Prof. Chang-Chieh Hang

Frequency Response Approach to Auto-Tuning and Adaptive Control in Industrial Process Control

Plenary Lecture

The Third Asian Control Conference

9:40-10:30, July 5, 2000

Auditorium

Shanghai Worldfield Convention Hotel

ASCC Asian Control Conference

FREQUENCY RESPONSE APPROACH TO AUTO-TUNING AND ADAPTIVE CONTROL IN INDUSTRIAL PROCESS CONTROL

C. C. Hang
Department of Electrical Engineering
National University of Singapore, Singapore 119260

Abstract On-line frequency response estimation based on relay oscillations or setpoint changes can be used to readily synthesize PID and other advanced controllers. This has led to the frequency response approach to auto-tuning and adaptive control. It is most relevant to industrial process control as it does not require much prior information about the process dynamics such as the exact order and the process dead time. The development of this practical approach for the auto-tuning and adaptive control of PID controllers for various types of single-variable processes is outlined in this paper. Their extensions to other advanced controllers and to multivariable systems are then reviewed. Simulations are used to substantiate and demonstrate the robustness and achievable performance of this approach to auto-tuning and adaptive control.

1. INTRODUCTION

The introduction of auto-tuning capabilities to PID controllers has shortened the time needed to commission control systems and to facilitate control optimization through regular retuning (Astrom and Hagglund, 1988; Hang and Sin, 1991; Hang *et. al.*, 1993a). The relay feedback auto-tuning method proposed by Astrom and Hagglund (Astrom and Hagglund, 1984a) was one of the first to be commercialized and has remained attractive owing to its simplicity and robustness (Astrom and Hagglund, 1988; Hang and Sin, 1991; Hagglund and Astrom 1991).

The original relay auto-tuning method is based on the estimation of one point on the Nyquist curve. It has been shown recently in a series of papers (Wang, Hang and Zou, 1997 a and b; Wang, Hang and Bi, 1997 c and d) that the relay oscillations could be further analyzed to generate estimates of other points on the Nyquist curve. This, when combined with one of the controller designs based on the frequency response information on the Nyquist curve, has given birth to a new approach for auto-tuning PID and other advanced controllers. The use of setpoint changes in place of relay oscillations extends the application of this approach to adaptive control in many practical situations. As the frequency response approach does not require much prior information about the process dynamics such as the exact order of the transfer function, the process dead time, presence of oscillatory poles, presence of non-minimum phase zeros, etc., it is most attractive to industrial process control where a general purpose controller is used. It is the intention of this paper to review these new developments of the frequency response approach.

The paper is organized as follows. In Section 2, the original relay feedback method is reviewed. The advantages and the limitations of the method are indicated. In Section 3, the frequency response estimation based on analysis of relay oscillations is highlighted. Refined controller tuning methods based on frequency response information are presented in Section 4. Extensions of the relay auto-tuning to processes with large dead-time and multivariable processes are summarised in Sections 5 and 6, respectively. Extensions of the frequency response approach to facilitate adaptive control are reviewed in Section 7. Concluding remarks are given in Section 8.

2. RELAY AUTO-TUNING

The majority of controllers used in industry are of the PID type. A large industrial process may have hundreds of these controllers. They have to be tuned individually to provide good and robust control performance (Astrom and Hagglund, 1984b). The tuning procedure, if done manually, is very tedious and time consuming; the resultant system performance depends mainly on the experience and the process knowledge the engineers have. It is recognized that in practice, many industrial control loops are poorly tuned. Automatic tuning is thus attractive to researchers and practicing engineers. By *automatic tuning* (or *auto-tuning*), we mean a method which enables the controller to be tuned automatically on demand from an operator or an external signal (Astrom and Hagglund, 1988; Astrom, *et. al.*, 1993). Industrial experience has clearly indicated that this is a highly desirable and useful feature.

Astrom and co-workers successfully applied the relay feedback technique to the auto-tuning of PID controllers for a class of common industrial processes (Astrom and Hagglund, 1988). The relay feedback auto-tuning technique has several attractive features. Firstly, it facilitates simple push-button tuning since the scheme automatically extracts the process frequency response at an important frequency and the information is usually sufficient to tune the PID controller for many simple processes. The method is time-saving and easy to use (Hang *et. al.*, 1993a). Secondly, the relay feedback auto-tuning test is carried out under closed-loop control so that with an appropriate choice of the relay parameters, the process can be kept close to the setpoint. This keeps the process in the linear region where the frequency response is of interest, which is precisely why the method works well on highly nonlinear processes (Astrom and Hagglund, 1988). Thirdly, unlike other auto-tuning methods, the technique eliminates the need for a careful choice of the sampling rate. This is very useful in initializing a more sophisticated adaptive controller (Lundh and Astrom, 1998). Fourthly, the relay feedback auto-tuning works well under disturbances and it is robust to process perturbation.

The critical point, *i.e.*, the process frequency response at the phase lag of π, has been employed to set the PID parameters for many years ever since the advent of the Ziegler-Nichols (Z-N) rule (Ziegler and Nichols, 1942). The point is traditionally described in terms of the ultimate gain k_u and the ultimate period T_u. The relay auto-tuning is based on the observation that a system with a phase lag of at least π at high frequency would oscillate with the period T_u under the relay control. To determine this critical point, the system is connected in a feedback loop as shown in Fig. 1.

Since the "describing function" of the relay is the negative real axis, the output $y(t)$ is then a periodic signal with the period T_u and the ultimate gain k_u is approximately given by (Astrom and Hagglund, 1984a; Hang and Astrom, 1988)

$$k_u = \frac{4d}{\pi a}, \tag{1}$$

where d is the relay amplitude and a is amplitude of the process output.

With the estimated information of the process critical point, the Z-N tuning rule or the modified Z-N rules (Astrom and Hagglund, 1988; Ziegler and Nichols, 1942; Hang *et. al.*, 1991a) can be used to tune the PID controller. The relay auto-tuning procedure is thus completed and the controller can be commissioned.

While the standard method is successful in many simple process control applications (Hagglund and Astrom, 1991; Astrom, *et. al.*, 1993), it also faces two problems. First, due to the adoption of the describing function approximation, the estimation of the critical point using the standard relay feedback method may not be accurate enough. Under some circumstances such as high order or long dead-time of the processes, the method could result in a significant error which would cause the system performance to deteriorate (Wang *et. al.*, 1997b). Second, only one frequency response point is obtainable from this method and it may be insufficient for describing the important dynamic characteristics of some processes or for designing advanced model based controllers.

3. FREQUENCY RESPONSE ESTIMATION FROM RELAY OSCILLATIONS

As mentioned above, the critical point estimation based on equation (1) is not always accurate. Furthermore, with the standard relay feedback auto-tuning, only one point on the process Nyquist curve is determined. For designing of the model-based controllers like Smith-Predictor (Plamor and Blau, 1994), (Hang *et. al.*, 1995), more points on the frequency response need to be extracted from the relay feedback experiment. It is possible, for example, to cascade a known linear dynamics to the relay in Fig. 1 to obtain a point other than the critical point. However, the testing time will increase proportionally when more frequency response point estimations are required, especially when high accuracy is required. This is particularly true when the process has a long dead-time.

It was shown in Hang *et al.* (Hang *et. al.*, 1995) that multiple points on the process frequency response could be obtained in a step test by first removing DC components from the input and output and then applying the Fourier Transform (FT) to the remaining signals. This has been further improved by Wang, *et al.* (1997c) who propose a method that can identify multiple points simultaneously under one relay test. For a standard relay feedback system in Fig. 1, the process input $u(t)$ and output $y(t)$ are recorded from the initial time until the system reaches a stationary oscillation. $u(t)$ and $y(t)$ are not integrable since they do not die down in finite time. They cannot be directly transformed to frequency response data meaningfully using FT. The solution is to introduce a modulation by a decay exponential $e^{-\alpha t}$ to form

$$\tilde{u}(t) = u(t)e^{-\alpha t}, \tag{2}$$

and

$$\tilde{y}(t) = y(t)e^{-\alpha t}, \tag{3}$$

such that $\tilde{u}(t)$ and $\tilde{y}(t)$ will decay to zero exponentially as t approaches infinity. Applying the Fourier Transform to (2) and (3) yields

$$\begin{aligned}\tilde{U}(j\omega) &= \int_0^\infty \tilde{u}(t)e^{-j\omega t}dt \\ &= \int_0^\infty u(t)e^{-\alpha t}e^{-j\omega t}dt = U(j\omega+\alpha),\end{aligned} \tag{4}$$

and

$$\begin{aligned}\tilde{Y}(j\omega) &= \int_0^\infty \tilde{y}(t)e^{-j\omega t}dt \\ &= \int_0^\infty y(t)e^{-\alpha t}e^{-j\omega t}dt = Y(j\omega+\alpha).\end{aligned} \tag{5}$$

For a process $G(s) = Y(s)/U(s)$, at $s = j\omega+\alpha$, one obtains

$$G(j\omega+\alpha) = \frac{Y(j\omega+\alpha)}{U(j\omega+\alpha)} = \frac{\tilde{Y}(j\omega)}{\tilde{U}(j\omega)}. \tag{6}$$

$\tilde{Y}(j\omega)$ and $\tilde{U}(j\omega)$ can be computed at discrete frequencies with the standard FFT technique (Bergland, 1969; Ramirez, 1985). Therefore, the shifted process frequency response $G(j\omega+\alpha)$

can be obtained from (6). To find $G(j\omega)$ from $G(j\omega+\alpha)$, we first take the inverse FT of $G(j\omega+\alpha)$ as

$$\tilde{g}(kT) := FFT^{-1}(G(j\omega+\alpha)) = g(kT)e^{-\alpha kT}. \quad (7)$$

It then follows that the process impulse response $g(kT)$ is

$$g(kT) = \tilde{g}(kT)e^{\alpha kT}. \quad (8)$$

Applying the FFT again to $g(kT)$ leads to the process frequency response:

$$G(j\omega) = FFT(g(kT)). \quad (9)$$

The method can identify accurate frequency response points as many as desired with one relay experiment. They may be very useful for improving performance of PID and model-based controllers. The required computations are more involved than the standard relay technique, especially if a large number of frequency response points are needed. The method has been applied to other non-decaying excitation test such as a step test and ramp test (Wang *et. al.*, 1997b). Note that the inverse FT computations of equations (7)-(9) are not needed in practice as the controller designs can be performed using the shifted frequency response data directly.

To illustrate the method, several different typical processes are considered. Fig. 2 shows the identified frequency responses for these processes using this method. The excellent results are self-explanatory.

4. REFINEMENT OF PID TUNING

In this section, we consider the tuning of a PID controller in the form of

$$u(t) = K_P(e_P + \frac{1}{T_I}\int_0^t edt + T_D\frac{de}{dt}) = K_P e_P + K_I\int_0^t edt + K_D\frac{de}{dt} \quad (10)$$

where $e_p = by_{sp} - y$, b is the setpoint weighting factor which is useful in reducing any large overshoot in the setpoint response. In the standard relay tuning case, the Z-N-like formulas are employed to tune PID controllers. These tuning rules are suitable only for those processes which can be accurately characterized by the critical point. To overcome this limitation, many modifications of the PID tuning rules have been reported.

4.1 Single-point Based Methods

It has been shown by Astrom, *et al.* (Astrom, *et. al.*, 1992) that for processes with monotone step responses, there exist quantities, such as the normalized dead-time and the normalized process gains, that are useful for assessing the achievable performance and choosing suitable controllers. Refined tuning formulae of the PID controller by incorporating heuristic knowledge of normalized dead-time to replace manual fine-tuning were developed (Hang *et. al.*, 1991a). A set of PI/PID controller tuning formulae for different normalized dead-time was given. They eliminate the need for manual fine-tuning and human expertise.

The above mentioned tuning rules depend on only one frequency point, the critical point, which may not be adequate to tune the PID controller to achieve an expected response. PID tuning rules that employ two or more points have thus been proposed.

4.2 Gain and Phase Margin Method

The gain and phase margins are very useful as measures of performance as well as robustness. Controller designs to satisfy gain and phase margin criteria are not new (Franklin *et. al.*, 1986). However, the solution is normally obtained by numerical methods or by trial and error using Bode plots. Such approaches are certainly not suitable for use in auto-tuning and adaptive control. The modified Ziegler-Nichols rule (Astrom and Hagglund, 1988) is a gain and phase margins tuning method. The solution is to achieve a compromise in phase and gain margins by moving the compensated Nyquist curve to pass through a specified design point. The method works well for processes with relatively small dead-time. When the dead-time is dominant, the actual phase margin may be very conservative although the prespecified gain margin is achieved.

An analytical method to tune the PID controller to pass through two design points on the Nyquist curve as specified by the gain margin A_m and phase margin ϕ_m was proposed (Ho *et. al.*, 1995). The method is based on the measurement of ultimate gain, ultimate period and the static gain of the process. For a first order plus dead-time process,

$$G(s)=\frac{Ke^{-Ls}}{Ts+1}, \tag{11}$$

the PI controller

$$K(s)=K_P(1+\frac{1}{sT_I}) \tag{12}$$

is given by

$$K_P=\frac{\omega_p T}{A_m K}, \tag{13}$$

$$T_I=\left(2\omega_p-\frac{4\omega_p^{\,2}L}{\pi}+\frac{1}{T}\right)^{-1}, \tag{14}$$

where

$$\omega_p=\frac{A_m\phi_m+\frac{1}{2}\pi A_m(A_m-1)}{(A_m^{\,2}-1)L}. \tag{15}$$

Example 1 Consider a process given by

$$G=\frac{1}{s+1}e^{-s}.$$

With the gain margin A_m and phase margin ϕ_m specified as 3 and $\frac{\pi}{3}$ respectively, the PI controller parameters in (13)-(14) are computed as

$$K_P=0.52$$

and

$$T_I=1.$$

For comparison, the PID controller tuned by the modified Z-N method (Astrom and Hagglund, 1988) is also considered, and given by $[K_P,T_I,T_D]=[0.80,2.38,0.59]$. The step responses obtained are shown in Fig. 5. The reason why the Ziegler-Nichols based method performs so poorly for this process is that the dead time is significant. It is well known that Ziegler Nichols methods give too small integral action in such cases.

The gain and phase margin method for the PI case can be extended to PID case using pole-zero cancellation. These simple PI/PID tuning formulae are particularly useful in the context of adaptive control and auto-tuning. The tuning method works well on processes that have the forms of (11). Simulation shows that the tuning rule, though simple, produces a much better system performance than the Z-N and modified Z-N tuning rules, especially for large dead-time processes.

4.3 Tuning via Frequency Response Fitting

A different but efficient solution to controller design using the estimated frequency response data was developed in (Wang *et. al.*, 1997b). It shapes the loop frequency response to optimally match the desired dynamics over a large range of frequencies.

Suppose that multiple process frequency response points $G(j\omega_i)$, $i=1,2,\cdots,m$, are available. The control specifications are formulated as a desirable closed loop transfer function

$$H_d = \frac{\omega_n^{\ 2}}{s^2 + 2\zeta\omega_n s + \omega_n^{\ 2}} e^{-Ls}, \tag{16}$$

where L is the apparent dead-time of the process, ω_n and ζ dominate the behaviour of the desired closed-loop response. The dead time can be read off from a process step response or be estimated by the knowledge-based approach (Lee, Wang and Tan, 1995). If the control specifications are given as the phase margin φ_m and gain margin A_m, ζ and ω_n in H_d are approximately determined by

$$\zeta = \sqrt{\frac{1-\cos^2\varphi_m}{4\cos\varphi_m}} \tag{17}$$

and

$$\omega_n = \frac{\tan^{-1}\left(\frac{2\zeta p}{p^2-1}\right)}{pL}, \tag{18}$$

where p is the positive root of equation

$$(A_m - 1)^2 = 4\zeta^2 p^2 + (1-p^2)^2. \tag{19}$$

The default settings for ζ and $\omega_n L$ values are $\zeta = 0.707$ and $\omega_n L = 2$, which imply that the overshoot of the objective set-point step response is about 5%, the phase margin is 60° and the gain margin is 2.2 (Wang *et. al.*, 1997b). The open-loop transfer function corresponding to H_d is

$$Q_d = \frac{H_d}{1-H_d}. \tag{20}$$

The design of the controller K is such that KG is fitted to Q_d in frequency domain as well as possible. Thus, the resultant system will have a desired performance.

For a PID controller in form of (10), we have

$$G(j\omega_i)\ \left[1\ \frac{1}{j\omega_i}\ \ j\omega_i\right]\ x = Q_d(j\omega_i), \quad i=1,2,\cdots,m, \tag{21}$$

where $x = [K_P\ \ \frac{K_P}{T_I}\ \ K_P T_D]^T$, m is chosen such that ω_m is bigger than the critical frequency of Q_d. Equation (21) can be rearranged into a set of linear equations. The least squares method can then be employed to obtain the PID parameters. If the solution satisfies the criterion

$$\max_i \left|G(j\omega_i)K(j\omega_i) - Q_d(j\omega_i)\right| \le \varepsilon, \tag{22}$$

where ε is the pre-specified fitting error threshold, then the design is finished. Otherwise, a high order controller may be considered or the control specifications may need to be relaxed. Then, we use the above procedure again to find a better fitting.

Simulations show that this is a simple and effective way of obtaining a desired response. The algorithm gives the optimal combination of PID settings that can achieve the desired transients. Its performance will be demonstrated by an example of an oscillatory process where other methods could not yield good results.

Example 2 Consider a plant with oscillatory dynamics:

$$G(s)=\frac{1}{s^2+0.2s+1}e^{-0.2s},$$

the PID controller designed by the method is

$$K(s)=0.59(1+\frac{1}{0.2s}+5.0s),$$

and the controller by the modified Z-N rule is

$$K(s)=0.36(1+\frac{1}{3.42s}+0.86s)$$

The system response is shown in Fig. 6.

5. PROCESSES WITH LONG DEAD-TIME

For a process with a long dead-time, a dead-time compensator is necessary for tight control. However, it requires a transfer function model. From the relay feedback test, the ultimate gain k_u and the ultimate frequency ω_u can be obtained. A primary PI controller tuned by Ziegler-Nichols formulae is then commissioned. With the system in closed-loop, the auto-tuner will wait for the next set point change to occur, and after the transient, the process static gain K can be calculated. It is well known that most of the industrial processes can be adequately approximated by a model of the form

$$G=\frac{K}{(Ts+1)^n}e^{-Ls}, \quad n=1,2. \tag{23}$$

The model can be recovered from k_u and ω_u by

$$T=\frac{\sqrt{(Kk_u)^{\frac{2}{n}}-1}}{\omega_u}, \tag{24}$$

$$L=\frac{(\pi-n\tan^{-1}(T\omega_u)}{\omega_u}. \tag{25}$$

The order of the model can be specified by the user based on prior knowledge of the process. The Smith predictor controller is the most popular dead-time compensator and can be auto-tuned (Hang *et. al.*, 1995) by combining the above relay identification and a primary controller design to be described below.

The Smith predictor scheme is shown in Fig. 7, where $G(s)=G_o e^{-Ls}$ is the process and $\hat{G}(s)=\hat{G}_0 e^{-Ls}$ is the model. Theoretically, the Smith predictor eliminates the dead-time from the closed-loop, and if the PI controller in the form of (12) is used as primary controller, it can be designed based only on the dead-time-free part of the model. The dead-time-free part is given by

$$G_0(s)=\frac{K}{(Ts+1)^n}, \qquad n=1, \text{ or } 2. \tag{26}$$

For a first order model, $n=1$, the design objective is such that

$$\frac{G_0(s)C(s)}{1+G_0(s)C(s)}=\frac{1}{1+T_d s}, \tag{27}$$

where T_d can be chosen as $T_d=\alpha T$ and a suitable range of α is 0.2 to 1. The PI controller is given by

$$T_I=T, \qquad K_P=\frac{T}{KT_d}. \tag{28}$$

For the second-order modeling, n=2, we choose PI parameters such that

$$\frac{G_0(s)K(s)}{1+G_0K(s)}=\frac{\omega_0^{\,2}}{s^2+2\zeta\omega_0 s+\omega_0^{\,2}}. \quad (29)$$

A simple solution is

$$T_I=T, \qquad K_P=\frac{1}{4\zeta^2 K}. \quad (30)$$

The only user-specified parameter is the damping factor ζ, which is chosen in the range of *0.5* to *1*.

Example 3 For a long dead-time process

$$G(s)=\frac{0.57e^{-18.70s}}{(8.60s+1)^2},$$

the identified model from relay feedback is

$$\hat{G}(s)=\frac{0.57e^{-18.80s}}{(7.99s+1)^2}.$$

The PI controller settings according to the method are

$$K_P=1.75, \quad T_I=7.99.$$

The pure PI controller without Smith predictor tuned by the Gain and Phase Margin method (Ho *et. al.*, 1995) is

$$K_P=0.50, \quad T_I=13.86.$$

The respective closed-loop responses are presented in Fig. 8.

This auto-tuning technique has been found to be effective even for high-order or non-minimum phase processes that exhibit apparent dead-time-like characteristics in their dynamics.

6. MULTIVARIABLE PROCESSES

Auto-tuning techniques for PID controllers are very successful when the process is essentially single-input single-output (SISO). For the commonly applied multi-loop controllers such as cascade control, the application of the frequency response method to their auto-tuning is quite straight forward (Hang *et.al.*, 1994). The extension of these techniques to multivariable processes is non-trivial and has attracted recent attention in the literature. Luyben (Luyben, 1986) has presented an iterative tuning procedure for multi-loop PID controller, where the stability of the whole system can only be guaranteed by introducing appropriate detuning factors on the PI/PID parameters. Hang *et al.* (Hang *et. al.*, 1994) propose two relay auto-tuning methods for multi-loop PI controllers. The first method, which adopts the sequential relay tuning method, tunes the multivariable system loop by loop, closing each loop once it is tuned, until all the loops are done. The Z-N rule is used to tune the PI controllers after the critical points are obtained. In the second method, all the loops of the multivariable are placed on the relay feedback in a multi-loop fashion, and the controllers are tuned simultaneously. The method is time-saving. However, several modes of oscillation may occur and should be treated individually. As in SISO cases, only one point information is usually used to tune the multi-loop controllers. In case of significant interaction, a fully cross-coupled multivariable controller rather than multi-loop PID controllers should be employed, and its auto-tuning becomes more difficult.

In principle, the relay identification techniques described in Section 3 can be applied for multivariable process modeling, if independent relay or the sequential relay test is adopted. In particular, it is straightforward (Wang *et. al.*, 1997b) to extend the method in Section 3 to the MIMO case, with the sequential relay. The arrangement is shown in Fig. 9. For a $m \times m$ multivariable process, its frequency response matrix $G(j\omega)$ is obtained with m relay tests. In what follows, we will briefly present two newly-developed tuning methods. One is for multi-loop controllers, the other for multivariable controllers, using the identified frequency response matrix. They can achieve performance improvement over other control schemes.

6.1 Multi-loop Controllers

The method to be described here is in fact a multi-loop extension of the original Astrom and Hagglund's modified Ziegler-Nichols method. In order to take into account the multivariable interactions, each loop is viewed as an independent equivalent process with all possible interactions lumped into it. For each loop, a controller is designed to meet the specifications that a given point on Nyquist curve be moved to a desired position for each equivalent process. A novel approach is developed to solve this nonlinear problem.

Consider a stable 2 by 2 process:

$$\begin{bmatrix} y_1(s) \\ y_2(s) \end{bmatrix} = \begin{bmatrix} g_{11}(s) & g_{12}(s) \\ g_{21}(s) & g_{22}(s) \end{bmatrix} \begin{bmatrix} u_1(s) \\ u_2(s) \end{bmatrix}. \tag{31}$$

The process is to be controlled by the multi-loop feedback controller:

$$K(s) = \begin{bmatrix} k_1(s) & 0 \\ 0 & k_2(s) \end{bmatrix}. \tag{32}$$

The resultant control system is shown in Fig. 10. Let $k_1(s)$ and $k_2(s)$ be of PID type, i.e.,

$$k_i(s) = K_{Pi}(1 + \frac{1}{T_{Ii}s} + T_{Di}s), \quad i=1,2, \tag{33}$$

which can be reduced to PI type when $T_{Di} = 0$.

The boxed portion in Fig. 10 can be viewed as an individual SISO process with an equivalent transfer function $g_1(s)$ between input u_1 and output y_1. It follows that $g_1(s)$ can be obtained as

$$g_1 = g_{11} - \frac{g_{12}g_{21}}{k_2^{-1} + g_{22}}. \tag{34}$$

Similarly, the equivalent process between u_2 and y_2 is given

$$g_2 = g_{22} - \frac{g_{21}g_{12}}{k_1^{-1} + g_{11}}. \tag{35}$$

Now, the modified Ziegler-Nichols method is applied to the equivalent transfer function $g_1(s)$ and $g_2(s)$, i.e., the controllers $k_i(s)$, $i = 1, 2$, are designed such that the given points on the Nyquist curve of $g_i(s)$, $i = 1, 2$, where

$$A_i = g_i(j\omega_i) = r_{ai}e^{(-\pi+\varphi_{ai})}, \quad i = 1,2, \tag{36}$$

is moved respectively to the points:

$$B_i = g_i(j\omega_i)k_i(j\omega_i) = r_{bi}e^{(-\pi+\varphi_{bi})}, \quad i = 1,2. \tag{37}$$

It should be pointed out here that, unlike the SISO case, due to the dependence of $g_1(s)$ (or $g_2(s)$) on $k_2(s)$ (or $k_1(s)$), ω_1 and thus $k_1(s)$ (or ω_2 and $k_2(s)$) cannot be determined until $k_2(s)$ (or $k_1(s)$) has been fixed. This is circular and causes a major design difficulty. A novel graphical method is presented by Wang, *et. al.* (1998) for finding $k_1(s)$ and $k_2(s)$.

***Example* 4** The 24 tray tower separating methanol and water has the following transfer function matrix:

$$G(s)=\begin{bmatrix}\dfrac{-2.2e^{-s}}{7s+1} & \dfrac{1.3e^{-0.3s}}{7s+1}\\ \dfrac{-2.8e^{-1.8s}}{9.5s+1} & \dfrac{4.3e^{-0.35s}}{9.2s+1}\end{bmatrix}.$$

As all its elements are of first-order in nature, we use a simple PI controller. It is designed with our method as

$$K(s)=diag\left\{-1.46(1+\frac{1}{4.30s}),\ 3.40(1+\frac{1}{5.84s})\right\}.$$

The step responses of the resultant feedback system to unit set-point changes followed by load disturbance changes are shown in Fig. 11 with solid lines, where load disturbance changes of 0.5 and 1 are applied directly on the two process inputs, u_1 and u_2, respectively. The proposed PID controller gives better loop and decoupling performance than the well-known BLT (Biggest Log Modulus) method (Luyben, 1986) (dashed lines).

6.2 Multivariable Controllers

The design method in Section 4.3 can be extended (Wang *et. Al.*, 1997b) to the multivariable case. Let $G(s)$ be the process transfer function matrix. The multivariable controller is chosen as PID type:

$$K(s)=K_P+\frac{1}{s}K_I+sK_D. \qquad (38)$$

Assume that the desired closed loop transfer function matrix H is

$$H(s)=diag\left\{\frac{\omega_{0i}{}^2e^{-L_is}}{s^2+2\zeta_i\omega_{0i}s+\omega_{0i}{}^2}\right\}. \qquad (39)$$

Matching GK to open-loop $H[I-H]^{-1}$ yields

$$GK=G\ [I\quad \frac{I}{s}\quad sI]\begin{bmatrix}K_P\\ K_I\\ K_D\end{bmatrix}=H[I-H]^{-1}, \qquad (40)$$

where K is the controller matrix. The parameters of the PID controller K can be computed by solving the above equation with the least squares method. This multivariable tuning method concerns a range of important frequencies instead of an individual frequency, and no iteration is needed. Extensive simulations have shown that the proposed method gives very satisfactory results for most processes. In some special case of large interaction, more than one stage of compensators may be employed to enhance the control performance.

Example 5 Consider the well known Wood & Berry's binary distillation column process:

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

The controller obtained with the method is

$$K(s)=\begin{bmatrix}0.156+0.053\frac{1}{s}+0.065s & -0.029-0.032\frac{1}{s}+0.046s\\ -0.021+0.024\frac{1}{s}+0.028s & -0.103-0.024\frac{1}{s}-0.094s\end{bmatrix},$$

and the controller by BLT Tuning (Luyben, 1986) is

$$K(s)=\begin{bmatrix} 0.375(1+\frac{1}{8.29s}) & 0 \\ 0 & -0.075(1+\frac{1}{23.6s}) \end{bmatrix}$$

The closed loop responses are illustrated in Fig. 12.

Recently, a method for auto-tuning fully cross-coupled multivariable PID controllers from decentralized relay feedback was proposed for multivariable processes with significant interactions (Wang *et. al.*, 1997b). It was shown (Wang *et. al.*, 1997b) that for a stable $m \times m$ process the oscillation frequencies would remain almost unchanged under relatively large relay amplitude variations. Therefore, m decentralized relay feedback tests are performed on the process and their oscillation frequencies would be close to each other so that the process frequency response matrix can be estimated at that point. A bias was further introduced into the relay to obtain the process steady state matrix. For multivariable controller tuning, a new set of design equations were derived under the decoupling conditions where the equivalent diagonal processes were independent of off-diagonal elements of the controller and used to design its diagonal elements first. The PID parameters of the controllers were determined individually by solving these equations at two points given above. The method has been successfully applied to various typical processes.

7. EXTENSION TO ADAPTIVE CONTROL

After a successful commissioning, the process will be under closed-loop control. When a new setpoint change occurs, the resultant output and input time responses may be employed to estimate a possible newly changed process frequency response and to re-tune the controller K with the same method used in auto-tuning. If this is done continuously, an adaptive control feature is provided. If the operator is alerted after a certain threshold of parameter variation has exceeded, an intelligent monitoring and auto-tuning feature is provided.

Example 6 Consider example 2 again. Suppose that the plant dynamics is changed to

$$G(s)=\frac{1}{s^2+0.8s+1}e^{-0.1s}.$$

When a setpoint change occurs (say, at $t=0$), the new model of the process can be identified by the method in Section 3 from the measured input and output data. The PID controller design method in Section 4.3 is used to automatically generate the new controller:

$$K(s)=0.7488(1+\frac{1}{0.7506s}+0.8122s).$$

The system responses before and after adaptation at $t=50$ are shown in Fig.13.

If no set-point change has occurred, any significant transient must be the result of some load disturbances. We shall approximately model all these disturbances as an equivalent disturbance d acting at the process output y through an unknown dynamic element G_d, as shown in Fig. 14. $y(t)$ and $u(t)$ are recorded from the time when the $y(t)$ starts to change, to the time when the system settles down. The process frequency response is re-estimated as follows.

We know

$$Y(s)=G(s)U(s)+G_dD(s) \tag{41}$$

If d is measurable, the FFT technique could be applied to compute $D(j\omega_i)$. If it is unmeasurable, then we could wait for the process input to reach a steady state and, then infer that

$$D(s)=\begin{cases}1 & if\, u(\infty)=u(0);\\ \frac{1}{s} & if\, u(\infty)\neq u(0).\end{cases} \tag{42}$$

We model G and G_d respectively as

$$G(s)=\frac{\beta s+1}{\alpha_1 s^2+\alpha_2 s+\alpha_3}e^{-Ls} \tag{43}$$

and

$$G_d(s)=\frac{\gamma s+1}{\lambda_1 s^2+\lambda_2 s+\lambda_3}. \tag{44}$$

Equation (41) can then be rearranged into

$$\begin{aligned}&a_1 s^4 Y(s)+a_2 s^3 Y(s)+\cdots+a_5 Y(s)\\ &=b_1 s^3 U(s)e^{-Ls}+\cdots+b_3 sU(s)e^{-Ls}+b_4 U(s)e^{-Ls}\\ &+c_1 s^3 D(s)+\cdots+c_3 sD(s)+D(s),\end{aligned} \tag{45}$$

Equation (45) is re-written as

$$\Phi(s)X=D(s), \tag{46}$$

where $\Phi=[s^4Y(s)\ s^3Y(s)\cdots Y(s)\ -s^3U(s)e^{-Ls}\cdots -U(s)e^{-Ls}\ -s^3D(s)\cdots -sD(s)]$, and $X=[a_1,a_2 \ldots a_5,b_1,b_2,\ldots,b_4,c_1,\ldots c_3]^T$ is the real parameters to be estimated. Assume first that the process dead-time L is known, then with frequency responses $Y(j\omega_i)$, $U(j\omega_i)$ and $D(j\omega_i)$, $i=1,2,\cdots,m$, computed via the FFT, (46) yields a system of linear algebraic equations. We can obtain the least squares solution X in (46). This solution in fact depends on L if L is unknown. The squared fitting error for (46) will be a scalar nonlinear algebraic equation in one unknown L only. The error is then minimized with respect to L in the given interval which is an iterative problem on one parameter L. Each iteration needs to solve a Least Squares problem corresponding to a particular value of L. The model parameters are obtained when the minimum J is achieved. To facilitate solution further, we next derive some bounds for L so that the search can be constrained to a small interval. This will greatly reduce computations, improve numerical property, and produce a unique solution. It is noted that the phase lag contributed by the rational part of the model, $G_0(s)=\frac{\beta s+1}{\alpha_1 s^2+\alpha_2 s+\alpha_3}$, is bounded, *i.e.*,

$$\arg G_0(j\omega)\in[-\pi\ ,\ \frac{\pi}{2}],\ \forall\omega\in(0,\infty), \tag{47}$$

so that we can impose an upper $\overline{L}$ and a lower $\underline{L}$ bound on L:

$$\overline{L}=\min\left\{-\frac{\arg G(j\omega_k)-\frac{\pi}{2}}{\omega_k}\right\},\ k=1,2,\cdots,m \tag{48}$$

and

$$\underline{L}=\max\left\{-\frac{\arg G(j\omega_k)+\pi}{\omega_k}\right\},\ k=1,2,\cdots,m \tag{49}$$

Actually, from a relay feedback or setpoint change, one can directly find out a gross estimate for dead time by measuring the time $\hat{L}$ between the control signal change to the output starting to move. Another possible bound may then be:

$$L \in \left[\ 0.5\hat{L} \quad 1.5\hat{L}\ \right]. \tag{50}$$

Remark 1. If the process dead time L is unchanged since the last identification of G, no iteration is needed to solve (46). This is a special case that the bound for L is specified as a zero interval. This case greatly simplifies the identification. It may be true in many practical cases as process dynamics perturbations are usually associated with operating point changes and/or load disturbance, which mainly cause time constant/gain changes. Furthermore, any small dead time change can be accommodated as in other parameters.

Remark 2. If the environment is noisy, all the measurements y, u and d can be filtered to yield y_f, u_f and d_f which are than used to produce Φ_f and D_f. Equation (46) then becomes $\Phi_f(s)X = D_f(s)$.

Example *7* As a demonstration, the process was chosen as

$$G(s) = \frac{1}{(5s+1)^2}e^{-5s}.$$

With the PID design method (Wang *et al.*, 1997d), the PID controller was obtained as

$$K(s) = 0.66(1 + \frac{1}{10.04s}). \tag{51}$$

The process response to a setpoint change at $t = 0$ is pretty good, which is shown in Fig. 15. Suppose that after the process settles down, the process gain was suddenly changed from *1* to 2 at $t = 66s$ as

$$G(s) = \frac{2}{(5s+1)^2}e^{-5s}.$$

The transients of the process input and output were recorded as shown in Fig. 15. As shown before, the process response under gain change is equivalent to a step load disturbance. The process model was re-estimated using the proposed method in Section 3 and the PID controller was re-tuned based on the re-estimation using the method in Section 4.3. This resulted in a new controller

$$K(s) = 0.45(1 + \frac{1}{11.67s} + 0.11s). \tag{52}$$

The next step set-point change at $t = 228s$ led to a satisfactory response, as shown in Fig. 15 (solid line). For comparison, if the controller in (51) was still used for the new process, the resultant response (dotted line) was also displayed in Fig. 15. The improved response due to adaptation was evident. Assume next that the process was further perturbed with a major change in process structure from 2nd order to 4th order and with a dead-time reduction to 2.5 as

$$G(s) = \frac{2}{(5s+1)^4}e^{-2.5s}. \tag{53}$$

An unknown load disturbance then occurred, which was generated by applying a step signal through an unknown disturbance channel:

$$G_d = \frac{1}{(2s+1)^2}. \tag{54}$$

The load disturbance responses were utilized to run the adaptation scheme. The process model was updated and the controller was adjusted to

$$K(s) = 0.33(1 + \frac{1}{15.32s}) \tag{55}$$

to reject the process perturbation. A subsequent setpoint change occurred at $t = 510s$ and the control performance of the new process under controller in (52) was shown in Fig. 15 (dotted line). The adaptation performance with the controller of (55) was also shown (solid line) and its improved response is evident.

8. CONCLUDING REMARKS

The relay feedback auto-tuning technique has been widely used to automatically tune single-loop PID controllers and to initialize adaptive controllers. The standard relay tuning technique has been successfully modified and extended by means of the frequency response approach to auto-tune advanced controllers. The relay-FFT technique that can be used to identify multi-points on the frequency response is most promising and impactful as it can be used to auto-tune dead-time compensators and multivariable controllers. The FFT technique can also be applied to setpoint changes or load responses to infer any significant changes in the process frequency response and hence provide information for on-line adaptation. This paper takes stock of these recent developments and extensions of the relay feedback auto-tuning technique. It is evident that this tuning and adaptation technique has become mature and ready for wider practical applications, in tune with the increasing demand for better control performance and also of new opportunities for implementation especially in modern distributed control.

REFERENCES

Astrom, K. J. and T. Hagglund (1984a), "Automatic Tuning of Simple Controllers with Specification on Phase and Amplitude Margins", *Automatica*, Vol. 20. No. 5, pp645-651.

Astrom, K. J. and T. Hagglund (1984b), "Automatic Tuning Of Simple Controllers", *Proceedings of the 9th IFAC World Congress*, Budapest. pp. 1867-1872.

Astrom, K. J. and T. Hagglund (1988), "Automatic Tuning of PID Controllers", *Instrument Society of America.*

Astrom, K. J., C. C. Hang, P. Person and W. K. Ho (1992), "Towards Intelligent PID Control", *Automatica*, Vol. 28, No.1, pp1-9.

Astrom, K. J., T. Hagglund, C. C. Hang and W. K. Ho (1993), "Automatic Tuning And Adaptation For PID Controllers-A Survey", *Control Eng. Practice*, Vol. 1, No.4, pp.699-714.

Bergland, G. D. (1969), "A Guided Tour of the Fast Fourier Transform", *IEEE Spectrum*, July, pp41-52.

Franklin, G. F., J. D. Powell and A. E. Baeini (1986), "Feedback Control of Dynamic Systems", *Addison-Wesley*, Reading, MA.

Hagglund, T. and K. J. Astrom (1991), "Industrial Adaptive Controllers Based On Frequency Response Techniques", *Automatica*, Vol. 27, pp599-609.

Hang, C. C. and K. J. Astrom (1988), "Practical Aspects Of PID Auto-Tuners Based On Relay Feedback", *Preprints IFAC Int. Symposium On Adaptive Control Of Chemical Processes, ADCHEM '88, Lyngby, Denmark.*

Hang, C. C. and K. K. Sin (1991), "A Comparative Performance Study Of PID Auto-Tuners", *IEEE Control Systems*, Vol.11, No.5, pp41-47.

Hang, C. C., K. J. Astrom, and W. K. Ho (1991a), "Refinements of the Ziegler-Nichols Tuning Formula", *IEE Proceedings-D*, Vol. 138, No.2, pp. 111-118.

Hang, C. C., K. J. Astrom, and W.K. Ho (1993a), "Relay Auto-tuning in the Presence of Static Load Disturbance", *Automatica*, Vol. 29(2), pp.563-564.

Hang, C. C., A. P. Loh and V.,U. Vasnani (1994), "Relay Feedback Auto-tuning of Cascade Controllers", *IEEE Trans. on Control Syst. Tech.*, Vol. 2, No.1, pp. 42-45.

Hang, C. C., Q. G. Wang and L. S. Cao (1995), "Self-tuning Smith Predictors for Processes with Long Dead Time", *Int. J. Adeptive control and Signal Processing*, Vol. 9, No. 3, pp. 255-270.

Ho, W. K., C. C. Hang and L. S. Cao (1995), "Tuning of PID Controllers Based on Gain and Phase Margin Specifications", *Automatica*, Vol. 31, No. 3, pp497-502.

Lee, T. H., Q. G. Wang, and K. K. Tan (1995), "A knowledge-based approach to dead time estimation for process control", *Int. J. Control*, Vol.56, No.5, pp1045-1072.

Lundh, M. and K. J. Astrom (1998), "Automatic Initialization and Robust Adaptive Controllers", *ISA Transactions*, Vol.37 , pp. 123-131.

Luyben, W. L., (1986) "A Simple Method for Tuning SISO Controllers in a Multivariable System", *Ind. Eng. Chem. Proc. Des. Dev.*, 25, 654-660.

Plamor, Z. J. and M. Blau (1994), "An Auto-tuner for Smith Dead Time Compensator", *Int. J. Control*, Vol. 60, No.1, 117-135.

Ramirez, R. W. (1985), "The FFT Fundamentals and Concepts", *Prentice-Hall*, Inc.

Schei, T. S. (1992), "A Method For Closed Loop Automatic Tuning Of PID Controllers", *Automatica*, Vol. 28, No.2, pp.587-591.

Wang, Q. G., C. C., Hang and B. Zou (1997a), "Low-Order Modelling From Relay Feedback", *Industrial and Engineering Chemistry Research*, Vol.36 , No.2, pp.375-381.

Wang, Q. G., C. C. Hang and B. Zou (1997b), "A Frequency Response Approach To Auto-tuning Of Multivariable Controllers", *Chemical Engineering Research and Design*, Vol.75, November, pp.797-806.
Wang, Q. G., C. C. Hang and Q. Bi (1997c), "Process frequency response estimation from relay feedback", Control Engineering Practice, Vol.5 , No.9, pp1293-1302.
Wang, Q. G., C. C. Hang and Q. Bi, (1997d), "A Frequency Domain Controller Design Method", *Chemical Engineering Research and Design*, Vol. 75, pp. 64-72.
Wang, Q. G., T. H. Lee and Y. Zhang (1998), "Multi-loop version of the modified Ziegler-Nichols Method for TITO processes", *Industrial and Engineering Chemistry Research*, Vol.37, No.12, pp.4725-4733.
Ziegler, J. G. and N. B. Nichols (1942), "Optimum settings for automatic controllers", *Trans. ASME*, 64, pp.759-768.

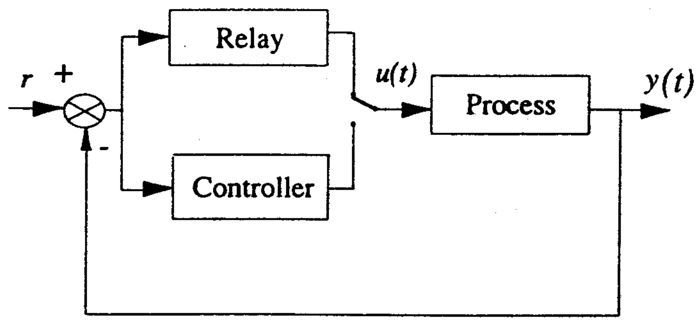

Fig. 1. Relay feedback system.

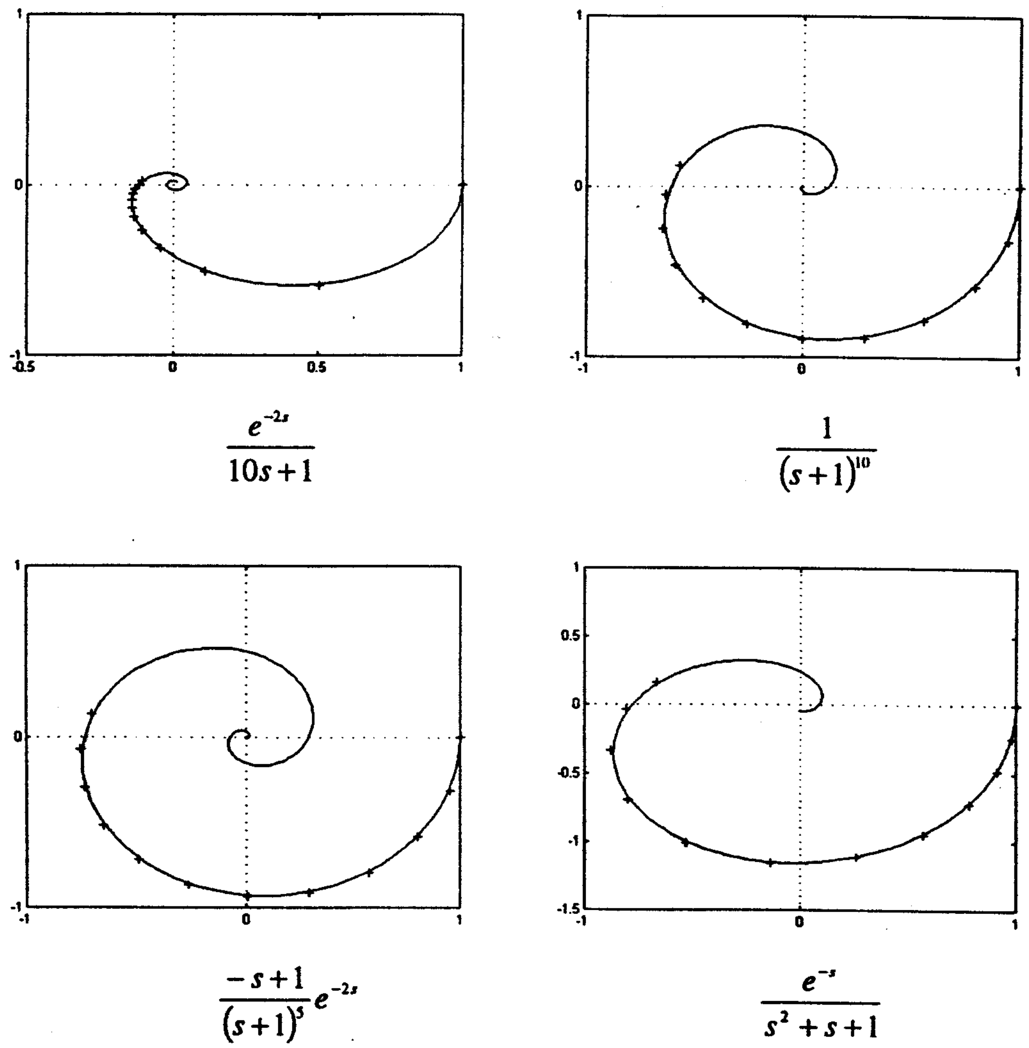

Fig. 2. Process Nyquist plots.
— actual, ++ estimated.

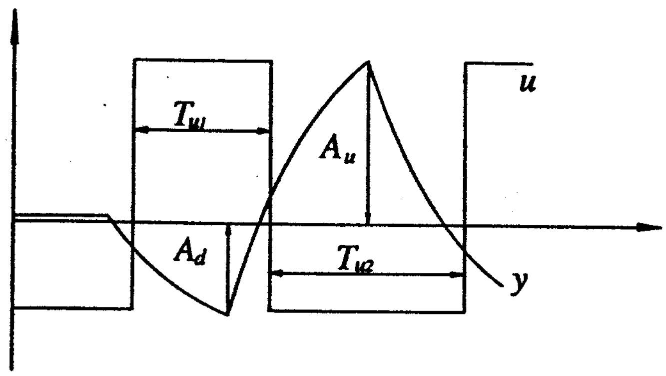

Fig. 3. Oscillatory waveforms under relay control.

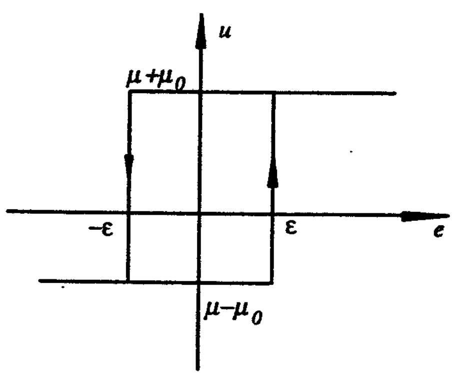

Fig. 4. The biased relay characteristics.

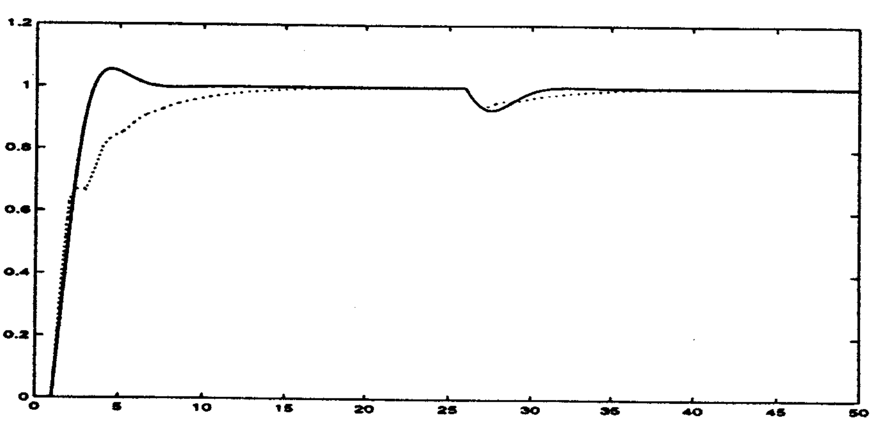

Fig. 5. PID tuning for $G = \frac{1}{s+1} e^{-s}$

— GPM, ··· Modified Z-N.

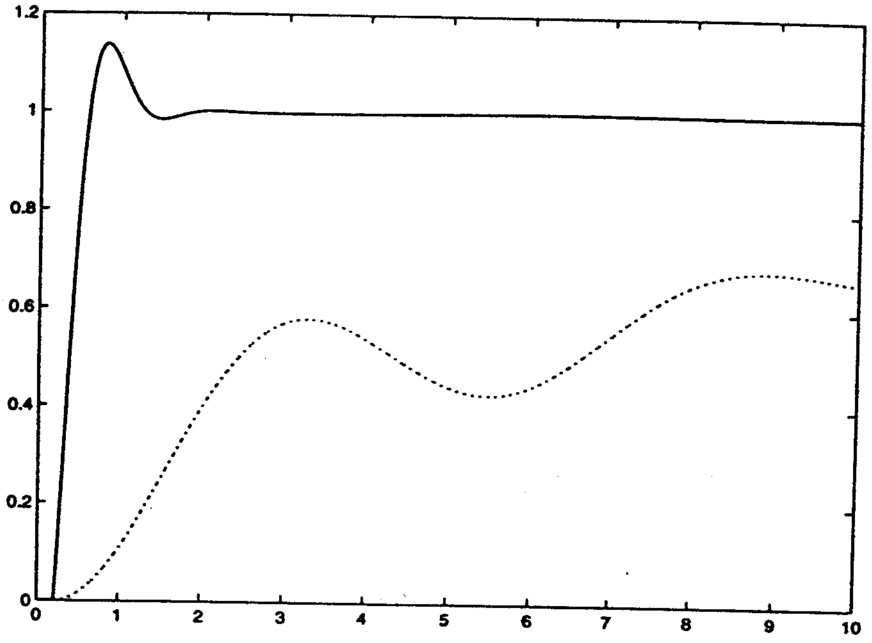

Fig. 6. Control performance for an oscillatory process.
— Proposed, ··· Modified Z-N.

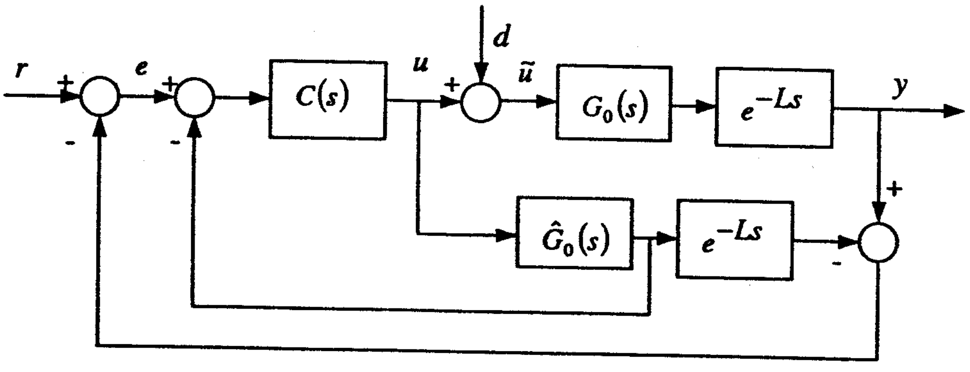

Fig. 7. Smith predictor scheme.

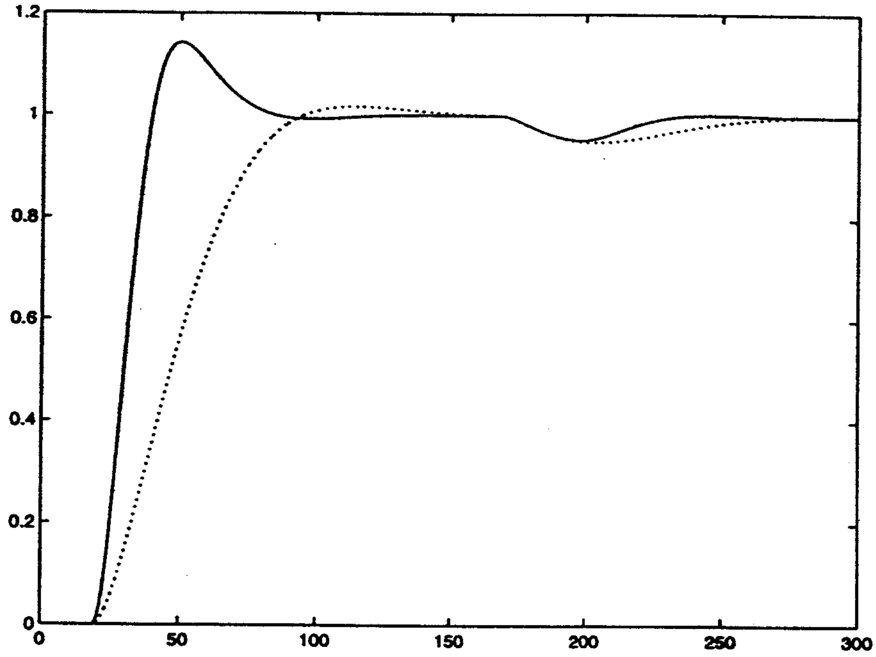

Fig. 8. Auto-tuning performance.
— Smith predictor, ··· Pure PI.

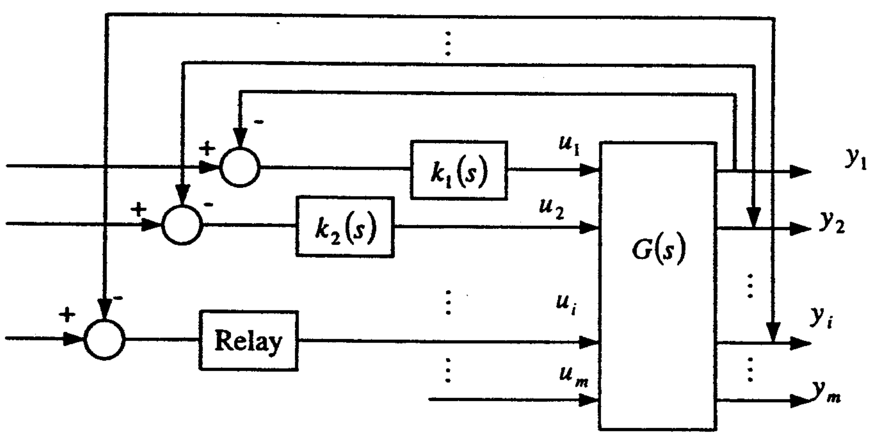

Fig. 9. Sequential relay identification.

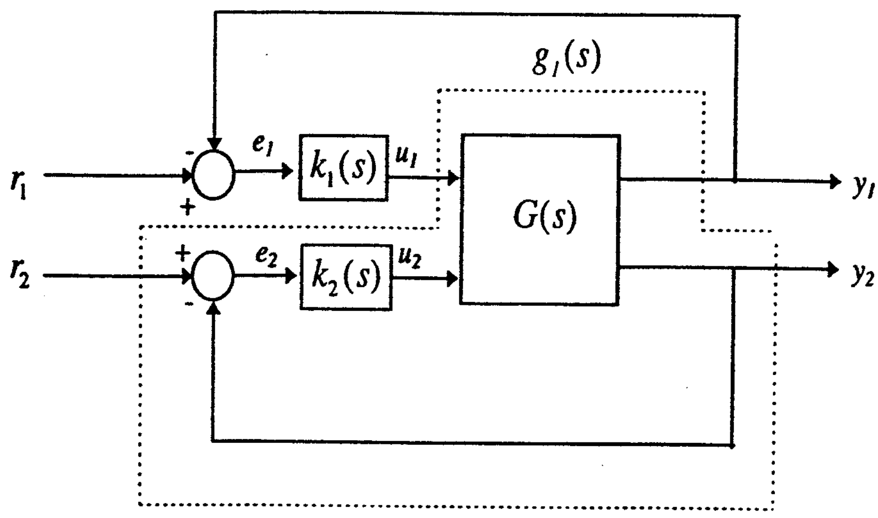

Fig 10. Multi-loop control system

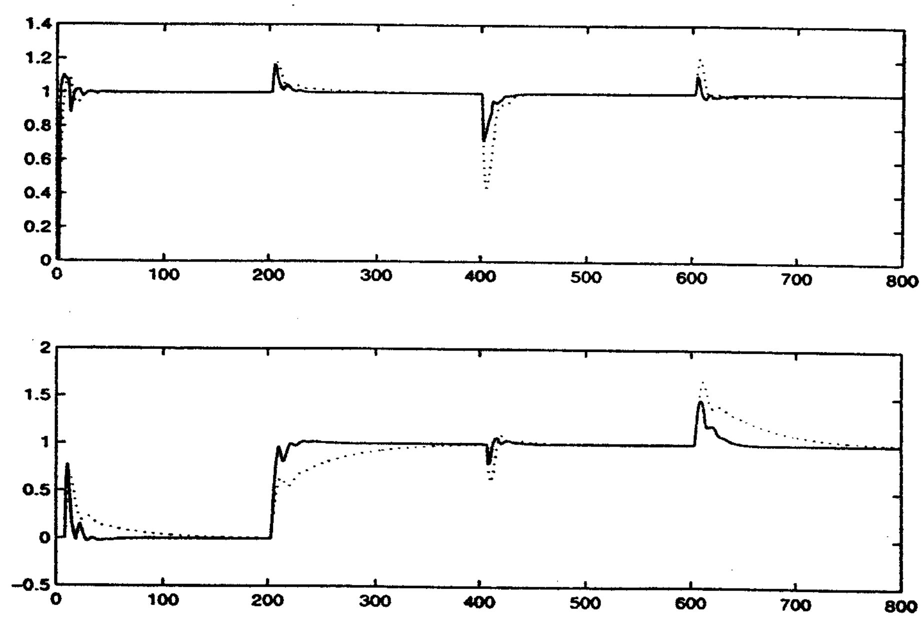

Fig. 11,Multi-loop control system step responses
—— Proposed method; ---- BLT method

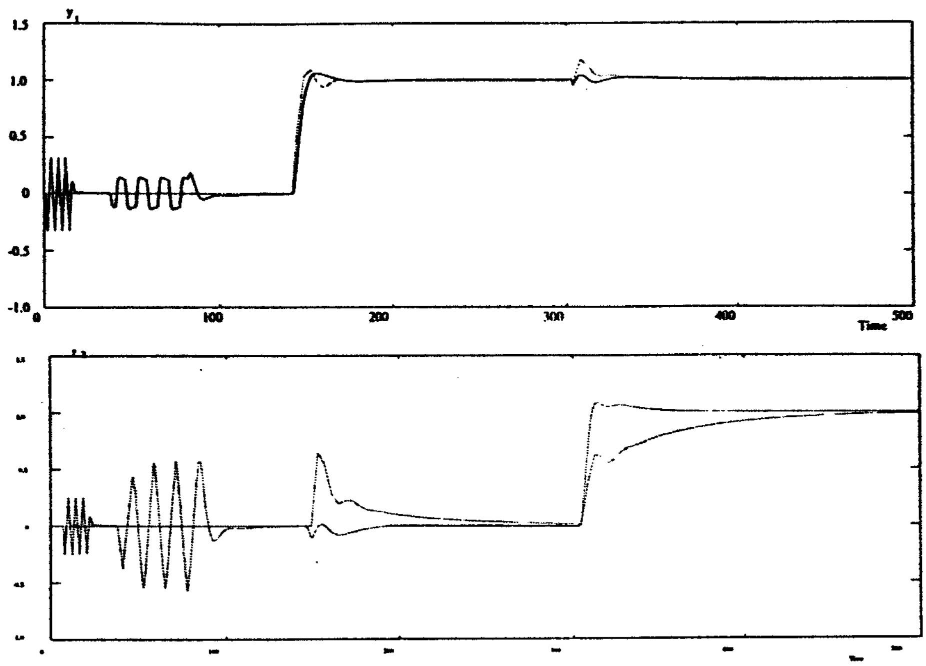

Fig. 12. System performance of multivariable control.
—— Proposed method; ---- BLT method

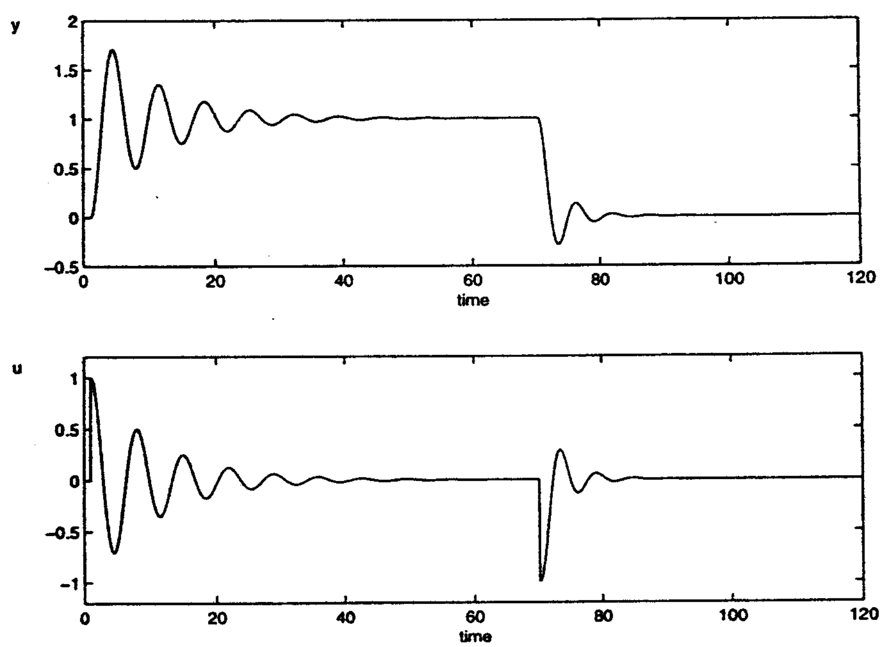

Fig. 13. Adaptation test.

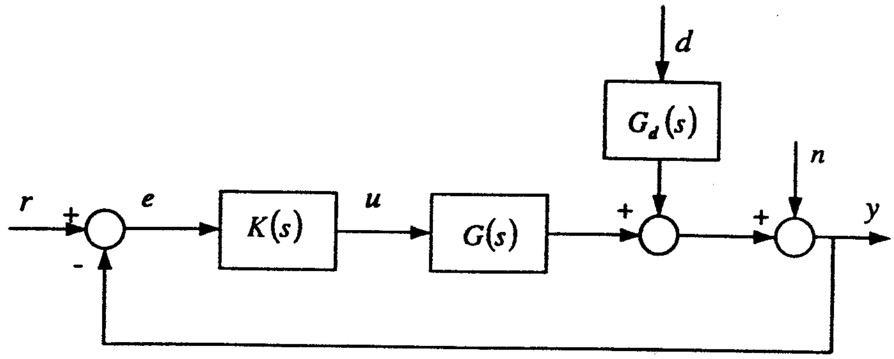

Fig. 14. A SISO control system with load disturbance.

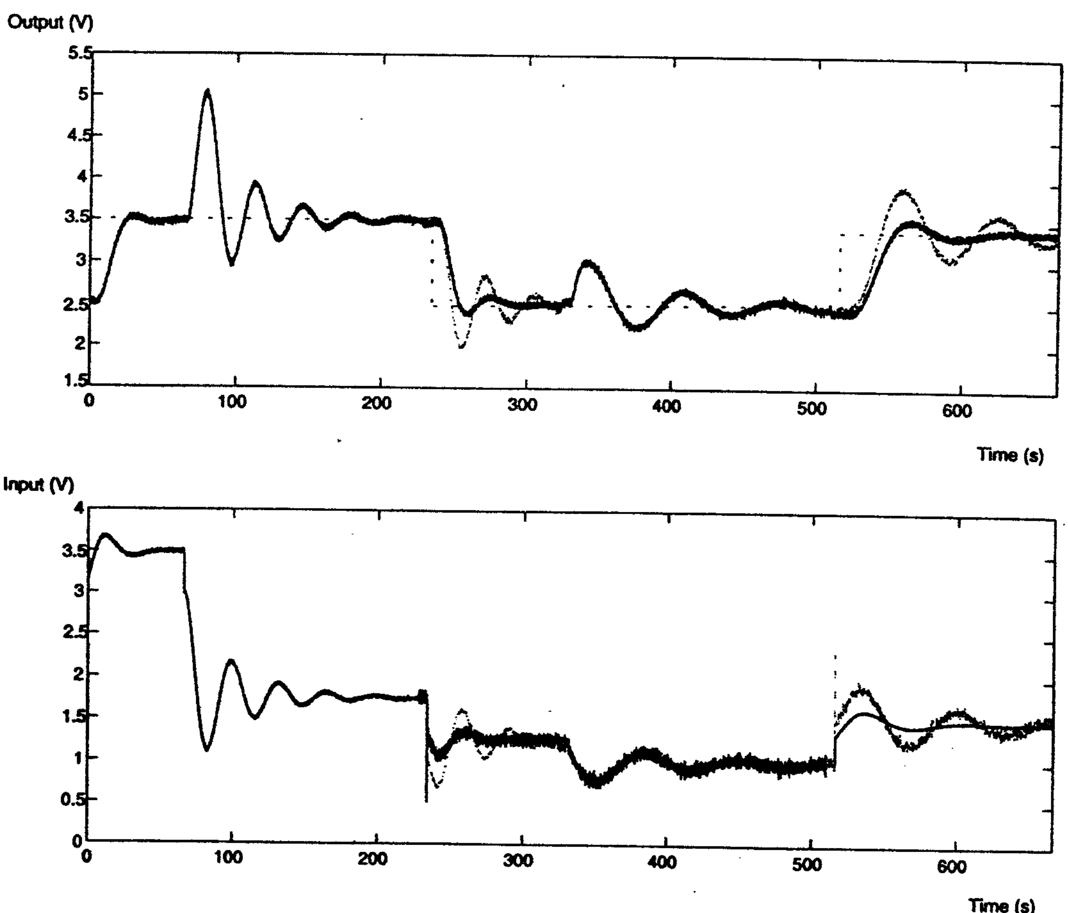

Fig. 15 Performance of adaptive control
—— with adaptation; ---- without adaptation