Practical Aspects of PID Auto-Tuners Based on Relay Feedback

Hang, Chang C.; Åström, Karl Johan

Published in:
2nd International IFAC Symposium on Adaptive Control of Chemical Processes 1988 (ADCHEM '88), Lyngby, Copenhagen, Denmark, 17-19 August

DOI:
10.1016/S1474-6670(17)53817-3

1988

Document Version:
Publisher's PDF, also known as Version of record

PRACTICAL ASPECTS OF PID AUTO-TUNERS BASED ON RELAY FEEDBACK

C. C. Hang

Dept of Electrical Engineering
National University of Singapore
10 Kent Ridge Crescent
Singapore 0511

K. J. Astrom

Dept of Automatic Control
Lund Institute of Technology
Box 118, S-221 00 Lund
Sweden

Abstract. The performance of an automatic PID controller tuning method based on relay feedback is studied in the presence of deterministic disturbances. It is found that the occurence of any static load disturbance could cause significant errors in the estimates of the ultimate gain and period. However, the resultant asymmetry of the relay switching intervals can be used as an error indicator, or used to compute a self-corrective bias to restore accuracy of the estimates. This corrective bias is found to be functional even in the presence of moderate nonlinearity. A reliable self-biasing auto-tuner is thus resultant. The effect of a less common sinusoidal load could be more serious since it may not be detectable and hence more prior knowledge about its presence is required.

Keywords. PID auto-tuner; relay feedback; load disturbance; ultimate gain and period; self-biasing.

1. INTRODUCTION

Automatic tuning of PID controllers has recently received much attention in the literature in view of its potential applications in reducing system start-up time, and in tightening process control through regular re-tuning (Bristol 1977, Krauss and Mayron 1984, Astrom and Hagglund 1984, Higham 1985, Balchen and Lie 1986, Hang, Lim and Soon 1986, Hess, Radke and Schumann 1987, Radke and Isermann 1987). Auto-tuning is particularly useful when the process time constants are long as valuable time spent by an instrument/control engineer on controller tuning can be saved by its automation. Several commercial products for auto-tuning have appeared in the market since 1981 as the required technology has been available to economically implement this function on stand-alone controllers and on control computers.

The relay-feedback auto-tuner (Astrom 1982, Astrom and Hagglund 1984) is one method for the automatic tuning of PID controllers. Its operational principle is extremely simple : it will switch to a relay feedback control mode during tuning; this will cause a controlled limit cycle to be automatically generated from which the ultimate gain and period, as required in the classical Ziegler-Nichols formula (1943) for optimum controller settings, can be easily measured. A hysteresis can also be introduced into the relay in which case a modified Ziegler-Nichols tuning rule based on phase or amplitude margin can be used. The greatest merits of this relay feedback method are that a priori information about the time scale or the structure of the process is not needed, and that a good excitation signal which is tuned to the process is generated automatically. Thus it can also be used to initialize more sophisticated auto-tuning, pre-tuning or adaptive control algorithms.

It is expected that this simple and yet effective auto-tuning method will be used increasingly in industry. The method, therefore, deserves a further study to explore its performance in more demanding situations. An important assumption made in the development of this auto-tuning method is that the input signal to the symmetrical relay has a zero mean. Under this condition, theoretical results are available on the existence of a stable limit cycle (Astrom and Hagglund 1984). In practice, especially in process control applications where the tuning durations are long, this condition may be easily violated owing to control loop interactions or unexpected load disturbances. A load disturbance could either quench the limit cycle or influence the estimation accuracy of the ultimate gain and period. The potential problem is studied in this paper and the possibility of error detection and self-correction is investigated. Particular attention is paid to the presence of static and sinusoidal load disturbances. The behaviour of the relay auto-tuner with and without hysteresis is studied using digital simulations.

2. STATIC LOAD DISTURBANCE

Static load disturbances are quite common in the process industry. In a typical relay-feedback auto-tuning set-up as shown in Fig. 1, a supervisory program will be activated to initialize the control error to zero and the bias

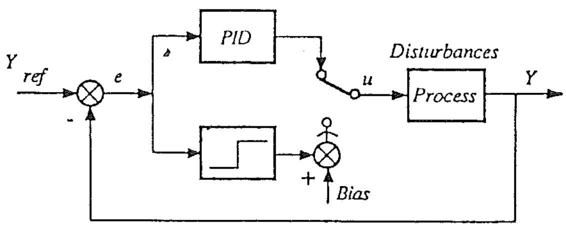

Fig. 1 Set-up for auto-tuning

to an appropriate level, and to disable any request for setpoint changes during tuning to ensure that the relay will receive a zero-mean input signal. Any unexpected static load disturbance may, however, upset this equilibrium if it occurs during tuning, the likelihood of which increases with the complexity and the time scale of the process. A large static disturbance will quench the limit cycle, as shown in an example in Fig. 2(a), in which case a warning message can be given to the operator, or the bias can be increased in the same direction as the current relay output until limit cycling is

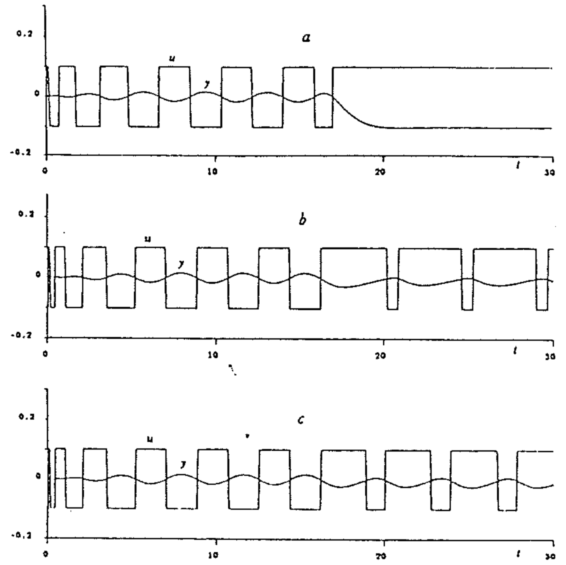

Fig. 2 Effects of static load disturbances
(a) load= -0.2; (b) load= -0.08; (c) load= -0.05.

resumed. At this stage, or when the static load is not large enough to quench the limit cycle, the accuracy of the tuning estimates may change significantly. This has been studied in extensive simulations and the results are summarized in the following. Without loss of generality and unless otherwise stated, the following low order, linear process will be used in all the digital simulations:

$$\frac{Y(s)}{U(s)} = \frac{1}{(1+s)^3} \quad (1)$$

2.1 Errors in Estimates of Ultimate Gain and Period

When there is no static load disturbance during tuning, the relay output would have a symmetrical positive and negative half-cycles, as shown in the first 16 seconds of the simulation result in Fig. 2. When a fresh static load occurs, the relay output will have to produce an average magnitude equal to an equivalent static input required to nullify the effect of the load on the process output. Fig. 2 shows this situation clearly for different values of static load. The asymmetry in the relay switching intervals causes errors in the estimates of the ultimate gain and period, of the order of +25% and +30% respectively in the case of Fig. 2(b), and +10% and +10% respectively in Fig. 2(c).

2.2 Error Detection and Self-Correction

It is evident from the above argument on the cause of the asymmetry in the relay switching intervals that the amount of bias required can be simply computed as the average value of the relay output. If t_1 and t_2 are the intervals of positive and negative relay outputs respectively, and d is the relay magnitude (0.1 being used in the example of Fig. 2), then we have:

$$\text{corrective bias} = \frac{t_1 - t_2}{t_1 + t_2} \cdot d \quad (2)$$

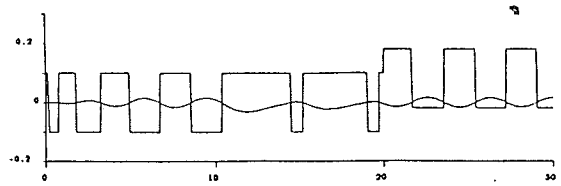

Fig. 3 Effectiveness of using corrective bias

Hence the supervisory program can monitor the switching intervals and give a warning message to request a re-start when the difference exceeds a pre-set tolerance, plus a suggestion of the corrective bias required as computed from equation (2). In a fully automated auto-tuner, the bias can be automatically added once the asymmetry condition is detected. This is demonstrated in Fig. 3 : at time t=10, a static load of -0.08 was introduced; the limit cycle sustained but significant errors in estimation were created; at t=20, automatic adjustment of bias computed using equation (2) was made and the estimation accuracy was restored.

2.3 Effects of Process Nonlinearity

A possible reservation about the use of the self-corrective bias may be that it will also attempt to correct the relay switching asymmetry caused by any process nonlinearity. This is usually not a major concern as the limit cycle has a small controlled amplitude which ensures that the normal process operating condition is not much disturbed and the process dynamics is linear. It is more likely that a static load may shift the operating point and the dynamics will become different due to the nonlinearity. The corrective bias will then ensure that the new operating point be established during tuning so that the correct set of tuning parameters is being estimated. In the special case of a large limit cycle, due for instance to the presence of large measurement noise or stochastic disturbance in which case the relay magnitude or hysteresis has to be large, the nonlinear mode may become prominant during the relay auto-tuning operation. However it will be shown in the following investigation that it is quite safe to apply the self-corrective bias even in the presence of moderate directional dependent nonlinearity.

First, the effects of a nonlinear gain are studied. A typical case, using the process of equation (1) but with a gain of 1.5 when the control signal is increasing and a gain of 1 when the control signal is decreasing, is shown in Fig. 4. Note that although the relay switching intervals are asymmetrical, the process output is nearly sinusoidal which is different from the vastly distorted waveform as in the case of a static disturbance; compare with Fig. 2. It is evident that the limit cycle period remains unchanged while the limit cycle amplitude assumes the average of those due to the low and high linear gains. This averaging property has been observed previously (Lim, Lim and Hang 1986) on a pilot plant experiment. When the self-corrective bias is applied to the nonlinear process using the formula of equation (2) but without a static load, symmetry is restored although this is not called for. Fortunately, as shown in Fig. 4(a), this correction has little influence on the amplitude and period of the limit cycle, which means that the averaging property is not sacrificed. In the presence of static load, the asymmetry is then due to the cummulative effects of the load and the nonlinearity. Fig. 4(b) shows that the bias correction in this case is still effective in restoring the accuracy of the tuning estimates. When a larger nonlinear gain of Fig. 5 is used, the bias will introduce a small change in the limit cycle amplitude as the average process gain has actually changed. This is shown in Fig. 6(a) and 6(b) respectively without and with a static load change of -0.2.

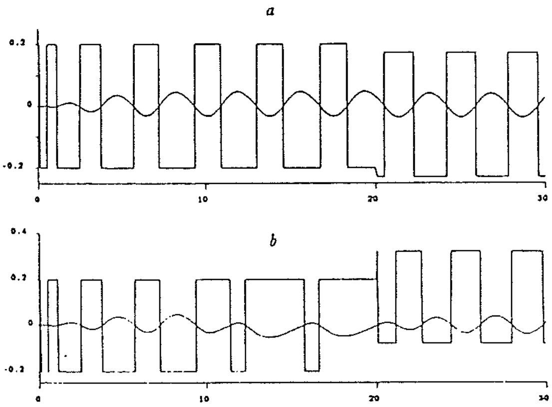

Fig. 4 Effects of nonlinear gains: (a) without load, correct bias = -0.029 ; (b) with load, corrective bias = 0.121.

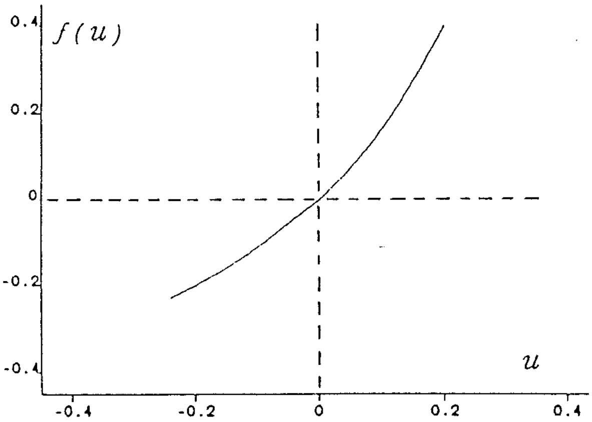

Fig. 5 Nonlinear gain charcteristics at process output

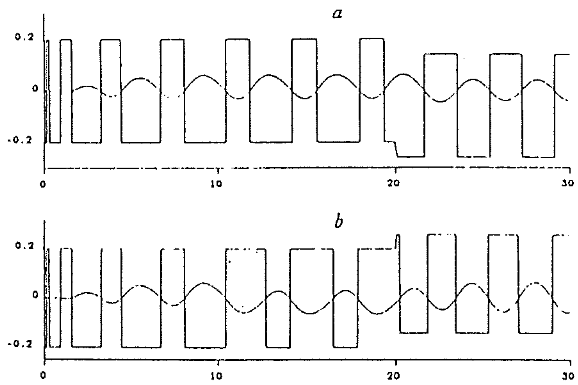

Fig. 6 Effects of large nonlinear gains: (a) without load, corrective bias =-0.06; (b) with load, corrective bias = 0.06

Second, the effects of a nonlinear time constant are studied. A typical case, using the process of equation (1) but with a time constant of 1.5 when the control signal is increasing and 1 when it is decreasing, is shown in Fig. 7. It is evident that the nonlinearity does not change the limit cycle amplitude, while the influence on the period is one of averaging. This property is preserved after the bias correction, with the nonlinearity alone as shown in Fig. 7(a), and with both the nonlinearity and static load effects as shown in Fig. 7(b).

2.3 Relay with Hysteresis or With Integrator

The basic principle of Ziegler-Nichols (1943) closed-loop tuning method is that it identifies the point on the Nyquist curve where it intersects the negative real axis, which then provides information on the ultimate gain and period. Astrom and Hagglund (1984) have extended this principle so that any one point on the Nyquist curve can be used. The tuning formula will have to be changed accordingly and corresponding gain-margin and phase-margin formulae have been developed. A relay with hysteresis, which is recommended for a process with significant measurement noise, will shift the limit cycling point on the Nyquist curve to a new location where the phase lag is less than 180°. A relay in cascade with an integrator will shift this point

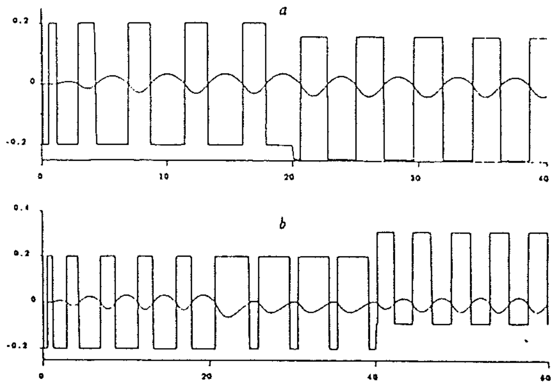

Fig. 7 Effects of nonlinear time constants : (a) without load, corrective bias = -0.045; (b) with load, corrective bias = 0.11

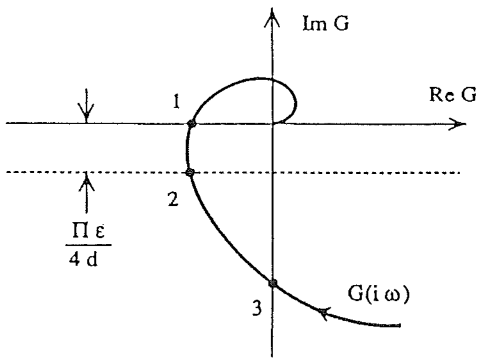

Fig. 8 Critical points for (1) a pure relay; (2) a relay with hysteresis; (c) a relay with cascade integrator

even further to the intersection on the imaginary axis where the phase lag is 90°. These are shown in Fig. 8. The modified tuning formulae (Astrom and Hagglund 1984) can be used for these two cases.

For a relay with hysteresis, the same problem of limit cycle quenching or estimation errors in the ultimate gain and period will occur when a static load disturbance is introduced during tuning. The same solution of using the resultant asymmetry in relay switching intervals for error detection or for computing a corrective bias can be used to overcome this problem. For a relay in cascade with an integrator, the static load will not cause the above problem. As shown in Fig. 9, the integrator will automatically provide the correct bias when a load change occurs at t=40 and there will be no error introduced in the estimates of the tuning parameters.

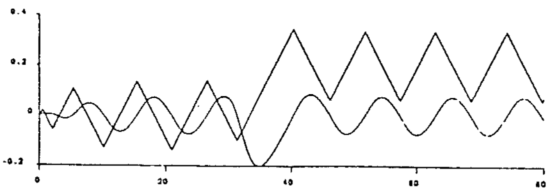

Fig. 9 Self-biasing effect of a relay with cascade integrator

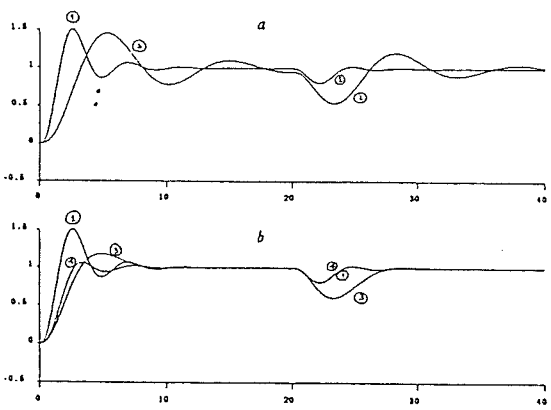

Fig. 10 Comparison of response rates: (1) Z-N tuning; (2) relay with integrator/phase margin of 30°; (3) same as 2) but with phase margin of 60°; (4) same as 1) but with b = 0.3.

The immunity to static load disturbances seems to suggest that a relay with a cascade integrator should always be used instead of the relay without an integrator. On a closer examination, however, this is found not to be valid for three reasons. First, a relay in cascade with an integrator cannot be used to tune a process which contains an integrator, such as in certain level control applications, as the Nyquist curve will not intersect the imaginery axis at low frequencies. Secondly, the tuning will take much longer time when the integrator is used since the limit cycle period and the transient to reach steady cycling are longer. In the case of a process given by equation (1), the tuning using a relay with integrator takes about three times longer than that using a pure relay. Third, the ultimate period obtained will be longer and this can result in a closed-loop bandwidth which is too small, as shown in Fig. 10. With the cascade integrator, the ultimate gain K_d and period t_d obtained can be used in the phase-margin design formula shown in App. I. With a phase margin of 30°, the tuned closed-loop response is much slower than that tuned by the Ziegler-Nichols formula using a pure relay. The damping may be improved by specifying a 60° phase margin but the bandwidth remains too small. It is noted that the Ziegler-Nichols tuning always produces a fast response to a load disturbance. For a setpoint change, the over-shoot for a low order process is usually too high; this is not due to poor tuning and can be easily overcome by introducing a weighting factor 'b' on the setpoint value in the proportional term (Hagglund and Astrom 1985) as described in App. II. With b = 0.3, the overshoot in the Ziegler-Nichols tuned closed-loop response is vastly improved as shown in Fig. 10. Notice that the factor 'b' does not affect the load regulation response.

3. SINUSOIDAL LOAD DISTURBANCE

A sinusoidal load disturbance is much less common than static load disturbances and hence may not be a major concern in most applications. With the normal commissioning procedure recommended for the auto-tuner where the process is first brought to steady state before tuning is started, the operator can also check that there is no periodic load disturbance in the process output. Nevertheless it is useful to know whether a sinusoidal disturbance would affect tuning accuracy and whether it is safe to use in the event that a sinusoidal load is significant.

3.1 Errors in Estimated Ultimate Gain and Period

It has been found from an extensive simulation study that the presence of a significant sinusoidal load disturbance could cause large errors in the estimates of the ultimate gain and period. The interference with the normal limit cycle is a nonlinear effect and either the magnitude or frequency of the disturbance could change the amplitude and period of the limit cycle.

In the process of equation (1), the effect of introducing a 1 rad/s sinusoidal load during a tuning exercise is shown in Fig. 11. Both the amplitude and period of the limit cycle vary as the load amplitude varies. With a load amplitude of 0.1, which produces an output amplitude about half of the relay amplitude, the estimated ultimate gain and period have an error of -330% and +60% respectively as shown in Fig. 11(a). When the load amplitude is halved, as shown in Fig. 11(b), the errors reduce to -70% and +25% respectively, which are still very significant for the purpose of controller tuning. The effects of different load frequencies are shown in Fig. 12. It is evident that the estimation errors are significant. The simulation results also indicate that the relay switching intervals are nearly symmetrical. Hence a simple error detection or self-correction scheme cannot be devised as in the case of the static load.

The addition of a hysteresis would improve the estimation quite significantly if a large enough dead zone is used. For instance, with a dead zone of 0.05, the estimation errors of the ultimate gain and period would reduce drastically from -330% and 60% to -40% and -20% respectively, as shown in Fig. 13. For good tuning accuracy an even larger dead zone may be needed and the

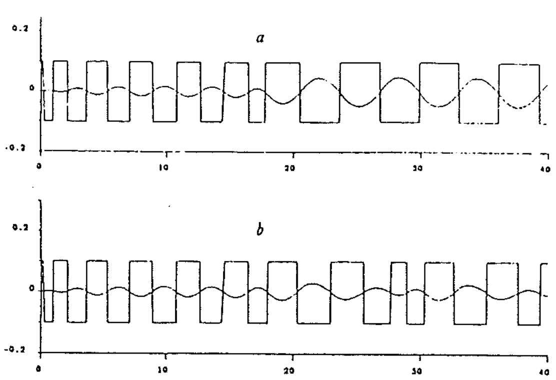

Fig. 11 Effects of a sinusoidal load of 1 rad/s (a) amp = -0.1 ; (b) amp = -0.05.

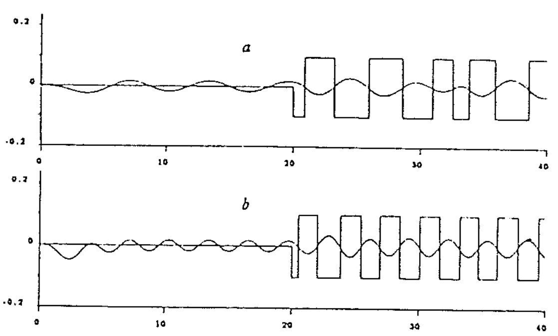

Fig. 12 Effects of varying the frequency of sinusoidal load : (1) 1 rad/s ; (2) 2 rad/s.

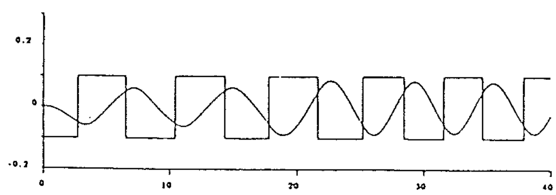

Fig. 13 Effects of using a dead zone of 0.05in the presence of a sinusoidal load of 1 rad/s.

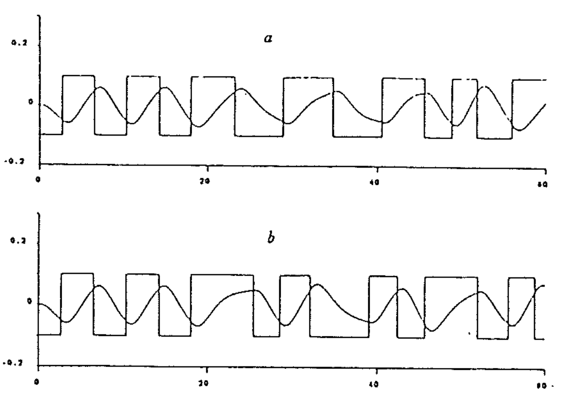

Fig. 14 Effects of a low frequency sinusoidal load (a) 0.5 rad/s ; (b) 0.25 rad/s.

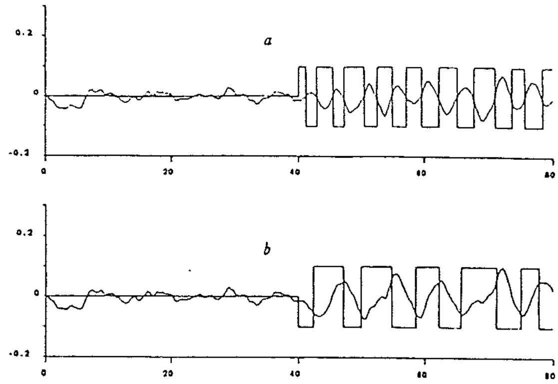

Fig. 15 Effects of a low frequency stochastic load (a) dead zone = 0 ; (b) dead zone = 0.05.

trade-off then is a longer duration of tuning and a larger oscillation of the process output, both of which may not be acceptable in practice. The use of a larger dead zone may also move the critical further away from the (-1 + j0) point and hence a somewhat slower closed-loop response may be resultant.

3.2 Error Detection

The simulation results shown in the previous section indicate that there may not be asymmetry or other clear indicator of possible error. A visual check by the operator or an automatic detection of significant presence of a sinusoidal load before tuning will thus be required. An exception, however, is found in the case of low frequency disturbance which can be intuitively regarded as a periodic offset. As shown in Fig. 14, the relay switching intervals could become asymmetrical when the load frequency is reduced to 0.5 rad/s or lower, and then automatic error detection becomes possible.

As a simple extension, the effect of low frequency stochastic disturbances, which may be more common than sinusoidal loads, could be related to static disturbances occuring at random intervals. The resultant asymmetry, an example being shown in Fig. 15, can thus be used to detect the significant influence of any low frequency stochastic load.

4. CONCLUSION

Auto-tuning of PID controllers using relay-feedback is known to be a robust technique which requires minimum prior knowledge of the process compared to other auto-tuning techniques. This paper has probed further into the performance of this auto-tuner in more demanding situations. It is shown that a static load disturbance could cause significant errors in the estimates of the ultimate gain and period. Fortunately the resultant asymmetry of the relay switching intervals could be used as an error detector, or alternatively used to compute a fast, corrective bias to restore accuracy. The corrective bias is functional even in the presence of moderate nonlinearity. A reliable self-biasing auto-tuner is thus resultant. For the less common occurrence of a sinusoidal load, significant errors in the estimates of tuning parameters could be produced and the detection of error may not be possible unless the load frequency is sufficiently low. A visual check by the operator or an automatic detection of significant presence of a sinusoidal load before tuning is therefore recommended as a commissioning procedure of the auto-tuner.

REFERENCES

Astrom, K.J. (1982). Ziegler-Nichols auto-tuners. Report TFRT-3167, Dept. of Automatic Control, Lund Inst. of Tech., Lund, Sweden.

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

Balchen, J.G. and B. Lie (1986). An adaptive controller based upon continuous estimation of closed loop frequency response. Proc. 2nd IFAC Workshop on Adaptive Systems in Control and Signal Processing. 31-36.

Bristol, E.H. (1977). Pattern recognition: an alternative to parameter identification in adaptive control. Automatica, 13, 197-202.

Hagglund, T. and K.J. Astrom (1985). Automatic tuning of PID controllers based on dominant pole design. Proc. IFAC Workshop on Adaptive Control of Chemical Processes, Frankfurt, FRG.

Hang, C.C. and K.J. Astrom (1988). Refinements of Ziegler-Nichols tuning formula for PID auto-tuners. To be presented at ISA annual Conf., Houston, October 1988.

Hang, C.C., C.C. Lim and S.H. Soon (1986). A new PID auto-tuner design based on correlation technique. Proc. 2nd Multinational Instrumentation Conf., China.

Hess, P., F. Radke and R. Schumann (1987). Industrial applications of a PID self-tuner used for system start-up. Proc. IFAC World Congress, Munich, FRG.

Higham, E.H. (1985). A self-tuning controller based on expert systems and artificial intelligence. Proc. Control 85, England, 110-115.

Hoopes, H.S., W.M. Hawk and R.C. Lewis (1983). A self-tuning controller. ISA Transactions, 22, 49-58.

Kraus, T.W. and T.J. Mayron (1984). Self-tuning PID controllers based on a pattern recognition approach. Control Engineering. 106-111.

Radke, F. and R. Isermann (1987) A parameter-adaptive PID controller with stepwise parameter optimization. Automatica, 23, 449-457.

Appendix I. Phase-margin Tuning Formula

If the ultimate gain K_d and period t_d are determined using a relay with cascade integrator, the corresponding point on the Nyquist curve will be the intersection on the imaginery axis at a frequency of $2/t_d$ and magnitude of $-1/K_d$. It is then straight forward to use this information, following Astrom and Hagglund (1984), to derive the following tuning formula to satisfy a phase margin specification of θ_m:

Proportional Gain $K = K_d \sin \theta_m$

Integral Time $$T_i = \frac{t_d (1 - \cos \theta_m)}{\pi \sin \theta_m}$$

Derivative Time $T_d = T_i /4$

Appendix II. Weighting On Setpoint

In the dominant pole design of Hagglund and Astrom (1985), a new tuning parameter 'b', which can be interpreted as a weighting factor on setpoint in the proportional term, has been introduced to modify the setpoint response independent of the load recovery response. This technique can be applied to the controller tuned by the Ziegler-Nichols formula which has been found to be near optimum for a step load change but producing an excessive overshoot for a step setpoint change. Using the symbol of Fig. 1 the PID controller becomes:

$$u = K \left[(b\, y_{ref} - y) + \frac{1}{T_i} \int e\, dt - T_d \frac{dy}{dt} \right]$$

It has been found experimentally that a small 'b' of 0.3 - 0.6 will reduce the large overshoot in setpoint response when the loop gain is high. Empirical formulae for computing 'b' are proposed in Hang and Astrom (1988).