DESIGN MODIFICATIONS FOR IMPROVED CONTROLLABILITY
- with application to design of buffer tanks

SIGURD SKOGESTAD *

Chemical Engineering Dept., University of Trondheim - NTH
N-7034 Trondheim, Norway

Presented at AIChE Annual Meeting, San Francisco, Nov. 13-18, 1994.

Abstract

A procedure for analyzing the input-output controllability of single-input single-output (SISO) systems is presented. This procedure is applied to the problem of designing buffer tanks. Although buffer tanks are often introduced for control purposes, they are usually sized in a rather *ad hoc* manner without explicitly considering the expected disturbances and desired control objectives. It is shown how to use a simple controllability analysis as a basis for a quantitative approach. The main steps are to consider the disturbances and scale the variables properly. Finally, these results are applied to the design of a pH-neutralization process.

*E-mail: skoge@kjemi.unit.no, phone: +47-7359-4154, fax: +47-7359-4080

1 Introduction

In process control courses the issues of controller design and stability analysis are often emphasized. However, in practice the following three issues are usually more important.

I. How well can the plant be controlled? Before attempting to start any controller design one should have some idea of how easy the plant actually is to control. Is it a difficult control problem? Indeed, does there even exist a controller which meets the required performance objectives?

II. What control strategy should be used? Another important question is to decide on the control strategy: What to measure, what to manipulate, how to pair? In textbooks one finds qualitative rules for this. For example in Seborg et al. (1989) one finds in a chapter called "The art of process control" the rules:

1. Control outputs that are not self-regulating
2. Control outputs that have favorable dynamic and static characteristics, i.e., there should exist an input with a significant, direct and rapid effect.
3. Select inputs that have large effects on the outputs.
4. Select inputs that rapidly effect the controlled variables

These rules are reasonable, but what is "self-regulating", "large", "rapid" and "direct". One objective of this paper is to quantify these terms.

III. How should the process be changed to improve control ? For example, one may want to find the required size of a buffer tank for damping a disturbance, or one may want to know how fast a measurement should be to get acceptable control.

Controllability analysis. All the above three questions are related to the inherent control characteristics of the process itself, that is, to what is denoted the *input-output controllability* of the process. We shall use the following definition:

(Input-output) controllability is the ability to achieve acceptable control performance, that is, to keep the outputs (y) within specified bounds or displacements from their setpoints (r), in spite of unknown variations such as disturbances (d) and plant changes, using available inputs (u) and available measurements (e.g.,y_m or d_m).

In summary, a plant is controllable if there *exists* a controller (connecting measurements and inputs) that yields acceptable performance for all expected plant variations.

Thus, controllability is independent of the controller, and is a property of the plant (process) only. It can only be affected by changing the plant itself, that is, by design modifications. These may include:

1. Change the apparatus itself, e.g., type, size, etc.
2. Relocate sensors and actuators
3. Add new equipment to dampen disturbances, e.g., buffer tanks.
4. Add extra sensors for measurement (to be used in feedforward and cascade control)
5. Add extra actuators (to be used for parallel control)
6. Change the control objectives
7. Change the structure of the lower levels of control already in place

(It may be argued whether it is appropriate to label the last two items as design modifications, but at least they address issues which come before the actual controller design.)

The focus in this paper is to consider the use of buffer tanks to improve the controllability.

Surprisingly, in spite of the fact that mathematical methods are used extensively for control system design, the methods available when it comes to controllability analysis are largely qualitative. In most cases the "simulation approach" is used. However, this requires a specific controller design and specific values of disturbances and setpoint changes. In the end one never really knows if a result is a fundamental property of the plant or if it depends on these specific choices The objective of the paper is to present a procedure for controllability analysis for scalar systems and to apply this procedure to a few examples. Earlier work on input-output controllability analysis includes that of Ziegler and Nichols (1943), Rosenbrock (1970), and Morari (1983) who made use of the concept of "perfect control".

One shortcoming with the controllability analysis presented in this paper is that all the measures are linear. This may seem very restrictive, but usually it is not. In fact, one of the most important nonlinearities, namely that of input constraints, can be handled well with a linear analysis. To deal with slowly varying changes one may perform a controllability analysis at several selected operating points. As a last step one may perform some nonlinear simulations to confirm the linear controllability analysis. Experience from a large number of case studies confirms that the agreement is generally very good.

Remarks on the definition of controllability. The above definition is in tune with most engineers' intuitive feeling about the term, and was also how the term was used historically in the control literature. For example, Ziegler and Nichols (1943) define controllability as *"the ability of the process to achieve and maintain the desired equilibrium value"*. Unfortunately, in the 60's the term "controllability"

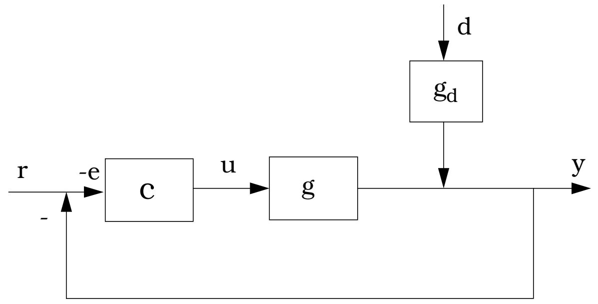

Figure 1: Block diagram of feedback control system.

became synonymous with the rather narrow concept of "state controllability" introduced by Kalman, and the term is still used in this restrictive manner in the system theory community. "State controllability" is the ability to bring a system from a given initial state to any final state (but with no regard to the quality of the response between these two states). This concept is of interest for realizations and numerical calculations, but as long as we know that all the unstable modes are both controllable and observable, it has little practical significance. For example, Rosenbrock (1970, p. 177) notes that "most industrial plants are controlled quite satisfactorily though they are not [state] controllable". To avoid confusion with Kalman's state controllability, Morari (1983) introduced the term "dynamic resilience". However, this term does not capture the fact that it is related to control, and instead it is proposed to use the term "input-output controllability" if one explicitly wants to make the distinction with "state controllability".

The outline of the paper is as follows: In Section 2 we describe general tools for controllability analysis, and in Section 3 we present a simple example to illustrate the use of these tools. The main new results are in Section 4 where we consider the design of buffer tanks. Thus, Sections 2 and 3 are really intended as an introduction to controllability analysis, and the reader may skip these sections if he/she is primarily interested in the design of buffer tanks. Finally, Section 5 describes an application to a pH neutralization process where the use of several buffer tanks is needed to get acceptable controllability.

2 Controllability analysis

Consider a linear process model in terms of deviation variables

$$y = g(s)u + g_d(s)d \tag{1}$$

Here y denotes the output, u the manipulated input and d the disturbance (including disturbances entering at the input which are frequently referred to as "load changes"). $g(s)$ and $g_d(s)$ are transfer function models which describe the effect on the output of the input and disturbance, and all controllability results in this paper are based on this information. The Laplace variable s is often omitted to simplify notation. The control error e is defined as

$$e = y - r \tag{2}$$

where r denotes the reference value (setpoint) for the output. In this paper we mostly consider feedback control as illustrated in Figure 1 where

$$u = c(s)(r - y) \tag{3}$$

and $c(s)$ is the controller. Eliminating u from equations (1) and (3) yields the closed-loop response

$$y = Tr + Sg_d d; \quad e = -Sr + Sg_d d \tag{4}$$

Here the sensitivity is $S = (1+gc)^{-1}$ and the complementary sensitivity is $T = gc(1+gc)^{-1} = 1-S$. The transfer function around the feedback loop is denoted L. In this case $L = gc$.

In this paper *bandwidth is defined as the frequency ω_B where the loop gain is one in magnitude*, i.e. $|L(j\omega_B)| = 1$ (or more precisely where the low-frequency asymptote of $|L|$ first crosses 1 from above). This frequency is also called the "gain crossover frequency". Other definitions of bandwidth are used, but the difference is small. At frequencies lower than the bandwidth ($\omega < \omega_B$) feedback is effective and will affect the frequency response. However, for sinusoidal input signals (for example, a disturbance) with frequencies higher than ω_B the response will not be affected much by the feedback.

The simplest interpretation of the frequency domain is that it represents the steady-state sinusoidal response. For example, if we send an input $u(t) = u_0 \sin(\omega t)$ through a stable system with transfer function $g(s)$, then the output as $t \to \infty$ is $y(t) = y_0(sin\omega t + \phi)$ where $y_0 = |g(j\omega)|u_0$ and $\phi = \angle g(j\omega)$. Here $g(j\omega)$ represents at each frequency ω a complex number obtained from $g(s)$ by setting $s = j\omega$. A common shorthand notation used in this paper to express the sinusoidal response is (phasor notation)

$$y(j\omega) = g(j\omega)u(j\omega) \tag{5}$$

where $y(j\omega)$ and $u(j\omega)$ are complex numbers (vectors) representing at each frequency the size and phase of a sinusoidal signal. For example, $u(j\omega) = 5$ means that $u(t) = 5\sin(\omega t)$. Thus $u(j\omega)$ is *not* equal to $u(s)$ evaluated at $s = j\omega$, nor is it equal to $u(t)$ evaluated at $t = j\omega$.

A frequency domain analysis, in particular at frequencies around the bandwidth, is very useful for systems under feedback control. This is the case even when the disturbances and setpoints entering the system are *not* sinusoids. One reason is that the effect of disturbances is usually largest around the bandwidth frequency; slower disturbances are attenuated by the feedback control, and faster disturbances are usually attenuated by the process itself.

Scaling. The interpretation of most measures presented in this paper assumes that the transfer functions g and g_d are in terms of scaled variables. The first step in a controllability analysis is therefore to scale (normalize) all variables (input, disturbance, output) to be less than 1 in magnitude (i.e., within the interval -1 to 1) by normalizing each variable by its maximum value, for example, $u = u'/u'_{max}$ where u' denotes the unscaled and u the scaled variable, and u'_{max} is the largest allowed input change (in unscaled variables). For the other variables we have $d = d'/d'_{max}$, $e = e'/e'_{max}$, $y = y'/e'_{max}$, and $r = r'/e'_{max}$, where d'_{max} is the largest expected disturbance and e'_{max} the largest allowed control error. In most cases the maximum values $(u'_{max}, e'_{max}, d'_{max})$ are assumed independent of frequency.

Thus, in the following *we assume that the signals are persistent sinusoids, and that g and g_d have been scaled, such that at each frequency the allowed input $|u(j\omega)| < 1$, the expected disturbance $|d(j\omega)| < 1$, the allowed control error $|e(j\omega)| < 1$, and the expected reference signal $|r(j\omega)| < r_{max}(j\omega)$.* Note that e and r are measured in the same units so $r_{max} = r'_{max}/e'_{max}$ is the magnitude of the largest expected setpoint change relative to the allowed control error. We will assume that $r_{max}(j\omega)$ is frequency dependent such that $|r_{max}(j\omega)| = R_{max}$ up to the frequency ω_r and is 0 above this frequency. In other words, for a setpoint change $r(t) = R_{max}\sin(\omega t)$, the tracking error $|e(t)|$ should be less than 1 up to the frequency w_r, and above this frequency there are no specifications on tracking. Throughout the paper we assume $R_{max} > 1$.

Summary of controllability rules for feedback control

Let ω_B denote the bandwidth of the system, $g(s)$ the process, and $g_m(s)$ the measurement device (i.e., the measured output is $g_m(s)y$). The following rules apply

Rule 1 *Speed of response to reject disturbances. Must at least require $\omega_B > \omega_d$.* Here ω_d is the frequency at which $|g_d(j\omega_d)|$ first crosses 1 from above.

Justification:. Without control $y = g_d d$. Scaling has been applied such that the largest disturbance at a given frequency is $d(t) = 1 \cdot \sin(\omega t)$ (i.e., $|d(j\omega)| = 1$). Thus, at frequencies $\omega < \omega_d$ the output y will be unacceptable ($|y| > 1$) for a disturbance $|d| = 1$, so control is needed at these frequencies, and we must require $\omega_B \geq \omega_d$.

More specifically, we must with feedback control require $|L| = |gc(j\omega)| > |g_d(j\omega)|$ at frequencies where $|g_d(j\omega)| > 1$. *Justification:* With feedback control $y = Sg_d d$ where $S \approx 1/L$ at frequencies where $|L| > 1$. Thus to have $|y| < 1$ for $|d| = 1$ we must require $|L| > |g_d|$.

Rule 2 *Speed of response to follow setpoints. Must at least require $\omega_B > \omega_r$ where ω_r is the frequency up to which tracking is required. More specifically, we must require $|L(j\omega)| > R_{max}$ up to frequency ω_r.*

Unless R_{max} is close to 1, the requirement $\omega_B > \omega_r$ is not tight, and a higher bandwidth is required in practice. The exact value depends on how sharply $|L(j\omega)|$ drops off in the frequency range from w_r (where $|L| > R_{max}$) to ω_B (where $|L| = 1$). For example, with $L(s) = \omega_B/s$ (first-order response) the required bandwidth is $\omega_B > \omega_r R_{max}$, while for $L(s) = \omega_B^2/s^2$ (not considering stability) the required bandwidth is $\omega_B > \omega_r \sqrt{R_{max}}$).

Justification: With feedback control $e = -Sr$ where $S \approx 1/L$ at frequencies where $|L| > 1$. Thus to have $|e| < 1$ for $|r| = |R_{max}|$ (up to frequency ω_r) we must require $|L| > |R_{max}|$.

Rule 3 *Input constraints for disturbances. Must require $|g(j\omega)| > |g_d(j\omega)|$ at frequencies where $|g_d(j\omega)| > 1$.* This is needed to avoid input constraints when perfectly rejecting a disturbance $d(t) = 1 \cdot sin(\omega t)$ (i.e., $d(j\omega) = 1$).

Justification. From $y = gu + g_d d = 0$ we get $u = -(g_d/g)d$ and with $d = 1$ we need $|u| = |g_d|/|g| < 1$ to avoid input constraints.

Strictly speaking, perfect control is not required, and the minimum input needed for "acceptable" control (namely $|y| < 1$) is $|u| = (|g_d| - 1)/|g|$. The difference is clearly small at low frequencies where $|g_d|$ is larger than 1. (However, for multivariable systems the difference may be large for ill-conditioned plants even at low frequencies).

Rule 4 *Input constraints for setpoints. Must require $|g(j\omega)| > R_{max}$ up to frequency ω_r where tracking is required.* This is needed to avoid input constraints for perfect tracking of $|r(j\omega)| = R_{max}$.

Justification. From $y = gu$ and $y = r$ (perfect control) we get $u = r/g$, and with $r = R_{max}$ (up to frequency ω_r) we need $|u| = R_{max}/|g| < 1$ to avoid input constraints.

Rule 5 *Time delay θ in $g(s)g_m(s)$. Must require $\omega_B < 1/\theta$ to have acceptable control performance.*

Justification. It is impossible to remove the effect of the delay and $L(s)$ must contain a term $e^{-\theta s}$. For example, the ideal controller which minimizes $J = \int_0^\infty |e(t) - r(t)|^2 dt$ when $r(t)$ is a step and there is no penalty on the inputs has complementary sensitivity $T = e^{-\theta s}$. The corresponding loop gain $L = T/(1-T)$ crosses 1 in magnitude at about the frequency $1/\theta$. In practice, the ideal controller cannot be realized so this value provides an upper bound on the bandwidth.

Rule 6 *Real RHP-zero z in $g(s)g_m(s)$. Must require $\omega_B < z/2$ to have acceptable control performance at low frequencies.*

Justification. Again, it is impossible to remove the effect of a RHP zero. The ideal controller which minimizes $J = \int_0^\infty |e(t) - r(t)|^2 dt$ when $r(t)$ is a step and there is no penalty on the inputs has complementary sensitivity $T = \frac{-s+z}{s+z}$. The corresponding loop gain $L = T/(1-T)$ crosses 1 in magnitude at about the frequency $z/2$. In practice, the ideal controller cannot be realized so this value provides an upper bound on the bandwidth.

Remark. Strictly speaking, a RHP-zero only makes it impossible to have tight control in the frequency range close to the location of the RHP-zero. If we do not need tight control at low frequencies, then we may reverse the sign of the controller gain, and instead achieve tight control at frequencies higher than z. One special example is for plants with a zero at the origin ($g(s)$ contains an isolated term s in the numerator) where one can achieve good transient control, but where there the control has no effect at steady-state.

Rule 7 *Phase lag constraint. Must require in most practical cases: $\omega_B < \omega_u$. Here the "ultimate" frequency ω_u is where the phase of $g(j\omega)g_m(j\omega)$ is -180^o.*

This rule is given by Balchen and Mumme (1988), but without any theoretical justification. In fact, the condition is *not* a fundamental limitation, since for minimum phase plants (no delays or RHP-zeros), any phase lag may in theory be counteracted (disregarding input constraints) by placing zeros in the controller (use of "derivative action"). However, in practice this is not possible, because the controller structure may be limited and because of model uncertainty.

Justification for PID-controller. With a PID-controller the maximum phase lead is 54.9^o for a controller with derivative action over one decade (the maximum phase lead for the term $\frac{\tau_d s+1}{0.1\tau_d s+}$ is 54.9^o at frequency $\sqrt{10}/\tau_d$). Thus, if we require a phase margin larger than 54.9^o we must require $|L| \leq 1$ at frequency w_u and the rule follows.

Rule 8 *Real open-loop unstable pole in $g(s)$ at $s = p$. We need high feedback gains to stabilize the system and must approximately require $\omega_B > p$.*

Justification. For example, to stabilize a plant $g(s) = 1/(s-p)$ with a constant gain controller $c(s) = K_c$ we need $K_c > p$, and we find that the asymptote of $|L|$ crosses 1 at frequency K_c, so we have $\omega_B > p$. Another justification follows from the fact that a strictly proper plant with a single unstable real pole (e.g., $g(s) = \frac{s-z}{(s-p)(\epsilon s+1)}$) can be stabilized by a stable controller if and only if $p < z$ (Youla et al., 1974).

In addition, for unstable plants we need $|g| > |g_d|$ up to the frequency $\omega_b > p$ (which may be larger than w_d). Otherwise, the input may saturate when there are disturbances, and the plant cannot be stabilized.

The above rules are necessary conditions ("minimum requirements") in order to achieve acceptable control performance. One reason they are not sufficient is that they are based on considering only "one effect at a time".

The rules quantify the qualitative rules from Seborg et al. (1989) given in the introduction. For example, the rule "Control outputs that are not self-regulating" may be quantified as: "Control outputs y for which $|g_d(j\omega)| > 1$ at some frequency" (Rule 1). The rule "Select inputs that have a large effect on the outputs" may be quantified as: "In terms of scaled variables we must have $|g| > |g_d|$ at frequencies where $|g_d| > 1$ (Rule 3), and we must have $|g| > R_{max}$ at frequencies where setpoint tracking is desired (Rule 4)".

Another important insight from the above rules is that a larger disturbance or a smaller allowed control error requires faster response (higher bandwidth).

In summary, Rules 1, 2 and 8 tell us that we need high feedback gain ("fast control") in order to reject disturbances, to track setpoints and to stabilize the plant. On the other hand, Rules 5, 6 and 7 tell us that we must use low feedback gains in the frequency range where there are RHP-zeros or delays or where the plant has a lot of phase lag. We have formulated these requirements for high and low gain as bandwidth requirements. If they somehow are in conflict then the plant is not controllable and the only remedy is to introduce design modifications. Often the problem is that the disturbances are too large such that we hit input constraints, or such that the required bandwidth is not achievable. To avoid the latter problem, we must at least require that the effect of the disturbance is less than 1 (in terms of scaled variables) at frequencies beyond the bandwidth, that is,

$$|g_d(j\omega)| < 1; \quad \omega \geq \omega_B \tag{6}$$

where as found above we must require (approximately) $\omega_B < 1/\theta$ and $\omega_B < z/2$ and $\omega_B < \omega_u$. Condition (6) may be used, as shown below, to determine the size of buffer tanks.

Feedforward control

Consider a feedforward controller $u = c_f(s)d_m$ where $d_m = g_{md}(s)d$ is the measured disturbance. The disturbance response with the feedforward controller in place is

$$y = g(s)u + g_d(s)d = \underbrace{(g(s)c_f(s)g_{md}(s) + g_d(s))}_{\hat{g}_d} d \tag{7}$$

We want to consider controllability (achievable performance) with feedforward control.

Rules 3 and 4 on input constraints apply directly to feedforward control, while Rule 8 does not apply since unstable plants can only be stabilized by feedback control. The remaining rules make use of the term "bandwidth" which we above defined as the frequency up to which the feedback loop gain $|L|$ is larger than one. However, if the term "bandwidth" (ω_B) is interpreted as "the frequency up to which control is effective" then the rules partly apply also to feedforward control. Rules 5 and 6 on time delay and RHP-zero must be modified by replacing $g(s)g_m(s)$ by $g_d(s)^{-1}g(s)g_{md}(s)$. This follows by considering the ideal feedforward controller which yields $\hat{g}_d = 0$ in (7). We get

$$c_f^{ideal}(s) = -g_d(s)g^{-1}(s)g_{md}^{-1}(s) \tag{8}$$

which should be stable and causal (contain no prediction) to be realizable. Note that a delay in $g_d(s)$ is an advantage for feedforward control (it gives the feedforward controller more time to make the right action).

Model uncertainty is a more serious problem for feedforward than for feedback control because there is no output measurement. Let the actual plant models be denoted g', g'_d and g'_{md}. Then the actual disturbance response with the ideal feedforward controller in (8) is (assuming that this controller is realizable)

$$y = g'u + g'_d d = \underbrace{g'_d\left(1 - \frac{g_d}{g'_d}\frac{g'}{g}\frac{g'_{md}}{g_{md}}\right)}_{\hat{g}'_d} d \tag{9}$$

The effectiveness of feedforward control is determined by the ratio $|\hat{g}'_d|/|g'_d|$. Ideally it is zero, but this requires accurate models of g and g_d as well as an accurate measurement. For example, a 10% error in each of these three may yield $|\hat{g}'_d|/|g'_d| = |1 - 1.1 \cdot 1.1 \cdot 1.1| = 0.33$, that is, because of uncertainty the ideal feedforward controller removes only 67% of the disturbance effect. If the ratio is larger than 1 at some frequency (which may easily happen) then feedforward control makes control worse.

Because of the sensitivity to model uncertainty and because of the presence of unmeasured disturbances, feedforward control is usually combined with feedback control. Assume that the feedforward controller has already been designed. Then the controllability of the remaining feedback problem can be analyzed using the above rules if $g_d(s)$ is replaced by $\hat{g}_d(s)$, where the latter denotes the effect of the disturbance with the feedforward controller in place. One should also include the expected model uncertainty when evaluating $\hat{g}_d$ as illustrated in Eq.9.

3 Application: Room heating

The objective of this section is to give a simple illustration of how controllability analysis may be applied to a practical example.

Consider the problem of maintaining a constant room temperature. A heat balance yields the following differential equation for the temperature T in the room

$$\frac{d}{dt}(C_V T) = Q + k(T_o - T) \tag{10}$$

Here Q [W] is the heat input, T_o is the outdoor temperature, and the term $k(T_o - T)$ [W] represents the heat loss due to heat conduction through the walls or due to inflow of fresh air [1]. Consider a case where the heat input Q is 2000W and the difference between indoor and outdoor temperature $T - T_o$ is 20K. Then the steady-state energy balance yields $k = 2000/20 = 100$ W/K.

Let the heat capacity be $C_V = 100$ kJ/K [2]. On introducing deviation variables and taking the Laplace transform we get

$$\Delta T(s) = \frac{1}{\tau s + 1}\left(\frac{1}{k}\Delta Q(s) + \Delta T_o(s)\right); \quad \tau = \frac{C_V}{k} \tag{11}$$

The time constant for this example is $\tau = 100 \cdot 10^3/100 = 1000\text{s} = 17$ min which seems reasonable (for a step increase in heat input it will take about 17 min for the temperature to reach 63% of its steady-state increase).

Problem statement. Feedback control should be used to maintain approximately constant room temperature. The measurement delay for T is $\theta = 100$s.

[1] The heat loss may be represented by $qc_P(T_o - T) + UA(T_o - T)$ where the first term represents the convective heat transfer (difference in energy of inflow and outflow of air) and the second term represents the heat loss through the walls and windows. Thus $k = qc_p + UA$, where q [kg/s] is the flowrate, c_p [J/ kg,K] is the heat capacity, U [W/m^2,K] is the heat transfer coefficient, and A [m^2] is the wall area.

[2] The value C_V=100 kJ/K corresponds approximately to the heat capacity of air in a room of about 100 m^3. Thus we neglect heat accumulation in the walls.

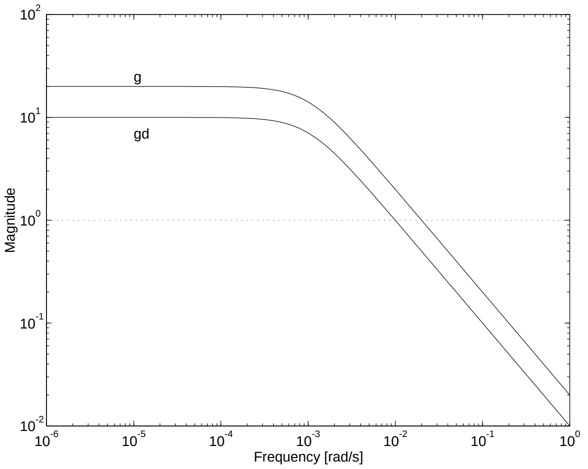

Figure 2: Frequency responses for room heating example

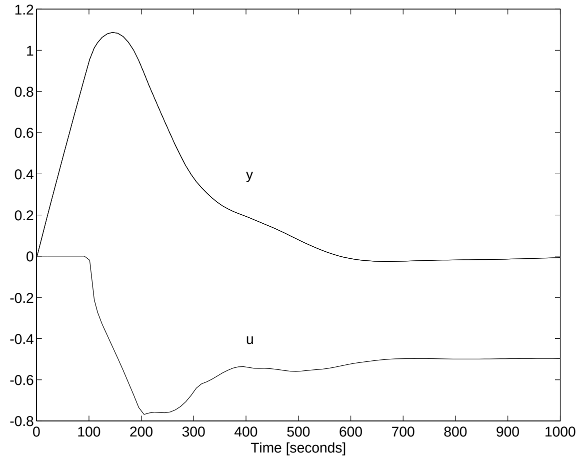

Figure 3: Feedback control for room heating example using PID controller. Step disturbance in outdoor temperature.

Assume the acceptable variations in room temperature are ± 1K, i.e., $T_{max} = 1$K. Furthermore, assume that heat input can vary between 0 W and 4000 W, i.e., the heat input is 2000 ± 2000 W so $Q_{max} = 2000$ W. Finally, the expected variations in outdoor temperature are ± 10 K, i.e, $T_{o_{max}} = 10$ K.

- Is the process controllable with respect to disturbances?
- Is the process controllable with respect to setpoint changes[3] of magnitude ± 3 K when the desired response time for setpoint changes is $\tau_r = 1000$ s (17min) ?

[3] The setpoint change may be due to a desired increase in temperature when we come home from work or get up in the morning.

Figure 4: Feedback control for room heating example using PID controller. Setpoint change $3/(150s+1)$.

Solution. A critical part of the controllability analysis is scaling, and we introduce the following scaled variables

$$y = \Delta T / 1\text{K}; \quad u = \Delta Q / 2000\text{W}; \quad d = \Delta T_0 / 10\text{K} \tag{12}$$

The model in terms of scaled variables then becomes

$$y = g(s)u + g_d(s)d \tag{13}$$

$$g(s) = \frac{20}{1000s+1}; \quad g_d(s) = \frac{10}{1000s+1} \tag{14}$$

The frequency responses of these transfer functions are shown in Fig. 2.

1. Disturbances. From Rule 1 feedback control is necessary up to the frequency $\omega_d = 10/1000 = 0.01$ rad/s, where $|g_d|$ crosses 1 in magnitude ($\omega_B > \omega_d$). This is exactly the same frequency as the upper bound given by the delay, $1/\theta = 0.01$ rad/s ($\omega_B < 1/\theta$). We therefore conclude that the system is barely controllable for this disturbance. From Rule 3 no problems with input constraints are expected since $|g| > g_d|$ at all frequencies. These conclusions are supported by the closed-loop simulation in Fig. 3 for a unit step disturbance (corresponding to a sudden 10 K increase in the outdoor temperature) using a PID-controller

$$c(s) = K_c \frac{1 + \tau_I s}{\tau_s} \frac{\tau_D s + 1}{0.1 \tau_D s + 1}$$

with $K_c = 0.4$ (scaled variables), $\tau_I = 200$s, $\tau_D = 60$s. The output error exceeds its allowed value of 1 for a very short time after about 100 s, but then returns quite quickly to zero. The input goes down to about -0.8 and thus remains within its allowed bound of ± 1.

2. Setpoints. The plant is also controllable with respect to the desired setpoint changes. First, the delay is 100s which is much smaller than the desired response time of 1000s, and thus poses no problem. Second, $|g(j\omega)| \geq R_{max} = 3$ up to about

$\omega = 0.007$ [rad/s] which is significantly higher than the required $w_r = 1/\tau_r = 0.001$ [rad/s]. This means that input constraints pose no problem. In fact, we should be able to achieve response times of about $1/0.007 = 150$s without reaching input constraints. This is confirmed by the simulation in Fig.4 for a desired setpoint change $3/(150s+1)$ using the same PID controller as above.

4 Design of buffer tanks.

Buffer tanks are frequently used in the process industry to dampen disturbances in temperature, concentration and flow. For "quality" (e.g., temperature and concentration) disturbances the idea is to dampen high-frequency disturbances by use of a well-mixed tank, and level control is not important. For flowrate disturbances the level control is used actively to dampen the disturbance and mixing is not important. Of course, it is possible to use the same tank for both kinds of disturbances - design of the tank must then be based on the most difficult disturbance from a control point of view.

Although buffer tanks are often introduced for control purposes, they are usually sized in a rather *ad hoc* manner without explicitly considering the expected disturbances and desired control objectives. Fortunately, the results on controllability with respect to disturbances presented in this paper, provide the basis for a quantitative approach.

To design the buffer tank consider the controllability of the plant when the disturbance transfer function $g_d(s)$ is replaced by

$$\hat{g}_d(s) = g_d(s)h(s) \tag{15}$$

where $h(s)$ represents the transfer function for the buffer tank(s). Presumably, the controllability is not acceptable without the buffer tank (i.e., with $h(s) = 1$), that is, the effect of the disturbance is too large such that, either the required speed of response is not achievable (typically due to a process delay θ), or the required inputs to reject the disturbance are too large.

The objective of the buffer tank is then to dampen the disturbance such that:

1. The required speed of response *is* achievable, for example, for a process delay θ we must require
$$|\hat{g}_d(j\omega_\theta)| \leq 1; \quad \omega_\theta \stackrel{\text{def}}{=} 1/\theta \tag{16}$$
More specifically, it must be possible to design a control system such that $|L(j\omega)| > |\hat{g}_d(j\omega)|$) for all frequencies where $|\hat{g}_d(j\omega)| > 1$. And conversely, if we have a control system with a given bandwidth ω_B, then we must have
$$|\hat{g}_d(j\omega)| \leq 1; \quad \forall \omega > \omega_B \tag{17}$$
where condition (16) is a special case of condition (17) which follows since $\omega_B < \omega_\theta$.

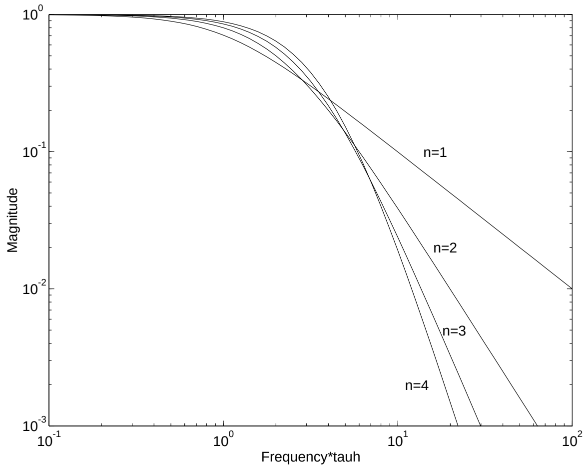

Figure 5: Frequency responses for n tanks in series with total residence time τ_h, $h_n(s) = 1/(\frac{\tau_h}{n}s+1)^n$.

2. Input constraints cause no problem, that is
$$|g(j\omega)| \leq |\hat{g}_d(j\omega)|, \quad \forall \omega \text{ where } |\hat{g}_d(j\omega)| > 1 \tag{18}$$
where $\hat{\omega}_d$ is the frequency where $|\hat{g}_d(j\hat{\omega}_d)| = 1$.

That is, $h(s)$ should be selected such that requirements (17) and (18) are satisfied. Although this is rather straightforward, we consider it in some detail because it yields some rather interesting results.

We shall first consider design of buffer tanks for "quality" disturbances (temperature and concentration) and then consider flow rate disturbances. The main difference between these cases is that for quality disturbances $h(s)$ has to be a series of first-order lags, whereas for flowrate disturbances one may use the level controller to get a desired $h(s)$.

4.1 Quality disturbances.

Consider a tank with constant volume V [m^3] and with an inlet and outlet flowrate q [m^3/s]. Let c_{in} denote the inlet concentration or temperature to the tank, and c the corresponding value in the outlet stream. A material or energy balance for a perfectly mixed tank yields

$$V\frac{dc}{dt} = qc_{in} - qc \tag{19}$$

The transfer function for one tank then becomes

$$c(s) = h_1(s)c_{in}(s); \quad h_1(s) = \frac{1}{\tau_h s + 1} \tag{20}$$

where $\tau_h = V/q$ [s] is the residence time in the tank (the subscript h denotes holdup). For n equal tanks in series with total residence time τ_h and total volume V, $h_1(s)$ is replaced by

$$h_n(s) = 1/(\frac{\tau_h}{n}s+1)^n \tag{21}$$

Typical frequency responses are presented in Figure 5. We have that $|h_n(j\omega)| \approx 1$ at frequencies $w <$

n/τ_h, that is, the buffer tanks are only effective for reducing the effect of disturbances at frequencies higher than the residence time of the individual tanks. The high-frequency asymptote is $|h(j\omega)| \approx (\frac{n}{\tau_h})^n$ so for rejecting high-frequency disturbances it is best to use many tanks if the objective is to minimize the total volume (actually, at intermediate frequencies we see from Fig.5 that there is a small frequency range where fewer tanks is slightly better). However, in practice we want to reduce the total cost and not the total volume, and this favors using fewer tanks.

The design of the buffer tank(s) now depends on which of the two requirements (17) and (18) is most difficult to satisfy.

The speed of response is the problem.

In this case the disturbance effect, $|g_d|$, is too large such that it is not possible to make $|L| > |g_d|$ at all frequencies where $|g_d|$ is larger than 1.

In most cases this means that $|g_d|$ is larger than 1 at frequencies beyond the (achievable) bandwidth, ω_B. Thus, in this case the objective of the buffer tanks is to make the curve for $|\hat{g}_d|$ cross 1 in magnitude before the bandwidth frequency ω_B. Thus, we need to select $h(s)$ such that

$$|h_n(j\omega_B)| \cdot |g_d(j\omega_B)| \leq 1 \qquad (22)$$

Introduce the factor by which the effect of the disturbance must be reduced

$$f = |g_d(j\omega_B)| \qquad (23)$$

We must *at least* require $|h_n(j\omega_B)| = 1/f$. This may be solved graphically using Fig.5. Alternatively, for n equal tanks in series Eq.(21) yields the required total residence time

$$\tau_h = n\sqrt{f^{2/n} - 1}/\omega_B \qquad (24)$$

The optimal number of tanks can then be found by taking into account cost for equipment, piping, control systems (each tank may require a level controller), etc.

If the bandwidth is limited by a delay we have

$$\tau_h = n\theta\sqrt{f^{2/n} - 1} \qquad (25)$$

where θ is the total delay in the feedback loop. As an example, for $f = 10$ we get

No. of tanks, n	1	2	3	4
Total residence time, τ_h	9.94θ	6.00θ	5.72θ	5.88θ

In this case the smallest total volume is obtained with 3 tanks, but with 2 tanks the required volume is only 4% larger and is clearly preferable. In practice one would probably prefer to use only 1 tank which has 66% larger total volume, but which saves additional equipment.

Remarks.

1. From (24) we find for large values of f (i.e., $f^{2/n} >> 1$) the following limiting value for the total residence time

$$\tau_h \approx n\theta f^{1/n} \qquad (26)$$

Thus, with one tank the residence time should be approximately equal to θf.

2. For many buffer tanks the resulting transfer function $\hat{g}_d(s) = g(s)h(s)$ may be of high order, and it may be difficult to have sufficiently high rolloff in the loop transfer function $L(s)$ to get $|L(j\omega)| > |g_d(j\omega)$ at frequencies lower than the bandwidth (although we are able to achieve this *at* the bandwidth, i.e., at $\omega = \omega_B$). The problem is that a high roll-off in $L(s)$ yields a large phase lag, and we get stability problems. This implies that the above analysis, which was based on considering the frequency ω_B only, may be optimistic, in particular when there are many tanks in series (n is large). For example, with $n = 2$ a step disturbance may affect the output in a parabolic fashion (increases proportionally to t^2 initially), or possible even worse, and we understand intuitively that it may be difficult for the controller to react sufficiently fast.

Constraints is the problem

In this case the disturbance effect, $|g_d|$, is too large such that we have $|g| < |g_d|$ at some frequencies where $|g_d|$ is larger than 1.

First, the buffer tanks do not affect the steady-state, so we must require that $|g(0)| > |g_d(0)|$. Now, consider higher frequencies and let ω_e be the frequency where $|g| = |g_d| > 1$. The objective of the buffer tanks in this case is to make $|\hat{g}_d|$ smaller than $|g|$ in the frequency range from ω_e to $\hat{\omega}_d$.

The following procedure may be used to achieve this: Let $n_e > 0$ be the difference in the slope of $|g_d|$ and $|g|$ (on a log-log plot) at frequency ω_e. Assume that the difference in slopes remains constant or decreases in the frequency range from ω_e to the frequency where $|g| = 1$. Select the number of tanks n equal to n_e, and select the holdup of the individual tanks as $1/\omega_e$, that is, select the overall residence time as $\tau_h = n_e/\omega_e$.

Example. Let

$$g(s) = \frac{200}{(55900s + 1)(89.4s + 1)}$$

$$g_d(s) = \frac{100(5000s + 1)}{(55900s + 1)(89.4s + 1)}$$

We find in this case $\omega_e = \frac{200}{100 \cdot 5000} = 0.0004$ [rad/s] (using asymptotic values) and $n_e = 0 - (-1) = 1$, and the difference in slopes remains constant at high frequencies. To reduce the effect of the disturbance to an acceptable level such that input constraints are avoided, we then need $n_e = 1$ buffer tank with residence time $1/\omega_e = 2500\text{s} = 0.7\text{h}$.

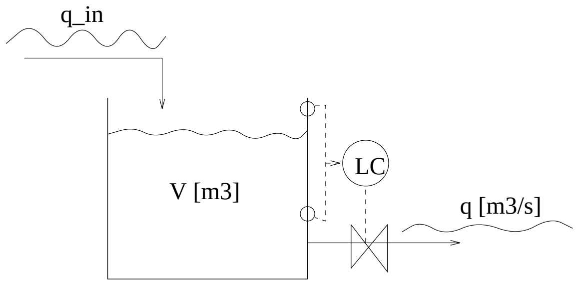

Figure 6: Use of slow level control to dampen flowrate disturbance.

Problems for the reader.

1. The effect of a concentration disturbance must be reduced by a factor of 100 at the frequency 0.5 rad/min. The disturbances should be dampened by use of buffer tanks and the objective is to minimize the total volume. How many tanks in series should one have? What is the total residence time?
2. The feed to a distillation column has large variations in concentration and the use of one buffer tank is suggest to dampen these. The effect of the feed concentration d on the product composition y is given by (scaled variables, time in minutes)
$$g_d(s) = e^{-s}/3s \qquad (27)$$
(that is, after a step in d the output y will, after an initial delay of 1 min, increase in a ramplike fashion and reach its maximum allowed value (which is 1) after another 3 minutes). Feedback control should be used and there is a additional measurement delay of 5 min. What should the residence time in the tank be?
3. Show that in terms of minimizing the total volume it is optimal to have buffer tanks of equal size.
4. Is there any reason to have buffer tanks in parallel (they must not be of equal size because then one may simply combine them) ?
5. What about parallel pipes in series (pure delay). Is this a good idea?

4.2 Flow rate disturbances

Flowrate disturbances may be dampened by use of a slow level controller as illustrated in Fig. 6. Let V [m^3] denote the volume of the buffer tank and let q_{in} and q [m^3/s] be the inlet and outlet flowrates. The dynamic model for the tank and the level control system is

$$V(s) = \frac{1}{s}(q_{in}(s) - q(s)); \quad q(s) = c(s)V(s) \qquad (28)$$

where $c(s)$ is the transfer function of the level controller (including measurement and actuator devices). We get

$$V(s) = \frac{1}{s + c(s)} q_{in}(s) \qquad (29)$$

and the transfer function of interest becomes

$$q(s) = h(s)q_{in}(s); \quad h(s) = \frac{c(s)}{s + c(s)} \qquad (30)$$

For flowrate disturbances we have more freedom in selecting $h(s)$ (as compared to quality disturbances) because we can select the algorithm for the level controller, $c(s)$. On the other hand, the level will vary so the size of the tanks must be such that the level does not reach constraints. The design of a buffer tank for flowrate disturbances then consists of two steps

1. Design the level *controller* $c(s)$ such that $h(s)$ has the desired shape, that is, such that (17) and (18) are satisfied.
2. Design the *size of the tank* such that the level remains within the allowed range for the expected disturbances.

First-order filtering

In many cases the desired $h(s)$ has the shape

$$h(s) = 1/(\tau s + 1) \qquad (31)$$

and we see from (30) that the required controller is a P-controller with gain $K_c = 1/\tau$. The response for the volume in the tank is given by (29), that is, we get $V(s) = \frac{\tau}{\tau s+1} q_{in}(s)$. This transfer function has its largest value equal to τ at low frequencies, and if the inlet flowrate varies within its full range $\pm q_{in}$ ($\pm 100\%$), we get that the volume will vary within $\pm \tau q_{in}$. This is in terms of deviation variables, and the total volume of the tank should be $2\tau q_{in}$. We then find, as one probably may expect, that the nominal residence time in the tank, τ_h, should be equal to the desired filter time constant τ.

Remark. In some cases one may want to add a slow integral action to the controller to reset the volume (level) to its nominal value, but this is not always desired. For example, if q_{in} is at its maximum value, then we may want V to stay at a large value to anticipate a possible large reduction in q_{in}.

Second-order filtering

Let the desired $h(s)$ have the shape $h(s) = \frac{1}{(\tau_2 s+1)^2}$. We get from (30) that the required controller is a lag

$$c(s) = 1/(2\tau_2)(\frac{\tau_2}{2}s + 1) \qquad (32)$$

The response for the volume in the tank is given by (29), and we get that the transfer function $V(s)/q_{in}(s)$ has its largest value equal to $2\tau_2$ at low frequencies. If the inlet flowrate varies within its full range $\pm q_{in}$ ($\pm 100\%$), we get that the volume will vary within $\pm 2\tau_2 q_{in}$. This is in terms of deviation variables, and the total volume of the tank should be $4\tau q_{in}$. We then find that the nominal residence time in the tank, τ_h, should be equal to $2\tau_2$.

An equivalent way of getting second-order filtering is to use two tanks, each with a nominal residence

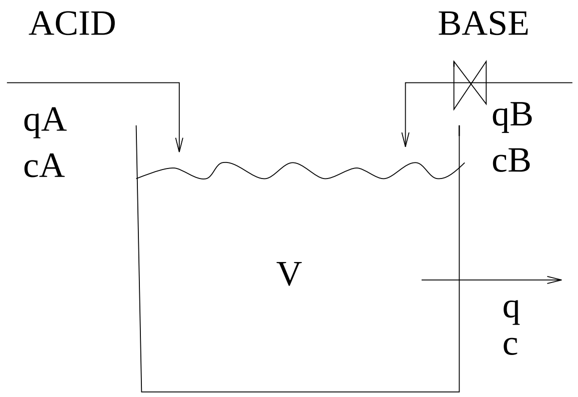

Figure 7: pH-neutralization process.

time of τ_2, and each with a P-controller with gain $1/\tau_2$. Clearly, this is not a good approach since we need two tanks with two level control systems, rather than one tank twice the size with one level control system. The only disadvantage in the latter case is that we need to use a slightly more complicated controller, as given by (32).

First-order versus second-order filtering

We have so far not discussed what shape $h(s)$ should have (first-order, second-order, etc.). The choice of $h(s)$ parallels the discussion we had in the subsection on quality disturbances, except that now we can get a high-order $h(s)$ with only one tank. This will favor the use of a high-order $h(s)$, at least for large values of f. However, as mentioned in Remark 2 in the section on quality disturbances, we must also take into account that if $h(s)$ is high order then we may have problems of achieving $|L| > |\hat{g}_d|$ at frequencies below the bandwidth.

5 Neutralization process

The derived controllability results are next applied to a neutralization process, and we find that more or less heuristic design rules given in the literature follow directly. The key point is to consider disturbances and scale the variables properly. The idea for this example came from the thesis of Walsh (1993).

One mixing tank. Consider the process in Figure 7 where a strong acid (pH= -1) is neutralized by a strong base (pH=15) in a tank with volume V= 10 m^3 to make q=0.01 m^3/s of "salt water". The pH in the product stream is adjusted to be in the range 7 ± 1 ("salt water") by manipulating the amount of base, q_B. The delay for the measurement of pH is $\theta = 10$s. Details about the dynamic model are given in Appendix 2. Introduce the excess of acid c [mol/l] defined as

$$c = c_H - c_{OH} \qquad (33)$$

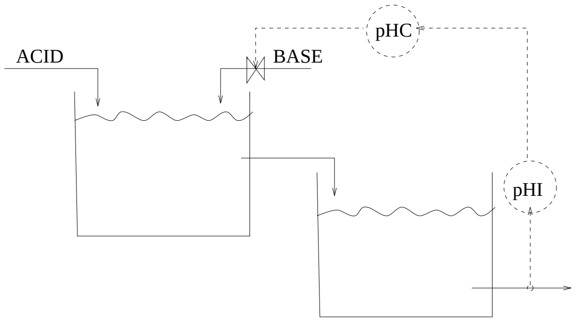

Figure 8: Control of neutralization process using two tanks.

Somewhat surprisingly, we find that in terms of c the dynamic model, which is usually believed to be strongly nonliner, is given by that of a simple mixing process

$$\frac{d}{dt}(Vc) = q_A c_A + q_B c_B - qc \qquad (34)$$

Introduce the following scaled variables

$$y = \frac{c}{10^{-6}}; \quad u = \frac{q_B}{q_B^*}; \quad d = \frac{q_A}{0.5 q_A^*} \qquad (35)$$

where superscript * denotes the steady-state value. The appropriately scaled linear model then becomes (see Appendix 2)

$$y = \frac{k_d}{1+\tau s}(-2u + d); \quad k_d = 2.5 \cdot 10^6 \qquad (36)$$

where $\tau = V/q = 1000$s. The output is extremely sensitive to both u and d and the large gain is easily explained: A change $d = 1$ corresponds to a 50% increase in the amount of acid which has a concentration of 10 mol/l of H^+ (pH=-1). This increases the amount of H^+ in the product from 0 to 2.5 mol/l, while the largest allowed amount of H^+ in the product is 10^{-6} mol/l (pH=6), thus the gain in terms of scaled variables is $k_d = 2.5/10^{-6}$.

Input constraints do not pose a problem since $|g| = 2|g_d|$ at all frequencies (Rule 3). T he main control problem is the high disturbance sensitivity, and from Rule 1 we find the frequency up to which feedback is needed

$$\omega_d \approx k_d/\tau = 2500 \text{ rad/s} \qquad (37)$$

This requires a response time of $1/2500 = 0.4$ millisecond. However, there is a delay $\theta = 10s$ so the bandwidth must be less than $\omega_B < 1/\theta = 0.1$ rad/s. From the controllability analysis we therefore conclude that acceptable control using a single tank is impossible.

Design change: Several tanks. The only way to improve the controllability is by design changes. The most useful change in this case is to do the neutralization in several steps. This can be considered as a special case of the buffer tank example considered above: The acid and base is mixed and is then send to one or more buffer tanks, and the measured pH of the final stream is used to adjust the addition of base, as is illustrated with two tanks in Figure 8. The mixing process itself is assumed immediate so in the following

$$g_d(s) = k_d \tag{38}$$

and the objective is to find an appropriate $h(s) = 1/(\tau s + 1)^n$ such that the "new" process $\hat{g}_d(s) = g_d(s)h(s)$ has acceptable controllability in terms of having acceptable self-regulation. Here τ is the residence time in each tank.

As noted the main control problem is a delay of θ= 10s, so we must design buffer tanks which reduces the effect of the disturbance by a factor $f = |g_d(j\omega_{1/\theta})| = k_d = 2.5 \cdot 10^6$ at frequency $\omega_\theta = 1/\theta$= 0.1 [rad/s]. The required total residence time, $\tau_h = n\tau$, is given by Eq.(24), and the corresponding total volume is

$$V = \tau_h q \tag{39}$$

where $q = 0.01$ m^3/s. From this we find that the following designs have the same controllability with respect to disturbance rejection:

No. of tanks n	Total volume V $[m^3]$	Volume each tank $[m^3]$
1	250000	250000
2	316	158
3	40.7	13.6
4	15.9	3.98
5	9.51	1.90
6	6.96	1.16
7	5.70	0.81
18	3.66	0.20
30	3.89	0.13

With one tank we need a volume corresponding to that of the worlds largest ship to get acceptable controllability. The minimum total volume is obtained with 18 tanks of about 203 l each - giving a total volume of 3.662 m^3. However, taking into the account the additional cost for extra equipment such as piping, mixing and level control, we would probably select a design with 3 or 4 neutralization tanks for this example.

Remarks.

1. Further remarks on some of the practical aspects and comparison with previous work are found in Skogestad (1994).

2. The use of several mixing tanks in series can be compared to playing golf: It is almost impossible to hit the hole with one stroke, but with 5 strokes or more almost anyone can do it.

3. Traditionally, a "feedforward" approach has been taken when considering controllability of such processes, and one key argument has been that control is difficult because on needs to adjust the amount of base extremely accurately to counteract the disturbance in the acid. This is a valid argument for feedforward control, but *not* for feedback control as the feedback control action will be able to adjust the input accurately. As demonstrated above the key problem for feedback control is that the output is extremely sensitive to disturbances (k_d and ω_d are large), which requires an extremely high bandwidth.

4. Of course, feedforward control based on measuring q_A and c_A can be used in addition to feedback to improve performance. According to McMillan (1984) one can typically save one buffer tank using a well designed feedforward controller.

5. The results given above compare well with results by other authors. A simple shortcut method given by McMillan (1984) is to use about one mixing tank for each 2 units change in pH. For example, with a pH change of 8, as in our example (from pH 15 to 7), four tanks is recommended.

6. For many tanks in series it is difficult with the control system in Figure 8 (one pH-controller) to achieve $|L| > |\hat{g}_d|$ at frequencies below the bandwidth, because $\hat{g}_d$ drops sharply with frequency. To get away from this problem one may select to control the pH in each tank as is often done in practice.

6 Conclusions

Although buffer tanks are often introduced for control purposes, they are usually sized in a rather *ad hoc* manner without explicitly considering the expected disturbances and desired control objectives. The simple results on controllability analysis with respect to disturbances presented above provide the basis for a quantitative approach.

To design the buffer tank consider the controllability of the plant when the disturbance transfer function $g_d(s)$ is replaced by $\hat{g}_d(s) = g_d(s)h(s)$ where $h(s)$ represents the transfer function for the buffer tank (including the level control). That is, $h(s)$ should be selected such that requirements (17) and (18) are satisfied.

We also discussed in detail a ph-example where the neutralization must be performed in several tanks to get acceptable controllability.

The tools presented in this paper may be generalized to multivariable plants where directionality becomes a further crucial consideration. Some results are given in Wolff et al. (1992) and Skogestad and Wolff (1992). A direct generalization to decentralized control of multivariable plants is given by Hovd and Skogestad (1992).

Acknowledgement. Manfred Morari was the first to consider a rigorous approach to controllability analysis. He also directed me to the paper of Ziegler and Nichols (1943) who first introduced the

term controllability in the control literature.

References

[1] Balchen, J.G. and K. Mumme, 1988, *Process Control. Structures and Applications*, Von Nostrand Reinhold, New York.

[2] Hovd, M. and S. Skogestad, 1992, "Simple frequency domain tools for control structure analysis, structure selection and design", *Automatica*, **28**, 989-996.

[3] McMillan, G.K., 1984, *pH Control*, Instrument Society of America.

[4] Morari, M., 1983. "Design of resilient processing plants III, A general framework for the assessment of dynamic resilience", *Chem. Eng. Sci.*, 38, 1881-1891.

[5] Rosenbrock, H.H., 1970. *State-Space and Multivariable Theory*, Nelson, London.

[6] Seborg, D.E., T.F. Edgar and D.A. Mellichamp, 1989, *Process Dynamics and Control*, Wiley, New York.

[7] S. Skogestad and E.A. Wolff, 1992, "Controllability measures for disturbance rejection", *Preprints IFAC workshop on Interactions between process design and process control, London, Sept. 1992*, Edited by J.D. Perkins, Pergamon Press, 23-29.

[8] Walsh, S., 1993, *Integrated design of chemical waste water treatment systems*, Ph.D. Thesis, Imperial College, UK.

[9] E.A. Wolff, S. Skogestad, M. Hovd and K.W. Mathisen, 1992, "A procedure for controllability analysis", *Preprints IFAC workshop on Interactions between process design and process control, London, Sept. 1992*, Edited by J.D. Perkins, Pergamon Press , 127-132.

[10] Youla, D.C., J.J. Bongiorno and C.N. Lu, 1974, "Single-loop Feedback Stabilization of Linear Multivariable Dynamical Plants", *Automatica*, **10**, 159-173.

[11] Ziegler, J.G. and N.B. Nichols, 1943, "Process Lags in Automatic Control Circuits", *Trans. ASME*, **65**, 433-444.

Appendix. Neutralization model

Derivation of model: Consider Fig.7. Let c_H [mol/l] and c_{OH} [mol/l] denote the concentration of H^+ and OH^--ions, respectively. Material balances for these two species yield

$$\frac{d}{dt}(Vc_H) = q_A c_{H,A} + q_B c_{H,B} - q c_H + rV$$

$$\frac{d}{dt}(Vc_{OH}) = q_A c_{OH,A} + q_B c_{OH,B} - q c_{OH} + rV$$

where r [mol/s,m^3] is the rate for the reaction $H_2O = H^+ + OH^-$ which for completely dissociated ("strong") acids and bases is the only reaction in which H^+ and OH^- participate. We may eliminate r from the equations by taking the difference to get a differential equation in terms of the excess of acid, $c = c_H - c_{OH}$:

$$\frac{d}{dt}(Vc) = q_A c_A + q_B c_B - qc$$

This is the material balance for a mixing tank without reaction. The reason is that the quantity $c = c_H - c_{OH}$ is not affected (invariant) by the reaction. Note that c will take on negative values when pH is above 7.

We are not interested in variations in the feed concentrations, c_A and c_B, so they are assumed constant. Linearization and Laplace transformation yields

$$c(s) = \frac{1}{1+\tau s}\left(\frac{c_A^* - c^*}{q^*} q_A(s) + \frac{c_B^* - c^*}{q^*} q_B(s)\right)$$

where $\tau = V/q^*$ is the residence time and * is used to denote steady-state values. To derive this we have made use of the total material balance $dV/dt = q_A + q_B - q$ (alternatively one may assume V is constant but this is not strictly necessary) and the corresponding steady-state balance $q_A^* + q_B^* = q^*$. We now introduce the following scaled variables

$$y(s) = \frac{c(s)}{c_{max}}; \quad d(s) = \frac{q_A(s)}{q_{Amax}}; \quad u(s) = \frac{q_B(s)}{q_{Bmax}}$$

and get

$$y(s) = \frac{1}{\tau s + 1}\left(\underbrace{\frac{c_A^* - c^*}{c_{max}} \cdot \frac{q_{Amax}}{q^*}}_{k_d} d(s) + \underbrace{\frac{c_B^* - c^*}{c_{max}} \cdot \frac{q_{Bmax}}{q^*}}_{k} u(s)\right)$$

We use the following numbers: V= 10 m^3, $q_A^* = q_B^* = 0.005$ m^3/s, $q^* = 0.01$ m^3/s, $c_{H,A}^* = 10$ mol/l (corresponding to pH= -1 and $c_A = 10 - 10^{-15} \approx 10$ mol/l), $c_{OH,B}^* = 10$ mol/l (corresponding to pH=15 and $c_B^* = 10^{-15} - 10 \approx -10$ mol/l), $c^* = 0$ mol/l (corresponding to pH=7), $c_{max} = 10^{-6} - 10^{-8} \approx 10^{-6}$ mol/l (i.e., pH= 7 ± 1), and $q_{Amax} = q_A^*/2 = 0.0025$ m^3/s, $q_{Bmax} = q_B^* = 0.005$ m^3/s. Note from the latter that the largest disturbance is $\pm 50\%$ of q_A^*, while the largest input is $\pm 100\%$ of q_B^*. With these values we get $\tau = 1000$ s, $k_d = 2.5 \cdot 10^6$ and $k = -5 \cdot 10^6$.