Proceedings of the 17th Conference on Recent Advances in Robotics (FCRAR 2004)
Orlando, Florida, May 6-7 2004

Autotune of PID Cryogenic Temperature Control Based on Closed-Loop Step Response Tests

David Sheats*, Zvi S. Roth, Mark Sheats*, and Elizabeth Strehlow*****

* Scientific Instruments, West Palm Beach, Florida
** Florida Atlantic University, Boca Raton, Florida
*** Purdue University, West Lafayette, Indiana

Abstract

A novel PID control autotune technique, developed empirically at Scientific Instruments for temperature control systems operating in the Cryogenic regime (that is between 4-100K) is described. The PID parameters tuning is based on a sequence of closed loop step response tests utilizing the measured peak overshoot values and the time between the first peak and first dip of the step response.

The paper describes the structure of typical cryogenic temperature control systems dynamic models, including typical nonlinear effects (such as actuator's saturation and uni-directional operation). It then outlines the step-by-step tuning method. The tuning method produces satisfactory results for set-point changes of up to 10K over the entire cryogenic temperature range. The paper does not present Controls Systems theoretical justification for the tuning method. This is deferred to a future paper.

The paper also includes the results of a set of open-loop step response tests done at different temperature levels. The tests provide small-signal modeling data, that allows in principle the creation of a dynamic simulator, featuring the process gain, time constant and pure delay time, all as functions of the temperature. Details of the simulator are not shown, and these too are deferred to a future paper.

The method has been applied successfully worldwide to three fundamentally different types of cryogenic environments and over one hundred cryogenic temperature control systems.

I. Introduction

A typical temperature control process set-up in a cryogenic environment is that of a heater (a strip heater wrapped around a metal block or a cartridge heater inserted into the block) and a temperature sensor mounted on the block. The block is cooled at all times by either a cryogenic gas environment, or by a refrigerator. The control objective is to bring the block's temperature to a required level and maintain it at that level.

The process that relates the block temperature (in K) to the heater power (in W) is typically modeled, for sufficiently small set-point changes, and without counting the controller's dynamics, as a third-order linear system, taking into account the heat capacity of the heater, block and sensor, and the thermal resistances of the heater-block connection, heater-sensor connection and heater-cryogenic bath path. These capacities and resistances translate into three time constants in an all-pole transfer function [1]. Typically, one of the three process time constants is large (i.e. in the order of magnitude of seconds, or even tens of seconds), whereas the other time constants are fast (i.e. in the order of magnitude of a few milliseconds).

The open-loop step response of such processes has the familiar S-shape featuring a DC gain K, an approximated pure time delay L and a time constant τ, where τ represents the dominant time constant, that of the block, and L represents the other time constants in the system. Measurement of K, L and τ constitutes the Ziegler-Nichols Step-Response Test. The Ziegler-Nichols' recommended PID control parameters are functions of K, L and τ [2,3]. Such tuning is valid only for processes that indeed feature an S-shaped step response. Auto-Tuning according to the Ziegler-Nichols step response test may involve the automated measurement of K, L and τ, followed by an automatic computation of the P, I and D control coefficients.

Control researchers, over the years, tried to optimize the selection of PI and PID control parameters, given the values of K, L and τ, according to various quantitative performance criteria. The most complete compilation of formulas for the PID gains may be found in [4].

The accuracy of any PID tuning, based on an open-loop step response test, depends on errors in measuring the time delay and time constant parameters. The delay parameter L, in particular, may be difficult to measure in systems where one time constant far dominates all the others. Indeed, in the cryogenic environment described above, the block's time constant is often by orders of magnitude larger than the time constants of the heater and the thermometer. Many such processes may, for all practical purposes, be modeled in terms of a first-order linear system, assuming L=0, which renders all Ziegler-Nichols type step response PID tuning methods useless.

The questions of when one can use a PI controller, and when a PID controller must be used, do not have an explicit answer in most literature sources. In [5] it is argued that the controller, for processes which exhibit a single dominant time constant, should not be more complicated than a PI. According to [5], a PID controller is needed whenever the system has two dominant time constants. According to that, a typical cryogenic temperature control system may be controlled by a PI controller. However, the analysis in [5] is based on an implicit assumption that the process is always linear. Amplifier saturation, uni-directional control action nonlinearity and parameter variation effects are shown to necessitate the use of a PID controller, over a PI controller, as a better practical controller.

The actual PID algorithm has several known variants [6], the most common being the analog control law

$$U(s) = P(1 + \frac{1}{Is} + Ds)E(s) \qquad (1.1)$$

where, using Laplace Transform of the respective time signals, U(s) is the Laplace Transform of the heater power u(t) (measured in [W]). Likewise, E(s) is the Laplace Transform of the error signal e(t). The error e(t) is the difference between the set-point temperature r(t) and the temperature measured by the sensor T(t). The control constants P, I and D are the proportional, integral and derivative control gains, s is a differentiation operator and 1/s is an integration operator. Equation (1.1) is equivalent to the following time-domain control law:

$$u = Pe(t) + \frac{P}{I}\int_0^t e(x)dx + PD\frac{de(t)}{dt} \qquad (1.2)$$

A digital control implementation of (1.2) involves the sampling of T(t) every T_s [s]: T(0), $T(T_s)$, $T(2T_s)$, …. For the SI M9700 system T_s was set to T_s=0.133s. Denoting the error samples as e(0), e(1),…, e(k),…, the implementation of (1.2) requires the accumulation of error values ($\int_0^t e(x)dx \approx T_s \cdot \sum_k e(k)$) and the derivative approximation by a finite difference ($\frac{de(t)}{dt} \approx \frac{e(k)-e(k-1)}{T_s}$):

$$u(k) \approx Pe(k) + \frac{PT_s}{I}\sum_k e(k) + PD\frac{e(k)-e(k-1)}{T_s} \qquad (1.3)$$

Actually, the derivative control action was implemented by a more complex algorithm, allowing a limited amount of signal smoothing. More specific details of the D control and study of various D implementation options are deferred to a future paper.

Cryogenic environment heating (or cooling) is typically characterized by the following nonlinear phenomena:

a) Temperature dependent block's specific heat - If we represent the small-signal reduced process open-loop transfer function, after removing all fast time constants, as

$$\frac{T(s)}{U(s)} \cong \frac{K}{1+s\tau} \qquad (1.4)$$

then both DC gain K and the time constant τ are functions of the temperature T. Typically, both increase with the temperature. If the approximated model includes an aggregate pure time delay L:

$$\frac{T(s)}{U(s)} \cong \frac{K}{1+s\tau}e^{-sL} \qquad (1.5)$$

then L too is a function of T, typically an increasing function.

b) Amplifier saturation – The heater's power u(t) cannot exceed a given maximum power U_m. Furthermore, heater power is always non-negative. That is, if, in the closed-loop system, the temperature overshoots beyond the commanded temperature, the cooling down must be done in open-loop by the cryogenic surroundings. There is no active control action in the negative direction. The following equation summarizes these actuator constraints:

$$0 \le u(t) \le U_m \qquad (1.6)$$

c) Temperature-dependent Holding Power - We shall assume throughout that the temperature set-point command is always a step function of size r. If the amount of set-point change r is positive (i.e. r equals the desired temperature minus the initial temperature) the controller applies a heating signal from t=0. If r<0, there is initially a zero control signal, but near steady-state a nonzero control signal resumes. If the PI (or PID) controller keeps the closed-loop system stable, the steady-state temperature equals $T_\infty = T_0 + r$, exactly as intended, and there exists a steady-state "holding" heater power u_∞ that is an increasing function of T.

An issuing of a new set-point implies the resetting of the clock to t=0 (or k=0). The "present" temperature becomes the "initial" temperature. The present integral control memory (the summation of all error terms from t=0 to the present time t) is always the beginning "holding power." This term is only reset to zero if the control action is terminated and then restarted.

II. Problem Formulation

A basic problem with all PID tuning methods, based on a single or finitely many real-time tests, is the slow testing time. Let us elaborate on that point. A test based on open-loop step response (such as Ziegler-Nichols first tuning method) necessitates that we wait till steady-state is attained, to be able to measure the open-loop gain. The real-time measurement of the open-loop time constant may or may not require the attainment of steady-state. A tuning test based on closed-loop oscillation is even more time consuming. Ziegler-Nichols P control oscillation experiment, in which P is gradually increased, till self oscillations arise, requires the observation of several oscillation periods, to verify that indeed these are sustained, rather than decaying, oscillations. Astrom's relay experiment [7] is faster (data collection time can be as short of one oscillation period) but it requires a symmetric relay (i.e. symmetric positive and negative control actions), and as mentioned earlier, in most temperature control applications done at the cryogenic regime, the control is uni-directional.

As is well known in the PID Autotuning literature, the period of sustained oscillations obtained in a tuning test, contains information that is important for tuning the I control parameter. What this paper tries to demonstrate is that the same type of information can be drawn from decaying oscillations as well.

In a closed-loop P control set-up there is a set of measured step response features that can be observed relatively quickly and accurately. Peaks and dips in the signal are very recognizable – such peak values and their time of occurrence can be measured in a fully automated manner.

This paper examines two specific closed-loop P control step response features – the first peak overshoot and first dip-undershoot that comes after the peak overshoot. Three parameters associated with these peak and dip parts of the response play a role in the tuning process: the amount of peak overshoot OS, the gain P_c at which specific prescribed peak overshoot value is obtained, and the time difference ΔT between the first peak time and the first dip time.

The paper claims that such information can be used effectively to tune the PID control parameters in a fast, repeatable and fully automated manner, for all small-signal applications that evolve in a specific given narrow temperature range.

There is no single "correct answer" or "perfect PID values" for a given control environment. Near a specific temperature range there exist many combinations of PID values that may produce acceptable results. Users' needs may also vary widely. Some users may only be concerned with stable control over a long period of time at a single temperature, while others may be changing temperature often and desire rapid change. Some may be very concerned that the temperature does not overshoot the desired setpoint at all, while others are willing to accept overshoot for increased speed of response.

For most cryogenic temperature control applications, a closed-loop step response which features some moderate peak overshoot (i.e. 15% or less) is considered acceptable. Keeping in mind that the control action is uni-directional, there is zero control signal whenever the output temperature overshoots above the set-point. A closed-loop underdamped step response has negative slope segments in which the cooling down is governed by open-loop cryogenic bath influence. Whatever the cooling down time is, there is no sense to require that the peak time be significantly shorter than the cooling time. In the "language" of Linear Control theory these design specifications loosely translate into specified desired (and not necessarily optimal) region for the phase margin and the open-loop crossover frequency.

III. PID Auto-Tuning Procedure

The iterative step-by-step PID tuning procedure, summarized below, was obtained through exhaustive experimentation. It is an empirical result. The authors have been working on a Control Systems theoretical explanation to this particular method, and related class of methods. Such analytical derivations are deferred to a future paper.

The step-by-step auto-tuning procedure is as follows:

Step 0) The system starts at a default pre-programmed set of PID parameters. In case of multiple tuning procedures (say, due to a move to a new temperature range) the system's initial PID settings is made the same as the last used set of PID parameters.

Step 1) *Need to allow the system to stabilize at any temperature*. [Due to the continuing cooling action of the cryogenic bath, there is a need to apply a control heater signal equal to the system's holding power at such temperature. This requires keeping the controller's I term at a nonzero constant value. We typically stabilize the system at a temperature near which the control actions are to be taken].

Step 2) *A set-point command of 2K above the existing set-point is initiated using "P only" control.* [The controller's I term is held constant, at the value of Step 1. No further integration action is taken at any step of the tuning process. The D control term is set to zero]

Step 3) *The amount of overshoot at the first peak time is measured.* [At this point there is no need to measure the peak time].

Step 4) *If the overshoot is below 1.2K,* [i.e. temperature peak is less than 3.2K above beginning temperature] *the P term is adjusted up by 50% and step response test is repeated. If the overshoot is above 1.2K, the P term is adjusted down by 50% and experiment is repeated.*

Step 5) *The peak overshoot at the second trial is measured and a linear interpolation between the two trials' observed peaks is carried to determine what P value is expected to produce an overshoot of 1.2K. This becomes the* $P=P_c$ *value for the subsequent test step, and for the PID tuning.* [Application of this value of P may not actually produce an exact 1.2K overshoot].

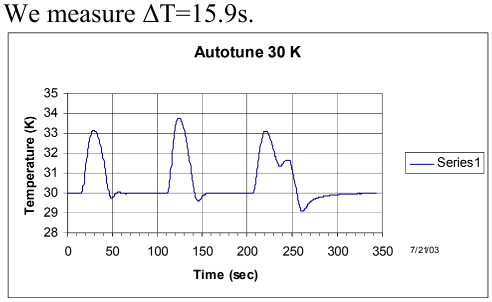
Step 6) *Using* $P=P_c$ *value, obtained at Step 5, a third trial [i.e. a 2K set-point shift] is performed, and the time* ΔT *between the first peak and the first dip [i.e. occurring after half cycle of the decaying oscillations] is recorded.*

Step 7) *The controller's I term is set to the oscillation period of Step 6:* $I = 2\cdot\Delta T$.

Step 8) *The controller's D term is set to D=I/4.* [The authors also experimented with D=I/7.

Comments: (a) The empirical search that led to the above tuning procedure revealed the special role played, during the tuning test, of a step response that has peak overshoot of around 60%. As the proposed tuning test does not advocate reaching a step response steady-state, it is clear that a search for an overshoot of 1.2K (which is 60% of a 2K set-point), is more easily implemented. Furthermore, the linear interpolation process done between the first two overshoot values predicts a 1.2K overshoot, but in reality the overshoot obtained in trial 3 is only in the vicinity of this desired value. The auto-tuning performance seems fairly robust to all these procedural inaccuracies, (b) The above tuning procedure, done with a set-point of 2K, works successfully for set-points that are as large as 10K. There are applications in which even larger set-point changes are sometimes issued, but in these cases the heater quickly reaches it upper saturation limit, and stays there for a while, till temperature approaches the desired level. This type of "large-signal control" is not discussed in this paper.

Experiment results, conducted on the SI M9700 Temperature Controller, are shown in Figure 3.1 and serve to illustrate the tuning procedure. At T=30K the first overshoot at t=29.1s is 1.149^0K. Shortly after the system is returned to T=30K. In trial 2, the overshoot (at t=124.5s) is 1.774K. Again, the system is reset. The actual peak overshoot, obtained in trial 3 is only 1.103K (at t=221.1s). The first dip is 0.657K below the set-point command of T=32K, occurring at t=237.0s. We measure ΔT=15.9s.



:

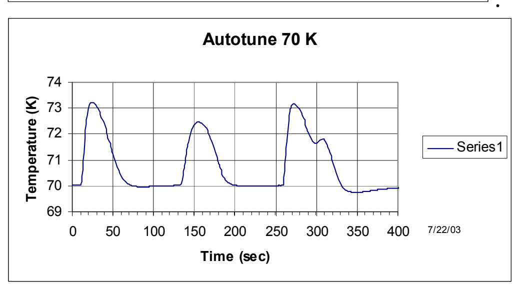


Figure 3.1: Auto-tuning experiments at two different temperatures.

The experiment at T=70K is similar, except for a larger than 1.2K overshoot occurring at trial 1, and therefore P was reduced by 50% for trial 2. Trial 3 produced an overshoot of 1.161K and ΔT=25.8s.

Of course, the auto-tuning at each temperature range produced two different sets of PID control parameters, each is valid for small signal controls in the vicinity of the respective temperature. In other words, if an application is conducted near $T=30^0K$, with a certain set of PID parameters, and there is a need to conduct another application near $T=70^0K$, then there is a need to re-tune at the higher temperature.

The resulting closed-loop step responses, with the PID parameters, set according to the tuning procedure, are shown in Figure 3.2:

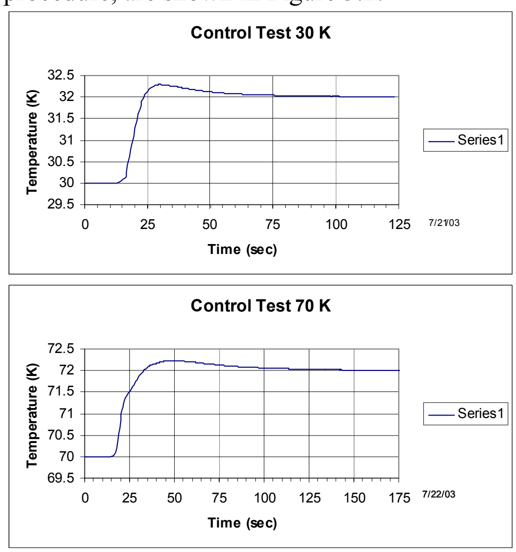


Figure 3.2: Closed-loop PID control performance.

In both cases shown, the resulting closed-loop peak overshoot is below 15%, and the peak time is in the same order of magnitude as the cooling-down time.

IV. Assessment of Control Performance

The first set of experiments compared PID control values, obtained by the new tuning method, to ones obtained via classic step-response and closed-loop oscillations Ziegler-Nichols tests. Figure 4.1 shows P and I values obtained by the three tuning methods, over the temperature range of 15-90K.

We see that the new tuning method developed at Scientific Instruments (SI) (thereafter referred to as the SI Method) recommends the same P control values as the ZN method, at the low temperature range (15-40K), and a significantly lower P value at the higher temperature range. On the other hand, the SI method tends to use relatively larger I (and consequently D) control values, compared to the ZN method.

For users who are interested in temperature control near a specific temperature, the SI tuning may be followed by a process of adjustments, aimed at optimizing better the response parameters. Figure 4.2 illustrates such optimization done to the resulting recommended SI PID parameters, via application of a qualitative set of adjustment guidelines listed in Table 4.1. Comparison of Figures 3.2 and 4.2 reveals that there is a tradeoff between overshoot size and speed of response

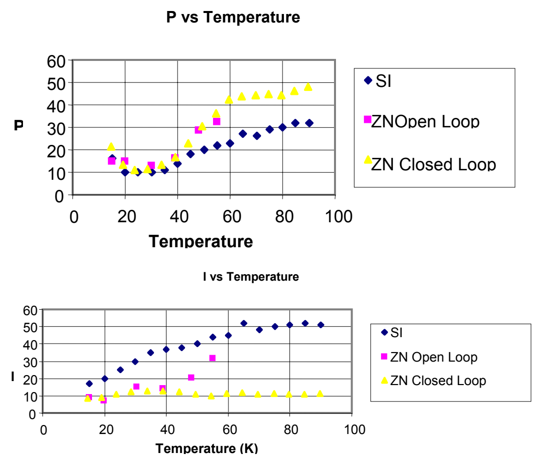


Figure 4.1 Comparison of PI parameters between new tuning method and Ziegler-Nichols (ZN) methods.

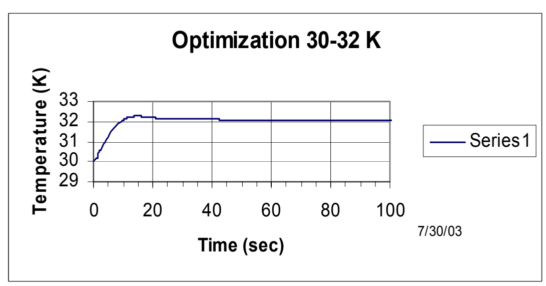


Figure 4.2: Optimized response obtained via adjustments of SI PID parameters at a specific temperature range.

Adjustments may be done aggressively (changing by doubling or halving a gain) or more conservatively, as desired.

Adding a D term to a PI controller provides for faster response with less overshoot. A particularly noisy environment or poor sensor resolution in a given temperature may render the D term of minimal usage. Users may opt to start with D=0 and gradually increase its value, observing the effect on performance.

Response time	Overshoot	I	P
Too slow	OK	Decrease	Increase
OK	Too large	Increase	Increase
Too slow	Too large	Hold	Increase
Too fast	Too large	Increase	Hold

Table 4.1: Qualitative PID adjustment guidelines

V. Modeling Results

The M9700 process is a typical representative of temperature controlled processes at the cryogenic temperature range. It is of interest to observe how the various system parameters vary as functions of the temperature T.

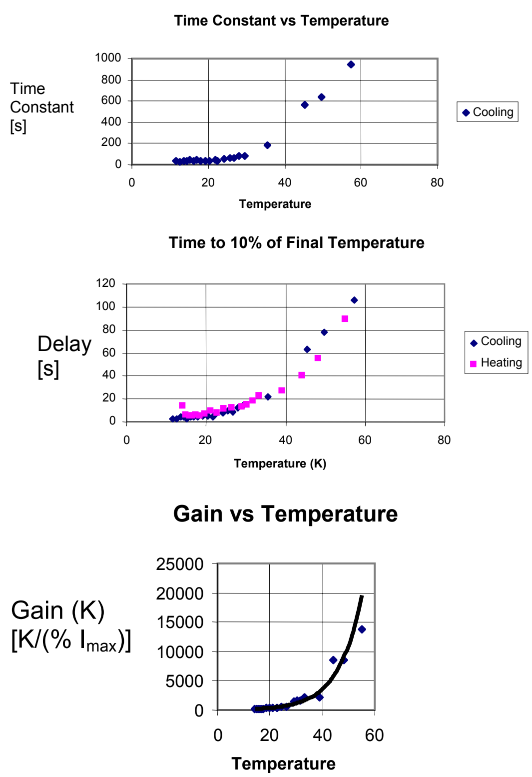


Figure 5.1 Time constant τ(T), Delay L(T) and Gain K(T) [Refer to formula (1.5)]

A set of open-loop small-signal step response tests were conducted, using very small set-point commands (0.1-0.3K) in both directions – heating up as well as cooling down. The sets of time constant and time delay parameters, as functions of T, obtained in the heating and cooling experiments were almost identical.

Figure 5.1 shows the variation of process small-signal time constant, gain and time delay, as functions of T.

In the literature of PID control tuning, a parameter that appears most frequently is the ratio of pure time delay L to the process time constant τ (L/τ). The results point to a typical ratio of between 1:8 and 1:10 at all temperatures. We should note though that the measured L is not really a pure time delay. It only represents the two fast time constants that exist in such processes. Two types of measurements – time to 10% and time to 0.1K were used and produced consistent results. Only the "time to 10%" is shown.

Figure 5.2 shows the measured steady-state Holding Power as function of the temperature T.

Least-squares curve fitting to the above M9700 system small-signal parameter produced the following results:

$$\tau = 11.238\exp(0.075T) \qquad R^2 = 0.9918$$

$$L_{10\%} = 2.1038\exp(0.0667T) \qquad R^2 = 0.8997$$

$$L_{0.1K} = 2.112\exp(0.0569T) \qquad R^2 = 0.9132$$

$$Holding\_Power = 12.235\ln T - 16.004$$

$$R^2 = 0.9724$$

$$K = 0.2483\exp(0.0772T) \qquad R^2 = 0.9506$$

The units of τ and L are seconds, temperature T is in K, holding power is in % of the maximum heater current, and gain K is in Kelvin/(% of maximum current).

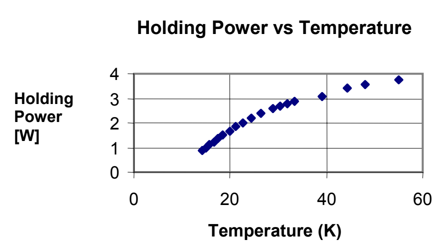


Figure 5.2: Holding Power as function of Temperature (K).

The above modeling results can have several applications: a) Construction of a dynamic simulator, and b) Development of large-signal control strategies.

Adaptive large-signal control requires, not only the absolute dependence of the above process parameters on T, but also data of the speed of parameters change, as a function of T. This too has been deferred to a future study.

VI. Summary and Future Work

The SI PID auto-tuning method has been applied successfully worldwide to three fundamentally different types of cryogenic environments and over one hundred cryogenic temperature control systems.

The tuning is done for set-points of 2K but the validity of the control parameters has shown to be holding for up to ±10K set-point changes. Larger set-point changes, if attempted, lead to heater saturation (i.e. maximum current limitation), and thus have not been carefully evaluated.

Future work will focus on finding theoretical justification to the success of the method. It is unclear, at this point, whether the closed-loop step response tuning test may prove successful in other industrial PID control application, or if its applicability may be limited to cryogenic temperature control environments only.

References

[1] E.M. Forgan, "On the use of temperature controllers in cryogenics", Cryogenics, Vol. 14, No. 4, April 1974.

[2] J.G. Ziegler and N.B. Nichols, "Optimum Settings for Automatic Controllers", Transactions of the ASME 1943

[3] K. Ogata, "Modern Control Engineering", Third Edition, Prentice Hall 1997, pages 669-674.

[4] M. Zhuang and D.P. Atherton, "Automatic tuning of optimum PID controllers", IEE Proceedings-D, Vol. 140, No. 3, May 1993, pp.216-224.

[5] A. Voda and I.D. Landau, "A Method for the Auto-Calibration of PID Controllers", Automatica, Vol. 31, No.2, February 1995.

[6] J.P. Gerry, "A Comparison of PID Control Algorithms", Control Engineering, March 1987, pp. 102-105.

[7] K.J. Astrom and T. Hagglund, "Automatic Tuning of Simple Regulators with Specifications on Phase and Amplitude Margins", Automatica, Vol. 20, No. 5, pp. 645-651, 1984.