $k_3 \approx 1$ below pH 3 rising linearly to 2 by pH 4. The saturation pH, used as pH_{ref} for calculating *cweak* was 5.5pH. This pH value was consistent with considerable supersaturation of the experimental solution with CO_2, and also suggests poor mixing. The fit achieved is shown in figure 4.8.

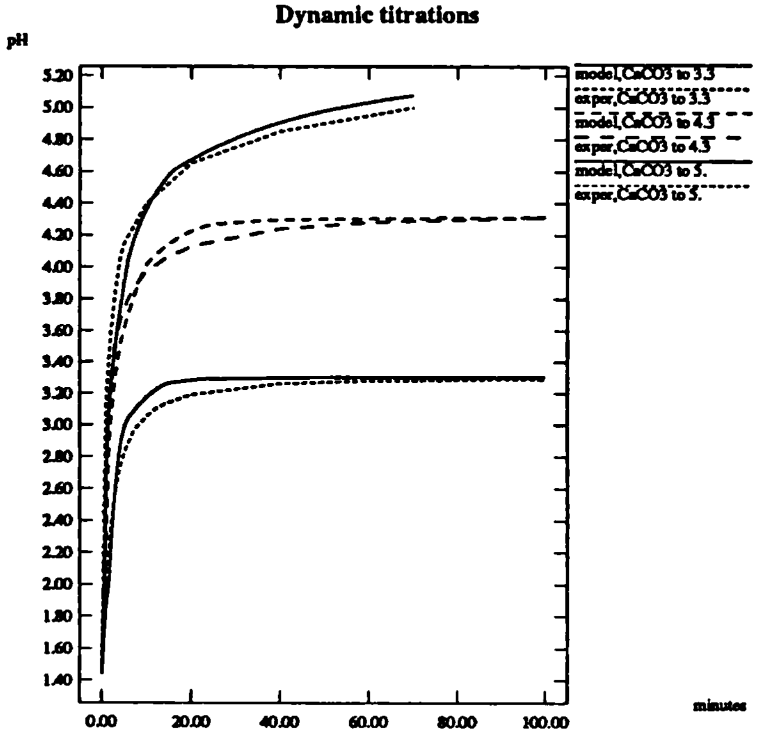


Figure 4.8: Batch experiments with $CaCO_3$

Model implementation, sources of uncertainty and experimental procedure are essentially the same as for $Ca(OH)_2$. Final pH values will be limited to between 5–7 pH depending on the degree of CO_2 supersaturation in the solution.

4.2.4 Other solid alkalis

The reaction model discussed for $Ca(OH)_2$ applies to quicklime, CaO, with the qualification that quicklime swells and fractures in reaction with water, so that the particle size distribution improves as the reaction progresses. The difference in reactivity between quicklime and $Ca(OH)_2$ appears to be due to the reduction in particle size by about a factor of 10 which can be achieved by reacting typical quicklimes with water under appropriate conditions (slaking).

Dolomitic slaked lime (MgO.CaO) reacts similarly to CaO up to about pH 7. Thereafter the Mg component fails to react and may precipitate out of solution (Haslam *et al.*, 1926). $Mg(OH)_2.Ca(OH)_2$ will probably behave similarly.

Dolomitic limestone ($MgCO_3.CaCO_3$) reacts at about one tenth the rate of $CaCO_3$. Its reaction characteristics are not very well understood (Chou *et al.*, 1989).

Reaction of magnesia ($Mg(OH)_2$) is not mass transfer controlled. Even if it was mass transfer controlled it would react about 100 times more slowly than $Ca(OH)_2$ due to lower solubility. This is only partially compensated for by the fact that typical magnesia products, manufactured by precipitation, have a narrower particle size distribution than typical $Ca(OH)_2$ products manufactured by pulverising, burning and slaking limestone. This counts strongly against its use.

The characteristics of waste alkali products may be very complex as they tend to have high and variable impurity levels and coarse and variable particle size distributions. These must be dealt with on a case by case basis. I examined one such reagent during this project (made up of varying proportions of calcium and magnesium oxides, hydroxides and carbonates together with sulphate impurities) but was unable to model it to a satisfactory accuracy.

4.3 Redox reactions

Redox reactions pose particular modelling challenges in two respects, firstly in determining the kinetics of the reactions and secondly in relating compositions of reactants and products to the redox measurement.

The kinetics of redox reactions are generally far from trivial. This is illustrated by the industrially important cyanide, CN^-, oxidation with hypochlorite, ClO^-. There is a clear consensus on the key reactions in the pH range 10–13 typically used for carrying out this reaction. The favoured reaction scheme (Eilbeck and Mattock, 1987) is that CN^-

is almost instantaneously oxidised to ClCN

$$CN^- + ClO^- + H_2O \rightarrow ClCN + 2OH^- \tag{4.34}$$

which is then hydrolysed to cyanate, NCO$^-$, according to the equation (Price *et al.*, 1947; Eden and Wheatland, 1950; Bailey and Bishop, 1973)

$$\begin{gathered} ClCN + 2OH^- \rightarrow NCO^- + Cl^- + H_2O \\ \frac{d[ClCN^-]}{dt} = -k[ClCN][OH^-],\ k \approx 8.10^9 exp(\frac{-6158}{T}) \end{gathered} \tag{4.35}$$

where k is in M^{-1}sec^{-1}. The rate coefficient given is an average of the data in the above references. Eden *et al.* note that this hydrolysis is catalysed by excess hypochlorite giving an increase in the reaction coefficient, k, of about 9.10^4M^{-1}sec^{-1}/M available chlorine at pH 11, with little variation with temperature. This compares to an uncatalysed rate of 8M^{-1}sec^{-1} at 25°C. Bailey and Bishop (1973) investigated the catalytic effect further and found that it only occurred consistently with fresh hypochlorite solution in the pH range 8.5–10.5. The catalytic effect observed with about .08mM available chlorine as hypochlorite diminished with the age of the solution. The explanation presented was that the catalytic effect was due to transient high concentrations of molecular chlorine which decayed slowly to a low equilibrium concentration. The decay appears to take several weeks. At pH values above 10.5 their results became "rather irreproducible", due to rapid hydrolysis of free chlorine. At pH values below 8.5 the reaction of cyanate with Cl$^-$ appears to eliminate the catalytic effect by removing the chlorine.

Weak bases such as borax or carbonate also accelerate the reaction (Bailey and Bishop, 1973).

NCO$^-$ is then oxidised further to CO$_2$ and N$_2$ according to the equation (Teo and Tan, 1987)

$$\begin{gathered} 2NCO^- + 3ClO^- + H_2O \rightarrow N_2 + 2CO_2 + 3Cl^- \\ \frac{d[CNO^-]}{dt} = k[CNO^-][ClO^-][H^+],\ k = 2.6e13exp(\frac{-3251}{T}) \end{gathered} \tag{4.36}$$

Note that the second stage reaction is favoured by acid conditions while the first stage reaction is favoured by alkali conditions so that two stage oxidation is normally carried out in two separate physical stages. The need for oxidation of NCO$^-$ is disputed as it is believed that cyanate is non-toxic and extremely stable at pH values above or near neutral. An alternative second stage oxidation is believed to occur as well generating NO$_3^-$ instead of N$_2$ and accounting, at least partially, for the observation that the second stage oxidation consumes up to 3.7 hypochlorite ions instead of the 3 ions indicated by equation 4.36. The reactions above are affected by the presence of metal ions which form complexes of varying stability with cyanide. Iron and, to a lesser extent, nickel form complexes which are not broken up under cyanide oxidation.

Mapstone (1978) carried out investigations of cyanide oxidation in the range pH 9–10. Per argues that both ClCN and NCO^- are destroyed with equal rapidity by this second stage reaction, making the hydrolysis of ClCN irrelevant to the rate of generation of second stage oxidation products. The second stage oxidation kinetics given by Mapstone are consistent with the results of Teo *et al.* (1987) for NCO^-, assuming a pH of 9.5. Teo's rate equation (4.36) is in turn consistent with that of Price (1947) for ClCN oxidation suggesting a common mechanism and bearing out Mapstone's contention that ClCN and NCO^- are oxidised at the same rate. This will be assumed in modelling work in this project.

Below pH 10, hydrolysis of ClCN is slow which is generally believed to give rise to a serious risk of its release as a gas (almost as toxic as HCN). HCN which is a weak acid is only half dissociated at pH 9.5 which if the initial cyanide oxidation was sufficiently slow would pose a risk of toxic gas emission as well. Mapstone attributes slow dynamics to the initial oxidation of cyanide which per treats as producing both ClCN and NCO^-. This conclusion is rebutted by Gerritsen and Margerum (1990) who attribute it to an experimental artefact and support the rapid conversion of cyanide to ClCN. Rapid conversion to ClCN is assumed in this project.

In summary, a reasonably clear picture of the reaction kinetics (equations 4.35 and 4.36) can be formed for moderately alkaline conditions with some uncertainty about the catalytic effect of excess hypochlorite on the hydrolysis of ClCN. The uncertainty in the rate of the hydrolysis reaction can be treated, to a first approximation, as up to a factor of 10 acceleration in the base rate. The rate of the second stage oxidation of $ClCN/NCO^-$ can be assigned a notional uncertainty of $\pm 20\%$ with the stoichiometry varying between 3 and 3.7 hypochlorite ions per 2 cyanide ions. This model and uncertainty description should encompass the true system in the absence of stable iron or nickel cyanide complexes. In the presence of such complexes alternative treatment methods may be needed.

Eilbeck (1987) discusses the complexity of electrode response

> Redox electrodes do not always respond rapidly; and indeed the less well defined the reaction system, or the greater the thermodynamic irreversibility the slower is the response. [For cyanide oxidation] with platinum and gold [electrodes] there is an immediate response but full equilibration takes some 10 minutes after following hypochlorite addition and approximately 3 minutes following the introduction of cyanide to lower the potential....Silver is much faster however.

The summarising conclusion is that for most reactions used in waste water

> treatment the redox curve does not accurately reflect the progress of the reaction. From a practical point of view this is not of major importance, however, because oxidative treatment of a pollutant demands that there be effectively no pollutant left ...and so inevitably a small amount of overdosing must occur. For this the redox curve as measured is normally an adequate measure of acceptable treatment.

The time response of the electrodes is noted by Eilbeck to be generally close to the time constants of the reaction implying that, as the residence time must allow the reaction to go virtually to completion, the probe lag should not prevent reasonable control of backmixed reactors. The uncertainty in the dynamics can be approximated as a range of first order probe lags. Problems with redox titration curve repeatability are most acute near the equivalence point of the reaction which is not used for control due to the need for overdosing. *Ratio* of oxidant to reductant, neglecting reaction, appears to be a suitable axis for redox curve tabulation (exact for ideal characteristics). Uncertainty in the redox curve seems best evaluated by repeated experiments, ideally with several effluent samples.

Construction of redox models for a given reaction scheme should be straightforward. The models can be validated to a first approximation against current design practice. The model of cyanide oxidation above forms the basis for design work discussed in section 6.6.

4.4 Measurement response

The response of electrodes used to measure pH or redox potential is commonly modeled as a first order lag ranging from one second to several minutes. The actual behaviour of electrodes is known to be very complex and a number of researchers have attempted to develop improved models.

McAvoy (1979) notes that responses of pH electrodes can be several times faster in buffered solutions than in unbuffered solutions, *usually* slower near neutral and that response is faster going from acid to water than from water to acid. The response is modeled using fast neutralisation models coupled with diffusion through a stagnant layer, which successfully predict different responses for increasing and decreasing pH though they do not give a good quantitative match to the experimental data. This work suggests that the probe dynamic response will vary with the particular composition of the mixture being measured.

Hershkovitch, McAvoy and Liapis (1978) extend this work, adding migration (motion under an electric field) effects to the diffusion model originally used by McAvoy. This work matches some important experimental observations, e.g. the salt effect in which mixing an acid with an acid and salt solution at the same pH generates an apparent pH transient before returning to the original pH value. Qualitative and quantitative discrepancies between experimental results and model predictions were ascribed to a combination of neglected flow effects and neglecting the electrode double layer potential. There is also a possibility that diffusion within the glass itself makes a significant contribution though *some* experimental work suggests not. This work supports the view that probe response is a function of the particular composition of the solution being tested. It does not however provide a model which successfully predicts probe response for the reasons below.

1. It requires a parameter, stagnant layer thickness, to be chosen experimentally to match the observed response.
2. Even for the case for which the model gives the best match, concentrated strong acids, the appropriate choice of this parameter varies substantially for moderate changes in pH (pH 3 — .12mm, pH 2 —.15mm), varies substantially with fluid velocity past the electrode (42cm/s — .12, 27cm/s — .16) and in some but not all cases varies significantly with direction of pH change.
3. The model breaks down fundamentally for the important case of weak acids, giving much too slow a response.
4. The model says nothing about how to deal with fouling of electrodes which may slow down the measurement response substantially.

Given the greatly increased complexity of the model, the considerable increase in data required to formulate the model (compositions, diffusion coefficients, dissociation constants), and the fact that the structural and parametric mismatch to the actual response has not been demonstrated to have been reduced relative to the first order lag approximation the use of this model is not justified.

Johansson and Norberg (1968) focus on the double layer effect, showing that its effect can be validly approximated by a first order lag in the electrode potential in response to a step in activity of H^+. They give an expression for the time-constant of the double layer effect which indicates that in aqueous solutions below pH 14 it gives a negligible contribution to the probe dynamics. For pH 14 the observed response was noted to be characteristic of mixed diffusion and phase boundary effects. These results are consistent with Hershkovitch *et. al.* (1978) but do not particularly support the electrode double potential as being the cause of anomalous results with weak acids at low pH. Even more so than the stagnant layer mass transfer model this model does not provide a suitable model for the probe response, as it predicts minimal lag in the pH range of interest.

Electrode response is an interesting technical area, but for the reasons discussed above none of the models proposed from theoretical analysis offer a clear benefit over assuming a first order lag with parametric uncertainty (e.g. a lag between 5 and 30 seconds) which has been used more conventionally in pH system modelling. 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, though the appropriate lag coefficient is expected to vary with operating conditions.

pH probe dynamic response 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 $\pm.25pH$ will also be taken as a default and it will be assumed that the pH setpoint lies between 2–12 pH to avoid the potentially large errors outside this region (see section 4.1). 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). The values chosen should be backed by a maintenance commitment to maintain the specified properties.

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.

4.5 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 ($t_{d_{mix}}$) and the rest of the reactor is taken as perfectly mixed (uniform concentrations) contributing a first order lag ($t_{c_{mix}}$) (Shinskey, 1973). The delay and lag in series are taken as describing the reactor residence time distribution (RTD).

As discussed in section 2.3 the delay for this model can be estimated for stirred tanks using the correlation of Hoyle (1976)

$$t_{d_{mix}} = (0.9 - 1.8)V^{0.85}/(Fagit) \tag{4.37}$$

where the variable parameter depends on tank geometry.

The effect of imperfect mixing on both the time response and on reaction extent needs to be considered for this project. A computational fluid dynamics (CFD) package, MIXFLOW/MIXREACT developed by ICI and the Department of Mechanical Engineering at Imperial College (Bolour-Froushan, 1986; Nouri, 1988) was used to investigate the validity of the fractional tubularity model and provide a comparison with the above correlation. The key assumptions of the package with regard to this work appear to be the boundary conditions at the liquid surface and at the impeller. The surface is assumed flat, i.e. effects such as vorticing and air entrainment are neglected. The impeller is modelled through the use of slip factors between agitator velocity and fluid velocity defining the time-averaged boundary conditions. All flows are solved as time-averages while a k-ϵ model is used to obtain the kinetic energy and intensity of turbulence distributions. Some less fundamental restrictions occur due to

1. inlets being assumed to be flush to the vessel walls;
2. the absence of good slip factors for pitched blade impellers;
3. the more accurate solution methods not being particularly robust.

A typical initial response to a step, generated using the package, is shown in figure 4.9.

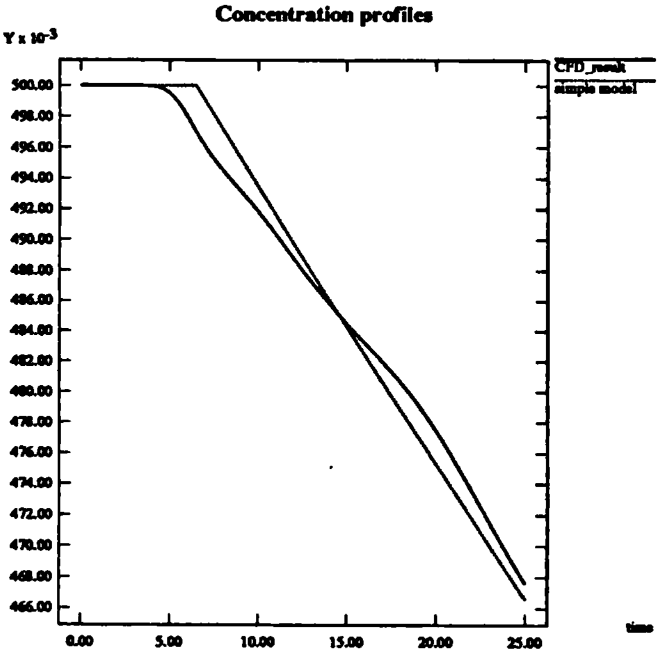


Figure 4.9: Predicted time response using a CFD model

This response is well matched by a simple delay-lag model, though with some fluctuation about the ideal response due to different flow streams reaching the exit at slightly different times. The delay observed is only about 60% of that predicted from Hoyle's correlation. This deviation may be due to the use of a side entry inlet below the liquid surface rather than a side entry inlet above the liquid surface as in Hoyle's experiments as the velocity of the inlet stream carries it rapidly across the top portion of the tank so that it is rapidly entrained into the recirculating flow due to the agitator. With the package used it was difficult to adjust the geometry to specify more typical geometries. CFD work was not pursued any further, as the limitations discussed above cast doubt on the usefulness of the particular package used for the problems of interest and there was insufficient time to pursue other packages.

The CFD results do support the validity of the basic delay and lag in series model, but it is not clear that CFD provides a better prediction of the delay than using Hoyle's correlation and standard correlations for the pumping rate (Uhl and Gray, 1986). Hoyle's correlation therefore remains the basis for estimating $t_{d_{mix}}$ in this project.

Further support for the basic delay-lag model comes from experimental RTD data. The data shown in figure 4.10 come from a three tank system. The data using a liquid tracer show an excellent fit to a model made up of a delay and three equal lags. The data using a solid tracer show significant deviation, probably due to incomplete suspension of the tracer. Due to uncertainty in the depth of liquid in connecting channels between the CSTRs it was not possible to validate the correlation for predicting mixing delay.

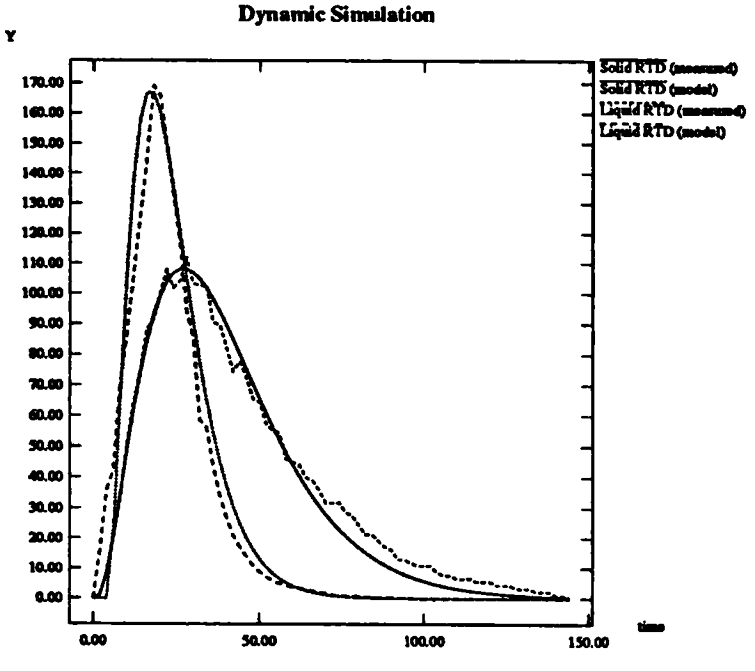


Figure 4.10: Predicted and measured residence time distributions

The discussion of mixers in section 2.3 noted that Hoyle's (1976) work could be interpreted as suggesting that there was a lower limit of about 9 seconds on the achievable delay in a stirred tank (while avoiding excessive splashing or air entrainment). This interpretation was noted to be potentially more appropriate than the more conventional interpretation that 3 minutes residence time is required for good mixing. This observation can be related to published mixing correlations. Combining correlations for a six-45°-pitched-blade turbine pumping rate (Uhl and Gray, 1986),

$$F_{agit} \approx 1.014 * (D_a/D_t)^{.21}(Z_t/D_t)^{-.17} N D^3 \tag{4.38}$$

maximum speed of agitation to avoid air entrainment (Greaves, 1981),

$$N_{max} \approx (.3 - .5)\frac{D_t^{\frac{2}{3}} H_t^{\frac{2}{3}}}{D_t^2}(1 - \frac{Z_t}{H_t})^{\frac{1}{3}} P_a^{-.13} \tag{4.39}$$

and delay 4.37,

$$t_{d_{mix}} \approx \frac{0.9V^{0.85}}{Fagit}$$

gives

$$t_{d_{min}} = \frac{(1.5 - 2.4)H_t^{0.183} D_t^{0.033}(Z_t/D_t)^{0.17} P_a^{0.13}}{(D_a/D_t)^{1.21}(1 - Z_t/D_t)^{.333}} \tag{4.40}$$

which for the default values below becomes

$$t_{d_{min}} \approx (5 - 8.5)D_t^{0.216} \tag{4.41}$$

All lengths are in metres and all times are in seconds. D_t is the tank diameter. D_a is the agitator diameter (default $D_t/3$). H_t is the liquid height (default D_t). Z_t is the agitator clearance from the base (default $H_t/3$), P_a is the pressure in atmospheres above the liquid (default 1). N is the agitator speed in revs/second. The expression obtained is consistent with air entrainment limiting achievable delay and with the limit being virtually independent of the tank volume ($V_t^{0.07}$). The higher values of minimum delay (lower values of N_{max}) correspond to solutions of concentrated electrolytes. This supports the use of a constraint on minimum dead-time of about 7 seconds. 10 seconds delay is a realistic design specification without detailed mixing analysis.

On the data available, I would not recommend going below 1 minute residence time without pilot plant mixing trials and a good reason for doing so.

The dynamic response of in-line mixers generally approximates a pure delay (Chemineer, 1988).

The effective mixing delay for jet-mixed tanks may be estimated very approximately from mixing correlations for t_{99} (time to 99% homogeneity) using $t_{d_{mix}} \approx t_{99}/10$. Once $t_{d_{mix}}$ is determined the mixing may be modelled in the same way as a stirred tank, using a delay-lag model.

To avoid the need for special procedures and modification of the integration algorithm (see appendix A), delays may be modelled 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 but 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 neutralisation reactions and reactions with first order kinetics this emphasis is correct but for many effluent treatment problems the reaction kinetics are complex combinations of second order reactions. The time response or residence time distribution (RTD) is then no longer sufficient to characterise the effect of mixing on the reaction. 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 to an ideal CSTR.

In the case of RTDs represented well 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 pre-mixed reactants. If the reactants are not premixed then 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 which gives a general-purpose conservative model. This model is used in the case-studies on solid reagents (chapter 6).

When trying to capture the interaction of mixing and reaction kinetics plug-flow elements cannot be approximated using general rational transfer functions or history arrays. A series lag approximation can be applied and is used in all the delay models in the work described in chapter 6.

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 should be 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 which 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 minimised by ensuring a good margin from the conditions for particles being just suspended. In some cases it may be desirable to exploit this effect to increase the average age of larger particles on leaving the reactor system. This requires care as it increases the risk of incomplete suspension and accumulation of unsuspended particles within the reactor. The effect for good or ill has not been modelled and is difficult to evaluate.

4.6 Precipitation reactions

While precipitation reactions are in principle amenable to mathematical analysis this analysis is often very difficult in practice. For steady-state properties, it might be hoped that solubility products would provide a useful basis for analysis. For all but the simplest cases, this would be overoptimistic. Heavy metals, for example, have very low solubility at pH values around neutral but in isolation remain in solution as electrically charged colloids except in a fairly narrow pH range about their characteristic "isoelectric point" for which the colloids are uncharged and coagulate reasonably readily. In the presence of other components however, they are often observed to precipitate more rapidly. This arises through a mixture of effects including surface adsorption onto other suspended solids and "coprecipitation" (if one component precipitates in large amounts it tends to take other components with it). Behaviour is further complicated by very slow approach to equilibrium by some metals, e.g. aluminium. Agglomeration of colloids can often be improved markedly by the addition of flocculants. To date the appropriate choice of flocculant and other operating conditions are determined experimentally using small scale "jar" tests (Cushnie, 1984).

For heavy metals, typical precipitation conditions would require pH adjustment with $Ca(OH)_2$ to close to 8.5pH, possibly with the addition of a flocculant. In this and similar cases precipitation becomes essentially a problem of experimentally determining the best operating conditions and then controlling pH to the required precision, typically ±.5pH around the optimal pH. The optimal pH on a full-scale plant may diverge from that in jar tests, so it is desirable to ensure some flexibility in the operating pH.

Precipitation will therefore be treated as a pH control problem with experimentally determined pH targets rather than as a distinct design problem.

4.7 Previous pH modelling environments

By way of comparison, I looked at two modelling packages for pH control system analysis. Both are limited to fast neutralisation reactions, i.e. no model for solid alkalis is included.

A modelling package produced by Jacobs and Badran (1987) provides a model of a CSTR neutraliser, which models imperfect mixing by a fixed series delay and characterises the pH response using a single titration curve tabulated against net concentration, *cnet*. Probe response is neglected. Noise can be injected at various points in the process as well as on the pH measurement. The package was developed primarily for testing modern control algorithms and provides a wide range of options in this respect. Its restriction to a single CSTR and a single titration curve, severely limits its applicability to system design but for single loop controller algorithm development it may well be useful.

Southwood-Jones (1990) provides an example of a recent attempt to develop a pH modelling environment. I differ with two of the choices made in this environment.

1. The model uses direct solution of concentration equilibria equations in preference to titration characteristics on the basis that the requirement for titration curves "restricts the designer if he [sic] does not have these at hand". The designer is much less likely to have a detailed composition analysis and the necessary physical properties at hand and as discussed above (4.1) titration curves can be used in a manner consistent with empirically fitted concentration equilibria models.

2. McAvoy's (1979) earlier model of electrode response is used in preference to the traditional first order lag approximation. McAvoy accepts that this model is inferior to that presented in Hershkovitch (1978) which does not itself appear suitable for the reasons discussed above.

In other respects, e.g. mixing, the model is consistent with that discussed above.

4.8 Summary

Models for the steady-state relationship between pH and reagent concentration have been explored. Both thermodynamic models and titration curve models are discussed. The thermodynamic models are shown to be useful in obtaining insights into complex behaviour, but to be difficult to apply as predictive tools. Models based on titration curves are discussed and methods of extrapolating titration curves to predict the effect of changes in the effluent are presented. These methods are shown to be approximately valid, though the need for experimental checking is noted. In general, titration curve models are more convenient and more readily determined and are used in the case study work described in chapter 6.

Models for reaction of the solid alkalis $Ca(OH)_2$ and $CaCO_3$ are developed using general mass-transfer analysis principles. The models have been validated against experimental data and results in the literature. Guidelines for experimental tuning of the model are given.

Measurement characteristics are reviewed and a default model including an uncertainty description has been defined.

Mixing properties are 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.

Cyanide oxidation is presented as an example of the modelling of redox reactions. It is shown that results in the literature allow a sound model to be developed but leave significant uncertainty in the reagent kinetics which must be considered in the design.

The limitations of modelling precipitation reactions are discussed, with the conclusion that, for the application of the design techniques in this project, the problem reduces to a pH control problem with a target pH determined by experimental investigation of the precipitation.

Previous pH modelling environments are examined in relation to the models developed for this work.

Chapter 5

Design procedures

This chapter presents procedures for tackling end-of-pipe chemical waste-water treatment designs. The guidelines presented explain how the tools discussed in chapter 3 and the models discussed in chapter 4 can be applied, in a well-coordinated manner, to the integrated design of waste-water treatment systems.

Design methods for the selection of structural or discrete design variables are reviewed to set the background for the specific procedures presented (5.1). Section 5.2 develops some useful results and insights using a generic neutralisation problem. The key procedures for design of systems for neutralisation and precipitation are then presented in section 5.3, building on the results of the generic analysis. Section 5.4 discusses the extension of this procedure to redox systems. It should be noted that the procedures presented were strongly influenced by the case study work described in chapter 6 but are presented prior to these examples for clarity.

5.1 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 presented have focussed on techniques for analysing performance of particular designs and for optimising continuous design variables.

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

1. The MINLP approach requires all the objectives and design options to be fully defined before solution can commence. In typical waste-water treatment 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 due 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 optimisation problem.

2. MINLP algorithms typically capture the results of NLP solutions for particular sets of discrete variables using some form of linearisation at the NLP solution. A design engineer can potentially extract much more information than this and bring judgement 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 then 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 takes hours to solve, as is often the case in dynamic optimisation or worst-case design, then 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 where good rules exist for modifying designs and 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 flowsheet modification strategy for modifying flowsheets to synthesise a flotation cell separation circuit and show that it gives reasonable results. Mizsey and Fonyo look at the use of heuristics in design more generally, and coin 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 this project. A range of tools are available which can be used for generating bounds on performance:

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

5.2 Exploration of a generic neutralisation control problem

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

5.2.1 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.

Disturbance conditions Disturbances involve both flow and concentration changes. Concentration changes are the most challenging as it is always possible to minimise the direct effect of flow by use of flow measurements, e.g. 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 which 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.

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. No assumption was made on relative sizing or the value $t_{d_{mix}}/t_{c_{mix}}$ for the CSTR as the review of mixing in section 2.3.1.3 indicated that the existing guidelines lacked clear justification. Most of the examples presented below use equal sized-tanks with $t_{d_{mix}} \approx .05V_T/F_T$ 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, though the associated lags have a similar effect on control performance 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 (5.2.2). 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.

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 (equation 4.5) 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 therefore 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 (0.0N 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 only need to be optimised 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 the concentration variations one hour after the disturbance to be less than 10% of the concentration variation which would cause violation of the 5–9 pH bounds. In terms of pH this corresponds to bounds of 6–8pH.

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 figure 5.1.

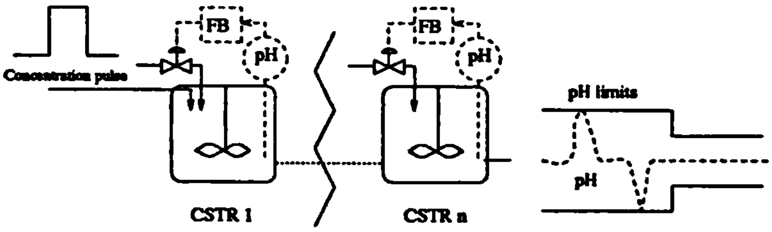


Figure 5.1: Generic pH control problem

5.2.2 Why pH control is difficult

A useful starting point is to examine the characteristics of the pH control problem which 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 3.4.2. Uncertainty imposes additional limitations. For example, if the measurement lag was accurately known it could be cancelled and would not limit control performance. As the measurement response is more complex than this and variable over time, it cannot readily be cancelled, and combines with the delay to limit controller performance (see sections 5.2.3 and 5.2.5). Limits on the reagent addition rate prevent the ideal delay-limited control bound being achieved in some cases even in the absence of uncertainty as infinite actuator range is required (see section 5.2.5). However, these limits are not observed to have a major effect on PI control performance as the degree of transient overshoot compared to the steady-state control output change required is moderate. Non-minimum phase characteristics other than delays are not observed in the generic system defined above.

Three effects which have a·less clear-cut impact on achievable performance are the pH nonlinearity, if uncancelled, the precision limits on the reagent addition, and measurement noise. These are investigated below, in the context of the generic problem of section 5.2.1, to complete the picture of why pH control is difficult.

Effect of nonlinearity Using a tank with 3 minutes residence time (V_T/F), $t_{d_{mix}} =$ 9 seconds and $t_{c_{mix}} = 171$ seconds, the achievable disturbance rejection was found by optimising a PI controller based on pH and a PI controller based on concentration. The tank parameters were chosen to match Hoyle's (1976) recommendations, as these are widely used. Secondary lags were neglected.

Delay-limited feedback analysis using equation 3.51

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

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 optimisation 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 minimise 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 non-linearity to be sufficiently compensated as to be negligible.

Effect of precision error in reagent addition McMillan (1984) notes that reagent control valve precision error is dominated by deadband error, where

> Deadband 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 $t_{d_{mix}} = 9$ seconds and $t_{c_{mix}} = 171$ seconds 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 behaviour. 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 one 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 behaviour, but sticking and lurching is also considered as it is not ruled out by the definition and is likely to give worse performance.

With an integral action time, *iat*, of 200 seconds, the system was found to have a critical gain of about 28.3 in the absence of deadband error. Holding *iat* constant, the gain was varied and the responses observed.

With the "stick and glide" model large limit cycles were observed when the gain exceeded 95% of the critical value, but degradation of performance below this gain was minimal for a single tank. The demanded valve position varied between about ± 1 for lower values of gain, while the actual valve position varied only slightly, but with the error being sustained for a considerable time. This behaviour would only cause difficulties if the load precision, δ_l (fractional change in reagent addition causing violation of pH limits), was very small and control was only applied to the first tank so that the small slow fluctuations caused by the deadband error were allowed to propagate unattenuated through the treatment system.

With the "stick and lurch" model the valve position was observed to cycle, as a square wave, over a wide range of gains with a peak-to-peak valve position amplitude of approximately 1. As the gain was reduced the period of oscillation increased relative to the natural period of the loop (about 40 seconds) with a corresponding increase in the output deviations. As the gain approached the critical gain the amplitude of the limit cycles increased with a corresponding increase in output deviations. A broad minimum in the peak to peak output deviation of about .1 was observed for gains in the range 15 to 25 with periods of oscillation between 40 and 80 seconds. With two identical tanks in series the output peak to peak error was about .007. With 3 tanks, the output error was about .0006. These observations indicate that the peak to peak deviation of the exit concentration can be bounded for fairly tightly tuned controllers by calculating the effect of the first harmonic of the cycling valve position on the system output. This effect can be calculated for a unit deadband error as

$$\frac{4\Delta_{cnet}}{\pi} \prod_{i=1,n} \frac{1}{(1+(\pi t_{c_{mix}}(i)/t_u)^2)^{1/2}} \tag{5.1}$$

where Δ_{cnet} is the inlet concentration change resulting from a unit change in valve position (1 in the examples above). This error would not be the limiting factor on transient performance unless the required control precision, δ_c, is much less than the load precision, δ_l, i.e. disturbances much smaller than total load, or the deadband error was much higher than the value of 1% of range achievable using a valve positioner. If the controller is detuned due to uncertainty in the process gain, then the period of oscillation increases by a similar factor to the gain reduction, making performance worse. The predicted degradation of performance increases with the number of downstream tanks, but downstream controllers should be able to improve the attenuation of the slower disturbances so the effect is still unlikely to become critical.

Equation 5.1 was confirmed by further simulations to be valid for a wide range of mixing conditions.

These observations contradict McMillan's (1984) assertion that the deadband error must be less than the allowable *steady-state* reagent error.

Effect of high frequency measurement noise In a well-designed pH measurement system, a moderate amount of high-frequency noise is usually present (about ± .1pH). This noise has not to my knowledge been analysed in the detail required to carry out a rigorous evaluation of its likely effect. In my 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 5.2.5).

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*.

5.2.3 Examination of bounds from delay-limited control analysis

The performance bound from delay-limited control analysis (section 3.4.2.7) may be computed for a number of series CSTR systems and compared to the performance obtained by optimising PI controllers. For simplicity and generality, this section assumes the pH nonlinearity to be cancelled at the controllers.

The CSTR response after the delay between disturbance effect and control action, t'_d, has elapsed may be modelled using equation 3.51 for equal-sized tanks or equation 3.52 for tanks of different sizes. The maximum disturbance amplitude is given by $10^{-5}/\delta_a$ N.

The results for some representative cases are tabulated below showing the discrepancy between the theoretical bound and the optimised PI controller performance with

n controlled tanks. A performance estimate based on observation of the results

$$\delta_a = \frac{3t_d'^n}{2\prod_{i=1,n} t_{c_{mix}}(i)} \tag{5.2}$$

is also tabulated to allow ready comparison. The CSTRs are defined in terms of the ratio of the tank residence times (V_T/F_T) and mixing delays ($t_{d_{mix}}$) to the minimum value of $t_{d_{mix}}$ (t_d'). It should be noted that the order of the CSTRs is irrelevant as the relationship between inlet and exit concentration for each CSTR may be represented by a linear transfer function so that the concentration variation at the treatment system exit is not affected by the order in which the CSTRs are placed. A "c" beside the entry for a CSTR indicates that the CSTR was controlled. 1,20(c);1,80 indicates a 2 CSTR system with only the first tank controlled, equal mixing delays in both tanks with the residence time of the first tank 20 times the minimum mixing delay and the residence time of the second tank equal to 80 times the minimum mixing delay. The disturbance levels are given in mN (N $\times 10^{-3}$).

Comparison of ideal control and optimised PI performance

case no.	configuration	ideal bound	PI feedback	estimated performance	PI/ estimate
1	1,20(c)	.195	.16	.13	1.23
2	1,20(c);1,20(c)	7.5	2.40	2.41	1.
3	1,20(c);1,20	7.5	1.60	2.41	.66
4	1,20(c);1,80(c)	29.5	11.1	10.	1.11
5	1,20(c);4,80(c)	29.5	6.91	10.	.69
6	1,7.9(c);1,7.9(c);1,7.9(c)	22.	2.9	2.2	1.31
7	1,20(c);1,20(c);1,20(c)	430	45	45.7	.98
8	1,100(c);1,100(c);1,100(c)	58900.	5620.	6470.	.87
9	As 8 except 5:1 gain variation on one tank	58900.	4107	6470.	.63

A number of conclusions can be drawn from these results.

1. For cases where all tanks are controlled and have the same minimum delay the estimated control performance has a good correspondence with the optimised PI performance (1,2,4,6,7,8)

2. If a single tank in a multiple tank system is uncontrolled (3) or has a delay much larger than the minimum delay (5) or has substantial uncertainty in process gain (9), then the disturbance rejection is reduced by about a third.

3. As the ratio of delay to residence time increases above about 1:20 (6) the estimate becomes conservative. For small ratios of delay to lag (8) the estimate is slightly optimistic.

4. The ratio between disturbance rejection with optimised PI control and ideal delay-limited control is about 1:1.5n!.

For preliminary design purposes the estimate from equation 5.2 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 rejection should be reduced by a third. The 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 3.4.2 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 four. The effective delay ($t_u/4$) was calculated for a typical tank with 3 minutes residence time, $t_{d_{mix}}$ = 9 seconds and 30 seconds measurement lag to be 24.7 seconds. Three identical tanks in series were considered. This gave a heuristic performance bound of 22.3×10^{-3}N (equation 3.51) and an estimate of the optimal PI performance of 2.2×10^{-3}N (equation 5.2). The performance of this system with optimised PI feedback was 3.4×10^{-3}N. The performance of a three tank system with a pure delay equal to this effective delay in place of the combined delay and measurement lag and the same $t_{c_{mix}}$ was 2.9×10^{-3}N (case 6). This example indicates the use of $t_u/4$ as an effective delay estimate to be appropriate, but slightly conservative.

Summary Estimating t'_d based on physical delays and using equations 3.51 and 3.52 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 which does not cancel the measurement dynamics and is therefore subject to the same bandwidth limitations). Estimating t'_d as $t_u/4$ and using equation 5.2 provides a realistic, though 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.

5.2.4 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. As discussed in section 2.3 there is a strong consensus in the literature that equally sized CSTRs are undesirable for control purposes. However, equally sized CSTRs are the industrial norm. Using the delay-limited control performance bound (equation 3.52)

$$\delta_a = 1 - \sum_{i=1}^{n} \frac{e^{-t'_d/t_{c_{mix}}(i)} t^{n-1}_{c_{mix}}(i)}{\prod_{j \neq i}(t_{c_{mix}}(i) - t_{c_{mix}}(j))} \approx \frac{1}{n!}(\frac{t'^n_d}{\prod_{i=1,n} t_{c_{mix}}(i)})$$

and a typical cost function of the form $k1 + k2V_T^x$ it is possible to evaluate the desirability of using tanks of different sizes. The control precision requirements can be approximated well by

$$\delta_c \geq \frac{1}{n!}(\frac{(F_T t'_d)^n}{\prod_{i=1,n} V_T(i)}) \qquad (5.3)$$

where F_T is the total flow through the system and $t'_d \ll t_{c_{mix}}$. The magnitude of the disturbance which 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 control precision, δ_c, at a flow F_T therefore takes the form

$$\min_{n, V_T(i), i=1,n} \{k1.n + k2 \sum_{i=1,n} V_T(i)^x\} \; s.t. \prod_{i=1,n} V_T(i) = (1/n!)(F_T t'_d)^n/\delta_c \qquad (5.4)$$

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

$$\min_{V_T(i), i=1,n} \{\sum_{i=1,n} V_T(i)^x\} \; s.t. \prod_{i=1,n} V_T(i) = (1/n!)(F_T t'_d)^n/\delta_c \qquad (5.5)$$

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

$$\min_n n(k1 + k2V^x_{T_{opt}}) \qquad (5.6)$$

with the optimal volume, $V_{T_{opt}}$, given by $F_T t'_d (n!\delta_c)^{-1/n}$. A similar analysis can be carried out using the PI performance estimate equation 5.2, 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 $V_{T_{opt}}$ 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 $(1.5n!)^{1/n}$ i.e., 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 control precision of .001 with a flow rate of $.03\mathrm{m}^3/\mathrm{sec}$ and $t_d' = 10$ seconds, using PI control. Using the cost function $20,000 + 2,000V_T^{0.7}$ (section 2.3.1.2) and equation 5.2 gives a minimum cost CSTR system with 2 tanks of volume $11.6\mathrm{m}^3$ 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 $13.6m^3$. It should be noted that the optimal volume can be expected to lie below $(k1/k2)^{1/x}$ ($27\mathrm{m}^3$ in this example) as the optimisation 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 $10\text{–}30\mathrm{m}^3$ if the cost expression used is appropriate.

For typical values of F_T between $30\text{-}300\mathrm{m}^3/\mathrm{hr}$, the predicted optimal volumes of $10\text{–}30\mathrm{m}^3$ correspond to residence times of between 2 and 60 minutes. It should be economically attractive to achieve mixing near the achievable bound (section 4.5) with tanks of this size so that $t_{d_{mix}} \approx 10$ seconds 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 seconds. This gives an effective $t_d' \approx 25\text{–}30$ seconds and a rigorous $t_d' \approx 10$ seconds, with both values independent of tank volume. This shows that the two main assumptions of the above analysis ($t_{d_{mix}} \ll t_{c_{mix}}$ 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 flow rates were much above $300\mathrm{m}^3/\mathrm{hr}$ or the CSTR cost function was quite different these assumptions might not hold.

The above analysis does not consider the economic benefits from choosing equal sized tanks arising from 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 control precision. 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 $30\mathrm{m}^3$ in size with residence times between 2 minutes and one hour. This is consistent with industrial practice and contradicts the recommendation in the literature that tank sizes should be split in a ratio of about 1:4 or greater.

5.2.5 Towards ideal control

In the light of the analysis in the rest of this section it is of interest to review what the cost of using PI control is in comparison to achieving the ideal delay-limited control performance bound and what measures can be taken to reduce this gap.

Section 5.2.4 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, $t_{d_{mix}}$ of 10 seconds and a probe lag of up to 30 seconds and a CSTR with $t_{c_{mix}} \gg t_{d_{mix}}$ is 28.7 seconds. The measurement lag may therefore imply almost a three-fold 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 characterise 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 which turns a reagent excess to 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 as the initial process gain to a pulse input is zero. Finite reagent delivery capacity therefore prevents the ideal control bound 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 constraints are symmetrical about the initial value. This means 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 which would allow closer approach to the ideal bound is reducing or cancelling the pH measurement lag. Cancellation 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 4.4 the probe response is very complex (though the effect on the control performance can be approximated by an uncertain first order lag).

A greatly improved understanding of probe behaviour would be needed to allow effective cancellation 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 and in particular decreases 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) so 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 minimising fouling (under-neutralisation rather than over-neutralisation) 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. 30 seconds has been used in most of the case studies presented in this chapter 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 section 6.2. Improved understanding of measurement response would be helpful in choosing the appropriate value.

If it is desired to evaluate advanced control schemes then the models used in this work should be regarded as generating an upper bound on practically achievable performance and 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

1. The greatest scope for moving performance towards the ideal control bound lies in reducing measurement lag and improving understanding of measurement response.
2. Constraints on reagent addition rate prevent ideal control being achieved for more than one tank with control on the first tank only.
3. Model errors and measurement noise obstruct the achievement of ideal control.

5.2.6 Summary

The exploration of the generic problem has allowed a number of issues to be clarified. The key conclusions are summarised 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 equation 5.2.
3. The optimal series CSTR configuration to minimise cost for a given control precision requirement and instantly reacting reagent is made up of equal-sized tanks for typical treatment systems.
4. The greatest potential for moving control performance towards the ideal delay-limited control bound lies in minimising or compensating for pH measurement lags.

5.3 Neutralisation of waste water

5.3.1 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 which allow higher level options to be considered without costly detailed design analysis or experimental work.

A complete characterisation of a neutralisation problem for design purposes might require

1. a set of titration curves 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 which may be used for neutralisation, 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 which 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 which 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 and they may be highly variable over time, so 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 chapter 4. Advice on experimental procedures is also given. Approximate costs for most options are given in section 2.3. 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 which 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.

The above 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 then the methods for worst-case design may be used to identify this condition given a model and uncertainty description.

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

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 maximising 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.

5.3.2 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.

5.3.2.1 Reagent choice

The review of current practice indicated that a good set of rules existed for guiding reagent choice. These were summarised in section 2.3.1.2 and will not be repeated here. The work described in this thesis has highlighted an additional rule for determining feasible reagents: sodium-based reagents should not be used for overneutralisation with high concentrations of acid and elevated temperatures at the treatment system exit, as the pH probe error due to sodium ion interference becomes excessive (see section 4.1.3). It has also highlighted two additional design considerations.

1. Impurities associated with lime-based reagents may provide a useful degree of buffering when neutralising high concentration acids (see section 4.1.3).

2. $Mg(OH)_2$ is likely to react at less than one tenth the rate of $Ca(OH)_2$ near neutral pH (see section 4.2.4).

The recommended procedure for using the rules and guidelines presented to select reagent is

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;
4. the total acid load.

5.3.2.2 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 3.4.2). 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 control precision δ_c (section 3.4.2.7). The acceptable δ_{cnet} can be computed from the titration curve 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 control precision is then given by $\delta_{cnet}/\Delta_{cnet}$.

The disturbance attenuation capability, δ_a, of a particular design can be calculated using equations 3.51 or 5.2 with appropriate choice of t'_d. The appropriate expression will depend on whether a rigorous bound, a heuristic bound or a PI performance estimate is of interest (see section 5.2.3). This attenuation can then be compared to the required control precision. 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 then 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 a given control precision by equating δ_a and δ_c and assuming all tanks are the same size (see section 5.2.4).

This analysis is a good starting point for preliminary design as it can give performance bounds and estimates for all neutralisation 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}. Considering the output variation due to the first harmonic of the square-wave input and assuming reagent flow is negligible compared to effluent flow gives the necessary condition for successful on-off control below

$$\delta_{cnet} \geq \frac{4\Delta_{cnet}}{\pi} \prod_{i=1,n} \frac{1}{(1 + (2\pi t_{c_{mix}}(i)/t_u)^2)^{1/2}} \tag{5.7}$$

If the worst-case disturbance is not evident then the ideal delay-limited control analysis is better approached using the worst-case design tools on a dynamic model as discussed in section 3.4.2. This optimisation can be solved in minutes (on a SPARC 2) for typical problems.

5.3.2.3 Solid reagent conversion

Given a design which 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 then 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 one for these reagents. Adding a plug-flow reactor stage early in the mixing scheme is likely to be beneficial and moving to an increased number of back-mixed 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 optimisation;

2. use the default models from section 4.2;

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 load precision requirements δ_l (permissible fractional carryover of reagent) computed. It should be noted that the worst-case load condition may be quite different to the worst-case disturbance condition for dynamic analysis. The worst-case load characteristics will normally be bounded by the maximum reagent concentration at the maximum flow and the maximum pH sensitivity and may be identified by worst-case design if desired.

The apparent reaction time constant, τ_r, may be estimated from a batch neutralisation 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 then 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. τ_r estimated in this way may be used as a cautious estimate of an effective first order kinetic time constant and the residual reagent, δ_{res} estimated as

$$\delta_{res} = \frac{e^{-\tau_{PFR}/\tau_r}}{\prod_{i=1,n}(1 + t_{c_{mix}}(i)/\tau_r)} \tag{5.8}$$

where τ_{PFR} is the total plug flow reaction time available. δ_{res} should be less than δ_l to ensure 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 then the reagent conversion may be analysed 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 analysing the conversions of reagent added to each tank in turn may be useful to provide insight into the design problem (see section 6.3 for an example).

5.3.2.4 Summary

At the conclusion of preliminary design analysis, there should be a clear picture of what the worst-case conditions are likely to be and a candidate design should have been obtained which 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.

5.3.3 Complete design

Complete design utilises 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 optimisation. If steady-state issues have not been fully explored in the preliminary design stage then 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 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.

Measurement noise, in a well-engineered system, is unlikely to affect the performance of a PI controller significantly (section 5.2.2), but may limit the use of derivative control action and will certainly limit the performance of advanced control schemes which attempt to approach ideal control (section 5.2.5). My approach to PID control has been to optimise 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 seconds) 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 but may be considered after the other design issues have been settled, as it is generally amenable to satisfactory solution by detailed design, without making a substantial difference to the overall design or to total cost, as discussed in section 2.3. In initial analysis, idealised valves can be considered which are deliberately oversized 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 used in the design obtained assuming ideal valves. If this sizing indicates extreme rangeability requirements then specialised equipment may be needed. If the rangeability requirement is moderately high, say 100:1, then 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 then use of standard equipment is acceptable. Equal-percentage valves, characterised 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, though this is likely to slow down the integration dramatically. Section 5.2.2 presents some evidence that reagent valve errors are not usually critical and an expression for estimating the exit deviations caused by dead-band errors (equation 5.1).

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 optimising the design then 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 necessary. For example, steady pH values outside the range 2 to 12 pH are associated with increased measurement bias and pH values above 7 encourage fouling of electrodes with carbonate. pH operating points between 2 and 7 are therefore to be preferred for good reliability.

Compliance to 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 such as pH limits must be satisfied x% of the time. One method of handling "x%" constraints would be by excluding worst-case combinations which 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.

5.3.4 Rules for modifying design

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

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

If steady-state reagent conversion of solid alkali is the limiting constraint then consider the options below.

1. Use a plug-flow reactor or a small CSTR as the first stage using under-neutralisation 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 6.4).
2. Provide acid reagent to downstream stages to allow compensation for carryover of reagent at intermediate stages in the treatment system, if 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 input conditioning in the initial design if there is a well-defined and significant nonlinearity.

If titration curves are highly variable outside the target pH range then consider the use of adaptive control or on-off control on the initial controlled stages. The conditions for adaptive control to be effective are discussed in section 2.4. On-off control may be effective if the detuning due to titration curve variability is severe (say 5:1 process gain variability or above) and the maximum disturbance is comparable to the maximum 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 then 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 specialised reagent addition equipment. It is not likely that such effects will require an increase in back-mixed 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 5.2.2).

5.4 Redox reactions in waste water

Much of the design methodology discussed for neutralisation transfers to redox reactions. The models required for the detailed analysis are quite different to neutralisation systems and each redox pair, e.g. CN^- and ClO^-, requires its own reaction scheme and kinetics. Redox systems often require pH control and pH and redox control will interact through both probe characteristics and reaction kinetics.

For redox reactions a redox curve replaces the titration curve of neutralisation reactions. Electrode dynamics show substantial dependence on the particular reactions taking place and on the electrode type. Differences in time response to added reductant and added oxidant may be observed. Probe response therefore needs to be characterised for the particular electrode and redox pair of interest.

Redox kinetics are typically more complex than the kinetics of neutralisation reactions, involving multiple competing reactions and intermediate products.

Redox treatment system design will often be dominated by steady-state considerations due to slow reaction kinetics. Dynamic effects may have a substantial impact on the optimal design in some cases.

Preliminary design may be based on a combination of steady-state analysis and ideal delay-limited control analysis to estimate dynamic effects.

The complete design uses the same tools as for pH control with the substitution of an appropriate model for the reaction kinetics and the measurement response.

Chapter 6

Design examples

The design examples presented in this chapter cover a wide range of neutralisation problems and a cyanide oxidation problem. The examples studied have confirmed the power of the design methods developed. The neutralisation examples are all based on industrial case studies and include

1. preliminary design of a central effluent plant(6.1);
2. pH control of a strong-acid/ strong-base system (6.2);
3. neutralisation and precipitation with $Ca(OH)_2$ in a central effluent plant with highly variable effluent characteristics (6.3);
4. neutralisation of a well-defined, highly-concentrated acid stream with $Ca(OH)_2$ (6.4);
5. neutralisation of a moderately variable effluent with $CaCO_3$ (6.5).

Section 6.6 presents some work on cyanide oxidation.

6.1 Preliminary design of a central effluent plant

This example illustrates the use of preliminary design tools in exploring design options efficiently. The design problem is represented schematically in figure 6.1.

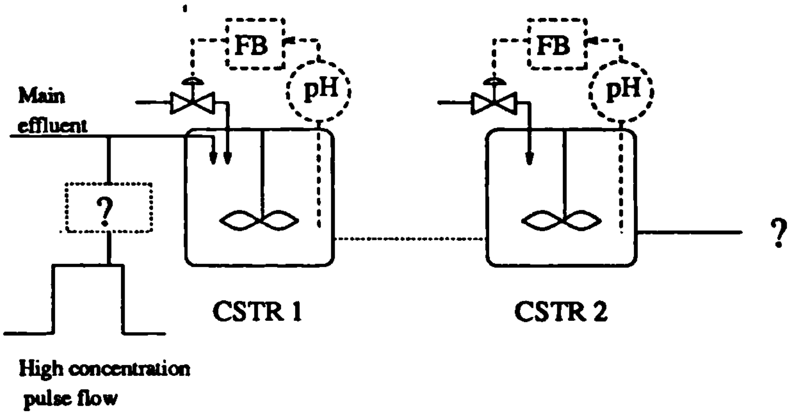


Figure 6.1: Design problem for preliminary design

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 to 9.5pH 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 5.3.2.2) was therefore used as the sole design tool. As two CSTRs with 10 minutes residence time at normal flow already exist, the analysis was carried out based on n CSTRs of this size. Applying Hoyle's correlation (equation 4.37) 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 3.4.2.7).

Inlet effluent concentration was normally about $\pm$.01N, but there was a regular pulse flow discharge of 30 minutes duration, at comparable flowrate to the main effluent streams and at between -1 and -3.5N 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 .0005N. This compares to the strong-acid/ strong-base concentration change of about .00003N, showing that significant buffering is present even in the worst case. The slope of the titration curve ($\frac{\partial pH}{\delta cnet}$) was observed to vary by more than 10:1 between experimental samples. The inlet concentration change, Δ_{cnet}, in the worst-case disturbance is 1.75N. The rising edge of the pulse is much more severe, because of the higher associated flow, so that the setpoint can be biased towards the upper consent level giving an allowable concentration deviation, δ_{cnet}, of approximately .0005N. The required control precision, δ_c is therefore about .0003 (.0005/1.75). The attenuation available from n tanks of the same design as those already available can be examined using the bounds on the performance computed using the pure delay and the effective delay (equation 3.51) and using the heuristic estimate of attenuation with PI control (equation 5.2). The results of this analysis are tabulated below.

Disturbance attenuation values for *n* CSTRs

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 at least 3 tanks would be required and the heuristic estimate indicates that a 4 tank system would be required when using PI control. Even a four tank system can be expected to find meeting the performance requirements difficult considering the variation in titration curve slope, and hence process gain, noted.

If local containment is applied to the worst-case disturbance then the new worst-case disturbance can be assumed to be bounded by start-up to maximum flow giving $\Delta_{cnet} = .01N$. This disturbance may be in either the acid or alkali direction so δ_{cnet} is reduced to .00025N. This corresponds to a required control precision of .025. The residence time has been doubled and the achievable precision decreases by a factor of 2^n compared to the table above. The rigorous bound suggests 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 5.3.4 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 .02N (to allow effluent concentration variations to be accommodated). The residence time is 10 minutes, and the natural period is about 2 minutes. Applying equation 5.7 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

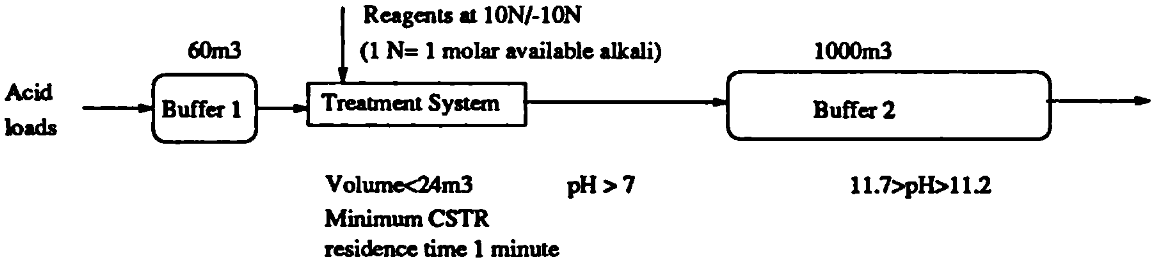
1. 2 CSTRs with PI or on-off control and local containment of the worst-case disturbance or
2. 4 CSTRs with careful controller design

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 5.3.2.2 quite closely, with some judgement being exercised 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 control precision 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 which could be used for worst-case design.

6.2 pH control of several strong acid streams with NaOH/HCl

6.2.1 Problem definition



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

Figure 6.2: Problem definition

The process requirements and constraints are summarised in figure 6.2. 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 modelled by a mixing delay, $t_{d_{mix}}$, of 7 seconds in series with a mixing lag, $t_{c_{mix}}$. The 7 seconds delay was based on avoiding both air entrainment and flow shortcircuiting 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 are unusually large compared to the mixing lags which are less than 2 minutes at maximum flow (section 4.5).

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.

6.2.2 Design

6.2.2.1 Preliminary design

Due to the virtually instantaneous neutralisation reaction any non-trivial screening test for this problem must be based on dynamics (section 5.3.2.2). The fraction of the worst-case disturbance controllable with ideal delay-limited feedback was calculated to provide a controllability measure (section 3.4.2.6). Disturbance concentrations were contracted towards 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 analysed 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, due to the delay in measurement response. Delay between reagent addition and the pH sampling point was assumed to be negligible, with the required micro-mixing to attenuate noise (section 2.3.1.3) being provided by turbulent diffusion in the sampling line. Optimisation 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 optimised 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 11.5pH due to the measurement bias. If no back-mixing is provided in the treatment system then 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 minimum 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 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 5.3.4). 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, two 12m^3 tanks with ratio control of reagent flow at the inlet to the treatment system becomes the starting point for the search for an effective plant.

The results of the disturbance fraction analysis are summarised in figure 6.3.

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.

Identifying the worst-case condition was less clear-cut in this example than in that presented in section 6.1. The worst-case disturbance combination varied with the process design and disturbance type considered. Using the optimisation formulation from section 3.4.2.6 to identify the worst-case was therefore very useful. The preliminary con-

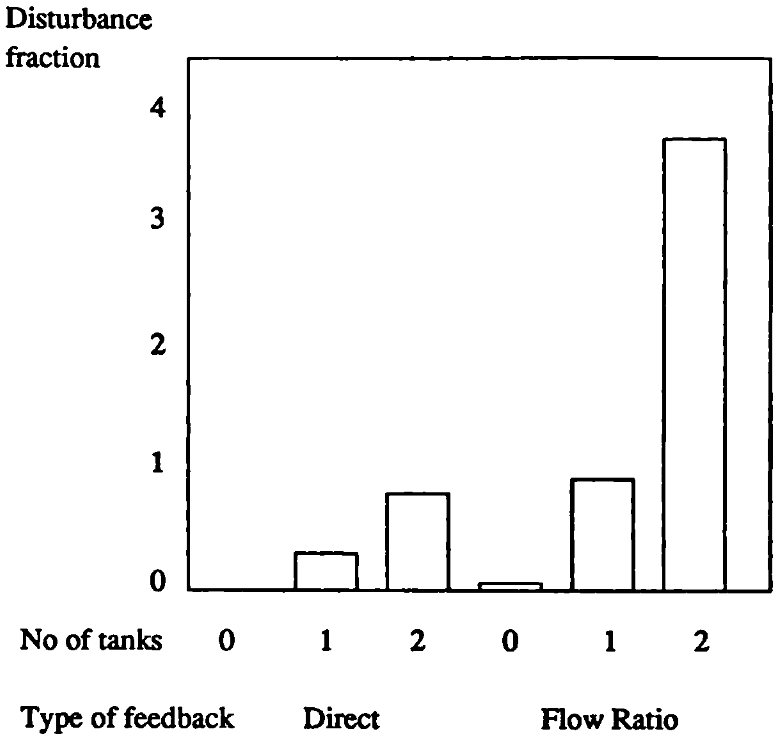


Figure 6.3: Relative performance of alternative process/control configurations

trol 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 usual norm of at least 3 minutes) and indicating the need for ratio control of reagent addition.

6.2.2.2 Complete design

PI controllers were used to implement feedback control. All controllers in the analysis below attempt cancellation 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, the optimisation objective becomes to minimise 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 figure 6.2. The disturbances were characterised 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

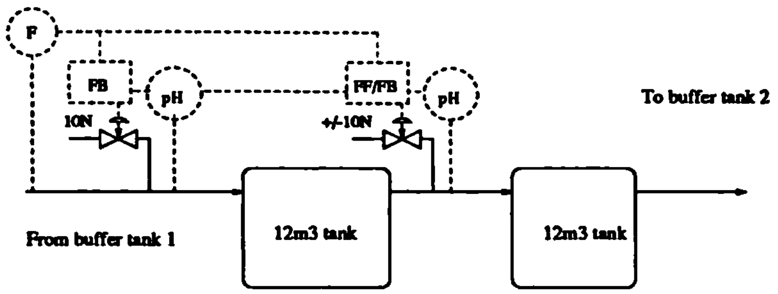


Figure 6.4: Final process and control scheme

between its minimum and maximum value. This gave a total of 11 uncertain parameters. In addition to the constraints specified in figure 6.2 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 ± 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 (figure 6.4) 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.

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 $m^3/(m^3N)$.

The outer approximation algorithm (section 3.3) took 6 iterations to identify this solution, with a projection factor, ϵ_p, of .05 on the disturbance amplitude. Both vertex and non-vertex constraint maximisers were identified confirming the need to consider non-vertex maximisers. The variables which contributed non-vertex maximisers 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 ($ivert = 1$, $nrand_{max} = 1$).

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 400m^3/hr acid flow at -.2N. The scheme in figure 6.4 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 3 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.

6.2.3 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 a feasible design to be developed 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 non-vertex constraint maximisers and indicates the use of a projection factor to trade off accuracy of solution and computation time to be effective.

6.3 $Ca(OH)_2$ neutralisation with highly variable titration characteristic (Central Effluent Plant)

6.3.1 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 (1.45N) 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 neutralisation 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 neutralisation tanks giving about 12 hours of concentration and flow equalisation at maximum flow. The scheme to be evaluated is shown in figure 6.5.

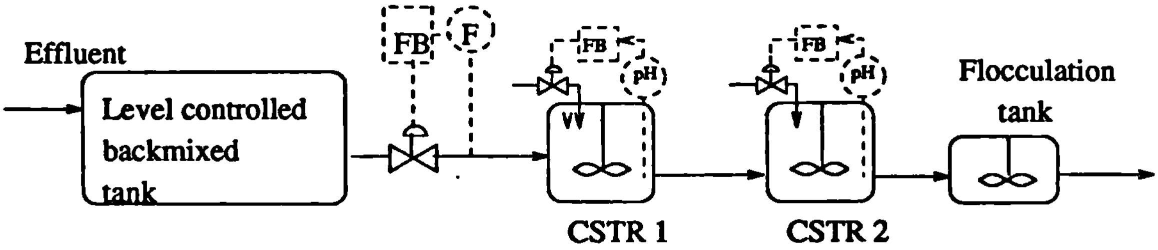


Figure 6.5: Treatment system to be evaluated

The aim is to be able to control with setpoints between 7.5 and 8.5pH ± .5pH. The ± .5pH 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 to 30 seconds. Biases of up to ±.25pH are assumed to be possible on pH measurement.

24-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 figure 6.6 with pH plotted against net concentration in normals, with pH 5 taken as zero net concentration.

The key characteristics of titration data for design purposes are the slope of the titration curve ($\frac{\partial pH}{\partial cnet}$) and the variability of this slope. The titration data are shown in figure 6.7 in the form of $\frac{\partial pH}{\partial cnet}$ as a function of pH. The titration curve slopes show a variability of between 4:1 to 10:1 above pH 3, e.g. the slope between 7 and 9 varies 5:1 with values between 1000pH/N and 200pH/N. The maximum slope of 2000pH/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 (7–9pH) was 1000pH/N. Inlet concentration ranged from .08M to .035M acid (-.08 to -.035N). For the analysis carried out, it is assumed that any titration curve lying within the range of sensitivities observed may occur, i.e. that there is no correlation between buffering at different pH levels. The "max buffering" and "min buffering" curves in figure 6.6 represent the extremes under this assumption and do not depart substantially from

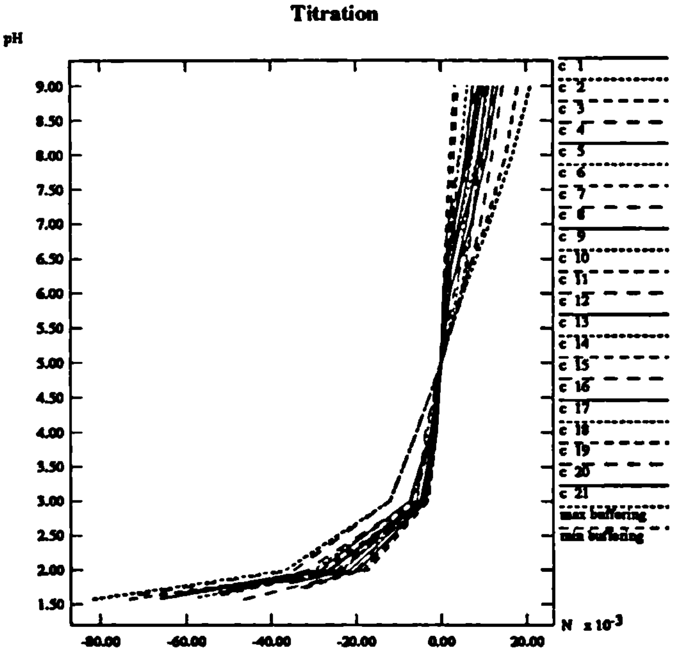


Figure 6.6: Titration curves

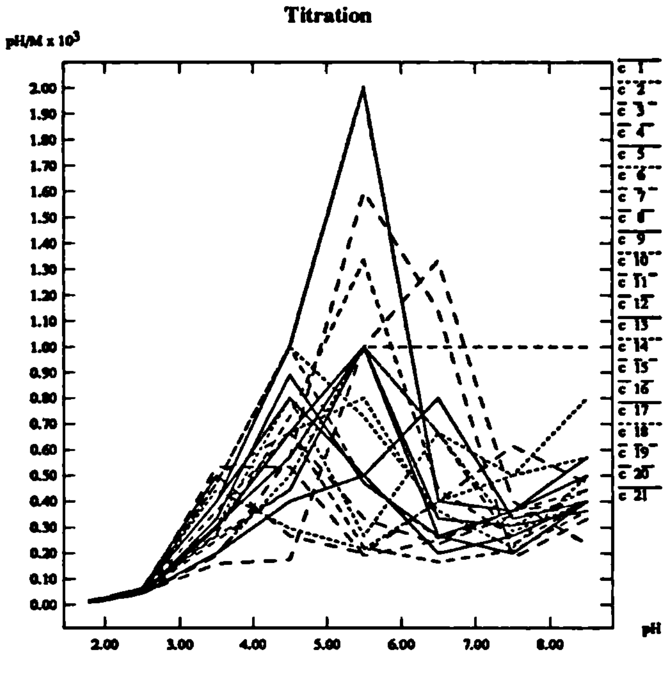


Figure 6.7: Variation of titration curve slope with pH

the observed curves.

The titration curves were confirmed to be reversible.

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 .035N to .08N, 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 modelled using the standard model (equation 4.32)

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

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 modelling error and k_{mix} was taken as 1 as there was no reason to expect atypical mixing. The batch titration curves for this application (figure 4.6) showed evidence of the "alkali inhibition" characteristic (section 4.2.2). This characteristic would have a favourable effect in slowing upward pH drift in the second and third tanks, particularly at the higher pH setpoint (8.5). This effect was neglected in the analysis, making the results potentially pessimistic.

The particle size distribution was assumed to be given by equation 4.33

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

6.3.2 Analysis of design

6.3.2.1 Preliminary control analysis

The maximum sensitivity of pH to concentration in the target pH range is 1000pH/N. Allowing for the measurement bias of ±.25pH the acceptable transient deviation is reduced from .5pH to .25pH corresponding to $\delta_{cnet} = .00025$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 .04N. This disturbance is worse than a step from 0 to 50% of flow as although this step has $\Delta_{cnet} = .08$N it has half the associated flow-rate, which with 3 tanks in the disturbance propagation path implies 8 times the disturbance attenuation available in the worst-case. The required control precision 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 3.4.2), is therefore about 41 seconds. The estimated disturbance attenuation with PI control is 8×10^{-5} (equation 5.2). Transient control performance is not expected to be a problem for this system and a basic control scheme should be adequate.

6.3.2.2 Steady-state analysis

To meet the performance requirements at steady-state, the reaction of $Ca(OH)_2$ downstream of the controlled tanks must not result in the pH rising by more than .5pH 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. It is also necessary that the reagent carrying over from the first tank to the second tank should 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 which 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 curves in figure 4.6 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 using equation 5.8.

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

The table indicates carryover of reagent added to one tank from the tank exits downstream. 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 then carryover of reagent added to this tank should be slightly lower due 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 problematical.

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. It should be noted that provided the required models and simulation tools are available the more precise calculation can be carried out very rapidly and efficiently.

6.3.3 Conclusions and review

The proposed design was verified to be able to meet the performance requirements quite comfortably without the need for a complete design to be carried out. The preliminary analysis (section 5.3.2) 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 .3pH) and may be dominated by measurement errors. These observations are consistent with the analysis above.

6.4 $Ca(OH)_2$ neutralisation of a single effluent stream at high intensity

6.4.1 Problem definition

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

The design problem is to produce a neutralisation system which 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. Examining the guidelines and costs in sections 2.3.1.2 and 5.3.2.1 indicates that dolomitic hydrated lime ($Mg(OH)_2.Ca(OH)_2$) is the lowest cost feasible reagent. Carbonates are eliminated due to the required final pH. Waste alkalis often contain carbonates and give large amounts of suspended solids and were therefore rejected as well. Dolomitic hydrated lime should be capable of meeting the requirements but is noted to react more slowly than high-calcium hydrated lime $Ca(OH)_2$ and to have less consistent properties. A good quality high-calcium hydrated lime was therefore used, despite the increased cost.

The reagent was to be delivered as a 20%w/w slurry (standard practice) giving a concentration of about 5.9N. The concentration of acid in the neutralised stream is therefore about -4N 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 the system to be able to startup 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 minimising the number of CSTRs (low marginal cost with volume and substantial cost per unit). Further 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 characterised 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 6-8pH. The effluent to be treated has a slowly varying acid concentration around -11.8N. Zero net acid/alkali concentration was taken as giving a pH of 7.

The batch titration data shown in figure 4.5 and discussed in section 4.2.2 are based on conditions close to those expected in this problem so the model developed in that section is taken as appropriate (equation 4.32)

$$\frac{\partial r_p}{\partial t} = -k_{mix}.max(1, \frac{10}{r.k_{mix}})(k_1 + k_2[H^+] + k_3cweak) \ \mu/second$$

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, $\frac{\partial r_p}{\partial t}$, of ± 40% was included to allow for these factors, with $k_1 = 0.1$, $k_2 = 7.5$ and $k_3 = 1.5$. 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 $1.5\mu/sec$ was assumed to allow for a possible change in reaction mechanism at high acid concentrations which was not ruled out by the experimental data. The particle size distribution was assumed to be given by equation 4.33

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

(same reagent source as previous example).

6.4.2 Design

6.4.2.1 Preliminary control analysis

The maximum sensitivity of pH to concentration is 250pH/N. Allowing for measurement bias the acceptable transient deviation is about .25pH, giving δ_{cnet} = .001N. 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 2N. The required control precision, δ_c, at maximum flow is therefore about .0005.

Ideal delay-limited feedback analysis with n equal size tanks (equation 3.51) 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 ±2% varying linearly with flow and ±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 control precision 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 equation 5.2 and an effective delay of about 41 seconds (as in section 6.3) 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.

6.4.2.2 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 upwards 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 optimisation 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 always to be adding reagent at steady-state. The pH setpoint for the first tank was set to 2, the allowable value most favourable to reagent conversion. The second tank setpoint was set to 6.75 as this is the lowest value consistent with meeting the specification in the face of measurement errors. Flow was set to 100% of design rate. The CSTR residence times were optimised. The measurement biases, the uncertainties in the rate expression and the uncertainty in the titration curve buffering were the relevant uncertain parameters. 10 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 pipeline-based plug flow reactor (PFR) before the first CSTR to achieve improved reagent conversion. A plug flow reactor 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 minimises 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 meeting steady-state requirements which emerged in this case-study has been incorporated into the design procedure (section 5.3.4) This scheme is shown in figure 6.8.

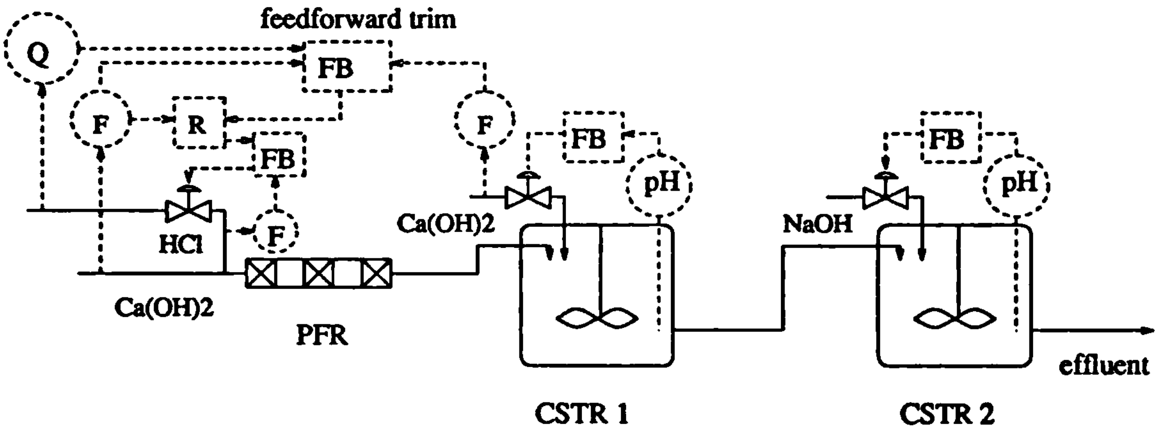


Figure 6.8: Final process and control scheme

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 plug flow reactor residence time went to its maximum value of 2 minutes. The fraction of total $Ca(OH)_2$ added to the PFR was optimised 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.

Using Shinskey's (1973) recommended scheme with a setpoint of 9.5 for the first CSTR and acid addition to the second CSTR (see section 4.2.2) gave a minimum residence time of 31 minutes due to the reduced rate of reaction in the first tank confirming that this recommendation is not generally optimal.

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

Each steady-state worst-case optimisation 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 upwards by .25pH and the minimum buffering. This was a relatively simple worst-case optimisation 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 maximisation. Most of the optimisation time was spent verifying this choice of the worst case by vertex enumeration and random search.

6.4.2.3 Complete design

At this point it becomes appropriate to carry out a full dynamic worst case design, as a design structure has been identified which 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 behaviour. 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 2pH 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 optimised along with the CSTR volume and the plug flow reactor volume.

The input conditioning function tabulated below was used on the first CSTR pH controller, as the steep titration curve slope between 3–6 pH poses a risk of limit-cycling (see section 5.3.4).

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

The optimisation generated a minimum residence time of 27.4 minutes with 2 minutes residence time in the plug flow reactor and 12.7 minutes in each CSTR. The controller tuning obtained was

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

The variables indicated by * were assigned *a priori*. The feedforward trim controller gain is dimensionless as both the input and the output are ratios of flows. The CSTR controller gains are in m^3/hr reagent per m^3/hr 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 re-run the optimisation.

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 optimisation 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 characterised by a fixed bias (± 2% of value) and biases at minimum and maximum flow (± 3.5% of value) with linear interpolation at intermediate flow, two pH measurement lags, two pH measurement biases, the titration curve slope between 6–8pH, the overall uncertainty in rate of reaction and the uncertainty in the rate of reaction in the plug flow reactor. Vertex enumeration and up to 200 random points were allowed for the global phase of the search and $\rho_{max} = 1$ 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 though 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 towards 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 towards the low pH limit. The pH responses associated with these two worst cases are shown in figure 6.9, illustrating the diversity of response which the treatment system is designed to accommodate.

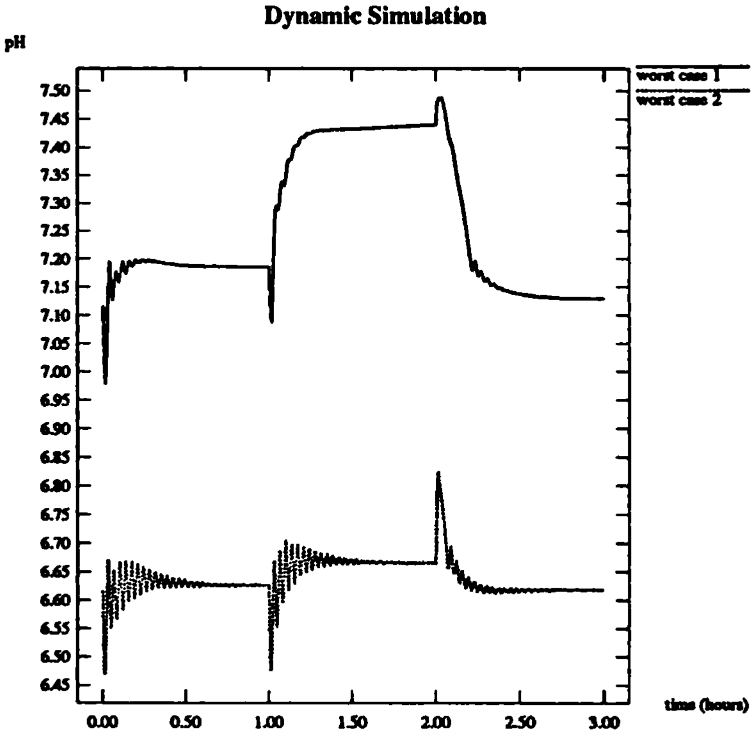


Figure 6.9: Worst-case downstream pH responses

The maximum transient deviation is actually somewhat less than expected from the earlier analysis (0.12pH versus 0.25pH). 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 3.3 to improve efficiency.

6.4.3 Conclusions and review

An unusual design meeting both control and steady-state requirements was developed. The design procedure (section 5.3) allowed a smooth refinement of the design; eliminating impractical options, identifying a core design suitable for control, modifying this design to get the required steady-state properties and finally refining the design using worst-case optimisation to obtain a design which met the actual performance requirements robustly.

Experimental trials within ICI have confirmed that the scheme proposed should work and a patent has been filed (Walsh *et al.*, 1992).

6.5 $CaCO_3$ neutralisation for a small number of effluent streams

6.5.1 Problem definition

The objective of this case study was to design an effluent treatment system for neutralisation of combinations of three effluent streams (A, B and C). The overall objective of the design was (as always) to obtain adequate performance at minimum cost. The central performance requirements are to meet the constraints on effluent discharge pH. Any effluent discharged should be kept above 4.5pH at steady state and above pH 4 at all times. For cost reasons the effluent is to be treated with $CaCO_3$ (see sections 5.3.2.1 and 2.3.1.2). Dolomitic limestone and locally available waste alkalis were found had unacceptable batch reaction characteristics (too slow or excessive downstream pH drift). The $CaCO_3$ reagent is to be delivered as a 30%w/w (7.4N) slurry (difficult to handle a higher concentration).

The treatment system is to be based on an existing continuous stirred tank reactor (CSTR) treatment plant. The plant consists of three series CSTRs. The working volume of each CSTR is about $20m^3$. The scope of the design was assumed to be limited to measurement and reagent addition systems, with the existing CSTR system being retained. A mixing delay of 15 seconds was applied to reagent addition to a CSTR based on residence time distribution data (figure 4.10, section 4.5). The slight beneficial effect of transport delays on reagent conversion is neglected in the design analysis as modelling it would greatly increase the integration time with minimal impact on the analysis.

The effluent is made up of three streams — A (100–140m^3/hr),B (40–46 m^3/hr) and C (about 8–12 m^3/hr). A and B are acid streams with concentrations of available acid around .135M. C is an alkali stream with concentration around .2M.

Titration characteristics are expected to be approximately constant for each stream except for buffering of B between pH 3–4 and are tabulated below.

A: pH range	1–2	2–3	3–4	4–5	5–6	6-8	8-9
Concentration change (N)	.1	.007	.015	.015	.004	.0025	.001

B: pH range	2–3	3–4	4–5	5–6	6-8	8–9	9–10
Concentration change (N)	.02	.05–.1	.04	.01	.01	.012	.01

C: pH range	1–2	2–3	3–4	4–5	5–6	6-8	8–9	9–12
Concentration change (N)	.01	.002	.001	.001	.001	.01	.04	.18

All the curves are based on titration with $Ca(OH)_2$. Titration with $CaCO_3$ is difficult due to the very slow dynamics around pH 5 or greater. Streams A and B contain a complex mixture of weak acids, while stream C has a substantial concentration of weak bases. A and C effluents are coupled so that if one is shutdown, the other will also be shutdown. The possible regimes are therefore A/B/C, A/C and B.

Each of these regimes encompasses a range of values of flows of each stream present and the values of measurement biases and lags may also vary. Strictly this requires consideration of three discrete conditions each with a set of uncertain parameters. This is possible, but problem analysis would be simplified substantially if it was not necessary to consider each combination of streams separately. This requires developing a *relaxed* problem definition which will give a tight but cautious approximation to the full problem definition. The total flow cannot fall below the minimum flow of B ($40m^3$/hr) or rise above the total maximum flow of all three streams (198 m^3/hr). Based on the method for predicting titration curves of mixtures discussed in section 4.1 the titration characteristics for acid loads can be approximated as lying between B and A (C makes only a minor contribution below pH 9), i.e. about 3:1 variation in buffering overall with an additional 2:1 variation between 3–4pH. The titration characteristic variation is not however independent of the flow variation as at high flows the overall characteristic is dominated by the less buffered A titration characteristic. To capture this the full variation in buffering is only allowed at low flows with the allowable range tapering into A's values as flow approaches the maximum value. The acid load is taken as .135N. The original multiple case problem definition is therefore replaced by a single flow range and a constant acid load with a titration characteristic ranging from that of A to that of B at low flow and converging to A at high flow.

The key disturbance to the system is a rapid surge of acid load of about .5 m^3/hr at -15N. This is taken as an instantaneous load change lasting several hours and provides a clear worst-case disturbance.

The pH measurements were assumed to exhibit slowly varying biases in the range ± 0.25pH and to exhibit a dynamic response characterised by a first order lag between 5 and 30 seconds.

All slurries are assumed to be prepared with sufficient accuracy to have less than a ±5% effect on controller gain and required valve sizing. This degree of uncertainty is negligible.

The model identified for the carbonate dissolution, based on the model discussed in section 4.2.3 was

$$\frac{\partial r_p}{\partial t} = -max(1, \frac{10}{r_p.k_{mix}})(10[H^+]+cweak)(1+max(0., min(1., pH-3)))\ \mu/second \quad (6.1)$$

where *cnet* is the concentration of weak acids which will dissociate below the saturation pH of carbonate (estimated as 6 pH based on expected residual CO_2 concentration).

There was particular concern about the variability of the $CaCO_3$ reagent particle size distribution. Five samples were analysed. The analyses of these samples showed this reagent's particle size distribution (p.s.d.) to be highly variable with the samples showing multiple peaks in the p.s.d. and the proportion of reagent associated with each peak varying considerably. Between 2-50% of the particles appear below 5 microns radius with a peak around 1 micron. Between 40–95% of particles appear between 5–50 microns radius with a peak around 30 microns. Between 3–40% of the particles are above 50 microns. One sample showed 10% of the reagent having particle size above 200 microns and the p.s.d. appeared to extend beyond the displayed range of up to 450 microns radius. The sample used in the experimental work showed the best particle size distribution of the 5 samples with a single peak around 30 microns and 95% of the reagent between 5–50 microns. This means that direct extrapolation of the experimental results would be optimistic in this respect with regard to rate of neutralisation and suspended solids. The observed variability of reagent particle size distribution must be considered in the analysis of expected treatment system behaviour. The carbonate p.s.d. is approximated as varying proportions of particles at 4 widely spaced radii.

radius (microns)	5.	30.	100.	300.
fraction	.1-.3	remainder	.1–.2	0.–.1

As the particle fractions must sum to one, only 3 of the fractions are varied, with the final fraction taking up the remainder.

6.5.2 Design

6.5.2.1 Preliminary control analysis

The allowable pH deviation is about .5pH. The titration curve slope is less than 66pH/N between 4 and 5pH (maximum of streams A and B). This gives an allowable concentration deviation, δ_{cnet}, of .0075N. The worst-case disturbance involves a load change of 7.5Nm^3/hr. The worst case will be at the maximum flow of 198m^3/hr (section 5.3.1) giving a inlet concentration change, Δ_{cnet}, of .04N. The control precision requirement is therefore about 0.2 which would be trivially achieved by the existing three CSTRs and a rapidly reacting reagent. Disturbance attenuation with PI control is estimated to be about .0014 with an effective delay of 35 seconds (mixing delay and measurement lag) and a minimum residence time of about 6 minutes per tank (equation 5.2). Problems if they arise will be associated with the slow reaction rate of $CaCO_3$ as the target pH around 4.5 is approached.

6.5.2.2 Steady-state analysis

In this case problems are not associated with upward downstream drift in pH. There is no upper limit on pH and the upward drift would be self-limiting to around 6pH. However the kinetics are so slow and variable (due to particle size variations) that they create a problem of control.

The reagent will not approach complete conversion unless it is added to the first tank and allowed to react in all three tanks. This can be inferred from the batch titration curves (figure 4.8) as the curve going to pH 5 does not approach its final value closely until about 20 minutes have elapsed which is comparable to the total residence time of the tanks at maximum flow. These curves were generated with a very good particle size distribution compared to the worst sample tested making this response substantially optimistic as an indicator of reagent performance, even though the experimental mixing was inferior to that expected in the CSTRs. It is therefore desirable to add all the reagent to the first tank to maximise conversion and minimise suspended solids and wasted reagent. Assuming the reagent is added in this way, it is then necessary to choose the appropriate feedback scheme to regulate reagent addition.

The worst-case reagent conversion was computed for integral feedback control of reagent addition to the first tank from a pH measurement in each of the tanks in turn. The results are tabulated below. In determining the worst case, bias in the pH measurement, variability in flow and titration characteristic and variability in the particle size distribution were considered. The setpoint of the controller was optimised to maximise the worst-case reagent conversion.

pH measurement in tank:	1	2	3
minimum conversion	20.8%	60.7%	76.3%
setpoint	4.59	4.74	4.75

Control of reagent addition directly from pH measurement in the first tank results in a major reduction in minimum reagent conversion compared to control from the second or third tank. The problem arises from the fact that the optimal setpoint in the first tank to just get the final pH above 4.5 varies substantially with flow, reagent p.s.d. and buffering. To guarantee acceptable final pH for low flow and good p.s.d. the first tank setpoint must be much higher than desirable for high flow and poor p.s.d. It should be noted that the conversions given are worst-case conversions and the conversion will normally be considerably better than this. Actual reagent conversion may be somewhat better than predicted as the larger particles in the reagent batches with poor p.s.d. may not be fully suspended and may therefore remain in the CSTR system for a longer time and react more completely, but insufficient information is available to model this effect. Reagent usage is optimised by controlling the reagent addition to give a nominal pH of 4.75 in the third tank (pH measurement errors may reduce this to a real pH of 4.5 which is the steady-state limit).

Controlling reagent addition directly from the third tank would maximise reagent conversion, but would not give a good dynamic response as control would be effectively around 3 large lags in series with the lags varying with flow. For good control, it is desirable to have a single first order lag response or as near to this as possible. Control considerations therefore point to control of reagent addition from the pH measured in the first tank.

There are two obvious compromises between these conflicting requirements. The first is to control the reagent addition directly from the pH in the second tank. The second is to use a cascade control scheme in which the reagent addition is controlled by feedback from the pH in the first tank while adjusting the setpoint of this controller based on feedback from the pH in the third tank.

The cascade control scheme will give better reagent conversion in steady-state and hence a clear economic benefit. This scheme is intuitively likely to give better control. Its only disadvantage is the need for an extra pH measurement. This scheme is therefore adopted for evaluation by worst-case dynamic optimisation.

6.5.2.3 Complete design

Both controllers were taken as being PI controllers. The setpoint of the primary controller (feedback from third tank pH) was set to 4.75 to give optimal steady-state reagent conversion. In addition to the variability considered for the steady-state analysis, uncertainty in the measurement lags was considered.

The controller tuning was optimised to minimise the maximum integral squared error between measurements and setpoints subject to keeping the pH above 4 at all times. Performance was evaluated over 7 hours simulated time with the acid disturbance applied after 5 minutes and maintained for 3 hours. The optimised response met the requirements, but was not attractively damped. The optimised tuning was therefore relaxed slightly and the performance checked by constraint maximisation to ensure this relaxation had not led to constraint violations. The typical and worst-case performances are shown in figures 6.10 and 6.11. Time is given in seconds.

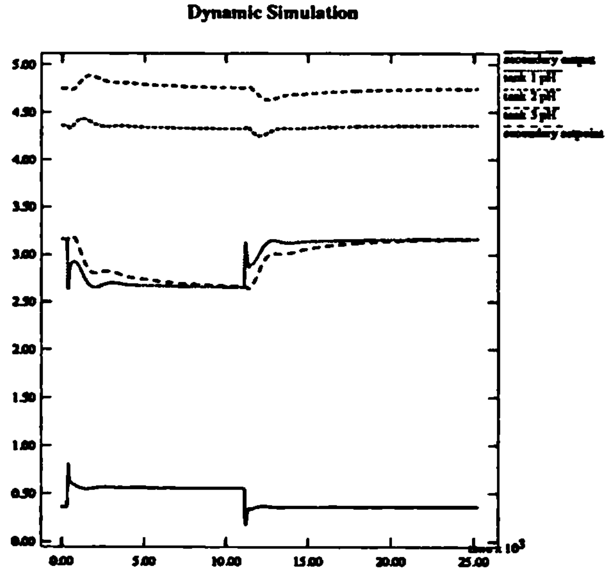


Figure 6.10: Typical response of cascade control scheme to acid disturbance

The design showed worst performance at low flows.

The final tuning for the primary controller was a gain of 1.9 pH/pH and an integral action time of 30 minutes. The tuning for the secondary controller is a gain of .0013 m^3/sec/pH and an integral action time of 60 minutes.

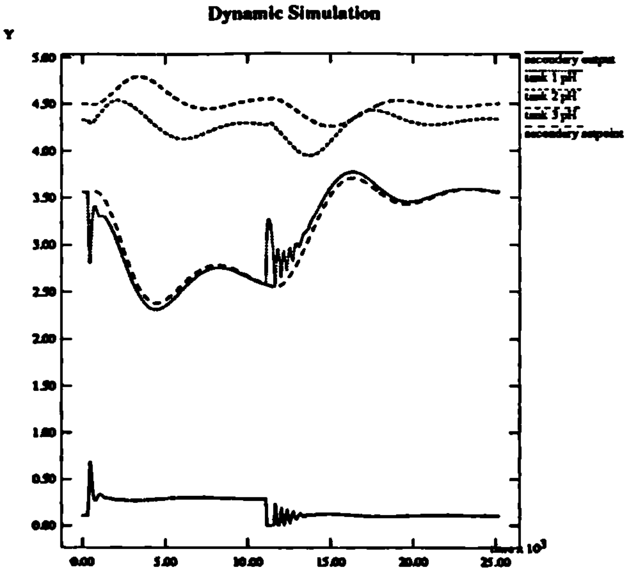


Figure 6.11: Worst response of cascade control scheme to acid disturbance

6.5.3 Conclusions and review

Due to the very slow reaction kinetics, the preliminary control analysis was not an effective tool for this problem. Steady-state analysis and qualitative consideration of the requirements for good control allowed identification of a candidate design. Worst-case design techniques verified that adequate performance was obtained.

This scheme has not yet been implemented so it is not possible to validate the design.

6.6 Oxidation of cyanide with hypochlorite

The work discussed in this section was carried out by an MSc student, Rodrigo Favela, and is documented in detail in Favela (1991). The model on which the work was based has been presented in section 4.3 and the general design issues were discussed in section 5.4. The work carried out by Favela identified a typical cyanide oxidation problem based on review of the published literature and used a mixture of nominal steady-state analysis and delay-limited control analysis to investigate a range of process options for oxidation of cyanide to nitrogen and carbon dioxide. The process options considered were reaction in a single CSTR or PFR and reaction in two CSTRs or two PFRs in series. pH and redox potential were controlled in each stage by the addition of NaOH, H_2SO_4 and NaOCl. Attempting to carry out the reaction in a single stage was shown by steady-state optimisation to be a very unattractive option, reflecting the fact the reaction proceeds in two stages for which the optimal pH conditions are very different (see section 4.3). The two PFR scheme was shown by delay-limited control analysis to be inadequate. The delay-limited control analysis also indicated a need to increase the size of the two CSTRs by about 50% compared to the steady-state optimum. The design which was shown to satisfy the nominal steady-state and delay-limited control requirements had a residence time of 12.5 minutes for the first tank and 38 minutes for the second tank which is quite close to the designs recommended in the literature — residence times of 30 minutes to 1 hour for each tank. A full worst-case design analysis is likely to increase the tank sizes significantly, but was not carried out due to lack of time. The good correspondence between the design developed from preliminary analysis and the recommended design suggests that the design approach used is valid. An industrial case study would be necessary to validate the design approach properly.

6.7 Conclusions and review

The case studies demonstrate the successful application of the procedure, tools and models developed to a wide range of real design problems.

The key differences between the design work carried out using the methods developed in this project, and the same design problems tackled using the established methods discussed in chapter 2 are summarised below.

1. Efficient preliminary design analysis may be carried out using rigorous design tools which allow designs to be eliminated *unambiguously*. This was not possible with the earlier design guidelines which were at best heuristic.
2. The worst-case design optimisation allows systematic development of robust low-cost designs.
3. The validated model of the reaction of $Ca(OH)_2$ allows model-based design to be applied with some confidence to the many neutralisation design problems for which this is the ideal reagent.
4. The design procedure developed allows the tools and models to be used in a well-coordinated way. It also presents a set of rules for guiding design choices which is more comprehensive than those given in the established guides.

These advances do not remove the need for engineering judgement (as is clear from the examples) but allow designs to be carried out more reliably and more efficiently.

Chapter 7

Conclusions and future research

This thesis has considered the application of integrated design to chemical waste-water treatment systems. This chapter summarises the results and conclusions from the work described in this thesis and identifies extensions of the work for future consideration.

7.1 Results and conclusions

The integrated design approach to waste water neutralisation developed has been shown to be effective in allowing the development of robust low cost designs efficiently.

The main results are reviewed below.

1. Previous work on waste water treatment system design has been reviewed and a critical summary presented.
2. A robust dynamic optimisation code has been developed allowing the optimisation of the nonlinear dynamic models required to fully describe waste-water treatment systems.
3. Preliminary control analysis may be carried out efficiently and effectively using measures of the achievable disturbance rejection ranging from rigorous bounds to effective but heuristic performance estimates.
4. An algorithm for robust (worst-case) design has been developed based on review of previous work and has been shown to solve realistic problems.
5. Validated models have been developed for the main solid alkali reagents, allowing a more systematic analysis of the design requirements for the use of these reagents.
6. Other aspects of modelling — steady-state pH characteristics, mixing properties, measurement properties, cyanide oxidation — have been reviewed and appropriate methods, models and default parameters identified.
7. A design procedure for integrated design of chemical waste-water treatment systems has been developed which allows well-coordinated use of the tools and models developed.

8. The procedure, tools and models have been extensively applied to industrial examples. Their effectiveness has been demonstrated by these examples, which include two cases in which industrial feedback has confirmed that the designs perform satisfactorily.

Many of the techniques developed provide a useful contribution to an integrated design "toolkit" for general process design problems. The design method developed could be applied, with some modification, to other problem areas. 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. I believe that appropriate design procedures for many problems would retain the basic structure of preliminary control analysis, steady-state analysis and full design with dynamic models and uncertainty.

7.2 Suggestions for future work

The algorithm 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 for general purposes and should be generalised.
2. Logical constraints could be introduced to the vertex search to identify and eliminate redundant vertices and make the search more efficient.

The relationship between mass-transfer coefficient and particle size for suspended particles and the dependence of this relationship on particle density stands out as a general modelling issue which deserves further attention.

Further development and validation of the solid reagent kinetic models is desirable, e.g. to allow prediction of the effect of temperature variations.

A suitable computational fluid dynamics package might be able to generate more useful information on mixing in stirred tanks than was achieved in this project. Multiphase CFD techniques may allow the development of models predicting the variation of residence time with particle size which might enable the development of appropriate shortcut models for use in design.

Modelling of electrode dynamics is still in need of some basic advances in understanding if advanced control schemes are to be able to compensate for their effect.

The use of the design techniques in other problem areas should be explored.

The techniques developed should be packaged for industrial use.

References

P. Armenante and D. J. Kirwan. Mass transfer to micro-particles in agitated systems. *Chemical engineering science*, 44:2781–2796, 1989.

S. Asai, Y. Konishi, and Y. Sasaki. Mass transfer between fine particles and liquids in agitated vessels. *Journal of Chemical Engineering of Japan*, 21:107–112, 1988.

W. A. Badran. *On designing algorithms for controlling pH.* PhD thesis, University of Oxford, 1984.

C. F. Baes and R. E. Mesmer. *The hydrolysis of cations.* John Wiley and sons, 1976.

P. L. Bailey and E. Bishop. Hydrolysis of cyanogen chloride. *Journal of Chemical Society, Dalton Transactions*, 9:912–916, 1973.

R. G. Bates. *Determination of pH: Theory and Practice.* John Wiley and sons, 1964.

M. S. Bazaara and C. M. Shetty. *Nonlinear Programming: Theory and Algorithms.* John Wiley and sons, 1979.

H. Becker and H. von Zander. The rate of neutralisation of wet or dry slaked hydrated limes as a function of their conditions of manufacture. *Zement,Kalk,Gips*, 29:236–238, 1976.

H. Becker and H. von Zander. The rate of solution of hydrated limes. *Zement,Kalk,Gips*, 30:188–191, 1977.

H. Becker. The effect of stirring and diffusion on the dissolving of lime hydrates. *Zement,Kalk,Gips*, 39:222–223, 1986.

L. T. Biegler and J. E. Cuthrell. Improved infeasible path optimization for sequential modular simulators - II: the optimisation algorithm. *Computers and Chemical Engineering*, 9:257–267, 1985.

L. T. Biegler. Strategies for simultaneous solution and optimization of differential–algebraic systems. In *Proceedings of the 3rd international conference on Foundations of Computer-aided Process Design, Colorado, July 1989*, pages 155–179, 1990.

I. Bjerle and G. T. Rochelle. Limestone dissolution from a plane surface. *Chemical Engineering Science*, 39:183–5, 1984.

A. H. Bolour-Froushan. *Prediction of single-phase turbulent flow in agitated mixing vessels.* PhD thesis, University of London, Imperial College of Science and Technology, 1986.

R. S. Boynton. *Chemistry and technology of lime and limestone.* John Wiley and sons, 1980.

P. L. T. Brian, H. B. Hales, and T. K. Sherwood. Transport of heat and mass between liquids and spherical particles in an agitated tank. *AIChE Journal*, 15:727–732, 1969.

A. Brucato, V. Brucato, L. Rizzuti, and M. Sanfilippo. Particle dissolution kinetics with batch stirred vessels. In *Fluid mixing IV: ICheMe symposium series No.121*, pages 327–341, 1990.

M. Caracotsios and W. E. Stewart. Sensitivity analysis of initial value problems with mixed ODEs and algebraic equations. *Computers and Chemical Engineering*, 9:359–365, 1985.

W. K. Chan and R. G. H. Prince. Heuristic evolutionary synthesis with non-sharp separators. In *Third International Symposium on Process Systems Engineering, Sydney, 28 Aug–2 Sept, 1988*, pages 300–312, 1988.

P. K. Chan and G. T. Rochelle. Limestone dissolution: Effects of pH, CO_2, and buffers modelled by mass transfer. ACS Symposium series 188. 1982.

Chemineer. Kenics static mixers: KTEK series. Technical Report K-TEK-4, Chemineer, 1988.

C. L. Chen. *A class of successive quadratic programming methods for flowsheet optimisation.* PhD thesis, University of London, 1988.

L. Chou, R. M. Garrels, and R. Wollast. Comparative study of the kinetics and mechanisms of dissolution of carbonate minerals. *Chemical Geology*, 78:269–282, 1989.

G. C. Cushnie. *Removal of metals from wastewater: Neutralisation and precipitation.* Noyes Publications, 1984.

G. C. Cushnie. *Electroplating wastewater pollution control technology; pollution technology review 115.* Noyes Publications, 1985.

P. J. Docherty. Automatic pH control: Neutralisation of acid wastes by addition of lime slurry. Master's thesis, Dartmouth College, Hanover, 1972.

A. M. Dunker. The decoupled direct method for calculating sensitivity coefficients in chemical kinetics. *Journal of chemical physics*, 81:2385–2393, 1984.

G. E. Eden and A. B. Wheatland. Effect of temperature and presence of hypochlorite on the rate of hydrolysis of cyanogen chloride in alkaline solution. *Society of Chemical Industry Journal*, 69:166–169, 1950.

T. J. Edwards, J. Newman, and J. Prausnitz. Thermodynamics of aqueous solutions containing volatile weak electrolytes. *AIChE Journal*, 21:248–259, 1975.

T. J. Edwards, G. Maurer, J. Newman, and J. M. Prausnitz. Vapor-liquid equilibria in multicomponent aqueous solutions of volatile weak electrolytes. *AIChE Journal*, 24:966–976, 1978.

W. J. Eilbeck and G. Mattock. *Chemical Processes in waste water treatment*. Ellis Horwood, 1987.

R. Favela. Design of a cyanide waste treatment plant. Master's thesis, Imperial College of Science and Technology, 1991.

F. N. Fritsch and J. Butland. A method for constructing local monotone piecewise cubic interpolants. *SIAM journal of Scientific and Statistical Computing*, 5:300–304, 1984.

C. W. Gear. Simultaneous numerical solution of differential-algebraic equations. *IEEE Transactions on Circuit Theory*, CT-18:89–95, 1971.

C. M. Gerritsen and D. W. Margerum. Non-metal redox kinetics: Hypochlorite and hypochlorous acid reaction with cyanide. *Inorganic Chemistry*, 29:2757–2762, 1990.

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

M. Greaves. Surface aeration in agitated vessels. *Institution of Chemical Engineers Symposium Series*, 64:H1–H22, 1981.

D. Gritsis. *The dynamic simulation and optimal control of systems described by index two differential algebraic equations*. PhD thesis, University of London, 1990.

I. E. Grossmann and C. A. Floudas. Active constraint strategy for flexibility analysis in chemical processes. *Computers and Chemical Engineering*, 6:675–693, 1987.

I. E. Grossmann and D. A. Straub. Recent developments in the evaluation and optimisation of flexible chemical processes. In *Proceedings of COPE-91, Barcelona, Spain*, pages 49–59, 1991.

I. E. Grossmann, K. P. Haleman, and R. E. Swaney. Optimization strategies for flexible chemical processes. *Computers and Chemical Engineering*, 7:439ff, 1983.

T. K. Gustafsson and K. V. Waller. Dynamic modelling and reaction invariant control of pH. *Chemical Engineering Science*, 38:389–398, 1983.

T. K. Gustafsson. An experimental study of a class of algorithms for adaptive pH control. *Chemical Engineering Science*, 40:827–837, 1985.

P. Harriott. Mass transfer to particles: Part I: Suspended in agitated tanks. *AIChE Journal*, 8:93–101, 1962a.

P. Harriott. Mass transfer to particles: Part II: Suspended in a pipeline. *AIChE Journal*, 8:101–102, 1962b.

R. T. Haslam, F. W. Adams, and R. H. Kean. The rate of solution and availability of commercial limes. *Industrial and Engineering Chemistry*, 18:19–23, 1926.

H. Z. Hershkovitch, T. J. McAvoy, and A. L. Liapis. Dynamic modelling of pH electrodes. *The Canadian Journal of Chemical Engineering*, 56:346–353, 1978.

B. R. Holt and M. Morari. Design of resilient processing plants—V: The effect of deadtime on dynamic resilience. *Chemical Engineering Science*, 40:1229–1237, 1985.

J. M. Howell. *The effect of energy integration on the switchability of chemical processes.* PhD thesis, University of London, 1984.

D. Hoyle. Designing for pH control. *Chemical engineering*, 83:121–126, 1976.

O. L. R. Jacobs, W. A. Badran, C. G. Proudfoot, and C. While. On controlling pH. *IEE Proceedings*, 134:196–200, 1987.

H. L. Jacobs. Neutralisation of acid wastes. *Sewage and Industrial wastes*, 23:900–906, 1951.

R. B. Jarvis and C. C. Pantelides. DASOLV — a differential-algebraic equation solver; version 1.1. Technical report, Center for Process Systems Engineering, Imperial College, 1992.

B. Jayadeva, Y. S. N. M. Rao, M. Chidambaram, and K. P. Madhavan. Nonlinear controller for a pH process. *Computers and Chemical Engineering*, 14:917–920, 1990.

G. Johansson and K. Norberg. Dynamic response of the glass electrode. *Journal of Electroanalytical Chemistry*, 18:239–250, 1968.

U. Kabatek and R. E. Swaney. Worst-case identification in structured process systems. *Computers and Chemical Engineering*, 16:1063–1071, 1992.

K. Kikuchi, T. Sugawara, and H. Ohashi. Correlation of liquid-side mass transfer coefficient based on the new concept of specific power group. *Chemical Engineering Science*, 43:2533–2540, 1988.

J. R. Leis and M. A. Kramer. Sensitivity analysis of systems of differential and algebraic equations. *Computers and Chemical Engineering*, 9:93–96, 1985.

O. Levenspiel. *Chemical reaction engineering*. John Wiley and sons, 1972.

D. M. Levins and J. R. Glastonbury. Application of Kolmogoroff's theory to particle-liquid mass transfer in agitated vessels. *Chemical Engineering Science*, 27:537–543, 1972a.

D. M. Levins and J. R. Glastonbury. Particle-liquid hydrodynamics and mass transfer in a stirred vessel, I —particle liquid motion. *Transactions of the Institution of Chemical Engineers*, 50:32–41, 1972b.

D. M. Levins and J. R. Glastonbury. Particle-liquid hydrodynamics and mass transfer in a stirred vessel, II — mass transfer. *Transactions of the Institution of Chemical Engineers*, 50:132–146, 1972c.

J. S. Logsdon and L. T. Biegler. A relaxed reduced space SQP strategy for dynamic optimization problems. *Computers and Chemical Engineering*, 17:367–372, 1993.

W. L. Luyben. *Process modeling, simulation and control for chemical engineers*. McGraw-Hill, 1990.

G. Mapstone. The hypochlorite oxidation of cyanide and cyanate. *Journal of applied chemical biotechnology*, 28:135–143, 1978.

G. Mattock. Automatic control in effluent treatment. *Transactions of the society of instrument technology*, pages 173–186, 1964.

D. Q. Mayne, H. Michalska, and E. Polak. An efficient outer approximations algorithm for solving infinite sets of inequalities. In *Proceedings of 29th Conference on Decision and Control, December 1990*, pages 960–965, 1990.

K. B. McAuley and J. F. MacGregor. Optimal grade transitions in a gas phase polyethylene reactor. *AIChE Journal*, 38:1564–1576, 1992.

T. J. McAvoy. Dynamic modelling of pH in aqueous systems. In *Proceedings of AIChE workshop in Industrial Process Control, 1979*, pages 35–39, 1979.

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

P. Mizsey and Z. Fonyo. Toward a more realistic overall process synthesis – the combined approach. *Computers and Chemical Engineering*, 14:1213–1236, 1990.

R. L. Moore. *Neutralisation of waste water by pH control.* Instrument Society of America, 1978.

K. R. Morison. *Optimal control of process systems described by systems of differential-algebraic equations.* PhD thesis, University of London, 1984.

T. J. Myron and F. G. Shinskey. Advanced control systems for waste pH processes. In *Proceedings of AIChE Workshop on Industrial Process Control, 1979*, pages 29–34, 1979.

L. T. Narraway and J. D. Perkins. Selection of process control structures based on economics. *Submitted to ESCAPE3*, 1993.

L. T. Narraway. *Selection of process control structure based on economics.* PhD thesis, Imperial College, 1992.

National Lime Association. Acid neutralisation with lime for environmental control and manufacturing processes. Technical Report Bulletin 216, National Lime Association, 1983.

A. W. Nienow. Agitated vessel particle-liquid mass transfer: A comparison between theories and data. *The Chemical Engineeering Journal*, 9:153–160, 1975.

N. Nishida, Y. A. Liu, and A. Ichikawa. Studies in chemical process design and synthesis, II. Optimal synthesis of dynamic process systems with uncertainty. *AIChE Journal*, 22:539–548, 1976.

D. K. Nordstrom, L. N. Plummer, D. Langmuir, E. Busenberg, H. May, B. F. Jones, and D. L. Parkhurst. Revised chemical equilibrium data for major water-mineral reactions and their limitations. In *Chemical modeling of aqueous systems II*, pages 398–413. 1990.

J. M. Nouri. *Single and two-phase flows in ducts and stirred reactors.* PhD thesis, University of London, Imperial College of Science and Technology, 1988.

D. R. Olander. Simultaneous mass transfer and chemical equilibrium. *AIChE Journal*, 6:233–239, 1960.

A. R. Oroskar and R. F. Turian. The critical velocity in pipeline flow of slurries. *AIChE Journal*, 26:550–559, 1980.

C-C. D. Pai and R. R. Hughes. Strategies for formulating and solving two-stage problems for process design under uncertainty. *Computers and Chemical Engineering*, 11:695–706, 1987.

J. R. Parrish and C. B. Brosilow. Nonlinear inferential control. *AIChE Journal*, 34:633–644, 1988.

J. D. Perkins and M. P. F. Wong. Assessing controllability of chemical plants. *Chemical Engineering Research and design*, 63:358–362, 1985.

L. R. Petzold. A description of DASSL: a differential/algebraic system solver. In *Proceedings IMACS World Congress, Montreal, Canada, August, 1982*, pages 65–68, 1982.

K. S. Pitzer. Thermodynamics of electrolytes. 1. Theoretical basis and general equations. *Journal of Physical Chemistry*, 77:268–277, 1973.

L. N. Plummer, T. M. L. Wigley, and D. L. Parkhurst. The kinetics of dissolution of calcite in CO_2-water systems at 5 to 60 deg C and 0.0 to 1.0 atm CO_2. *American Journal of Science*, 278:179–216, 1978.

E. Polak and D. M. Stimler. Majorisation: A computational complexity reduction technique in control system design. *IEEE Transactions on Automatic Control*, 31:1010–1021, 1988.

E. Polak. An implementable algorithm for the optimal design centering, tolerancing, and tuning problem. *Journal of optimization theory and applications*, 37:45–67, 1982.

D. M. Prett and C. E. Garcia. *Fundamental Process Control.* Butterworths, 1988.

D. M. Prett, C. E. Garcia, and B. L. Ramaker. *The Second Shell Process control workshop.* Butterworths, 1990.

C. T. Price, T. E. Larson, K. M. Beck, F. C. Harrington, L. C. Smith, and L. Stephanoff. Hydrolysis and chlorinolysis of cyanogen chloride. *Journal of American Chemical Society*, 69:1640–1644, 1947.

C. G Proudfoot. *Industrial implementation of on-line computer control of pH.* PhD thesis, University of Oxford, 1983.

R.C. Reid, J. M. Prausnitz, and B. E. Poling. *The properties of gases and liquids, 4th edition.* McGraw-Hill Book Company, 1988.

H. Renon. Electrolyte solutions. *Fluid Phase Equilibria*, 30:181–195, 1986.

O. Rosen and R. Luus. Evaluation of gradients for piecewise constant optimal control. *Computers and Chemical Engineering*, 15:273–281, 1991.

R. W. H. Sargent and G. R. Sullivan. Development of feed changeover policies for refinery distillation columns. *Industrial and Engineering Chemistry, Process Design and Development*, 18:113, 1979.

F. G. Shinskey. *pH and pIon control.* Wiley environmental science and technology series, 1973.

D. M. Southwood-Jones and M. O. Tade. Simulation and experimental evaluation of pH controllers. In *Proceedings The Institution of Engineers Australia Fourth conference on control engineering, Gold Coast, August, 1990*, pages 28–32, 1990.

K. E. Spells. Correlations for use in transport of aqueous suspensions of fine solids through pipes. *Transactions of the Institution of Chemical Engineers*, 33:79–84, 1955.

G. Sposido. *The environmental chemistry of aluminium.* CRC Press, Florida, 1989.

R. E. Swaney and I. E. Grossmann. An index for operational flexibility in chemical process design. Part I: Formulation and theory. *AIChE journal*, 31:621–630, 1985a.

R. E. Swaney and I. E. Grossmann. An index for operational flexibility in chemical process design. Part II: Computational algorithms. *AIChE journal*, 31:631–641, 1985b.

W. K. Teo and T. C. Tan. Hypochlorite oxidation of cyanate under mildly alkaline conditions. *Water Research*, 21:677–682, 1987.

A. L. Tits. On the optimal design centering, tolerancing, and tuning problem. *Journal of Optimization Theory and Applications*, 45:487–494, 1985.

A. Törn and A. Žilinskas. *Global Optimisation.* Springer-Verlag, 1989.

V. L. Trevanthan. Characteristics of pH control. In *Proceedings of AIChE workshop in Industrial Process Control, 1979*, pages 24–28, 1979.

R. E. Treybal. *Mass-transfer operations.* McGraw-Hill International, 1980.

A. H. Truesdell and B. F. Jones. WATEQ, a computer program for calculating chemical equilibria of natural waters. *Journal of research of the U.S. Geological Survey*, 2:233–248, 1974.

V. W. Uhl and J. B. Gray. *Mixing, Vol.3.* Academic Press, 1986.

S. Vasantharajan and L. T. Biegler. Simultaneous strategies for optimisation of differential–algebraic systems with enforcement of error criteria. *Computers and Chemical Engineering*, 14:1083–1100, 1990.

V. Vassiliadis. *Computational solution of dynamic optimisation problems with general differential-algebraic constraints.* PhD thesis, Imperial College, 1993.

F. W. Velguth and R. C. Anderson. Determination of minimum capacities for control applications. *ISA Journal*, 1:33–38, 1954.

V. Visweswaran and C. A. Floudas. A global optimisation algorithm (GOP) for certain classes of nonconvex NLPs .1. Theory. *Computers and Chemical Engineering*, 14:1419–1417, 1990.

A. G. Volosov, I. L. Khodakovskiy, and B. N. Ryzhenko. Equilibria in the system SiO_2-H_2O at elevated temperatures along the lower three-phase curve. *Geochemistry International*, 9:362–377, 1972.

M. Wallin and I. Bjerle. Rate models for limestone dissolution: a comparison. *Geochimica et Cosmochimica Acta*, 53:1171–1176, 1989.

S. Walsh, D. Annells, and J. Harkin. *GB patent application 9217962.1.* 1992.

G. L. Williams, R. R. Rhinehart, and J. B. Riggs. In-line process-model-based control of waste water pH using dual base injection. *Industrial and Engineering Chemistry Research*, 30:1254–1259, 1990.

H. Yagi, T. Motouchi, and H. Hikita. Mass transfer from fine particles in a stirred vessel. Effect of specific surface area of particles. *Industrial and Engineering Chemistry, Process Design and Development*, 23:145–150, 1984.

J. F. Zemaitis, D. M. Clark, M. Rafal, and N. C. Scrivner. *Handbook of aqueous electrolyte thermodynamics*. American Institute of Chemical Engineers, 1986.

TH.N. Zwietering. Suspending of solid particles in liquid by agitators. *Chemical Engineering Science*, 8:244—253, 1958.

Appendix A

Solution of DAEs to generate gradients using the sensitivity approach

Differential algebraic equation representations allow equations to be defined in a natural way and in particular accommodate implicit algebraic equations systematically, unlike ordinary differential equation representations in which such equations must be eliminated or masked by special solution procedures. The general expression for the differential algebraic equations is given by:

$$\begin{aligned} f(\dot{x}, x, z, \theta, t) &= 0 \\ g(x, z, \theta, t) &= 0 \end{aligned} \tag{A.1}$$

where x represents the vector of state variables, z the vector of algebraic variables and θ the vector of parameters.

The Backward Difference Formula (BDF) solution technique developed by Gear (1971) has become a standard technique for DAE solution. Implementations of the technique vary slightly but the main differences between packages lies in the supporting facilities such as initialisation and discontinuity handling. DASSL (Petzold, 1982) is quite weak in these latter respects. The SPEEDUP package provides robust initialisation facilities and discontinuity handling but is unwieldy as a development tool. The package chosen for this work was therefore DASOLV (Jarvis and Pantelides, 1992), a stand-alone FORTRAN code based on the SPEEDUP numerical solution techniques. This software includes routines to evaluate the sensitivity of the states to the parameters, $\frac{\partial x}{\partial \theta}$, which may be used to calculate gradients of performance indices for the optimisation routines. These techniques are not as widely known as BDF techniques so an outline of their development is given below.

The development of the results given below follows Leis and Kramer (1985). The implementation in DASOLV is based on Caracotsios and Stewart (1985) and further details can be found in Jarvis and Pantelides (1992). Differentiating the general differential algebraic equation system with respect to θ gives:

$$\begin{aligned} f_{\dot{x}}\dot{S}_x + f_x S_x + f_z S_z + f_\theta &= 0 \\ g_x S_x + g_z S_z + g_\theta &= 0 \end{aligned} \tag{A.2}$$

where S_x represents the matrix $\frac{\partial x}{\partial \theta}$ and S_z represents the matrix $\frac{\partial z}{\partial \theta}$. $\dot{S}_x$ ($S_{\dot{x}}$) is given by the *kth* order backwards differentiation formula (BDF) $h\beta \dot{S}_x^{n+1} = S_x^{n+1} - \sum_{i=0}^{k-1} a_i S_x^{n-i}$ Rearranging these equations gives

$$\begin{pmatrix} f_{\dot{x}} + h\beta f_x & h\beta f_y \\ g_x & g_y \end{pmatrix} \begin{pmatrix} S_x^{n+1} \\ S_z^{n+1} \end{pmatrix} = \begin{pmatrix} f_{\dot{x}}(\sum_{i=0}^{k-1} a_i S_x^{n-i}) - h\beta f_\theta \\ -g_\theta \end{pmatrix} \tag{A.3}$$

where all partial derivative matrices of f and g are evaluated at the converged solution of the last step of the state trajectory solution.

It may be noted that the left hand side jacobian matrix is the iteration matrix used during the integration of the DAE's and is therefore, in principle, available at no extra computational cost. However, while the jacobian matrix need not be evaluated on every integration step in order to solve the DAE system, solving the sensitivity equations accurately requires this. This is the main computational overhead of this approach. The method requires $8.neq.np$ double precision elements as workspace, where neq is the number of equations and np is the number of parameters for which gradients are required.

It is necessary to provide initial conditions for the sensitivity equations. If the number of parameters is np and the number of differential equations in the original system is n, then the number of initial conditions required is $np \times n$. The initial conditions are evaluated as follows. The system of DAEs at the initial condition is:

$$\begin{aligned} f(\dot{x}_0, x_0, z_0, \theta) &= 0 \\ g(x_0, z_0, \theta) &= 0 \\ R(\dot{x}_0, x_0, z_0)^T - \alpha &= 0 \end{aligned} \tag{A.4}$$

where α is an n dimensional set of initial values and R is a 'selection matrix' which contains 1's and 0's only, and selects which n of the values $\dot{x}_0$, x_0 and y_0 are chosen to be specified as the initial values.

Differentiation of this system with respect to each parameter, θ, provides the following system of equations which may be solved to provide the initial condition for the sensitivity equations.

$$\begin{pmatrix} f_{\dot{x}} & f_x & f_z \\ 0 & g_x & g_z \\ & R & \end{pmatrix} \begin{pmatrix} \frac{\partial \dot{x}_0}{\partial \theta} \\ \frac{\partial x_0}{\partial \theta} \\ \frac{\partial z}{\partial \theta} \end{pmatrix} = \begin{pmatrix} -\frac{\partial f}{\partial \theta} \\ -\frac{\partial g}{\partial \theta} \\ \frac{\partial \alpha}{\partial \theta} \end{pmatrix} \tag{A.5}$$

It may be noted that the matrix on the left hand side of this expression is the one used in solving the initialisation problem for the original DAE system, and so is available at no extra cost. If variables which are set on initialisation are dependent on parameters then

$\frac{\partial \alpha}{\partial \theta}$ will be non-zero. Initial values of sensitivities must then be placed in the appropriate history array on the first call to DASOLV.

Use of this gradient evaluation technique in case studies indicates that a gradient evaluation takes about four times as long as a function evaluation. The cost of gradient evaluation is discussed in more detail in Vassiliadis (1993).

For discontinuities which occur at times independent of the parameters the sensitivity values can be reinitialised following the discontinuity taking the values of the state sensitivities prior to the discontinuity as the initial values of the state sensitivities after the discontinuity, $S_+(t_d) = S_-(t_d)$, so no special modification is required. For general discontinuities this is no longer true as the sensitivity after the discontinuity is based on two components, the sensitivity of the state prior to the discontinuity and the sensitivity of the time the discontinuity occurs, $S_+(t_d) = S_-(t_d) + (\frac{\partial x_-}{\partial t} - \frac{\partial x_+}{\partial t})\frac{\partial t_d}{\partial \theta}$. The implementation of sensitivity evaluation in DASOLV would require modification to deal with this type of discontinuity which occurs, for example, with integral desaturation of controllers. It is easier to smooth discontinuities to avoid this limitation than to systematically overcome the problem (see section 3.2.2.5).

If the equation system involves variables subject to a pure time delay, $x' = x(t - t_d)$, further modifications would be necessary as the right hand side of the equation for updating the sensitivity values becomes

$$\begin{pmatrix} f_{\dot{x}}(\sum_{i=0}^{k-1} a_i S_x^{n-i}) - h\beta f_\theta - h\beta f_{x'} \frac{\partial x'}{\partial \theta} \\ -g_\theta \end{pmatrix} \tag{A.6}$$

where $\frac{\partial x'}{\partial \theta} = \frac{\partial x(t-t_d)}{\partial \theta}$. This problem therefore requires the storage of sensitivity values along with the delayed values.

Neither general implicit discontinuity handling nor delay handling have been implemented in this project.

Appendix B

Worst-case design algorithm

```
1: Initialisation
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 effort in the constraint maximisation algorithm.

For each $v_i$ define $pvert_i$

Set $\epsilon_p$

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

Proceed to multi-model design (2:)
```

2: Multi-model 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 optimisation subject to constraints from active models
If the optimisation 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 maximisation

LOCAL SEARCH:
    Local vertex search
    lv1: Choose a random initial value of $v$
    lv2: Update $v$ using
        $v_i = v_i^l$ if $\frac{\partial c_{max}}{\partial v_i} \leq 0$
        $v_i = v_i^u$ 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 $\Delta c_k$ 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 $\rho$ is less than $\rho^{max}$
    or no new vertices are identified then local vertex search is complete
    Else select the new vertex value of $v$ giving the maximum value of $\rho$
    and go back to lv2:
    If no constraint violations have been found set $\rho^{max}$ to $\rho^{max}_{min}$
    to increase the depth of search and go to back to lv3:

    Carry out local searches into the interior of $V$ from up to $nlocl$ 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 $nlocl$ until either
    $nlocl = nlocl_{max}$ 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 maximiser
        with an ascent direction into $v$
    Multi-start random search
    Carry out random trials until the number of trials exceeds $nrand$
        Generate trial points using
            $v_i = v_i^l + (v_i^u - v_i^l)\min(1, max(0, \frac{r_i}{1-pvert_i} + .5))$
            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 $nloc2$ local searches
            have been used in the global search increment $nloc2$ up to a maximum of $nloc2_{max}$
            Carry out local search unless the number of local searches in the global search
             phase exceeds $nloc2$
        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 multi-model design (2:)
else
    Optimum of worst-case design has been found, STOP
```