Modeling Aspects of Temperature Controlled Cryogenic Processes

David C. Sheats
Scientific Instruments Inc. West Palm Beach, FL 33407
dsheats@scientificinstruments.com

Zvi S. Roth
Florida Atlantic* University, Boca Raton, FL 33431
rothz@fau.edu

Joseph W. Snyder
Scientific Instruments Inc. West Palm Beach, FL 33407
jsnyder@scientificinstruments.com

ABSTRACT

The Scientific Instruments, Inc. (SI) feedback control PID tuning is based on a sequence of P control 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. This autotuning technique, developed empirically at Scientific Instruments, Inc. for temperature control of systems operating in the cryogenic regime (i.e. - below 100K), and its associated set of experiment results were described in detail in an earlier paper [1].

This paper focuses on modeling studies aimed at identifying the relationship between the process time constants and specific physical system entities such as specific heat and thermal resistances.

Keywords

PID, Temperature control, Autotuning, Cryogenic, SI method.

1. MODEL FACTS

A typical application of temperature control in a cryogenic environment would be 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 through contact with either a cryogenic gas or a refrigeration stage. The control objective is to bring the block to thermal equilibrium at a user-specified temperature and maintain it at that level.

The process that relates the block temperature (in Kelvin) to the heater power (in Watts) is typically modeled, for sufficiently small set-point changes, and without counting the controller's dynamics, as a third-order linear system. This model must take 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 [2]. Typically, one of the three process time constants is large (i.e. in the order of magnitude of tens or hundreds of seconds), whereas the other time constants are fast (i.e. in the order of magnitude of a few seconds to a fraction of a second). Indeed, the block's time constant is often orders of magnitude larger than the time constants of the heater and the thermometer.

This paper focuses on the physical components of these three time constants and the process DC gain.

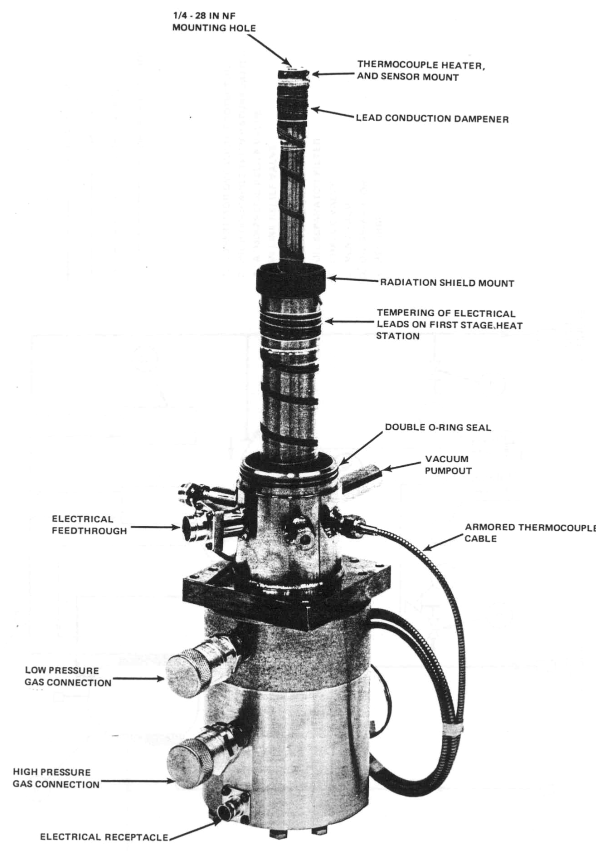

Figure 1.1: The SI M9700 process.

The simplest analog form of the PID control algorithm is:

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

where U(s) is the Laplace Transform of the heater power u(t) (measured in [W]) and E(s) is the Laplace Transform of the error signal e(t). The error e(t) is the difference between the set-point temperature T_{ref} and the temperature measured by the sensor T(t). The control constants P, I and D are the proportional, integral and derivative control gains.

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

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

$$\frac{T(s)}{U(s)} \cong \frac{K}{1+st}$$

then both the DC gain K and the time constant τ are functions of the temperature. Typically, both increase with the temperature. Assuming that all fast time constants contribute an aggregate pure time delay L to the model:

$$\frac{T(s)}{U(s)} \cong \frac{K}{1+st} e^{-sL} \qquad (1.2)$$

then L is also a function of T, typically an increasing function. Experimental curves of K(T), t(T) and L(T) for the SI M9700 process were presented in [1].

b) Amplifier saturation – The heater's power u(t) cannot exceed a given maximum power U_m. Furthermore, heater power is always non-negative. In the closed-loop system this means that if the temperature overshoots beyond the commanded temperature the cooling down is 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:

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

c) Temperature-dependent Holding Power - If the amount of set-point change r is positive (r being equal to 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 at a fixed temperature of $T_\infty = T_0 + r$ then there exists a steady-state "holding" heater power u_8 that is an increasing function of T.

An issuing of a new set-point implies the resetting of the clock to t=0. The "present" temperature becomes the "initial" temperature. The present integral control memory (i.e. 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.

PID control references listed in [1] and elsewhere are geared towards systems such as described by (1.2). Fast time constants are normally part of the process' "un-modeled dynamics" making it convenient for control researchers to work with an aggregated pure-time delay L. The very essence of PID process control is a "black box" view of the process: "A feedback control is found based on a single or a few tests, without knowing the detailed model parameters and without attempting to identify the process' mathematical model." As was pointed in [2] the simulation studies for the SI PID control method required an alternate equation to (1.2) in which the fast time constants were explicitly taken into account. Let t_1 denote the slow time constant of the open-loop process, and let $t_{f1}, t_{f2},\ldots, t_{fN}$, denote N fast time constants. Without loss of generality (in the simulations conducted in [2]) all individual fast time constants were replaced with a single multiple pole of multiplicity N located at $s = -1/t_f$. N is an integer greater or equal to 2.

$$\frac{T(s)}{U(s)} \cong \frac{K}{(1+st_1)(1+st_{f1})\ldots(1+st_{fN})} \approx \frac{K}{(1+st_1)(1+st_f)^N} \qquad (1.3)$$

Cryogenic temperature control processes typically have N=2 for the number of fast time constants. The SI PID autotuning method has been implemented successfully on over hundred systems worldwide. In one isolated case the tuning method provided suboptimal performance when the user attempted to connect two DC power supplies in series to power the control heater. The two power supplies and the thermometer all constituted three fast time constants (N=3).

2. SUMMARY OF THE SI PID AUTOTUNING ALGORITHM

The iterative step-by-step PID tuning procedure, summarized below, was discovered through exhaustive experimentation. It is an empirical result. The simulation studies reported in this paper are in part aimed at obtaining better insight to eventually allow for Control Systems theoretical explanation of this particular method, and related class of methods.

The step-by-step auto-tuning procedure presented in [1] is detailed below, with some additional modifications:

Step 0) The system starts at a default pre-programmed set of PID parameters. In the 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) 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 issued 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 output peak $T_{peak,1}$ at the first peak time is measured and the initial value of $P=P_1$ is recorded. [At this point there is no need to measure the peak time].

Step 4) If peak temperature $T_{peak,1}$ is less than 3.2K above the beginning temperature the P term is adjusted up by 50% (that is, between the two trials' observed peaks is carried to determine what P value is expected to produce a peak of $T_{peak,3}$= 3.2K.

$$P_3 = P_1 + \frac{T_{peak,3} - T_{peak,1}}{T_{peak,2} - T_{peak,1}} \cdot (P_2 - P_1) \quad (2.1)$$

The value P_3 becomes P for the subsequent test step, and it is also taken to be the control value of P for the PID tuning. [Application of $P=P_3$ may not actually produce an exact 3.2K peak, as overshoots vs. P does not obey a linear model]. If $T_{peak,2}$=3.2K we go to Step 6, setting $P=P_3=P_2$ and allow the response to continue to the first dip before being reset to the original set-point.

$P_2 = 1.5P_1$) and step response test is repeated. If the temperature peak is larger than 3.2K, the P term is adjusted down by 50% (that is $P_2 = 0.5P_1$) and the experiment is repeated . In either case the system is reset to its original set-point prior to starting the second test. If peak temperature $T_{peak,1}$ equals 3.2K, we jump to the end of Step 6 by setting $P=P_1$ and allowing the response to continue to the first dip following the first peak.

Step 5) The peak overshoot at the second trial $T_{peak,2}$ is measured and if different than 3.2K a linear interpolation

Step 6) Using the P value obtained at Step 5, a third trial [i.e. a 2K set-point shift] is performed, and the time ?T 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·?T.

Step 8) The controller's D term is set to D=I/4.

In [2] it was demonstrated that the SI PID autotuning method performs well only when the number of fast time constants is N=2.

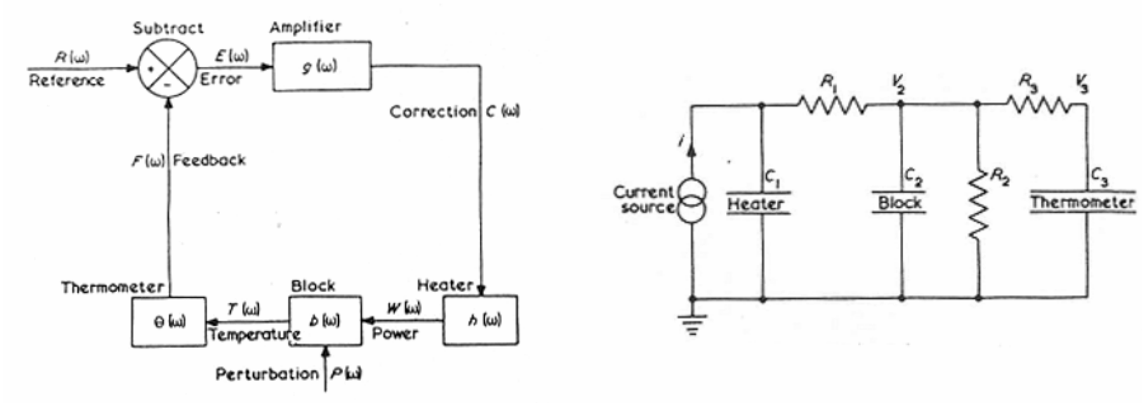

Figure 3.1: Functional structure and electrical analog model of the temperature control system operating in the cryogenic range.

3. PROCESS THERMAL MODELING

Refer to Figure 3.1 taken from Forgan [3]. The components' indices follow the order of Heater → Block → Thermometer. That is, C_1 is the heat capacity of the heater, R_1 is the thermal resistance between the heater and the block, C_2 is the heat capacity of the block, R_2 is the thermal resistance between the block and the reference level (that is, a "ground" zero voltage in the electrical analog system represents the temperature of the cryogenic bath), R_3 is the thermal resistance between the block and the thermometer and C_3 is the heat capacity of the thermometer.

Assumptions:

(A1) The block can be represented by a lumped model. We neglect thermal propagation effects within the block system itself.

(A2) The heat capacity of the block is much larger than those of the heater and thermometer. That is, $C_2 >> C_1, C_3$.

(A3) The thermal resistance of the block is much larger than those of the heater and thermometer. That is, $R_2 >> R_1, R_3$.

We define three thermal time constants as $t_i = R_i C_i \quad i = 1,2,3$, and loosely refer to these as the heater's, block's and thermometer's thermal time constants. Consequently of (A2) and (A3) we have that $t_2 >> t_1, t_3$ (Assumption (A3')).

In the electrical analog model voltage V_3 represents the temperature of the thermometer. The input current source represents the input heat generated by the heater.

By a straightforward circuits theory calculation the exact expression for the transfer function $V_3(s)/I(s)$ (where s is the Laplace Transform variable), in terms of all heat capacities and thermal conductance parameters, is found to be (see Appendix A for the calculation details):

$$\frac{V_3}{I}(s) = \frac{G_1 G_3}{C_1 C_2 C_3 s^3 + a_2 s^2 + a_1 s + G_1 G_2 G_3}$$
$$a_2 = C_1 C_2 G_3 + C_1 C_3 G_1 + C_1 C_3 G_2 + C_1 C_3 G_3 + C_3 C_2 G_1 \tag{3.1}$$
$$a_1 = C_3 G_1 G_2 + C_3 G_1 G_3 + C_2 G_1 G_3 + C_1 G_1 G_3 + C_1 G_3 G_2$$
$$G_i = R_i^{-1} \quad i = 1,2,3$$

Using assumptions (A2)-(A3) to neglect some of the terms in the coefficients a_1 and a_2, approximates transfer function (3.1) by (see details in Appendix B):

$$\frac{V_3}{I}(s) \approx \frac{R_2}{t_1 t_2 t_3 s^3 + (t_1 t_2 + t_2 t_3) s^2 + t_2 s + 1} \tag{3.2}$$

Assumption (A3') is used to artificially add a few negligible terms to the coefficients of s^2 and s such that:

$$\frac{V_3}{I}(s) \approx \frac{R_2}{(1 + s t_1)(1 + s t_2)(1 + s t_3)} \tag{3.3}$$

This intuitively plausible separation of heater, block and thermometer time constants is somewhat unexpected (from a circuits theory viewpoint – see Appendix B). It is valid though only whenever the block's time constant is much larger than those of the heater and thermometer.

The thermal resistance of the block which appears as the numerator of $V_3/I(s)$ is exact. This is the gain K of the system's open-loop transfer function.

4. SUMMARY AND CONCLUSIONS

We developed simple and physically intuitive approximated formulas to the gain and time constants of typical temperature controlled cryogenic processes. In an earlier paper [1] curve-fitting to open-loop step response tests yielded the temperature functions K(T) = R_2(T) and the slow time constant t_2(T) = R_2(T)C_2(T). In addition, curves of "time-to-10%" (of the step response steady-state value), provided some partial indirect information about the fast time constants t_1(T) and t_3(T). The results of this paper pave the way for better understanding of the physical explanation of the above experimentally obtained curves.

In [3] several exceptional cases, such as those in which the temperature sensor is located far away from the block, are discussed. In such cases the simple model (3.3) may no longer be valid. These cases were not included in our paper.

It is important to note that the gain and time constants are only thermal steady-state features valid in control tuning involving small set-point changes (up to 10 K). Large set-point change control set-ups may involve rapid temperature changes over a large range of temperatures. This may necessitate more sophisticated control design, in which some understanding of the thermal convergence time ("mixing time") for the gain and time constant parameters may be necessary.

5. REFERENCES

[1] David Sheats, Zvi S. Roth, Mark Sheats, and Elizabeth Strehlow, "Autotune of PID Cryogenic Temperature Control Based on Closed-Loop Step Response Tests", Proceedings of the 17th Conference on Recent Advances in Robotics (FCRAR 2004), Orlando, Florida, May 6-7 2004.

[2] David C. Sheats, Zvi S. Roth, Jon K. Burford and Joseph W. Snyder, "Simulation Studies of the SI PID Control Autotuning", Proceedings of FCRAR 2005, Florida Conference on Recent Advances in Robotics, May 5-6 2005, University of Florida, Gainesville Florida.

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

Appendix A: Computation of (3.1)

Referring to the circuit diagram in Figure 3.1, we apply Kirchhoff's Current Law at the nodes of V_1, V_2 and V_3 and obtain the three equations:

$$I = sC_1 V_1 + G_1 (V_1 - V_2)$$
$$G_1 (V_2 - V_1) + sC_2 V_2 + G_2 V_2 + G_3 (V_2 - V_3) = 0 \tag{a1}$$
$$G_3 (V_3 - V_2) + sC_3 V_3 = 0$$

One may solve these equations directly to obtain V_1, V_2 and V_3 as functions of I.

Alternatively, (a1) may be rearranged as:

$$V_1 = \frac{1}{sC_1 + G_1} I + \frac{G_1}{sC_1 + G_1} V_2$$
$$V_2 = \frac{G_1}{sC_2 + G_1 + G_2 + G_3} V_1 + \frac{G_3}{sC_2 + G_1 + G_2 + G_3} V_3 \qquad \text{(a2)}$$
$$V_3 = \frac{G_3}{sC_3 + G_3} V_2$$

Equations (a2) provide the gains for a Signal Flow Graph having an input node I and output nodes V_1, V_2 and V_3.

By Mason's formula $\frac{V_3}{I} = \frac{\sum_i P_i \Delta_i}{\Delta}$, P_i represent all "forward path gains" leading from I to V_3. In this case there is only one forward path $I \rightarrow V_1 \rightarrow V_2 \rightarrow V_3$. Its gain equals:

$$P_1 = \frac{1}{sC_1 + G_1} \cdot \frac{G_1}{sC_2 + G_1 + G_2 + G_3} \cdot \frac{G_3}{sC_3 + G_3} \qquad \text{(a3)}$$

The signal flow graph forms two touching loops $V_1 \rightarrow V_2 \rightarrow V_1$ and $V_2 \rightarrow V_3 \rightarrow V_2$. Therefore:

$$\Delta = 1 - \frac{G_1}{sC_1 + G_1} \cdot \frac{G_1}{sC_2 + G_1 + G_2 + G_3} - \frac{G_3}{sC_3 + G_3} \cdot \frac{G_3}{sC_2 + G_1 + G_2 + G_3} \qquad \text{(a4)}$$

Furthermore

$$\Delta_1 = 1 \qquad \text{(a5)}$$

Substitution of (a3)-(a5) into Mason's formula yields (3.1).

Appendix B: Approximations to (3.1)

Starting with the coefficients of s^2 in the denominator of (3.1): By assumption (A3), $C_1C_3G_2 << C_1C_3G_1$. We thus neglect $C_1C_3G_2$.

By assumption (A2), $C_1C_3G_1 << C_2C_3G_1$. We thus neglect $C_1C_3G_1$.

Similarly, by assumption (A2), $C_1C_3G_3 << C_1C_2G_3$. We thus neglect $C_1C_3G_3$.

In the coefficients of s (in the denominator of (3.1)) we first observe that due to Assumption (A3) $C_3G_1G_2 << C_3G_1G_3$. We thus neglect $C_3G_1G_2$. Similarly, we neglect $C_1G_3G_2$ with respect to $C_1G_3G_1$. In the remaining three terms we use Assumption (A2) to neglect the two terms $C_3G_1G_3$ and $C_1G_1G_3$ with respect to $C_2G_1G_3$.

We then divide the numerator and denominator of (3.1) by $G_1G_2G_3$. The resulting expression is (3.2).

Assuming now that the poles of (3.2) are real, $s_i = -1/t_i$ for i=A,B,C, we compare the polynomial $(1+st_A)(1+st_B)(1+st_C)$ to the denominator of (3.2). Assuming that $t_A = t_1$, $t_B = t_2$ and $t_C = t_3$, it is observed that the two polynomials differ by a few negligible terms that can be added to make the two polynomials equal. The final result (equation (3.3)) is now evident.

The result may come as a surprise to circuit analysts who are used to the practice of estimating time constants by sort of "superposition" (for each capacitor, we find the equivalent resistance "seen" in parallel to the capacitor, assuming that the other capacitors are short circuits). This method will erroneously lead to the three time constants: C_1R_1 (this one is correct!), $C_2(R_1||R_2||R_3) \approx C_2(R_1||R_3)$ (which is wrong) and C_3R_3 (which is correct).

The "DC gain" R_2 of V_3/I is obtained directly from observation of the circuit diagram, assuming that all capacitors are open circuits.