Pergamon

PII: S0005-1098(96)00177-X

Automatica, Vol. 33, No. 3, pp. 319–330, 1997

Printed in Great Britain
0005-1098/97 $17.00 + 0.00

Auto-tuning of Multivariable PID Controllers from Decentralized Relay Feedback*

QING-GUO WANG,† BIAO ZOU,‡ TONG-HENG LEE† and QIANG BI†

A method for automatically tuning fully cross-coupled multivariable PID controllers from decentralized relay feedback is developed with new techniques for process frequency-response matrix estimation and multivariable decoupling design.

Key Words—Auto-tuning; multivariable systems; decentralized relay feedback; PID control; frequency response.

Abstract—A method for auto-tuning *fully cross-coupled multivariable* PID controllers from *decentralized relay feedback* is proposed for multivariable processes with significant interactions. Multivariable oscillations under decentralized relay feedback are first investigated, and, in particular, it is shown that for a stable $m \times m$ process the oscillation frequency will remain almost unchanged under relatively large relay amplitude variations. Therefore m decentralized relay feedback tests are performed on the process, with their oscillation frequencies close to each other so that the process frequency-response matrix can be estimated at that frequency. A bias is further introduced into the relay to additionally obtain the process steady-state matrix. For multivariable controller tuning, a new set of design equations are derived under the decoupling condition where the equivalent diagonal plants are independent of off-diagonal elements of the controller, and are used to design the controller diagonal elements first. The PID parameters of the controllers are determined individually by solving these equations at the oscillation and zero frequencies. The proposed method is applied to various typical processes, where significant performance improvement over the existing tuning methods is demonstrated. © 1997 Elsevier Science Ltd.

1. INTRODUCTION

The paper, *for the first time,* presents a decentralized relay auto-tuner for fully cross-coupled multivariable PID controllers. It is well known that PID controllers have dominated applications for 50 years, though there has been a lot of interest in research into and implementation of advanced controllers. Relay feedback is a simple and reliable test that keeps the process output under closed-loop control and makes it close to the operating point. Åström and Hagglund (1984) combined the strengths of both PID and relay and invented the relay auto-tuner for SISO PID controller. The tuner has been widely and successfully applied in industry for over 10 years (Åström and Hagglung, 1984). This has naturally led to attempts (Zgorzelski *et al.*, 1990; Loh *et al.*, 1993; Palmor *et al.*, 1993; Zhuang and Atherton, 1994) to develop its counterpart for multivariable PID controllers, since many industrial processes are inherently of multivariable nature and need multivariable control to enhance performance. The demand for such a tuner is evidenced by the latest survey (Kong *et al.*, 1995) from the leading controller manufacturers such as Fish-Rosemount, Yokogawa and Foxboro, all of whom ranked poor decoupling as the principal common control problem in industry. A multivariable PID controller is readily implementable within the existing control hardware and software and incurs no more cost. With proper tuning, it may well improve productivity and profitability of plants.

Multivariable control has developed over a number of years (Rosenbrock, 1969; Macfarlane and Kouvaritakis, 1977; Maciejowski, 1989). One may note that most designs, such as state-space, optimization, inverse Nyquist array and characteristic locus methods, need a full model of the process in a form of a state-space description or a transfer-function matrix or a frequency-response matrix over the entire working frequency range, and such models are usually not available in the content of relay auto-tuning, where in general only partial dynamic information on the process in the form of a frequency-response matrix at one or two points may be obtained. These methods can hardly lead

* Received 7 November 1995; revised 1 June 1996; received in final form 27 August 1996. This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor Guy Dumont under the direction of Editor C. C. Hang.
Corresponding author Professor Qing-Guo Wang. Tel. +65 772 2282; Fax +65 779 1103; E-mail elewqg@nus.sg.

† Department of Electrical Engineering, National University of Singapore, 10 Kent Ridge Crescent, Singapore 119260.

‡ Honeywell Hi-Spec Solutions Asia Pacific, Singapore 307591.

to PID-type controllers. PID control is always preferred to more advanced controllers in practical applications (Åström and Hang, 1991; Luyben, 1990) unless evidence is given or accumulated showing that PID control is not adequate and cannot meet specifications. For multivariable PID control, Koivo and Tanttu (1991) gave a recent survey of its tuning techniques. These techniques mainly aim at decoupling the plant at certain frequencies. A software package that includes several methods for tuning MIMO PI/PID controllers as well as an expert system to assist the user has also been developed (Lieslehto *et al.*, 1993). However, these methods are both manual and time-consuming in nature. In addition, they also require detailed dynamic models of the multivariable process, and are therefore not suited for on-line tuning. In the context of relay auto-tuning of multivariable PID controllers, Luyben (1986) presented a tuning procedure for decentralized PID controllers from independent single-loop relay tests, where the stability of the whole system can only be guaranteed by introducing appropriate detuning factors on the PI/PID parameters. Loh *et al.* (1993) proposed a tuning method that is a combination of sequential loop closing and relay tuning. Palmor *et al.* (1993) and Zhuang and Atherton (1994) proposed design algorithms for 2×2 plants from decentralized relay feedback, with the decentralized PID controllers then being obtained from the resulting oscillations.

Decentralized relay feedback (DRF) is a *complete* closed-loop test while *independent single-relay feedback* (IRF) and *sequential relay feedback* (SRF) are only partial closed-loop tests. Closed-loop testing is preferred to open-loop testing (Åström and Hogglund 1988), since the former causes less perturbation to the process and makes a linear model valid. Decentralized relay feedback can be employed to effectively excite and identify multivariable process interaction, and to make decoupling design possible, while IRF and SRF may not excite multivariable interaction directly. To the best of our knowledge, all the existing works using DRF (Plamor *et al.*, 1993; Zhuang and Atherton 1994) can be used to tune *only* a multiloop (i.e. decentralized) PID controller, and thus they are suitable for processes with modest interactions. *Fully cross-coupled multivariable controllers are necessary for processes with significant interactions where the performance by decentralized control may be too poor to meet specifications, or the processes may not even be stabilized by decentralized control.*

In this paper, a method for auto-tuning fully cross-coupled multivariable PID controllers from decentralized relay feedback is proposed for multivariable processes with significant interactions. Multivariable oscillations under decentralized relay feedback are first investigated, and, in particular, it is shown that for a stable $m \times m$ process the oscillation frequencies will remain almost unchanged under relatively large relay amplitude variations. Therefore m decentralized relay feedback tests are performed on the process, with their oscillation frequencies close to each other so that the process frequency-response matrix can be estimated at that frequency. A bias is further introduced into the relay to additionally obtain the process steady-state matrix. For multivariable controller tuning, a new set of design equations are derived under the decoupling condition where the equivalent diagonal plants are independent of off-diagonal elements of the controller and are used to design the controller's diagonal elements first. The PID parameters of the controllers are determined individually by solving these equations at the oscillation and zero frequencies. The proposed method has been applied to various typical processes, and significant performance improvement over the existing tuning methods substantiated.

The paper is organized as follows. In Section 2, the properties of multivariable process under decentralized relay feedback are investigated. Section 3 deals with frequency-response identification from the decentralized relay feedback tests, and the design method for multivariable PID controllers is presented in Section 4. Examples are given in Section 5, followed by conclusions in Section 6.

2. DECENTRALIZED RELAY FEEDBACK

Auto-tuning of controllers requires some information on the process dynamics, and this may be obtained by injecting a test signal into the process. Åström and Hagglund (1984, 1988) proposed a relay feedback auto-tuning technique that can approximately determine the ultimate gain and frequency of the process. Their technique is greatly welcome in applications, since it is a closed-loop method. The process will not drift away from the set point, the test time is short, and it is easy to control the amplitude of the output oscillation.

When the relay technique is extended to a MIMO system, there are three possible relay feedback schemes.

(i) *independent single-relay feedback*: only one loop at a time is subject to relay feedback, while all others are kept open;

(ii) *sequential relay feedback*: a loop is closed with a simple controller once a relay test had been done to that loop; this will be repeated until all the loops are performed;

(iii) *decentralized relay feedback*: loops are placed on relay feedback simultaneously, as shown in Fig. 1.

Of the three relay feedback schemes, decentralized relay feedback is the most desirable and will be used as our test for the process frequency-response matrix estimation in the next section. Note that DRF is a *complete* closed-loop test, meaning that for an $m \times m$ plant at any instant during a test, all the m outputs are simultaneously under feedback control, while IRF and SRF are only partial closed-loop tests. For IRF, only one loop is closed, with the other $m-1$ open. For *sequential relay feedback* (SRF), at the ith test, i loops are closed, with $m-i$ loops open. Closed-loop testing is preferred to open-loop testing (Åström and Hagglund, 1988), since a closed-loop test keeps outputs close to the set points, so that it causes less perturbation to the process and makes a linear model (such as frequency response or transfer function) valid. In addition, we employ DRF to effectively excite and identify multivariable process interaction, and also use the interaction to design a decoupling controller, while IRF and SRF may not excite multivariable interaction directly, and it is difficult to tune a fully cross-coupled multivariable PID tuning with IRF or SRF.

If an $m \times m$ process is controlled by decentralized relay feedback, its outputs usually oscillate in the form of limit cycles after an initial transient. Each output has its own oscillation frequency ω_{ic}, $i = 1, 2, \ldots, m$, and they may be different. For instance, a 2×2 process consisting of two independent (or very little coupled) but different loops has different output oscillation frequencies. However, it was found by Atherton (1975) that for typical coupled multivariable processes, m outputs normally have the same oscillation frequencies, that is, $\omega_{1c} = \omega_{2c} = \ldots \omega_{mc}$, but different phases. For ease of reference later, we call this kind of multivariable oscillations *oscillations of a common frequency* and the frequency as a *process critical frequency*, denoted by ω_c.

The describing-function method was extended by Loh and Vasnani (1994) to analyze multivariable oscillations under decentralized relay feedback. In this context, it is assumed that the m-input, m-output process has low-pass characteristics in each element of its transfer-function matrix and one of its characteristic loci has at least 180° phase lag. Let the relay output amplitudes be d_i, the inputs of the relays have amplitudes a_i, and $N(a, d) = \text{diag}\{4d_i/\pi a_i\}$ be the describing-function matrix of the decentralized relay controller.

Lemma 1. (Loh and Vasnani (1994).) If the decentralized relay feedback system oscillates at a common frequency then at least one of the characateristic loci of $N(a, d)G(j\omega)$ crosses the point $-1 + j0$ on the complex plane, and the oscillation frequency corresponds to the frequency at which the crossing occurs. Further, if the process is stable then the limit cycle oscillation is stable, the outermost characteristic locus of $N(a, d)G(j\omega)$ passes through the point $-1 + j0$ and the process critical frequency is the same as the critical frequency of the outermost characteristic locus.

It should be noted that the crossing condition and the oscillation frequency in Lemma 1 are related to $N(a, d)$, and the latter cannot be calculated until the oscillations are observed and the amplitudes a_i of relay inputs measured from the oscillation waveforms. It would be useful if the frequency were given in terms of the information on the process only but independently of the relay controller. To this end, consider an $m \times m$ multivariable process $G(s)$ with row Gershgorin bands as shown in Fig. 2. For each band, let $c_{i1} = g_{ii}(\omega_{i1})$ and $c_{i2} = g_{ii}(\omega_{i2})$ be the centers of circles that are tangential to the negative real axis, and let $-\beta_{i1} + j0$ and

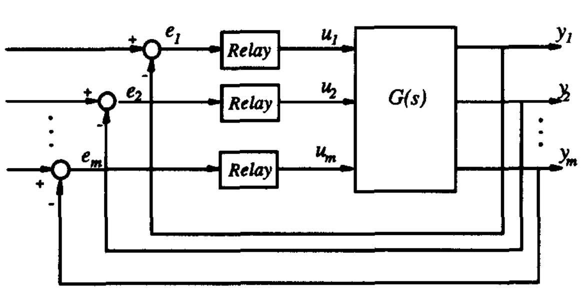

Fig. 1. Decentralized relay test.

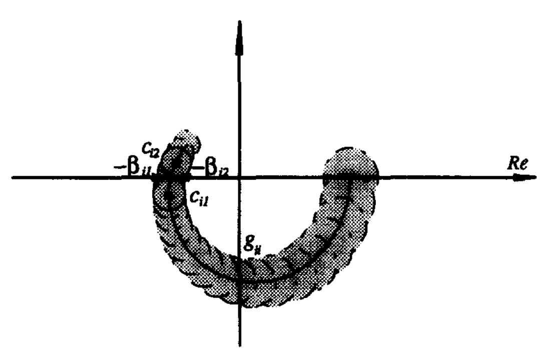

Fig. 2. Gershgorin band.

$-\beta_{i2}+j0$ be the points at which the outer rim and inner rim respectively of the ith Gershgorin band intersect the negative real axis. If the ith Gershgorin band does not intersect the negative real axis, $[\omega_{i1}, \omega_{i2}]$ is defined to be empty. The following result gives an estimate of ω_c in terms of ω_{i1} and ω_{i1}.

Proposition 2. If the decentralized relay feedback system oscillates at a common frequency ω_c, then there exists a $k \in \{1, 2, \ldots, m\}$ such that $\omega_c \in [\omega_{k1}, \omega_{k2}]$.

Proof. By the Gershgorin theorem (Maciejowski, 1989), we know that the characteristic loci of $G(s)$ lie in the union of its Gershgorin bands. The point at which the ith characteristic locus $\lambda_i(j\omega)$ of G crosses the negative real axis, if it exists, can only lie in the union of circles with centers from c_{i1} to c_{i2}. It follows that for $\lambda_i(j\omega)$ the critical frequency ω_{ic} at which the crossing occurs is in the range $[\omega_{i1}, \omega_{i2}]$. Suppose now that that the transfer-function matrix $G(s)$ is multiplied by a diagonal constant matrix $K = \operatorname{diag}\{k_i\}$ as

$$Q = KG = [k_1 g_{1.} \quad \ldots \quad k_i g_{i.} \quad \ldots \quad k_m g_{m.}]^{\mathrm{T}}$$

where $g_{i.}$, $i = 1, \ldots, m$ are the row vectors of $G(s)$. The centers of the circles for the ith Gershgorin band of Q have now been shifted to $k_i g_{ii}$, with their radii magnified k_i times, as shown in Fig. 3. Since k_i is constant, the center $k_i g_{ii}(\omega_{i1})$ has the same phase as $c_{i1} = g_{ii}(\omega_{i1})$ and is on the straight line drawn through the origin and $g_{ii}(\omega_{i1})$. Furthermore, the magnitude $|k_i g_{ii}(\omega_{i1})|$ differs from $|g_{ii}(\omega_{i1})|$ by a factor of $|k_i|$. Therefore, the distance between the point $k_i g_{ii}(\omega_{il})$ and the negative real axis is $|k_i|$ times as large as that between the point $g_{ii}(\omega_{i1})$ and the axis, which is exactly the radius of the circle with the center $k_i g_{ii}(\omega_{i1})$. This implies that this circle is still tangential to the negative real axis, and thus $\tilde{\omega}_{i1}$ for Q is equal to ω_{i1} for G. The same can be said for c_{i2} and $\tilde{\omega}_{i2} = \omega_{i2}$. It follows that the critical frequency $\tilde{\omega}_{ic}$ for the ith characteristic locus of $Q(s)$ is still in $[\omega_{i1}, \omega_{i2}]$. Since the describing matrix $N(a, d)$ is also a constant diagonal matrix, the critical frequency for the ith characteristic locus of $N(a, d)G(s)$ is thus in $[\omega_{i1}, \omega_{i2}]$. By Lemma 1, the limit-cycle oscillation frequency must be in one of $[\omega_{i1}, \omega_{i2}]$, $i = 1, 2, \ldots, m$, and our result follows. □

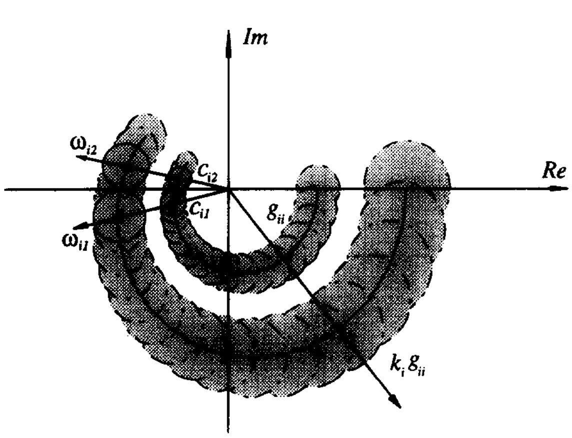

Fig. 3. Expansion of a Gershgorin band.

In view of Lemma 1 and Proposition 2, the oscillation frequency ω_c for a stable process depends on which characteristic locus of $G(s)$ is moved outermost by the multipliction of the corresponding relay element describing function $N_i = 4d_i/\pi a_i$. In general, one can enlarge the gain N_i by increasing the ratios of the relay amplitudes in the ith loop to those in other loops. We call this outermost loop the *dominant* loop. It should be noted that the dominant loop remains dominant and the critical frequency varies very little with a fairly large change of relay amplitude ratios unless an inner characteristic locus becomes a new outermost. As an example, consider the following typical process (Wood and Berry, 1973):

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

Let d_1 and d_2 be the relay amplitudes in loops 1 and 2 respectively. When $r := d_2/d_1$ varies from 1 to 2 by 100%, the process exhibits oscillations with a common frequency, and the process critical frequency ω_c changes from 0.494 to 0.496 by 0.4%. This feature is addressed in the following proposition.

Proposition 3. If the decentralized relay feedback system for a stable process oscillates at a common frequency and for some k, $N_k > N_i\beta_{i1}/\beta_{k2}$, $i = 1, 2, \ldots, m$, $i \neq k$, then only the kth characteristic locus of $N(a, d)G(j\omega)$ crosses the point $-1 + j0$ and the oscillation frequency satisfies $\omega_c \in [\omega_{k1}, \omega_{k2}]$.

Proof. The conditions $N_k > N_i\beta_{i1}/\beta_{k2}$, $i = 1, 2, \ldots, m$, $i \neq k$, guarantee that the kth Gershgorin band of $N(a, d)G(j\omega)$ is the outermost of all the m Gershgorin bands. Since the kth characteristic locus of $N(a, b)G(j\omega)$ is in this band, it is the outermost locus of

$N(a, b)G(j\omega)$. it follows from Lemma 1 that the kth characteristic locus of $N(a, d)G(j\omega)$ crosses the point $-1 + j0$ and the oscillation frequency ω_c is equal to ω_{kc}, which is in $[\omega_{k1}, \omega_{k2}]$. □

By Proposition 3, if we vary relay amplitudes such that the resulting describing-function gain matrix $N'(a, d)$ still satisfies $N'_k > N'_i\beta_{i1}/\beta_{k2}$, $i = 1, 2, \ldots, m$, $i \neq k$, then the resulting limit-cycle oscillation frequency is expected to be in the range $[\omega_{k1}, \omega_{k2}]$, and thus close to the previous value if the interval $[\omega_{kl}, \omega_{k2}]$ is small. In general, the conditions $N_k > N_i\beta_{i1}/\beta_{k2}$, $i = 1, 2, \ldots, m$, $i \neq k$, remain true if one increases the relay amplitude of the dominant loop or decreases one or more relay amplitudes among other loops. It is the feature of approximate invariance of oscillation frequency under substantial relay amplitude changes that enables us in the next section to conduct a series of decentralized relay tests on the process and estimate the multivariate process frequency response at the critical point.

3. ESTIMATION OF PROCESS FREQUENCY RESPONSE

An $m \times m$ multivariable process can be described in the frequency domain as

$$\begin{bmatrix} y_1(j\omega) \\ \vdots \\ y_m(j\omega) \end{bmatrix} = \begin{bmatrix} g_{11}(j\omega) & \cdots & g_{1m}(j\omega) \\ \vdots & \ddots & \vdots \\ g_{m1}(j\omega) & \cdots & g_{mm}(j\omega) \end{bmatrix} \times \begin{bmatrix} u_1(j\omega) \\ \vdots \\ u_m(j\omega) \end{bmatrix}. \quad (1)$$

We here want to estimate the process frequency response $G(j\omega)$ at the critical oscillation frequency ω_c for controller tuning. To identify the steady-state gain matrix of the process additionally, a biased relat as shown in Fig. 4 instead of a standard relay is used in the dominant loop to make the process inputs and outputs have non-zero means. Thus, a test in Fig. 1 with a biased relay in the dominant loop and

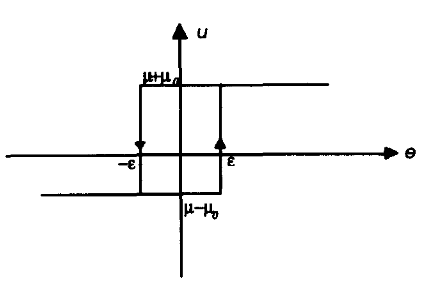

Fig. 4. Biased relay.

standard relays in other loops is applied to the process. One waits for the process to reach stationarity. The process stationary inputs $u_i(t)$ and outputs $y_i(t)$, $i = 1, 2, \ldots, m$, are all periodic, and they can be expanded in Fourier series. If the oscillations in m loops have a common frequency ω_c then the direct-current components and the first harmonics of these periodic waves are extracted (Ramirez, 1985) as

$$U^1(0) := \begin{bmatrix} \int_0^{T_c} u_1(t)\,dt \\ \vdots \\ \int_0^{T_c} u_m(t)\,dt \end{bmatrix}, \quad Y^1(0) := \begin{bmatrix} \int_0^{T_c} y_1(t)\,dt \\ \vdots \\ \int_0^{T_c} y_m(t)\,dt \end{bmatrix} \quad (2)$$

$$U^1(j\omega_c) := \begin{bmatrix} \int_0^{T_c} u_1(t)e^{-j\omega_c t}\,dt \\ \vdots \\ \int_0^{T_c} u_m(t)e^{-j\omega_c t}\,dt \end{bmatrix}, \quad Y^1(j\omega_c) := \begin{bmatrix} \int_0^{T_c} y_1(t)e^{-j\omega_c t}\,dt \\ \vdots \\ \int_0^{T_c} y_m(t)e^{-j\omega_c t}\,dt \end{bmatrix}. \quad (3)$$

We have

$$Y^1(0) = G(0)U^1(0), \quad (4)$$

$$Y^1(j\omega_c) = G(j\omega_c)U^1(j\omega_c). \quad (5)$$

Since (4) and (5) are vector equations, they are not sufficient to determine $G(j\omega_c)$ and $G(0)$ from Y^1 and U^1 only. Next, we slightly increase the relay amplitude of the dominant loop or decrease that of another loop, and repeat the above procedure until m tests have been completed. According to Proposition 3, the process is likely to have all the oscillation frequencies close to each other for the m tests. $Y^2(0)$, $U^2(0)$, $Y^2(j\omega_c)$, $U^2(j\omega_c), \ldots, Y^m(0)$, $U^m(0)$, $Y^m(j\omega_c)$, $U^m(j\omega_c)$ are obtained subsequently, so that

$$[Y^1(0) \;\; \ldots \;\; Y^m(0)] = G(0)[U^1(0) \;\; \ldots \;\; U^m(0)], \quad (6)$$

$$[Y^1(j\omega_c) \;\; \ldots \;\; Y^m(j\omega_c)] = G(j\omega_c)[U^1(j\omega_c) \;\; \ldots \;\; U^m(j\omega_c)]. \quad (7)$$

While (6) is accurate for any decentralized relay test, (7) is only approximate, since ω_c is not exactly same for all m tests. The U^i, $i = 1, 2, \ldots, m$, are linearly independent, since there is always a relay amplitude change for each test. It follows from (6) and (7) that the steady-state gain matrix $G(0)$ and frequency-response matrix $G(j\omega_c)$ are determined as

$$G(0) = [Y^1(0) \ \ldots \ Y^m(0)] \times [U^1(0) \ \cdots \ U^m(0)]^{-1}, \tag{8}$$

$$G(j\omega_c) = [Y^1(j\omega_c) \ \cdots \ Y^m(j\omega_c)] \times [U^1(j\omega_c) \ \ldots \ U^m(j\omega_c)]^{-1}. \tag{9}$$

Our tuning experiment thus consists of m decentralized relay test, and switches directly from one another. To design this experiment, one needs to specify relay amplitudes for each test. The following design parameters are recommended for use, and have been obtained from our extensive case studies. For the first test, the relay amplitude for each loop is set as in the single-variable case (Åström and Hagglund, 1988). In most circumstances, stationary oscillations of a common frequency will occur in the system (Atherton, 1975). For the subsequent tests, either the relay amplitude in the dominant loop is increased or the relay amplitude in one of other loops is decreased by 5–20%. This usually leads to oscillations with frequencies close to the previous ones.

It should be pointed out that m decentralized relays in our test scheme are reasonable and even necessary to tune an $m \times m$ system. Our test scheme may actually need *less* time than those for IRF and SRF. To see this, our scheme uses m *non-stop* relays, while both IRF and SRF also contain m relays to tune an $m \times m$ system, and furthermore, between their m relays, there are *additionally* $m - 1$ PID control transients to bring outputs back to set-points before the next relays can be performed. In the context of resonance approximations, the number of relays should be at least m in order to identify an $m \times m$ frequency-response matrix $G(j\omega)$, as explained above, and the interaction represented by off-diagonal elements of G must be available to design a decoupler and thus tune a fully cross-coupled multivariable controller. Third, the information extracted from one decentralized relay test is not adequate to tune multivariable PIDs, and only a decentralized PID controller can be obtained (Zhuang and Atherton 1994), with which the achievable performance is limited. In our opinion, the main shortcoming of the decentralized relay test is that it may cause complicated multivariable oscillations (Atherton, 1975; Loh and Vasnani 1993; Zhuang and Atherton 1994), where three modes of multivariable oscillations have been observed. If there are no oscillations or the oscillations have different frequencies at different outputs, our method cannot be used, and this is a restriction on it. But oscillations with a common frequency are the mode that is most likely to occur (Atherton, 1975) when the process has significant interaction, which is the case considered in this paper. In fact, oscillations with a common frequency are also assumed to be the case (Palmor *et al.*, 1993; Zhuang and Atherton, 1994).

Noise is an important issue in identification problem. Åström and Hagglund (1984) pointed out that a hysteresis in the relay is a simple way to reduce the influence of the measurement noise. The width of the hysteresis should be greater than the noise band (Åström and Hagglund, 1988) and is usually chosen as 1–2 times that of the noise band. Filtering is another possibility (Åström and Hagglund, 1984). The measurement noise is usually of high frequency, while for controller design the process frequency response of interest is usually in the low-frequency region. Therefore a low-pass filter can be employed to reduce the measurement noise further. Still another measure to overcome the noise problem is to adopt more oscillation periods in calculating the static gain and the critical point of the process (Shen *et al.*, 1996). Accurate estimation of process frequency response from relay oscillations may then be achieved with these measures (Shen *et al.*, 1996).

4. CONTROLLER DESIGN

With the process frequency response $G(j\omega)$ available at two points, $\omega = 0$, ω_c, we now consider a multivariable control systems as shown in Fig. 5, where $G(s)$ is an m-input, m-output process and $K(s)$ is a fully cross-coupled multivariable controller to be designed. The objectives of control design here, as many applications require, are to make the closed-loop control system decoupled and the resultant independent loops have good transient and accuracy in the usual sense of single-variable systems. For ease of presentation, some notation is needed. Let A be a matrix. Its (i, j) element, ith row and jth column are denoted by a_{ij}, $a_{i.}$ and $a_{.j}$ respectively. For the system in Fig. 5 to be decoupled, it is shown by Wang (1992) that the

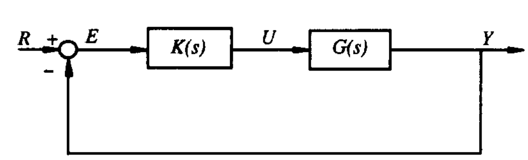

Fig. 5. Multivariable control system.

open-loop transfer matrix $Q = GK$ must be diagonal and nonsingular, that is,

$$GK = \operatorname{diag}\{q_{ii}\}, \quad i = 1, 2, \ldots, m. \tag{10}$$

For each column of GK, we have

$$\begin{bmatrix} g_{1.} \\ g_{2.} \\ \vdots \\ g_{m.} \end{bmatrix} k_{.i} = [0 \ \ \ldots \ \ 0 \ \ q_{ii} \ \ 0 \ \ \ldots \ \ 0]^{\mathrm{T}}, \qquad i = 1, 2, \ldots, m, \tag{11}$$

which is equivalent to

$$\begin{bmatrix} g_{1.} \\ \vdots \\ g_{i-1.} \\ g_{i+1.} \\ \vdots \\ g_{m.} \end{bmatrix} k_{.i} = 0, \quad i = 1, 2, \ldots, m, \tag{12}$$

$$g_{i.}k_{.i} = q_{ii} \neq 0, \quad i = 1, 2, \ldots, m. \tag{13}$$

For any i, we can solve (12) to obtain $k_{1,i}, \ldots, k_{i-1,i}, k_{i+1,i}, \ldots, k_{m,i}$, in terms of k_{ii} as

$$\begin{bmatrix} k_{1,i} \\ \vdots \\ k_{i-1,i} \\ k_{i+1,i} \\ \vdots \\ k_{m,i} \end{bmatrix} = f_{.i}k_{ii}, \quad i = 1, 2, \ldots, m, \tag{14}$$

where

$$f_{.i} := -\begin{bmatrix} g_{1,1} & \cdots & g_{1,i-1} \\ \vdots & \ddots & \vdots \\ g_{i-1,1} & \cdots & g_{i-1,i-1} \\ g_{i+1,1} & \cdots & g_{i+1,i-1} \\ \vdots & \ddots & \vdots \\ g_{m,1} & \cdots & g_{m,i-1} \end{bmatrix}
\begin{bmatrix} g_{1,i+1} & \cdots & g_{1,m} \\ \vdots & \ddots & \ddots \\ g_{i-1,i+1} & \cdots & g_{i-1,m} \\ g_{i+1,i+1} & \cdots & g_{i+1,m} \\ \vdots & \ddots & \vdots \\ g_{m,i+1} & \cdots & g_{m,m} \end{bmatrix}^{-1}
\begin{bmatrix} g_{1,i} \\ \vdots \\ g_{i-1,i} \\ g_{i+1,i} \\ \vdots \\ g_{m,i} \end{bmatrix}. \tag{15}$$

Equation (13) then becomes

$$\tilde{g}_{ii}k_{ii} = q_{ii}, \quad i = 1, 2, \ldots, m, \tag{16}$$

where

$$\tilde{g}_{ii} = g_{ii} + [g_{i,1} \ \ \cdots \ \ g_{i,i-1} \ \ g_{i,i+1} \ \ \cdots \ \ g_{i,m}]f_{.i} \tag{17}$$

One notes that for a given i, (16) is independent of the controller off-diagonal elements, but contains only the controller diagonal element k_{ii}. This k_{ii} can be designed for the equivalent single-variable plant $\tilde{g}_{ii}$ with a single-variable design method, and afterwards all the off-diagonal elements are determined from (14).

To the best of our knowledge, (14)–(17) for decoupling design have not previously appeared in the literature. Usually, decoupler design needs to specify the decoupled open-loop transfer matrix Q a priori before the decoupler can be calculated. The choice of Q may not be an easy task (Maciejowski, 1989) since it is related to the plant characteristics and controller structure, which are yet not known. This kind of design often leads to a high-order or unrealizable decoupler. With our new design equations, there is no need to specify Q, and the individual k_{ij} are determined with scalar methods. One may also note that most designs, such as the state-space, optimization, inverse nyquist array and characteristic locus methods, need a full model of the process in a form of either a state-space description or a transfer-function or frequency-response matrix over the entire working frequency range, and such models are usually not available in the content of relay auto-tuning, where in general only partial dynamic information on the process in a form of frequency-response matrix at one or two points can be obtained, but our design can work with such incomplete information. Furthermore, we can easily choose simple controllers of PID type for the individual k_{ij} to obtain good approximate solutions to (14) and (16), which will be shown below.

The exact solutions for k_{ij} given by (16) and (14) are usually too complicated to use in practical applications, as one can see from the expressions for $f_{.i}(s)$ and $\tilde{g}_{ii}(s)$. Instead, simple controllers of PID type are preferred in the context of auto-tuning. In addition, we cannot calculate $k_{ij}(s)$ from (16) and (14), since $f_{.i}(s)$ and $\tilde{g}_{ii}(s)$ are not available from our process test and identification in Section 3. What we actually know is the two points, at $\omega = 0$, ω_c, on the frequency responses for $\tilde{g}_{ii}(j\omega)$ and $f_{.i}(h\omega)$. It is thus desired to make full use of them to design simple controllers of PID type so that good approximations to (16) and (14) can be achieved. For each diagonal element k_{ii}, we regard $\tilde{g}_{ii}$ as a generalized process. Given the information on the process as $\tilde{g}_{ii}(0)$ and $\tilde{g}_{ii}(j\omega_c)$, there are many SISO methods available to tune k_{ii} of PID type. It turns out that the gain and phase margin rule (Ho *et al.*, 1995b) is most suitable for our application here, since performance and robustness of the system thus tuned is very promising for most cases (Ho *et al.* 1995a).

With the method in Ho et al. (1995b), the two points $\tilde{g}_{ii}(0)$ and $\tilde{g}_{ii}(j\omega_c)$ are fitted to a first-order plus dead-time model

$$\tilde{g}_{ii}(s) = \frac{\tilde{k}_{ii}}{1 + s\tilde{\tau}_{ii}} e^{-s\tilde{L}_{ii}}, \tag{18}$$

where

$$\tilde{k}_{ii} = \tilde{g}_{ii}(0), \tag{19}$$

$$\tilde{\tau}_{ii} = \frac{1}{\omega_c}\sqrt{\frac{\tilde{k}_{ii}^2}{|\tilde{g}_{ii}(j\omega_c)|^2} - 1}, \tag{20}$$

$$\tilde{L}_{ii} = \frac{1}{\omega_c}\{-\arg[\tilde{g}_{ii}(j\omega_c)] - \tan^{-1}(\omega_c\tilde{\tau}_{ii})\}. \tag{21}$$

$k_{ii}(s)$ is taken as $k_{ii}(s) = k_{Pii}(1 + 1/T_{Iii}s)$, and its parameters are given by

$$k_{Pii} = \frac{\omega_{pii}\tilde{\tau}_{ii}}{A_m\tilde{k}_{ii}}, \tag{22}$$

$$T_{Iii} = \left(2\omega_{pii} - \frac{4\omega_{pii}^2\tilde{L}_{ii}}{\pi} + \frac{1}{\tilde{\tau}_{ii}}\right)^{-1}, \tag{23}$$

where

$$\omega_{pii} = \frac{A_m\phi_m + \frac{1}{2}\pi A_m(A_m - 1)}{(A_m^2 - 1)\tilde{L}_{ii}}, \tag{24}$$

A_m and ϕ_m are the specified gain and phase margins (e.g. $A_m = 3$ and $\phi_m = \frac{1}{3}\pi$) respectively. In general, we choose relative large gain and phase margins for fast loops and choose small ones for slow loops.

For the off-diagonal elements $k_{ij}(s)$, $j \neq i$, we choose $k_{ij}(s)$ of PID type as

$$k_{ij}(s) = k_{Pij}\left(1 + \frac{1}{T_{Iij}s} + T_{Dij}s\right),$$

such that (14) is satisfied at $\omega = 0$ and $\omega = \omega_c$; that is,

$$\lim_{s\to 0} sk_{ij}(s) = \lim_{s\to 0} sf_{ij}(s)k_{jj}(s), \tag{25}$$

$$k_{ij}(j\omega_c) = f_{ij}(j\omega_c)k_{jj}(j\omega_c) := \gamma_{ij}e^{j\varphi_{ij}}. \tag{26}$$

A simple derivation gives the controller parameters as

$$k_{Pij} = \gamma_{ij}\cos\varphi_{ij}, \tag{27}$$

$$T_{Iij} = \frac{k_{Pij}T_{Ijj}}{k_{Pjj}f_{ij}(0)}, \tag{28}$$

$$T_{Dij} = \frac{1}{\omega_c}\left(\tan\varphi_{ij} - \frac{1}{\omega_c T_{Iij}}\right). \tag{29}$$

5. CASE STUDIES

In this section, three typical examples are given to illustrate the proposed method.

Example 1. Consider the well-known Wood and Berry (1973) binary distillation column plant:

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

It is a typical MIMO plant with strong interaction and significant time delays. For tuning test, the relay in loop 1 is set as an ideal symmetric relay with switching levels 1.00 and −1.00, and a biased relay with switching levels 1.50 and −1.00 is used in loop 2. The system exhibits limit-cycle oscillations of a common frequency with frequency $\omega_c^1 = 0.485$. The switching levels of relay in loop 2 are then changed to 1.80 and −1.20. The system exhibits limit-cycle oscillations of a common frequency, with $\omega_c^2 = 0.484$ in this case. The steady-state gain matrix $\hat{G}(0)$ and the frequency-response matrix $\hat{G}(\omega_c)$ are computed from (8) and (9) as

$$\hat{G}(0) = \begin{bmatrix} 12.8 & -18.9 \\ 6.60 & -19.4 \end{bmatrix},$$

$$\hat{G}(j\omega_c) = \begin{bmatrix} 1.56e^{-1.92j} & 1.86e^{0.221j} \\ 1.21e^{1.46j} & 2.79e^{0.260j} \end{bmatrix},$$

where $\omega_c = \frac{1}{2}(\omega_c^1 + \omega_c^2) = 0.485$. They are very accurate compared with their true values:

$$G[0] = \begin{bmatrix} 12.8 & -18.9 \\ 6.6 & -19.4 \end{bmatrix},$$

$$G(0.485j) = \begin{bmatrix} 1.57e^{-1.93j} & 1.85e^{0.21j} \\ 1.23e^{1.50j} & 2.75e^{0.26j} \end{bmatrix}.$$

The frequency responses of the generalized process $\tilde{g}_{11}(s)$ for loop 1 at $\omega = 0$ and $\omega = 0.485$ are obtained from (17) as

$$\tilde{g}_{11}(0) = 6.37, \quad \tilde{g}_{11}(0.485j) = 2.35e^{-1.85s}.$$

They are fitted to the following first-order plus dead-time model:

$$\tilde{g}_{11}(s) = \frac{6.37}{1 + 5.19s} e^{-1.36s}.$$

The gain margin and phase margin are chosen as $A_{m1} = 5$ and $\phi_{m1} = \frac{1}{3}\pi$; (22)–(24) then yield

$$k_{11}(s) = 0.184\left(1 + \frac{1}{3.92s}\right).$$

Similarly, we can obtain

$$\tilde{g}_{22}(s) = \frac{-9.65}{1+4.25s} e^{-3.49s}.$$

The second diagonal element in K to achieve $A_{m2} = 3$ and $\phi_{m2} = \frac{1}{2}\pi$ is given by

$$k_{22}(s) = -0.0660\left(1 + \frac{1}{4.25s}\right).$$

The off-diagonal controllers are calculated from (27)–(29) as

$$k_{12}(s) = -0.0102\left(1 + \frac{1}{0.445s} - 0.804s\right),$$

$$k_{21}(s) = -0.0674\left(1 - \frac{1}{4.23s} + 0.796s\right).$$

The tuning process and the resulting performance are shown in Fig. 6. The relay test is performed from $t = 0$ to $t = 73.0$, the cross-coupled PID controller is commissioned at $t = 73.0$, and the unit-step set-point changes of loops 1 and 2 occur at $t = 141.5$ and $t = 301.5$ respectively. For comparison, the BLT method (Luyben, 1986) gives the parameters of multi-loop PI controllers as $K_c = \text{diag}\{0.375, -0.075\}$ and $T_i = \text{diag}\{8.29, 23.6\}$. Its set-point responses are also shown in Fig. 6. One sees from Fig. 6 that the proposed method gives significant improvement in both the system decoupling and the diagonal loop performances. In addition, we have compared other advanced multivariable design methods such as the inverse Nyquist array method (Waller *et al.*, 1982; Deshpande, 1989) and the simplified model predictive control (SMPC) method (Deshpande, 1989) for the plant, and it turns out that the proposed method gives the best result.

Example 2. This example is adopted from Palmor *et al.* (1993) as

$$G(s) = \begin{bmatrix} \dfrac{0.5}{(0.1s+1)^2(0.2s+1)^2} & \dfrac{-1}{(0.1s+1)(0.2s+1)^2} \\ \dfrac{1}{(0.1s+1)(0.2s+1)^2} & \dfrac{2.4}{(0.1s+1)(0.2s+1)^2(0.5s+1)} \end{bmatrix}.$$

There exist large interactions in this process, and the performances of Ziegler–Nichols tuned decentralized PID controllers based on the critical point derived by Palmor *et al.* (1993) are not satisfactory, as can be seen from Fig. 7. For our method, two centralized relay tests are performed for this process. The relay in loop 1 is an ideal one with unit switching levels, and switching levels of the relay in loop 2 are 1.40 and −0.933 in the first test, and are then changed to 1.50 and −1.00 in the second. Both tests result in limit-cycle oscillations with the same fre-

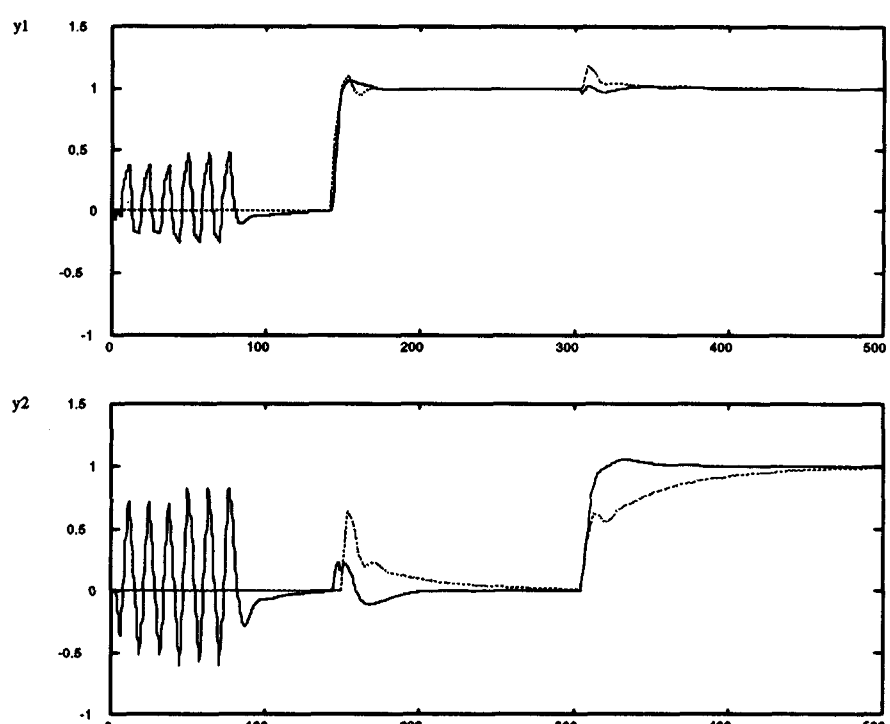

Fig. 6. Autotuning process of example: ——, proposed method; - - -, BLT tuning method.

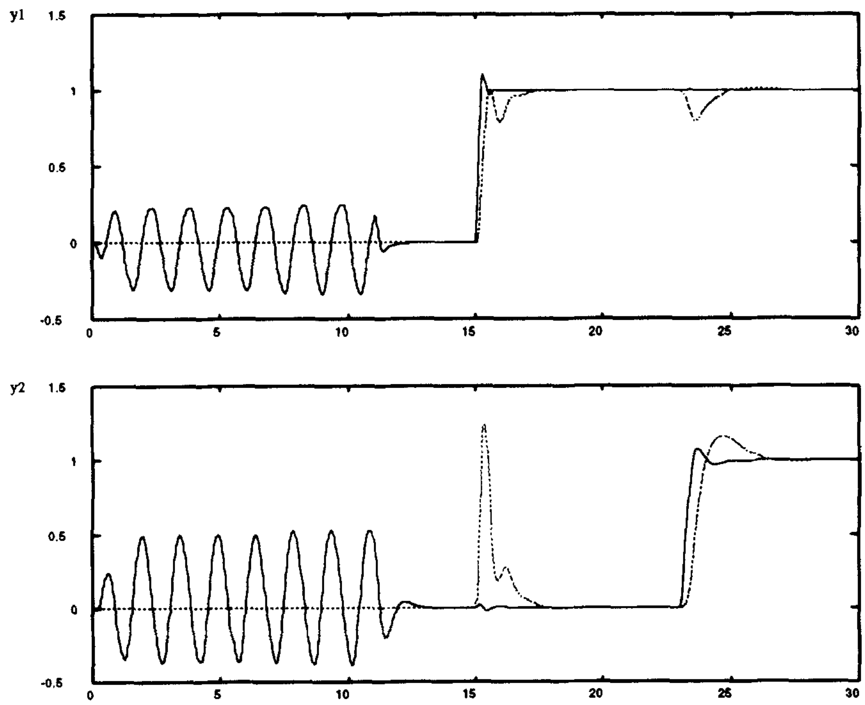

Fig. 7. Autotuning process for Example 2: ——, proposed method; - - -, decentralized PID control.

quency $\omega_c = 4.29$. The estimated steady-state gain matrix $\hat{G}(0)$ and the frequency-response matrix $\hat{G}(j\omega_c)$ are

$$\hat{G}(0) = \begin{bmatrix} 0.500 & -1.00 \\ 1.00 & 2.40 \end{bmatrix},$$

$$\hat{G}(4.29j) = \begin{bmatrix} 0.243e^{-2.25j} & 0.529e^{1.31j} \\ 0.529e^{-1.84j} & 0.537e^{-2.98j} \end{bmatrix},$$

while the true values are

$$G(0) = \begin{bmatrix} 0.5 & -1 \\ 1 & 2.4 \end{bmatrix},$$

$$G(4.29j) = \begin{bmatrix} 0.24e^{-2.23j} & 0.53e^{1.32j} \\ 0.53e^{-1.82j} & 0.54e^{-2.96j} \end{bmatrix}.$$

The design method in Section 4 yields

$$K(s) = \begin{bmatrix} 2.83\left(1 + \dfrac{1}{0.285s}\right) & 1.51\left(1 + \dfrac{1}{0.865s} + 0.0911s\right) \\ -3.25\left(1 + \dfrac{1}{0.785s} + 0.182\right) & 0.667\left(1 + \dfrac{1}{0.776s}\right) \end{bmatrix}.$$

The tuning response is shown in Fig. 7, where the set-point changes in loops 1 and 2 occur at $t = 18$ and $t = 24$ respectively. The results show that the closed loop is nearly decoupled and the diagonal loops have excellent performances.

Example 3. Consider the solid-fuel boiler plant

$$G(s) = \begin{bmatrix} \dfrac{-e^{-2s}}{1+10s} & \dfrac{-1}{1+10s} \\ 0 & \dfrac{e^{-10s}}{1+60s} \end{bmatrix},$$

for which a multivariable controller was designed by Johansson and Koivo (1984) using the inverse Nyquist array (INA) method. For our method, the switching levels of the relays in loops 1 and 2 are first set at 1.00, −1.00 and 1.50, −1.00 respectively, and the switching levels of the relay in loop 2 are later increased to 1.80, −1.20. Both decentralized relay feedback produce stationary oscillations of the same frequency $\omega_c = 0.164$, and it is equal to that of the limit cycle of g_{22}, since loop 2 is dominant in this case. Applying the method in Section 3 results in the very accurate steady-state gain matrix $\hat{G}(0)$ and frequency-response matrix $\hat{G}(j\omega_c)$ as

$$\hat{G}(0) = \begin{bmatrix} -1.00 & -1.00 \\ 0.00 & 1.00 \end{bmatrix}, \quad \left(G(0) = \begin{bmatrix} -1 & -1 \\ 0 & 1 \end{bmatrix}\right),$$

$$\hat{G}(0.164j) = \begin{bmatrix} 0.520e^{1.79j} & 0.520e^{2.12j} \\ 0.00 & 0.101e^{-3.11j} \end{bmatrix},$$

$$\left(G(0.164j) = \begin{bmatrix} 0.52e^{1.79j} & 0.52e^{2.12j} \\ 0.00 & 0.10e^{-3.11j} \end{bmatrix}\right).$$

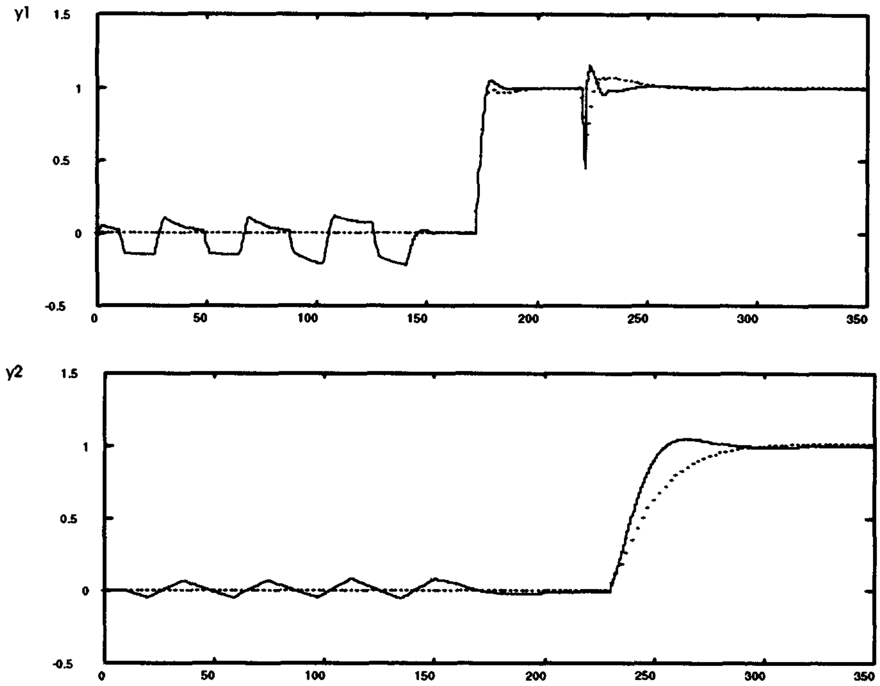

Fig. 8. Autotuning process for Example 3: ——, proposed method; - - - , INA method.

The controller is designed as

$$K(s)=\begin{bmatrix} -2.61\left(1+\dfrac{1}{10.0s}\right) & -3.08\left(1+\dfrac{1}{58.8s}+2.04s\right) \\ 0.00 & 3.14\left(1+\dfrac{1}{60.0s}\right) \end{bmatrix}.$$

The tuning response is shown in Fig. 8, where the set-point changes in loops 1 and 2 occur at $t=170$ and $t=220$. Compared with the INA design in Johansson and Koivo (1984), the proposed method gives a better result.

6. CONCLUSIONS

The tuning of multivariable process controllers on line has always been a difficult problem. In this paper, a new auto-tuning method for multivariable PID controllers from decentralized relay feedback has been presented. The proposed method has several advantages. Firstly, it is a complete closed-loop tuning method for multivariable control system. Secondly, the low-order transfer function model of a multivariable process can be obtained from the identified two-point frequency responses of the process, which may be needed in other model-based multivariable design methods. Thirdly, the designed controllers achieve approximate decoupling near the critical and zero frequencies, and the calculations of controller parameters are quite easy and are therefore suitable for on-line tuning. Finally, very little prior knowledge of the process is required, and our method can be applied to a large class of general MIMO systems. Various typical processes have been employed to illustrate the effectiveness of the method.

REFERENCES

Åström, K. J. and T. Hagglund (1984). Automatic tuning of simple regulators with specifications on phase and amplitude margins. *Automatica*, **20**, 645–651.

Åström, K. J. and T. Hagglund (1988). *Automatic tuning of PID Controllers*. Instrument Society of America, Research Triangle Park, NC.

Åström, K. J. and C. C. Hang (1991). Towards intelligent PID control. *Automatica*, **28**, 1–9.

Atherton, D. P. (1975). *Non-linear Control Engineering*. Van Nostrand Reinhold, London.

Deshpande, P. B. (1989). *Multivariable Process Control*. Instrument Society of America, Research Triangle Park, NC.

Ho, W. K., C. C. Hang and J. H. Zhou (1995a). Performance and gain and phase margins of well-known PI tuning formulas. *IEEE Trans. Control Syst. Technol.*, **3**, 245–248.

Ho, W. K., C. C. Hang and L. S. Cao (1995b). Tuning of PID controllers based on gain and phase margin specifications. *Automatica*, **31**, 497–502.

Johansson, L. and H. Koivo (1984). Inverse Nyquist array technique in the design of a multivariable controller for a solid-fuel boiler. *Int. J. Control*, **40**, 1077–1088.

Koivo, H. N. and J. T. Tanttu (1991). Tuning of PID controllers: survey of SISO and MIMO techniques. In *Preprints of IFAC International Symp. on Intelligent Tuning and Adaptive Control, Session* 1, Singapore.

Kong, K. Y. *et al.* (1995). Feasibility report on a frequency-domain adaptive controller. Department of Electrical Engineering, National University of Singapore.

Lieslehto, J., J. T. Tanttu and H. N. Koivo (1993). An expert system for multivariable controller design. *Automatica*, **29**, 953–968.

Loh, A. P. and V. U. Vasnani (1994). Describing function matrix for multivariable systems and its use in multiloop PI design. *J. Process Control*, **4**, 115–120.

Loh, A. P., C. C. Hang, C. K. Quek and V. U. Vasnani (1993). Autotuning of multiloop proportional–integral controllers using relay feedback. *Ind. Engng Chem. Res.*, **32**, 1102–1107.

Luyben, W. L. (1986). A simple method for tuning SISO controllers in a multivariable system. *Ind. Eng. Chem. Proc. Des. Dev.*, **25**, 654–660.

Luyben, W. L. (1990). *Process Modeling Simulation and Control for Chemical Engineers.* McGraw-Hill, New York.

Macfarlane, A. G. J. and B. Kouvaritakis (1977). A design technique for linear multivariable systems. *Int. J. Control*, **25**, 737–874.

Maciejowski, J. M. (1989). *Multivariable Feedback Design.* Addison-Wesley, Workingham, U.K.

Palmor, Z. J., Y. Halevi and N. Krasney (1993). Automatic tuning of decentralized PID controllers for TITO processes. In *Proc. of 12th IFAC World Congress*, Sydney, Vol. 311–314.

Ramirez, R. W. (1985). *The FFT: Fundamentals and Concepts.* Prentice-Hall, Englewood Cliffs, NJ.

Rosenbrock, H. H. (1969). Design of multivariable control systems using the inverse Nyquist array. *Proc. IEE, Pt D*, **116**, 1926–1936.

Shen, S. H., J. S. Wu and C. C. Yu (1996). Use of biased-relay feedback for system identification. *AIChE J.*, **42**, 1174–1180.

Wang, Q. G. (1992). Decoupling with internal stability for unity output feedback systems. *Automatica*, **28**, 411–416.

Waller, K. V., K. E. Wikman and S. E. Gustafsson (1982). Design and tuning of dual-composition distillation control systems by the inverse Nyquist array method. In *Proc. 32nd Canadian Chemical Engineering Conf.*, Vancouver, BC, Vol. 9, p. 11.

Wood, R. K. and M. W. Berry, (1973). Terminal composition control of a binary distillation column. *Chem. Engng Sci.*, **28**, 1707–1717.

Zgorzelski, P., H. Unbehauen and A. Niederlinski (1990). A new simple decentralized adaptive multivariable regulator and its application to multivariable plants. In *Proc. 11th IFAC World Congress*, Tallin, Estonia, pp. 226–231.

Zhuang, M. and D. P. Atherton (1994). PID controller design for a TITO system. *Proc. IEE, Pt D*, **141**, 111–120.