Journal of Process Control 12 (2002) 143–162
www.elsevier.com/locate/jprocont

Review

# Relay feedback auto-tuning of process controllers — a tutorial review

C.C. Hang[a], K.J. Astrom[b], Q.G. Wang[a,*]

[a]Department of Electrical Engineering, National University of Singapore, Singapore 119260, Singapore
[b]Department of Automatic Control, Lund Institute of Technology, Lund, Sweden

Received 12 August 1999; received in revised form 3 March 2001; accepted 22 March 2001

**Abstract**

The PID relay auto-tuner of Astrom–Hagglund is one of the simplest and most robust auto-tuning techniques for process controllers and has been successfully applied to industry for more than 15 years. This tuner is based on an approximate estimation of the critical point on the process frequency response from relay oscillations. Many developments have recently been reported to extend its applications. It turns out that more and accurate information on process dynamics can be obtained from the same relay test with the help of new identification techniques, and used to tune PID controllers better. Extensions are also made to tune model-based advanced controllers and multivariable controllers. The present paper reviews these developments and shows the state-of art in relay auto-tuning of process controllers. 

**Contents**



* Correspondence author. Tel.: +65-874-2282; fax: +65-779-1103.
*E-mail address:* elewqg@nus.edu.sg (Q. G. Wang).

0959-1524/02/$ - see front matter 
PII: S0959-1524(01)00025-7

## 1. Introduction

The introduction of auto-tuning capabilities to PID controllers has shortened the time to commission control system and facilitated control optimization through regular retuning [1–3]. The relay feedback auto-tuning method proposed by Astrom and Hugglund [4] was one of the first to be commercialized and has remained attractive owing to its simplicity and robustness [1,2,5]. The recent studies on PID controller are reported by Astrom et al. [6] and Leva and Colombo [7].

Many research works on modifying the relay feedback auto-tuning method have been reported in recent years. Improvements of the relay identification accuracy and efficiency have been proposed [8–12] by reducing high-order harmonic terns or using the Fourier analysis instead of the describing function method. The PID tuning formulae are refined to improve the controller performance for diverse processes such as long dead-time processes and oscillatory processes [12–17]. Auto-tuning for processes with varying time delay was proposed by Leva [14]. The relay auto-tuning method has been extended to advanced controllers such as the cascade controllers [18], the Smith Predictor [19–21] and the finite spectrum assignment controller [22]. It has also been incorporated in knowledge-based and intelligent systems as integrated initialization and tuning modules [23–27]. There have also been attempts to extend the method to multivariable systems [12,28,29]. It is the intention of this paper to review these new developments and extensions of the relay auto-tuning technique.

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, new relay based identification methods are highlighted. Refined controller tuning methods are presented in Section 4. Extensions of the relay auto-tuning to oscillatory, dead-time and multivariable processes are shown in Sections 5–7, respectively. Conclusions are given in Section 8.

## 2. Relay auto-tuning

The majority of the 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 match the process dynamics in order to provide good and robust control performance [30]. The tuning procedure, if done manually, is very tedious and time consuming; the resultant system performance mainly depends 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 techniques thus draw more and more attention of the 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 [1,25]. Typically, the user will either push a button or send a command to the controller. Industrial experience has clearly indicated that this is a highly desirable and useful feature. Earlier authors [31–34] proposed different auto-tuning methods which have great practical values. However, they all suffer from some major limitations. The Cohen–Coon method requires an open-loop test on the process and is thus inconvenient to apply. The disadvantage of the Yuwana and Seborg method and the Bristol method is the need of large setpoint change to trigger the tuning which may drive the process away from the operating point. Self-tuning controllers based on minimum variance, pole placement or LQG design methods may also be configured to give PID control [35,36]. These controllers have the disadvantage that a priori information about the time scale of the process dynamics must be provided to determine the suitable sampling intervals and filtering. Besides, conventional self-tuning controllers based on the recursive estimation of a parametric model requires a computer code of few kilobytes [37]. Relay auto-tuning method does not have these shortcomings.

Relay was mainly used as an amplifier in the fifties and the relay feedback was applied to adaptive control in the sixties [38]. Astrom and co-workers successfully applied the relay feedback technique to the auto-tuning of PID controllers for a class of common industrial processes [1]. 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 processes. The method is time-saving and easy to use [3]. 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 [1]. Thirdly, unlike other auto-tuning methods, the technique eliminates the need for a careful choice of the sampling rate from the a priori knowledge of the process. This is very useful in initializing a more sophisticated adaptive controller [39]. Fourthly, the relay feedback auto-tuning can be modified to cope effectively with disturbances and perturbations to the process [3,40,41].

The critical point, i.e. the process frequency response at the phase lag of $\pi$, has been employed to set the PID parameters for many years since the advent of the Ziegler–Nichols (Z–N) rule [42]. 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 $\pi$ at high frequency may oscillate with the period $T_u$ under relay control. To determine the 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 [4,43]

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

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

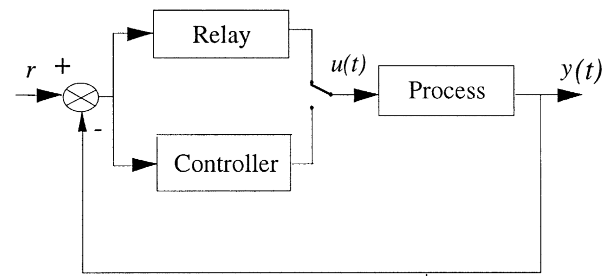



Fig. 1. Relay feedback system.

With the information of the process critical point, Z–N tuning rule or the modified Z–N rules [1,42,44] can be used to tune the PID controller. The relay auto-tuning procedure is then completed and the controller can be commissioned.

While the standard method is successful in many process control applications [5,25], 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 processes, the method could result in a significant error which would cause the system performance to deteriorate [12]. Second, only one frequency response point is obtainable from this method and it may be insufficient for describing some processes or for designing model based controllers.

## 3. Refinement of relay identification

As mentioned above, the critical point estimation from (1) is not always accurate. Some theoretical works [45] have investigated the estimation validity and accuracy of the describing function method, but they have not arrived at any results of true practical significance. An acceptable approach is to formulate a distortion criterion which checks the accuracy of the earlier assumptions to validate the information obtained [46]. However, such an approach will not improve the accuracy of the critical point. A great deal of effort has been directed at identifying desired frequency information below the ultimate frequency [17,47,48] by adding a time delay or an integrator to the process. Another problem with the standard relay feedback auto-tuning is that only one point on the process Nyquist curve is determined. For designing of the model-based controllers like the Smith–Predictor [19,21] and the finite spectrum assignment controllers [22,49], more points on the frequency response are needed 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 estimations are required, especially when high accuracy is required. This is particularly true when the process has a long dead-time. To obtain *more* and *accurate* points on process frequency response, several modified relay feedback based methods have been proposed [11,12,50–54] by using relay transients instead of only stationery oscillations or devising new types of relay functions with effective excitations at multiple frequencies instead of a single one around the process critical frequency.

### 3.1. Use of relay transient

It was shown in Hang et al. [21] 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 fast Fourier transform (FFT) to the remaining signals. This has been further improved by Wang et al. [52] 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 stationery 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 meaningfully using FFT. A decay exponential $\mathrm{e}^{-\alpha t}$ is then introduced to form

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

and

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

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

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

and

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

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

$$G(j\omega+\alpha) = \frac{Y(j\omega+\alpha)}{U(j\omega+\alpha)} = \frac{\tilde{Y}(j\omega)}{\tilde{U}(j\omega)}. \tag{4}$$

$\tilde{Y}(j\omega)$ and $\tilde{U}(j\omega)$ can be computed at discrete frequencies with the standard FFT technique [55,56]. Therefore, the shifted process frequency response $G(j\omega+\alpha)$ can be obtained from (4). To find $G(j\omega)$ from $G(j\omega+\alpha)$, we first take the inverse FFT of $G(j\omega+\alpha)$ as

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

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

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

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

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

The method can identify accurate frequency response points as many as desired with one relay experiment. They may be very useful for improving the 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 [12]. Note that for the purpose of controller tuning alone, the shifted frequency response may be used without the need to computer the actual one [57].

To illustrate the method, several different typical processes are considered in simulation. Fig. 2 shows the identified frequency responses for these processes using this method.

It is interesting to note that Sung and Lee [53] also made use of relay transients and presented a method for transfer function estimation based on the new regression equation derived from some integral transform. It should be pointed out that in the presence of static disturbance, biased relay feedback should be used for robustness. Otherwise, the model performance could deteriorate severely.

### 3.2. Use of the biased relay

A large number of processes can be characterized by the first order plus dead-time model [16,58]

$$G(s) = \frac{K\mathrm{e}^{-Ls}}{Ts+1}. \tag{5}$$

For these kinds of processes, Wang et al. [11] have used a biased relay feedback test and derived the formulae that could precisely yield the critical point and the static gain simultaneously with a single relay test. The biased relay is shown in the Fig. 3.

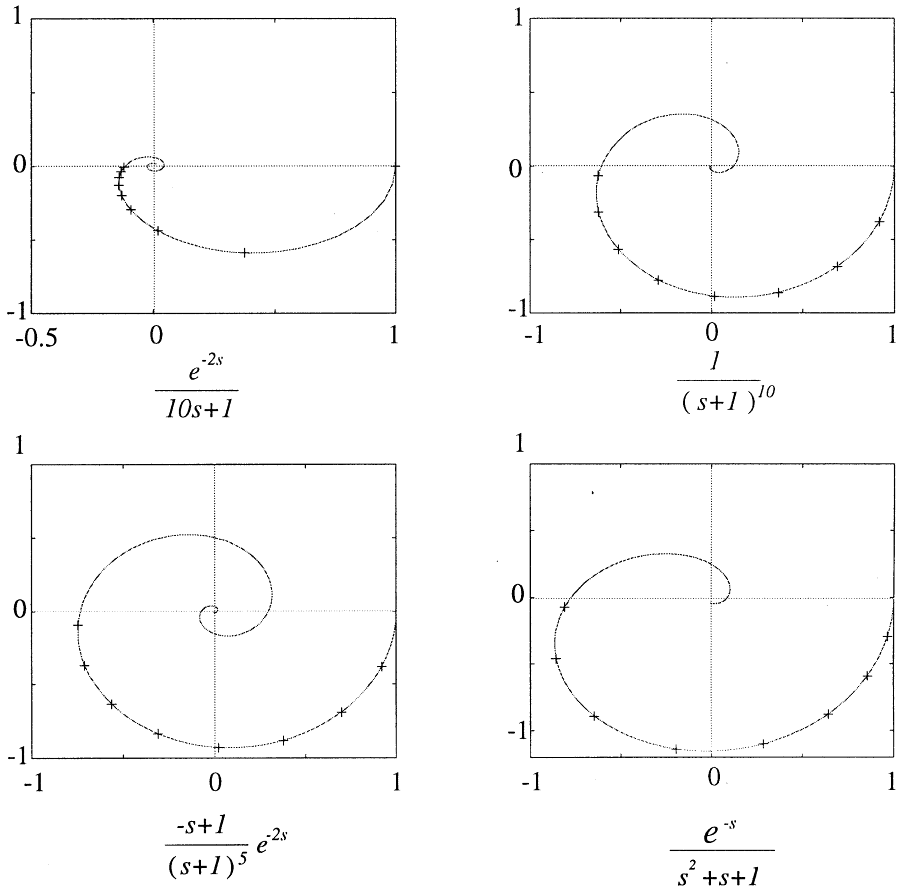



Fig. 2. Process Nyquist plots. — Actual, + Estimated.

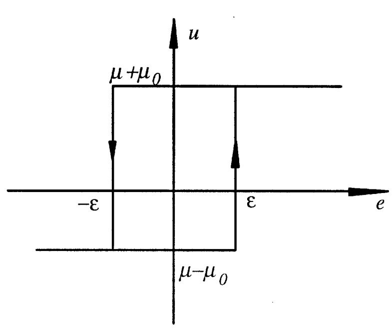



Fig. 3. The biased relay.

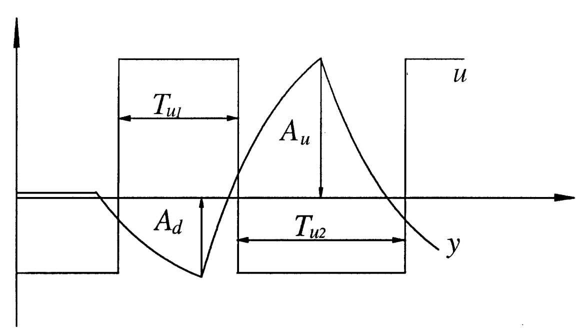



Fig. 4. Oscillatory waveforms.

Under the biased relay feedback, the process input $u$ and the process output $y$ is shown in Fig. 4. It is shown that for the process in (5), the output $y$ converges to the stationary oscillation in one period ($T_{u1} + T_{u2}$), and the oscillation is characterized by

$$A_u = (\mu_0 + \mu)K\left(1 - e^{-\frac{L}{T}}\right) + \varepsilon e^{-\frac{L}{T}}, \tag{6}$$

$$A_d = (\mu_0 - \mu)K\left(1 - e^{-\frac{L}{T}}\right) - \varepsilon e^{-\frac{L}{T}}, \tag{7}$$

$$T_{u1} = T\ln\frac{2\mu K e^{\frac{L}{T}} + \mu_0 K - \mu K + \varepsilon}{\mu K + \mu_0 K - \varepsilon}, \tag{8}$$

and

$$T_{u2} = T\ln\frac{2\mu K e^{\frac{L}{T}} - \mu K - \mu_0 K + \varepsilon}{\mu K - \mu_0 K - \varepsilon}. \tag{9}$$

The above four equations are the accurate expressions for the period and the amplitude of the limit cycle oscillation of the first order plus dead-time. By measuring any three of $A_u$, $A_d$, $T_{u1}$ and $T_{u2}$, the parameters of the model $K$, $T$ and $L$ can be calculated from (6)–(9). Solving these four equations is not an easy task. To simplify the computation, $K$ may be alternatively computed as the ratio of DC components in the output and input:

$$K = \frac{\int_0^{T_{u1}+T_{u2}} y(t)\mathrm{d}t}{\int_0^{T_{u1}+T_{u2}} u(t)\mathrm{d}t}.$$

The normalized dead time of the process $\theta = \frac{L}{T}$ is obtained from (6) or (7) as

$$\theta = \ln\frac{(\mu_0+\mu)K-\varepsilon}{(\mu_0+\mu)K-A_u},$$

or

$$\theta = \ln\frac{(\mu-\mu_0)K-\varepsilon}{(\mu-\mu_0)K+A_d}.$$

It then follows from (8) or (9) that

$$T = T_{u1}\left(\ln\frac{2\mu K e^{\theta}+\mu_0 K-\mu K+\varepsilon}{\mu K+\mu_0 K-\varepsilon}\right)^{-1},$$

or

$$T = T_{u1}\left(\ln\frac{2\mu K e^{\theta}-\mu_0 K-\mu K+\varepsilon}{\mu K-\mu_0 K-\varepsilon}\right)^{-1}.$$

The dead time is thus

$$L = T\theta.$$

The method produces two accurate process frequency points in just one relay test. The two points can then be used to design the controller. The method is simple for implementation and suitable for the processes that can be characterized by the first order plus dead-time model.

Simulation is carried out for processes with different normalized dead-times to illustrate the accuracy of the proposed method. The outputs of biased relay are 1.3 and −0.7 respectively, and the hysteresis of relay is 0.1. The resultant limit cycles and model parameters are presented in Table 1. For comparison, this table also shows the parameters obtained by the autotune variation (ATV) method [59], where it is assumed that the steady-state gain is known and the dead time is read exactly. The method has also been applied to processes which are not in the form of (5), and frequency responses of the identified first order plus dead-time models are close to the true ones [11].

It is noted that Chang et al. [50] gave analytical expressions for the period and amplitude of limit cycles and used them to derive process transfer function models. Huang et al (1996) proposed the formulae for obtaining a FOPDT model from relay oscillations.

### 3.3. Use of the parasitic relay

Another modification of the standard relay for multipoint identification of frequency response is to superimpose a parasitic relay to the standard relay. The parasitic relay on–off period is twice as large as that of the standard relay. This arrangement is shown in Fig. 5. The excitations at $0.5\omega_c, 1.5\omega_c, \cdots$ are now available in addition to the standard excitations at $\omega_c, 3\omega_c, \cdots$. Consequently, frequency response estimation at all these frequencies can be obtained.

Let the standard relay operate as usual and its output be $u_1(k)$ with the amplitude $h$. The parasitic relay is realized by

$$\begin{cases} u_2(0) = \alpha h; & \\ u_2(k) = -\alpha h\cdot\mathrm{sign}(u_2(k-1)), & \text{if } u_1(k-1) > 0 \text{ and } u_1(k) < 0; \\ u_2(k) = u_2(k-1), & \text{otherwise;} \end{cases}$$

where $\alpha$ is a constant coefficient. $\alpha$ Should be large enough to have sufficient stimulation on the process while it should also be small enough such that the parasitic relay will not change the period of oscillation generated by the main relay too much.

Based on extensive simulation, $\alpha$ is recommended to be $0.1-0.3$. The output of the modified relay test is thus given by $u(k) = u_1(k) + u_2(k)$, and is sent as the input to process. The resultant process output $y$ from the modified relay test is shown in Fig. 7 and will reach a stationary oscillation with the period being $2T_c$. Due to two excitations in $u$, $y$ consists of frequency components at $\frac{2\pi}{T_c}$, $\frac{\pi}{T_c}$ and their odd harmonics $\frac{6\pi}{T_c}$, $\frac{10\pi}{T_c}, \cdots$, and $\frac{3\pi}{T_c}$, $\frac{5\pi}{T_c}, \cdots$, respectively. For a linear process, the process frequency response can be obtained by

$$G(j\omega_i) = \frac{\int_0^{2T_c} y_s(t)e^{-j\omega_i t}\mathrm{d}t}{\int_0^{2T_c} u_s(t)e^{-j\omega_i t}\mathrm{d}t}, \qquad i = 1,\ 2, \cdots, \tag{10}$$

Table 1
Parameter estimation from biased relay

| Case | Process | | | Biased relay | | | | New method | | | ATV method | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $K$ | $T$ | $L$ | $T_{u1}$ | $T_{u2}$ | $A_u$ | $A_d$ | $K$ | $T$ | $L$ | $K$ | $T$ | $L$ |
| 1 | 1.0 | 2.0 | 2.0 | 2.79 | 3.91 | 0.859 | −0.480 | 1.000 | 1.999 | 2.002 | 1.0 | 1.658 | 2.0 |
| 2 | 1.0 | 1.0 | 3.0 | 3.50 | 4.18 | 1.241 | −0.670 | 1.000 | 0.999 | 3.006 | 1.0 | 1.042 | 3.0 |
| 3 | 1.0 | 5.0 | 2.0 | 3.44 | 5.46 | 0.497 | −0.299 | 0.999 | 4.990 | 2.009 | 1.0 | 4.068 | 2.0 |

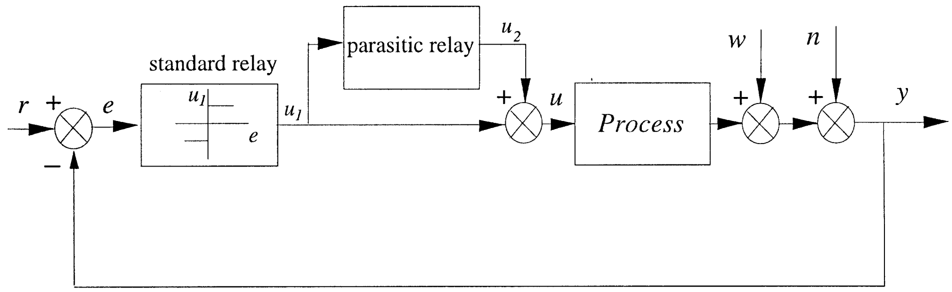



Fig. 5. Modified relay feedback system.

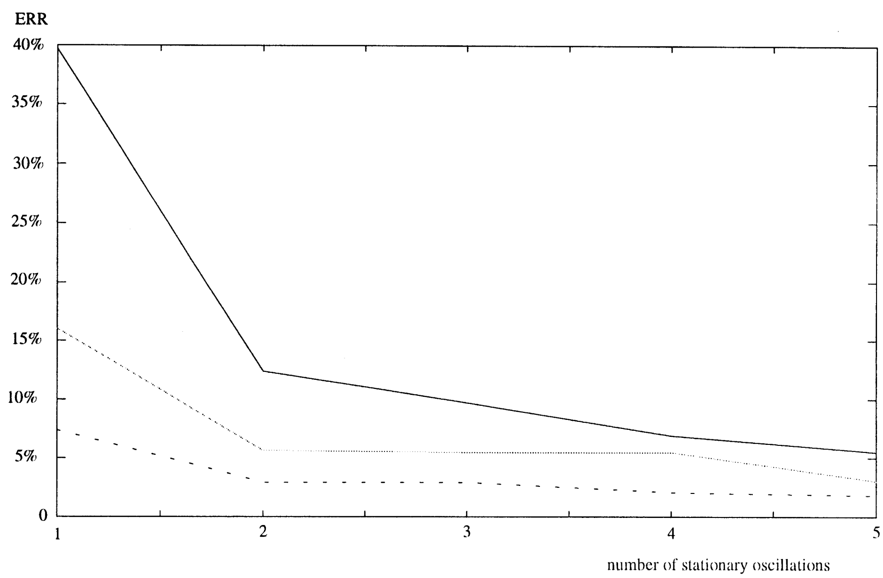



Fig. 6. *ERR* vs number ($N$) of stationary oscillations adopted. - - - $N_1 = 0\%$, $N_1 = 1\%$ and — $N_1 = 10\%$.

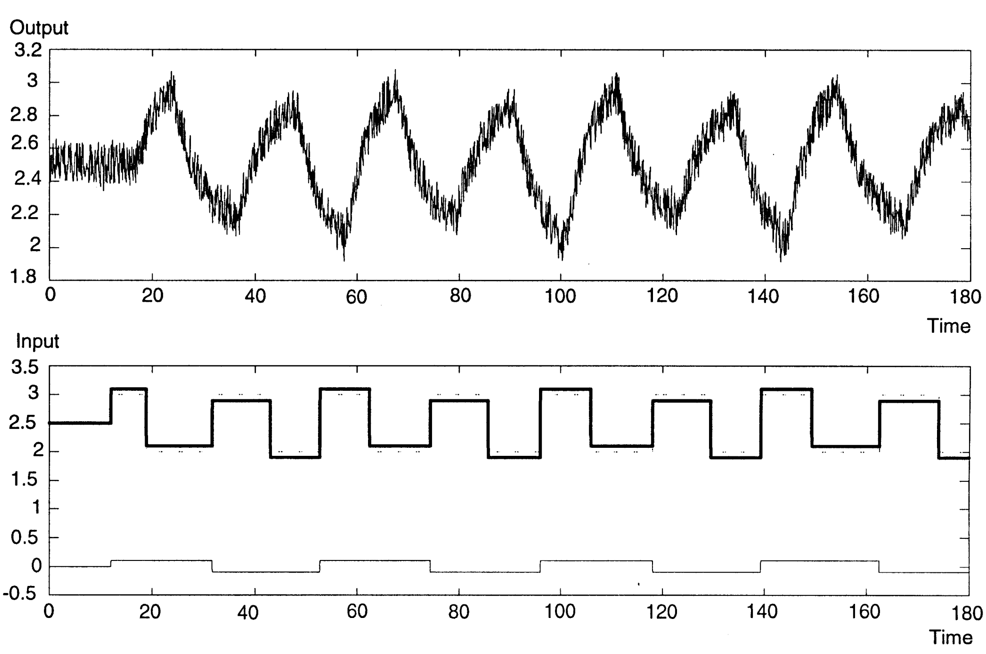



Fig. 7. Process input and output in the modified relay test. - - - - $u$, - - - - $u_1$ and — $u_2$.

where

$$\omega_i = \frac{(2i-1)2\pi}{2^l T_c}, \qquad l = 0, \quad 1,$$

are the basic and its odd harmonic frequencies in $u_s$ and $y_s$, $u_s$ and $y_s$ are a period ($2T_c$) of the stationary oscillations of $u(k)$ and $y(k)$ respectively. $G(j\omega_i)$ in (10) can be computed using the FFT algorithm [77] as

$$G(j\omega_i) = \frac{FFT(y_s)}{FFT(u_s)}.$$

**Example 1.** Consider a first order plus dead-time process

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

The amplitude of the standard relay is chosen as 0.5 and that of the parasitic relay $20\% \times 0.5$. Without additional noise, the noise-to-signal ratio $N_1$ of the inherent noise in our test environment is 0.025% ($N_2 = 4\%$). The identification error $ERR$ is 2.57%. To see noise effects, extra noise is introduced with the noise source in the *Simulator*. Time sequences of $y(t)$ and $u(t)$ in a relay test under $N_1 = 10\%$ ($N_2 = 31\%$) are shown in Fig. 7. The first part of the test in Fig. 7 ($t = 0-12$) is the "listening period", in which the noise bands of $y(t)$ and $u(t)$ at steady-state are measured. The hysteresis is chosen as 0.3. With averaging 4 periods of stationary oscillations, the estimated frequency response points under this noise level are shown in Fig. 8(a). The result is pretty good.

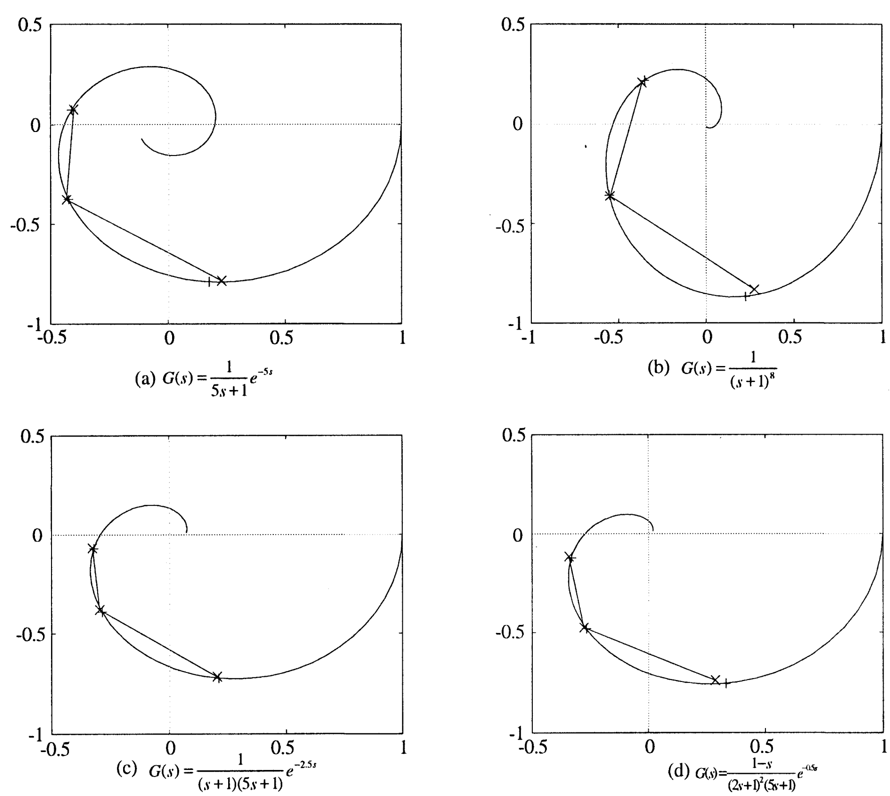
To ensure estimation accuracy under different noise levels, the number of stationary oscillation periods adopted in average calculation should be different. The estimation error $ERR$ vs the number of stationary oscillation periods adopted in average is plotted in Fig. 6, which can be used as a guide in deciding how many periods are enough to achieve a certain estimation accuracy under a given noise level.

## 4. Refinement of PID tuning

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

$$u(t) = K_P\left(e + \frac{1}{T_I}\int_0^t e(\tau)d\tau + T_D\frac{de}{dt}\right) = K_P e + K_I\int_0^t e(\tau)d\tau + K_D\frac{de}{dt}$$

where $e = b \cdot y_{SP} - y$, and $b$ is set-point weighting. In the standard relay tuning case, the Z–N-like formulas are



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

(b) $G(s) = \frac{1}{(s+1)^8}$

(c) $G(s) = \frac{1}{(s+1)(5s+1)}e^{-2.5s}$

(d) $G(s) = \frac{1-s}{(2s+1)^2(5s+1)}e^{-0.5s}$

Fig. 8. Nyquist plots - - + - - actual.

employed to tune PID controllers. These tuning rules are suitable only for those processes which can be 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. [24] 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. The idea of using normalized dead-time to improve controller tuning has been around for a long time. One of the early proposals was made by Cohen and Coon [31]. The earlier tuning formulas are, however, not very good. Refined formulae of the PID controller by incorporating heuristic knowledge of normalized dead-time to replace manual fine-tuning were developed [44]. 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.

All of 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. The kappa–tau method

This method can be regarded as an extension of the Cohen–Coon method. It was proposed in Astrom and Hagglund [60], and is also described in Astrom and Hagglund [61]. The idea is the same as for the Cohen–Coon method, namely to characterize the process by three parameters. There are two versions of the method, a time domain method and a frequency domain method.

The time domain method uses parameters derived from the model

$$G(s) = \frac{K}{Ts+1}e^{-sL}$$

The parameters chosen for tuning are $L$ and

$$a = \tfrac{KL}{T} \qquad \tau = \tfrac{L}{L+T}$$

This choice makes it easy to compare with the Ziegler–Nichols method which uses the parameters $a$ and $L$.

The frequency domain method characterizes the process by ultimate gain $k_u$, ultimate period $T_u$ and gain ratio

$$\kappa = \left|\frac{G(jw_u)}{G(0)}\right| = \frac{1}{k_u K}$$

This choice makes it easy to compare with the Ziegler–Nichols method which uses parameters $k_u$and $T_u$. The kappa–tau method also gives the value of the set point weighting $b$.

The tuning method is developed by computing controllers for a wide range of systems using a method which maximizes integral gain $k_i$ subject to a constraint on the maximum sensitivity $M_s$. It has been found empirically that the controller parameters can be well approximated by formulas of the type

$$K_P = \frac{1}{a}f_{k\tau}(\tau)$$

$$T_I = Lf_{T_i\tau}(\tau)$$

$$T_D = Lf_{Td\tau}(\tau)$$

$$b = f_{b\tau}(\tau)$$

for the time domain method and

$$K_P = k_u f_{kk}(\kappa)$$

$$T_I = T_u f_{T_i k}(\kappa)$$

$$T_D = T_u f_{T_d k}(\kappa)$$

$$b = f_{b\kappa}(\kappa)$$

for the frequency domain method. The functions $f_i$ are given in Astrom and Hagglund [61]. Since the method is based on a technique which has a constraint on the robustness the responses obtained are much better damped than those obtained by the Cohen–Coon method.

### 4.3. Gain and phase margin method

The gain and phase margins are very useful as they can serve as the measures of performance as well as robustness. Controller designs to satisfy gain and phase margin criteria are not new [62,63]. However, the solution is normally obtained by numerical methods or trial and error graphically using Bode plots. Such approaches are certainly not suitable for use in adaptive control and auto-tuning. The modified Ziegler–Nichols rule [1] 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. Otherwise, especially when the dead-time is dominant, the 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 $\phi_m$ was proposed [15,16]. The method is based on the measurement of the ultimate gain, ultimate period and the static gain of the process. For a process in the form of (5), the parameters of the PI controller

$$K(s) = K_P\left(1 + \frac{1}{sT_I}\right) \tag{11}$$

are given by

$$K_P = \frac{\omega_p T}{A_m K}, \tag{12}$$

$$T_I = \left(2\omega_p - \frac{4\omega_p^2 L}{\pi} + \frac{1}{T}\right)^{-1}, \tag{13}$$

where

$$\omega_p = \frac{A_m\phi_m + \frac{1}{2}\pi A_m(A_m - 1)}{\left(A_m^2 - 1\right)L}.$$

**Example 2** Consider a process given by

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

With the gain margin $A_m$ and phase margin $\phi_m$ specified as 3 and $\frac{\pi}{3}$ respectively, the PI controller parameters in (12) and (13) are computed as

$$K_P = 0.52$$

and

$$T_I = 1$$

For comparison, a PID controller tuned by the modified Z–N method [1] is also considered, and given by $[K_P, T_I, T_D] = [0.80, 2.38, 0.59]$.

The relative dead time is $\tau = 0.5$ and the step response version of the kappa–tau method with $M_s = 2$ gives the following parameters for a PID controller $[K_P, T_I, T_D, b] = [0.801,\ 1.198,\ 0.298,\ 0.308]$. We have $\kappa = 0.442$, $k_u = 2.261$ and $T_u = 3.097$, the frequency response version of the kappa–tau method gives the parameters $[K_P, T_I, T_D, b] = [1.015, 1.108, 0.279, 0.313]$.

The step responses obtained are shown in Fig. 9. The reason why the Ziegler–Nichols based methods perform so poorly for this process is that the relative time delay is large. It is well known that Ziegler–Nichols methods give too small an integral action in such cases.

The gain and phase method can be extended to the 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 form of (5). Simulations show 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.

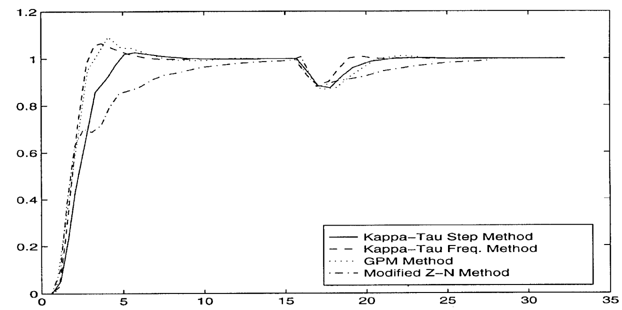


Fig. 9. PID tuning for $G = \frac{1}{S+1}e^{-s}$.

### 4.4. Enhanced Astrom–Hagglund controller design

To improve the modified Z–N method, an enhanced Astrom-Hagglund-based PID/PI controller design procedure was proposed [17]. Assume that the gain and phase margins of the closed-loop system are specified as $(A_m, \phi_m)$. and the critical point $(k_u, \omega_u)$ and another point $G(j\omega_\phi)$ with $G(j\omega_\phi) = -\pi + \phi_m$ are known. Define $k_\phi = 1/\left|G(j\omega_\phi)\right|$.

For a *PID controller*, by moving the second Nyquist point $G(j\omega_\phi)$ to a desired position specified by the combined gain and phase margin and letting the frequency response of the compensated system attenuate at $\omega = \omega_u$ with an attenuation factor $r$, the PID settings are given by

$$K_P = \frac{k_\phi}{A_m},$$

$$T_I = \frac{rk_\phi}{\omega_u\sqrt{k_u^2 - r^2k_\phi^2}}\left(\frac{\omega_u}{\omega_\phi}\right)^2 - 1,$$

and

$$T_D = \frac{1}{\omega_\phi^2 T_I},$$

where $r$ is recommended as

$$r = 1 + 0.9\left(\frac{k_u}{k_\phi} - 1\right).$$

For a process of the first-order dynamics or a process with a long dead-time, a *PI controller* in (11) is sufficient.

Suppose that the normalized dead-time is $\theta$. The PI controller parameters that can achieve the specified performance are given by

$$T_{\mathrm{I}} = \frac{1}{r\omega_{\phi}\tan\left(\arg G_{\mathrm{p}}\left(jr\omega_{\phi}\right) - \phi_{\mathrm{m}}\right)},$$

and

$$K_{\mathrm{P}} = \frac{rT_{\mathrm{I}}\omega_{\phi}}{A_{\mathrm{m}}\left|G_{\mathrm{p}}\left(jr\omega_{\phi}\right)\right|\sqrt{\left(rT_{\mathrm{I}}\omega_{\phi}\right)^2+1}}. \tag{14}$$

where $r$ is given by

$r = 0.8, \quad \text{for } n = 1, \ \theta < 0.5,$

$r = 0.5, \quad \text{for } \theta > 0.5.$

The enhanced auto-tuning technique incorporates the process multiple Nyquist points identification technique. It uses the information of two process frequency points which are near the critical point and it does not require knowledge of the process structure. It is applicable to a wider range of the processes.

**Example 3** Consider a process described by

$$G(s) = \frac{1}{(s+1)^2}e^{-0.5s}.$$

The specifications of $A_{\mathrm{m}} = 1$ and $\phi_{\mathrm{m}} = \frac{\pi}{3}$ are chosen. The two identified points are $(k_{\mathrm{u}}, \omega_{\mathrm{u}}) = (4.68, 1.92)$ and $\left(k_{\phi}, \omega_{\phi}\right) = (2.03, 1.02)$. The normalized dead-time $\theta$ is 0.33, hence the PID controller in (14) is selected and its parameters are computed as $[K_{\mathrm{P}}, T_{\mathrm{I}}, T_{\mathrm{D}}]$ =[2.03, 2.78, 0.53]. The PID controller calculated by the modified Z–N method is $[K_{\mathrm{P}}, T_{\mathrm{I}}, T_{\mathrm{D}}] = [1.67, 2.50, 0.63]$. In the kappa–tau method, we have $\kappa = 0.213$, $k_{\mathrm{u}} = 4.688$ and $T_{\mathrm{u}} = 1.041$, the frequency response version of the kappa–tau method gives the $[K_{\mathrm{P}}, T_{\mathrm{I}}, T_{\mathrm{D}}, b]$=[2.543, 1.488, 0.373, 0.280]. The step responses obtained are shown in Fig. 10, and the enhancement obtained is evident.

## 5. Processes with oscillatory dynamics

Because of the simple structure of the PID controller, most of the PID auto-tuning rules utilize only one or two points of process frequency response, which may not describe some process dynamics well, such as very oscillatory dynamics. For such processes, a frequency region rather than one or two points should be considered when designing the controller. Along this line, a

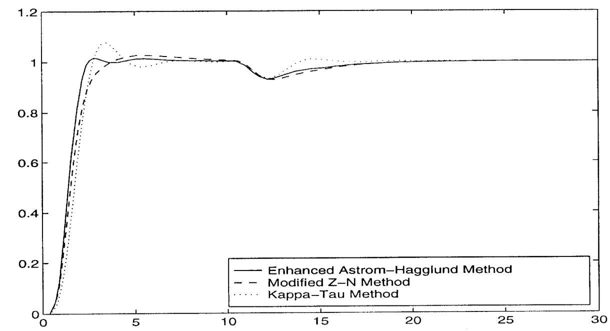



Fig. 10. PID tuning for $G = \frac{1}{(S+1)^2}e^{-0.5s}$.

number of methods have been reported [12,64–69]. Sung et al [66] used a relay test as well as a P control to identify a second-order plus dead time model and the ITAE rule to tune a PID control. In the sequel, we present frequency response fitting and model reduction methods in details.

### 5.1. Tuning via frequency response fitting

A simple but efficient solution to this kind of processes was developed [12]. It shapes the loop frequency response to optimally match the desired dynamics over a large range of frequencies. Thus the closed-loop performance is more guaranteed than one or two points PID or PI tuning laws.

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

$$H_{\mathrm{d}} = \frac{\omega_{\mathrm{n}}^2}{s^2 + 2\zeta\omega_{\mathrm{n}}s + \omega_{\mathrm{n}}^2}e^{-Ls} \tag{15}$$

where $L$ is the apparent dead-time of the process, $\omega_{\mathrm{n}}$ and $\zeta$ dominate the behavior of the desired closed-loop response. If the control specifications are given as the phase margin $\phi_{\mathrm{m}}$ and gain margin $A_{\mathrm{m}}$, $\zeta$ and $\omega_{\mathrm{n}}$ in $H_{\mathrm{d}}$ are approximately determined by

$$\zeta = \sqrt{\frac{1 - \cos^2\varphi_{\mathrm{m}}}{4\cos\varphi_{\mathrm{m}}}}$$

and

$$\omega_{\mathrm{n}} = \frac{\tan^{-1}\left(\frac{2\zeta p}{p^2 - 1}\right)}{pL},$$

where $p$ is the positive root of equation

$$(A_{\mathrm{m}} - 1)^2 = 4\zeta^2p^2 + \left(1 - p^2\right)^2.$$

The default settings for $\zeta$ 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 [12]. The open-loop transfer function corresponding to $H_d$ is

$$G_d = \frac{H_d}{1 - H_d}.$$

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

For a PID controller in the form of (15), we have

$$G(j\omega_i)\left[1 \quad \frac{1}{j\omega_i} \quad j\omega_i\right]x = G_d(j\omega_i), \quad i = 1, 2, \cdots m, \tag{16}$$

where $x = \left[K_P \quad \frac{K_P}{T_I} \quad K_P T_D\right]^T$, $m$ is chosen such that $\omega_m$ is greater than the critical frequency of $G_d$. Eq. (16) 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

$$\max_i \left|G(j\omega_i)K(j\omega_i) - G_d(j\omega_i)\right| \leqslant \varepsilon,$$

where $\varepsilon$ 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 be reduced. Then, the above procedure is repeated 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.

**Example 4** Consider an oscillatory dynamics:

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

the PID controller designed by the method is

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

and the controller by the modified Z–N rule is

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

The system responses are shown in Fig. 11. The effectiveness of this frequency response fitting method is clearly superior.

An alternative approach to tuning PID controllers using the desired closed-loop transfer function was proposed by Lee et al. [67]. They first design a IMC controller, then get its feedback equivalent and approximate it with a PID controller by matching Maclaurin series of their transfer functions.

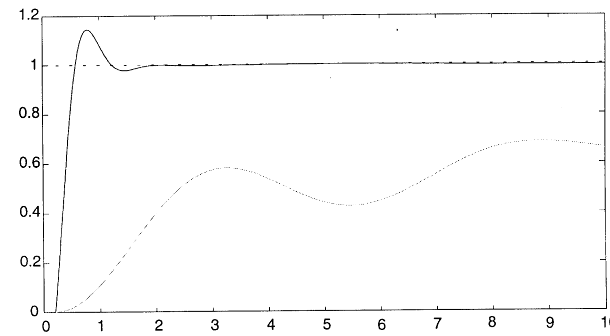


Fig. 11. Control performance for an oscillatory process. — Proposed, ··· Modified Z–N.

### 5.2. Tuning via model reduction

Model reduction was employed to tune PID controller in many works [65,69]. In this method [69], a real process $G(s)$ is first approximated by a second order plus dead time model with the following structure:

$$\hat{G}(s) = \frac{e^{-sL}}{as^2 + bs + c}, \tag{17}$$

which can represent either a monotonic or oscillatory process and yet is of sufficiently low order. Then the PID controller is designed for this process.

To determine the four unknowns, four real equations are needed and can be constructed by fitting $G(s)$ at two non-zero frequency points into (17). The two points, $s = j\omega_c$ and $s = j\omega_b$, are chosen here to satisfy $G(j\omega_c) = -\pi$ and $G(j\omega_b) = -\frac{\pi}{2}$. $L$ is then a smaller absolute root of

$$p\left(\omega_c^2 - \theta \ \omega_b^2\right)L^2 + (q\omega_c - \theta \ r\omega_b)L - \theta = 0,$$

where $p = \frac{8}{\pi^2}\left(1 - \sqrt{2}\right)$, $q = \frac{2}{\pi}\left(2\sqrt{2} - 1\right)$, $r = \frac{2}{\pi}\left(2\sqrt{2} - 3\right)$ and $\theta = \frac{\omega_c\left|G(j\omega_c)\right|}{\omega_b\left|G(j\omega_b)\right|}$.The coefficients $a$, $b$ and $c$ are obtained as

$$a = \frac{1}{\omega_c^2 - \omega_b^2}\left[\frac{\sin(\omega_b L)}{\left|G(j\omega_b)\right|} + \frac{\cos(\omega_c L)}{\left|G(j\omega_c)\right|}\right],$$

$$b = \frac{\sin(\omega_c L)}{\omega_c\left|G(j\omega_c)\right|},$$

$$c = \frac{1}{\omega_c^2 - \omega_b^2}\left[\frac{\omega_c^2 \sin(\omega_b L)}{\left|G(j\omega_b)\right|} + \frac{\omega_b^2 \cos(\omega_c L)}{\left|G(j\omega_c)\right|}\right].$$

The PID settings turn out to depend on the characteristics of (17), measured in terms of the equivalent time constant $\tau_o$ defined by

$$\frac{1}{\tau_o} = \begin{cases} \dfrac{c}{\sqrt{b^2-2ac}}, & b^2-4ac \geqslant 0; \\ \dfrac{b}{2a}, & b^2-4ac < 0, \end{cases}$$

and the damping ratio $\varsigma_o$ defined by

$$\varsigma_o = \begin{cases} \dfrac{b}{2\sqrt{ac}}, & b^2-4ac < 0; \\ 1, & b^2-4ac \geqslant 0. \end{cases}$$

Pole-zero cancellation is employed to tune the PID controller:

$$K(s) = k\left(\frac{as^2+bs+c}{s}\right),$$

where

$$k = \begin{cases} \dfrac{0.5}{L}, & \text{if } \varsigma_0 < 0.7071, \text{ or } 0.05 < \dfrac{L}{\tau_0} < 0.15, \quad \text{or } \dfrac{L}{\tau_0} > 1; \\ \min\left\{\dfrac{1}{\tau_0}e^{-\frac{L}{\tau_0}}, \dfrac{1}{eL}\right\}, & \text{otherwise} \end{cases} \tag{18}$$

**Example 5** Consider an oscillatory high order process:

$$G(s) = \frac{1}{(s^2+2s+3)(s+3)}e^{-0.3s}.$$

The process frequency response at $G(2.105j)$ = -0.061 and $G(1.101j)$ = -0.110$j$ is fitted into the second order model:

$$\hat{G}(s) = \frac{1}{3.825s^2+6.903s+9.573}e^{-0.523s}.$$

The controller adopts the structure:

$$K(s) = k\frac{3.825s^2+6.903s+9.573}{s}.$$

Since this model has $\varsigma_o = 0.570 < 0.707$, applying (18) yields

$$K(s) = 3.885 + \frac{5.388}{s} + 2.153s.$$

Ho's method gives

$$K(s) = 5.064 + \frac{5.920}{s} + 1.083s.$$

The step responses of the controllers are given in Fig. 12. The proposed method results in an improved response with smaller overshoot and shorter settling time.

This new PID controller tuning method is based on the fitting of the process frequency response to a particular second order plus dead time structure and it works for a general class of self-regulating linear processes with different dynamics. With the help of pole-zero cancellations in the model and controller, closed-loop poles are easily assigned by the conventional root locus method.

## 6. 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. For modeling, the process is first put under relay feedback control. From the test, the ultimate gain $k_u$ and the ultimate frequency $\omega_u$ 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 in form of

$$G = \frac{K}{(Ts+1)^n}e^{-Ls}, \quad n = 1, 2. \tag{19}$$

The model can be recovered from $k_u$ and $\omega_u$ by

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

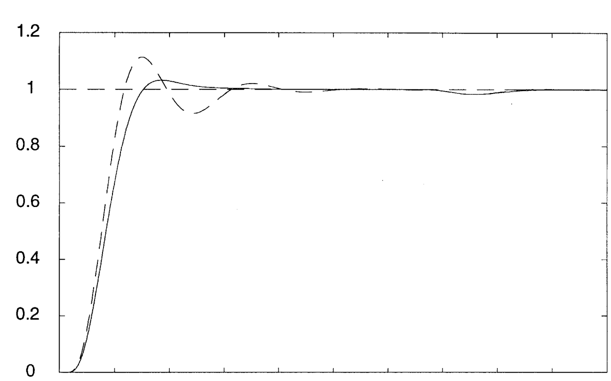



Fig. 12. Step response of the process $G(s) = \frac{1}{(s^2+2s+3)(s+3)}e^{-0.3s}$. (Solid line: Proposed, dashed line: Ho's.).

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

The order of the model can be specified by the user based on the prior knowledge of the process.

## 6.1. Smith predictor

The Smith predictor controller can be auto-tuned [21] by combining the above relay identification and a primary controller design to be described below.

The scheme of Smith predictor is shown in Fig. 13, where $G(s) = G_o e^{-Ls}$ is the process and $G(s) = G_0 e^{-Ls}$ is the model. Theoretically, the Smith predictor eliminates the dead-time from the closed-loop, and if the PI controller in the form of (11) is used as the primary controller, it can be designed based only on the delay-free part of the model $G_p$. The delay-free part is given by

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

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

$$\frac{G_0(s)K(s)}{1 + G_0(s)K(s)} = \frac{1}{1 + T_d s},$$

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

$$T_I = T, \quad K_P = \frac{T}{KT_d}$$

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

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

A simple solution is

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

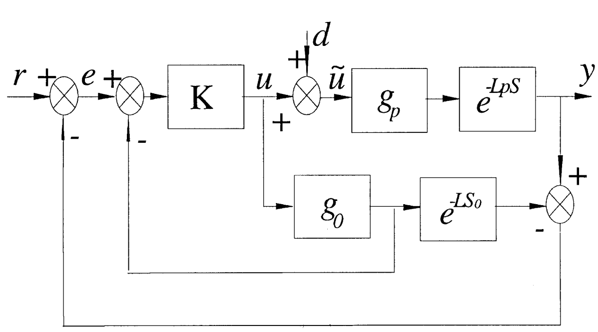



Fig. 13. Smith predictor scheme.

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

**Example 6** Consider a high vacuum distillation column which is a typical long dead-time process,

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

The model identified from relay control is

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

The Smith predictor PI controller settings according to the method are

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

The pure PI controller without dead-time compensation, which is tuned by the gain and phase margin method [16], is

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

The closed-loop responses of both methods are presented in Fig. 14. It is evident that the auto-tuned Smith predictor performs well.

This auto-tuning technique has been found to be effective even for high-order or non-minimum phase processes that exhibit apparent dead-time characteristics in their dynamics. In the presence of model uncertainty, simple approaches for robust tuning of Smith predictor can be found in Palmor and Blau [19] and Lee et al. [70].

## 6.2. FSA controller

For the less common case of unstable long dead-time processes, Smith predictor control will yield unstable systems. Then, finite spectrum assignment (FSA) method may be used, if dead-time compensation is still needed. To achieve asymptotic tracking and regulation, a modified FSA (MFSA) scheme was proposed by Wang et al. [22] and is shown in Fig. 15.

Theoretically, the FSA will eliminate the dead-time from the closed-loop characteristic equation, and the control performance is dominated by the desirable closed-loop polynomial $p(s)$. If the process has the first-order modeling (19) for $n = 1$, the generalized process is

$$\bar{g}(s) = \frac{a_0}{s(s + b_1)} e^{-Ls},$$

where $a_0 = -\frac{K}{T}$ and $b_1 = \frac{1}{T}$. Since it is a second-order system, we choose $p(s)$ of degree 2 as

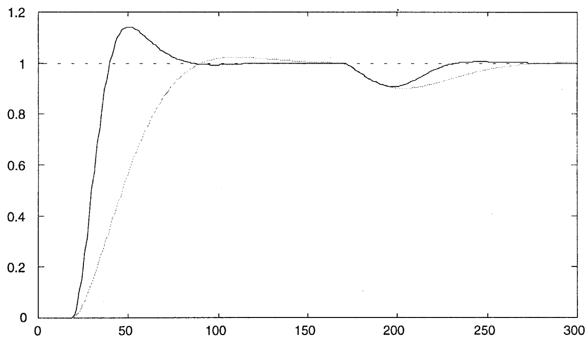



Fig. 14. Extension to Smith predictor. — Proposed, ··· pure PI.

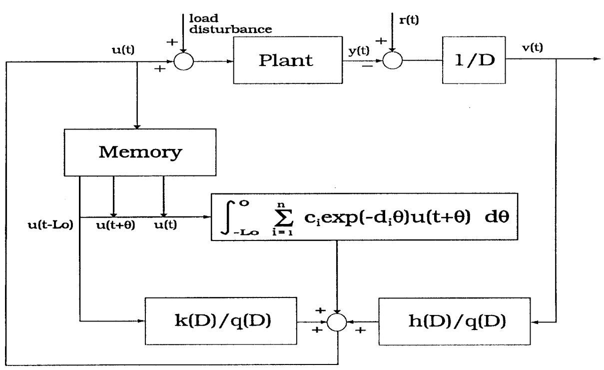



Fig. 15. FAS control scheme.

$$p(s) = s^2 + 2\varsigma\omega_0 s + \omega_0^2,$$

where $\omega_0$ and $\varsigma$ are the natural frequency and the damping factor of the desired closed-loop response, respectively. $\omega_0$ and $\varsigma$ are specified as $\frac{1}{\varsigma\omega_0} = \lambda T$ where $\lambda$ reflects the closed-loop response speed relatively to the process time constant, and it usually takes its value between 0.5 and 1.0. $\varsigma$ is chosen according to the required closed-loop overshoot and is usually in the range of 0.5–1.

In order to simplify design, we choose the observer polynomial $q(s)$ same as the process denominator:

$$q(s) = s + \frac{1}{T},$$

and set $h(s)$as $h(s) = h_1(s + b_1)$. Then, the solution to the polynomial equation

$$k(s)b(s) + h(s)a(s) = q(s)f_L(s),$$

is

$$k(s) = \left(\frac{1}{T} - 2\omega_0\varsigma + \omega_0^2 T\right)e^{-\frac{L}{T}} - \omega_0^2 T,$$

and

$$h(s) = \frac{T\omega_0^2}{K} + \frac{\omega_0^2}{K}.$$

**Example 7** Consider the non-minimum-phase system

$$\tilde{g}(s) = \frac{1 - \alpha s}{(1 + s)^3}, \qquad \alpha = 1, \ 1.5.$$

Using auto-tuning, the identified models are, respectively,

$$g(s) = \frac{1}{1 + 1.61s} e^{-2.25s}$$

for $\alpha = 1$ and

$$g(s) = \frac{1}{1 + 1.01s} e^{-2.89s}$$

for $\alpha = 1.5$. The auto-tuning and subsequent performances are shown in Fig. 16.

## 7. Multivariable processes

Auto-tuning techniques for PID controllers are very successful when the process is essentially single-input single-output (SISO). The extension of these techniques to multivariable processes is non-trivial and has attracted much attention in the literature. Luyben [71] 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. [18] and Vasani [29] proposed two relay auto-tuning methods for multi-loop PI controllers. The first method, which adopts the sequential relay tuning approach [72,73], 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. Like in SISO cases, only one point information is usually used to tune the multi-loop controllers [29,71,74,75]. 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

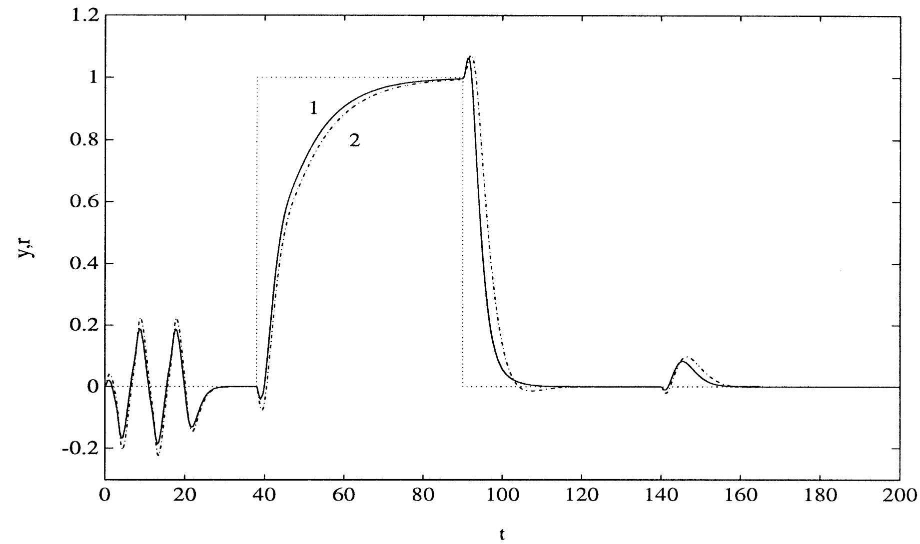



Fig 16. MFSA auto-tuning for a non-minimum phase process.

relay test is adopted. In particular, it is straightforward [12] to extend the method in Section 3.1 to the MIMO case, with the sequential relay, The arrangement is shown in Fig. 17. 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 recently developed tuning methods. One for multi-loop controllers, the other for multivariable controllers, using the identified frequency response matrix. They can achieve performance improvement over other control schemes.

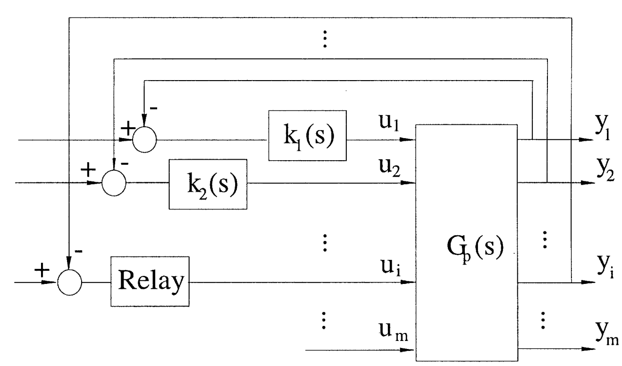



Fig. 17. Sequential relay identification.

### 7.1. 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 move to a desired position for each equivalent process. A novel approach has been developed to solve this nonlinear problem.

Consider a stable 2 by 2 process:

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

The process is to be controlled in a negative feedback configuration by the multi-loop controller:

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

The resultant control system is shown in Fig. 18. Let $k_1(s)$and $k_2(s)$ be of PID type, i.e.

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

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

The boxed portion in Fig. 18 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

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

Similarly, the equivalent process between $u_2$ and $y_2$ is given

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

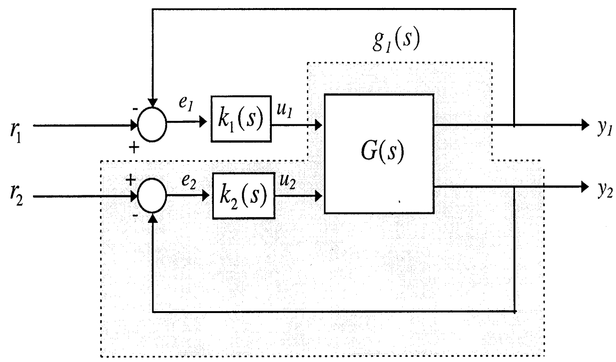



Fig 18. Multiloop control system.

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

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

is moved respectively to the points:

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

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 $k_1(s)$), $\omega_1$ and thus $k_1(s)$ (or $\omega_2$ 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. [76] for finding $k_1(s)$ and $k_1(s)$.

**Example 8** The 24 tray tower separating methanol and water has the following transfer function matrix:

$$G(s) = \begin{bmatrix} \dfrac{-2.2e^{-s}}{16.7s+1} & \dfrac{-1.3e^{-0.3s}}{21.0s+1} \\ \dfrac{-2.8e^{-1.8s}}{9.5s+1} & \dfrac{4.3\mathrm{e}^{-0.35s}}{9.2s+1} \end{bmatrix}.$$

As all its elements are of first-order in nature, we use a simple PI controller. The above method gives

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

The step responses of the resultant feedback system to set-point changes followed by load disturbance changes are shown in Fig. 19. Load disturbance changes of 0.5 and 1 are applied directly on the two process inputs, $u_1$ and $u_2$, at different time. The above PID controller gives better loop and decoupling performance than the well known BLT method [71].

## 7.2. Multivariable controllers

The design method in Subsection 5.1 can be extended [12] to the multivariable case. Let $G(s)$ be the process transfer function matrix,. The multivariable controller is chosen as PID type:

$$K(s) = K_{\mathrm{P}} + \frac{1}{s}K_I + sK_{\mathrm{D}}.$$

Assume that the desired closed loop transfer function matrix $H$ is

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

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

$$GK = G\begin{bmatrix} I & \frac{1}{s} & sI\end{bmatrix}\begin{bmatrix} K_{\mathrm{P}} \\ K_{\mathrm{I}} \\ K_{\mathrm{D}}\end{bmatrix} = H[I-H]^{-1},$$

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 this method gives very satisfactory results for most processes. In some special cases of large interaction, more than one stage of compensators may be employed to enhance the control performance.

**Example 9** Consider again the Wood and Berry's binary distillation column process:

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

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 biggest log modulus tuning (BLT) [71] is

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

The closed loop responses in Fig. 20 illustrate the decoupling effectiveness of the auto-tuned multivariable

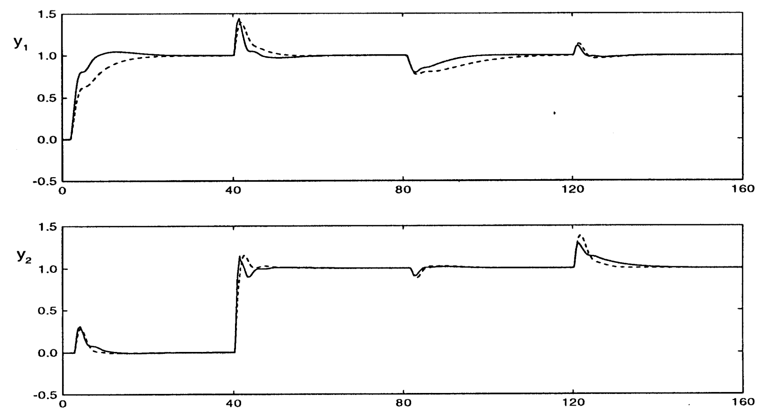



Fig. 19. Multiloop control system step responses.

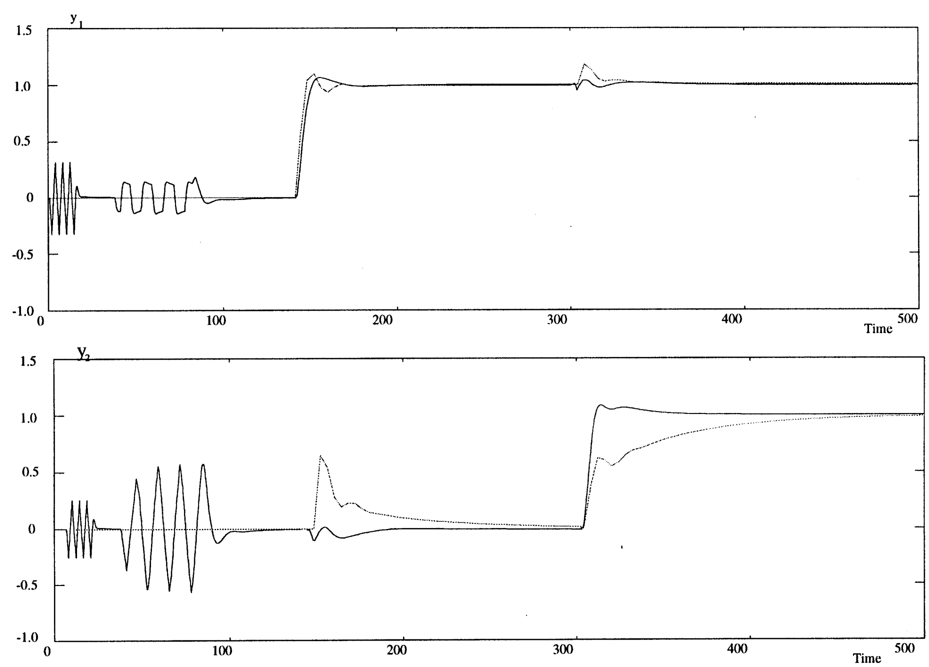



Fig. 20. System performance of multivariable control. — Proposed method, - - - - BLT method.

controller. The relay tuning parts of the responses are also shown in this figure to give a more complete account of the entire auto-tuning process.

Recently, a method for auto-tuning fully cross-coupled multivariable PID controllers from decentralized relay feedback was proposed for multivariable processes with significant interactions [12]. It was shown [12] 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 was 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.

## 8. Conclusions

The relay feedback auto-tuning technique has been widely used to automatically tune PID controllers and initialize adaptive controllers. The standard relay tuning technique has been successfully modified and extended to auto-tune advanced controllers. The relay-FFT technique that can be used to identify multi-points on the frequency response is most promising as it can be used to auto-tune dead-time compensators and multivariable controllers. This paper takes stock of these recent developments and extensions of the relay feedback auto-tuning technique. It is evident that this tuning 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 such as the drive towards field-bus and distributed control.

## Acknowledgements

The authors are grateful to Qiang Bi, Xueping Yang and Chang-En Li for their assistance in preparing this paper.

## References

[1] K.J. Astrom, T. Hagglund, Automatic tuning of PID controllers, Instrument Society of America, 1988.
[2] C.C. Hang, K.K. Sin, A comparative performance study of PID auto-tuners, IEEE Control Systems 11 (5) (1991) 41–47.
[3] C.C. Hang, K.J. Astrom, W.K. Ho, Relay auto-tuning in the presence of static load disturbance, Automatica 29 (2) (1993a) 563–564.
[4] K.J. Astrom, T. Hagglund, Automatic tuning of simple controllers with specification on phase and amplitude margins, Automatica 20 (5) (1984) 645–651.
[5] T. Hagglund, K.J. Astrom, Industrial adaptive controllers based on frequency response techniques, Automatica 27 (1991) 599–609.
[6] K.J. Astrom, K.H. Panagopoulos, T. Hagglund, Design of PI controllers based on non-convex optimisation, Automatica 34 (5) (1998) 585–601.
[7] A. Leva, A.M. Colombo, Method for optimising set-point weights in ISA-PID, IEE Proc. Pt. D 146 (2) (1999) 137–146.
[8] T.H. Lee, Q.G. Wang, K.K. Tan, A modified relay-based technique for improved critical point estimation in process control, IEEE Trans. Control Systems Tech. 3 (3) (1995c) 330–337.
[9] S.W. Sung, J.H. Park, I. Lee, Modified relay feedback method, Ind. Eng. Chem. Res. 34 (1995) 4133–4135.
[10] S. Shen, H. Yu, C. Yu, Use of saturation-relay feedback for autotune identification, Chem. Eng. Sci. 51 (1996a) 1187–1198.
[11] Q.G. Wang, C.C. Hang, B. Zou, Low-order modelling from relay feedback, Industrial and Engineering Chemistry Research 36 (2) (1997) 375–381.
[12] Q.G. Wang, C.C. Hang, B. Zou, A frequency response approach to auto-tuning of multivariable controllers, Chemical Engineering Research and Design 75 (1997) 797–806.
[13] T.S. Schei, A method for closed loop automatic tuning of PID controllers, Automatica 28 (2) (1992) 587–591.
[14] A. Leva, PID auto-tuning algorithm based on relay feedback, IEE Proc. Pt. D 140 (5) (1993) 328–338.
[15] W.K. Ho, C.C. Hang, L.S. Cao, Tuning of PID controllers based on gain and phase margin specifications, Proceedings of the 12th IFAC World Congress 5 (1993) 267–270.
[16] W.K. Ho, C.C. Hang, L.S. Cao, Tuning of PID controllers based on gain and phase margin specifications, Automatica 31 (3) (1995) 497–502.
[17] K.K. Tan, T.H. Lee, Q.G. Wang, An enhanced automatic tuning procedure for PI/PID controllers for process control, AIChE Journal 42 (9) (1996) 2555–2562.
[18] C.C. Hang, A.P. Loh, V.U. Vasnani, Relay feedback auto-tuning of cascade controllers, IEEE Trans. on Control Syst. Tech. 2 (1) (1994) 42–45.
[19] Z.J. Plamor, M. Blau, An auto-tuner for Smith dead time compensator, Int. J. Control 60 (1) (1994) 117–135.
[20] T.H. Lee, Q.G. Wang, K.K. Tan, Automatic tuning of the Smith-predictor controller, J. Systems Engng. 5 (1995a) 102–114.
[21] C.C. Hang, Q.G. Wang, L.S. Cao, Self-tuning Smith predictors for processes with long dead time, Int. J. Adeptive control and Signal Processing 9 (3) (1995) 255–270.
[22] Q.G. Wang, T.H. Lee, K.K. Tan, Automatic tuning of finite spectrum assignment controllers for delay systems, Automatica 31 (3) (1995) 477–482.
[23] K.J. Astrom, J.J. Anton, K.E. Arzen, Expert control, Automatica 22 (3) (1986) 227–286.
[24] K.J. Astrom, C.C. Hang, P. Person, W.K. Ho, Towards intelligent PID control, Automatica 28 (1) (1992) 1–9.
[25] K.J. Astrom, T. Hagglund, C.C. Hang, W.K. Ho, Automatic tuning and adaptation for PID controllers — a survey, Control Eng. Practice 1 (4) (1993) 699–714.
[26] K.J. Astrom, T.J. McAvoy, Intelligent control, J. Proc. Cont. 2 (3) (1993) 115–127.
[27] T.H. Lee, Q.G. Wang, K.K. Tan, Knowledge-based process identification using relay feedback, Journal of Process Control 5 (6) (1995b) 387–397.
[28] C.C. Hang, T.T. Tay, V.U. Vasnani, Auto-tuning of multivariable decoupling controllers, IFAC Symposium on Intelligent Tuning and Adaptive Control, 1991, pp. 115-120.
[29] Vasani, V. Towards Relay Feedback Auto-tuning of Multiloop Systems, *PhD. thesis,* Department of Electrical Engineering, National University of Singapore, 1994.
[30] K.J. Astrom, T. Hagglund, Automatic tuning of simple controllers, in: Proceedings of the 9th IFAC World Congress, Budapest, 1984, pp. 1867–1872.
[31] G.H. Cohen, G.A. Coon, Theoretical consideration of retarded control, Trans. ASME 75 (1953) 827–834.
[32] E.H. Bristol, Pattern recognition: an alternative to parameter identification in adaptive control, Automatica 13 (1977) 197–202.
[33] E.H. Bristol, The design of industrially useful adaptive controllers, ISA Trans. 22 (3) (1983) 17–25.
[34] M. Yuwana, D.E. Seborg, A new method for on-line controller tuning, AIchE J. 28 (1982) 434.
[35] B. Wittenmark, K.J. Astrom, Simple self-tuning controllers, in: H. Unbehauen (Ed.), Methods And Applications In Adaptive Control. Springer, Berlin, 1980.
[36] P.J. Gawthrop, Self-Tuning PI and PID Controllers. Proceeding Functions and Nonlinear Systems Design, McGraw Hill, New York, 1982.
[37] K.J. Astrom, B. Wittenmark, Adaptive Control, Addison-Wesley, 1989.
[38] Y.Z. Tsypkin, Relay Control Systems, Cambridge University Press, Cambridge, UK, 1984.
[39] M. Lundh, K.J. Astrom, Automatic initialization and robust adaptive controllers, ISA Transactions 37 (1998) 123–131.
[40] S. Shen, J. Wu, C. Yu, Autotune identification under load disturbance, Ind. Eng. Chem. Res. 35 (1996b) 1642–1651.
[41] J.H. Park, S.W. Sung, I. Lee, Improved relay auto-tuning with static load disturbance, Automatica 33 (1997) 711–715.
[42] J.G. Ziegler, N.B. Nichols, Optimum settings for automatic controllers, Trans. ASME 64 (1942) 759–768.
[43] C.C. Hang, K.J. Astrom, Practical aspects of PID auto-tuners based on relay feedback, in: Preprints IFAC Int. Symposium On Adaptive Control Of Chemical Processes, ADCHEM '88, Lyngby, Denmark, 1988.

[44] C.C. Hang, K.J. Astrom, W.K. Ho, Refinements of the Ziegler–Nichols tuning formula, IEE Proceedings-D 138 (2) (1991) 111–118.
[45] R.W. Bass, Mathematical legitimacy of equivalent linearization by describing function, in: Proc. IFAC, Moscow, 1960, pp. 2074-2080.
[46] D.P. Atherton, Nonlinear Control Engineering — Describing Function Analysis and Design, Van Nostrand Reinhold, Wokingham, Berkshire, 1975.
[47] Y.H. Kim, PI controller tuning using modified relay feedback method, J. of Chem. Eng. of Japan 28 (1995) 118–121.
[48] M. Friman, K.V. Waller, A two-channel relay for autotuning, Ind. Eng. Chem. Res. 36 (1997) 2662–2671.
[49] K.K. Tan, Q.G. Wang, T.H. Lee, Finite spectrum assignment control of unstable time delay processes with relay tuning, Industrial and Engineering Chemistry Research 37 (4) (1998) 1353–1357.
[50] R. Chang, S. Shen, C. Yu, Derivation of transfer function from relay feedback systems, Ind. Eng. Chem. Res. 31 (1992) 855–860.
[51] H. Huang, C. Chen, C. Lai, G. Wang, Autotuning for model-based PID controllers, AIChE J. 42 (1996) 2687–2691.
[52] Q.G. Wang, C.C. Hang, Q. Bi, Process frequency response estimation from relay feedback, Control Engineering Practice 5 (9) (1997) 1293–1302.
[53] S.W. Sung, I. Lee, New process identification method for automatic design of PID controllers, Automatica 34 (1998) 513–520.
[54] J.H. Park, I. Lee, An enhanced process identification method for automatic tuning of PID controller, Proceedings of the 14th IFAC World Congress, Beijing, China, 1999, pp. 721-276.
[55] G.D. Bergland, A guided tour of the fast fourier transform, IEEE Spectrum (July 1969), 41-52.
[56] R.W. Ramirez, The FFT fundamentals and concepts, Prentice-Hall, 1985.
[57] Q.G. Wang, C.C. Hang, Q. Bi, A frequency domain controller design method, Chemical Engineering Research and Design 75 (1997) 64–72.
[58] C.C. Hang, D. Chin, Reduced order process modelling in self-tuning control, Automatica 27 (3) (1991) 529–534.
[59] W.L. Luyben, Derivation of transfer functions for highly non-linear distillation columns, Ind. Eng. Chem. Res. 26 (1987) 2490–2495.
[60] K.J. Astrom, T. Hagglund, New tuning methods for PID controllers, in: Proc Europen Control Conference, Rome, Italy, 1995, pp. 2456-2462.
[61] K.J. Astrom, T. Hagglund, PID Controllers: Theory, Design, and Tuning, 2nd edition, Instrument Society of America, Research Triangle Park, NC, 1995.
[62] K. Ogata, Modern Control Engineering, 2nd Edition, Prentice-Hall, Englewood Cliffs, NJ, 1990.
[63] G.F. Franklin, J.D. Powell, A.E. Baeini, Feedback Control of Dynamic Systems, Addison-Wesley, Reading, MA, 1986.
[64] T.J. Barnes, L. Wang, W.R. Cluett, A frequency domain design method for PID controllers, in: Proc. of American Control Conference, San Francisco, California, 1993, pp. 890-894.
[65] S.W. Sung, I. Lee, Limitations and countermeasures of PID controllers, Ind. Eng. Chem. Res. 35 (1996a) 2596–2610.
[66] S.W. Sung, J. O, I. Lee, J. Lee, S. Yi, Automatic tuning of PID controller using second-order plus time delay model, J. Chem. Eng. Japan 29 (1996b) 990–999.
[67] Y. Lee, M. Lee, S. Park, C. Brosilow, PID controller tuning for desired closed-loop responses for SI/SO systems, AIChE J. 44 (1998) 106–115.
[68] Y. Lee, S. Park, PID controller tuning to obtain desired closed-loop responses for cascade control systems, Ind. Eng. Chem. Res. 37 (1998) 1859–1865.
[69] Q.G. Wang, H.W. Fung, T.H. Lee, PID tuning for improved performance, accepted by IEEE Trans. Control Systems Tech., 1999.
[70] D. Lee, M. Lee, S.W. Sung, I. Lee, Robust PID tuning for Smith predictor in the presence of model uncertainty, J. Process Control 9 (1999) 79–85.
[71] W.L. Luyben, A simple method for tuning SISO controllers in a multivariable system, Ind. Eng. Chem. Proc. Des. Dev. 25 (1986) 654–660.
[72] A.P. Loh, C.C. Hang, C.K. Quek, Vsanani, Autotuning of multiloop proportional-integral controllers using relay feedback, Ind. Eng. Chem. Res. 32 (1993) 1102–1107.
[73] S. Shen, C. Yu, Use of relay-feedback test for automatic tuning of multivariable systems, AIChE J. 40 (1994) 627–646.
[74] Z.J. Plamor, Y. Halevi, N. Krasney, Automatic tuning of decentralized PID controllers for TITO processes, Proceedings of IFAC 12th World Congress, Sydney, Australia 2 (1993) 311–314.
[75] M. Friman, K.V. Waller, Autotuning of multiloop control systems, Ind. Eng. Chem. Res. 33 (1994) 1708–1717.
[76] Q.G. Wang, T.H. Lee, Yu Zhang, Multi-loop version of the modified Ziegler–Nichols method for TITO processes, Industrial and Engineering Chemistry Research 37 (12) (1998) 4725–4733.
[77] N. Morrison, Introduction to Fourier Analysis. John Wiley & Sons Inc., New York, 1994.