By Peter Galan, OCM Technology Inc

TODAY'S INDUSTRIAL APPLICATIONS DEMAND BOTH SPEED AND PRECISION IN TEMPERATURE CONTROL. AN IMPROVED TEMPERATURE CONTROLLER MEETS THOSE DEMANDS BY ADDING FEATURES YOU DON'T FIND IN TRADITIONAL PID CONTROLLERS.

Enhanced temperature controller is both fast and precise

Despite the common perception of temperature control as a mature and largely unchanging area of technology, applications such as injection-molding processes increasingly require not only precise temperature control but also fast response to heat disturbances and minimal overshoot and undershoot when temperature setpoints change. Traditional PID (proportional-integral-derivative) control techniques are inadequate; extra capabilities are necessary.

Figure 1's continuous-time (analog) PID controller, which is typical of most closed-loop control systems, calculates a regulation error, e(t), as the difference between a desired value, or setpoint, and an actual value, x(t). The proportional component of the controller multiplies the error signal by a proportional constant, K_P. The integral component calculates an integral of e(t) and multiplies it by an integral constant, K_I. The derivative member calculates a derivative of e(t) and multiplies it by a derivative constant, K_D. The sum of these three components is the actuating value, y(t), which controls the process:

$$y(t) = K_P e(t) + K_I \int e(t)dt + K_D \frac{de(t)}{dt}. \qquad (1)$$

More and more, today's controllers exist in digital form, so, instead of using continuous-time signals, they work with digitized samples. **Equation 1**'s representation of **Figure 1**'s analog PID controller thus becomes a discrete-time implementation:

$$y_i = K_P e_i + TK_I \sum e_i + \frac{K_D}{T}(e_i - e_{i-1}), \qquad (2)$$

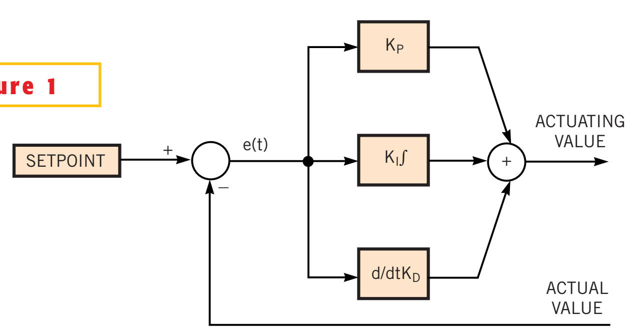
where y_i is the actuating value at the current time, i; e_i is the regulation error at time i; e_{i-1} is the regulation error at the previous sample time, i_{-1}; and T is the time period of the sampling and processing.

For practical application, **Equation 2** requires certain modifications, beginning with the integral member. **Equation 2** adds each value of the regulation error to a sum and then multiplies this sum by the time constant and the integral constant. If the value of the time constant or the integral constant changes, which can happen, especially during a system tuning process, the actuating value abruptly changes and causes problems. A better approach is to multiply the regulation error by these two constants before accumulating the product. You can achieve higher precision by using trapezoidal integration instead of rectangular integration to reduce errors.

The derivative member of **Equation 2** is a second source of problems. In its simple form, this member tends to be rather noisy. To reduce the noise, you can use more than two (say four) suitable filtered samples of the regulation error. The modified PID formula then appears as:

$$y_i = K_P e_i + \sum TK_1\left(\frac{e_i + e_{i-1}}{2}\right) + \frac{K_D}{6T}(e_i - e_{i-3} + 3(e_{i-1} - e_{i-2})). \qquad (3)$$

Figure 1



A PID controller calculates an actuating value from proportional, integral, and derivative components.

Heating processes exhibit their own behavior, however, and you have to take their special characteristics into account. Consider, for example, a barrel in a plastic-molding injection machine. **Figure 2**'s thermal model of this typical heating system is a simplified model that ignores heat-transport delays.

Assume that a constant power source, W, when switched on, powers an electrical heating element with a heat capacity C_h. The temperature of the heating element, $T_h(t)$, rises with time, t. The heat continuously propagates, by direct conduction, to the barrel. R_{hb} represents a thermal resistance between the electrical element and the barrel, and C_b represents the barrel's thermal capacity. The barrel temperature, $T_b(t)$, also rises with time. R_b thus represents a thermal resistance between the barrel and surrounding environment (with ambient temperature $T_a(t)$), which tends to cool the barrel. Ideally, without any cooling from the outside, both temperatures, $T_h(t)$ and $T_b(t)$, would rise forever. In practice, however, natural cooling prevents this situation from occurring. So, after a certain time period, both temperatures stabilize at certain constant values.

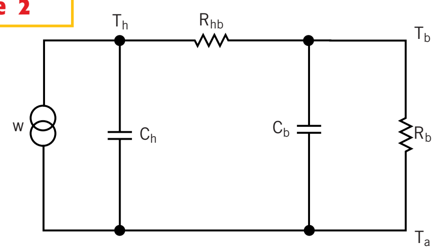
Figure 2's model represents a second-order system. However, in practice, the output, or barrel, temperature follows a simple exponential curve when rising or falling, especially if R_{hb} is relatively small.

Increasing the temperature of a typical heated system from the surrounding temperature by a couple of degrees takes substantially less time than to let it cool down to the surrounding temperature. On the contrary, increasing the temperature of the same system by a couple of degrees when it is close to its maximum temperature takes substantially longer than to bring the temperature back down. **Figure 3** shows situations for three setpoints. One is close to the ambient temperature, one is at the balanced temperature, and one is close to the maximum temperature. Temperatures are rising and decreasing exponentially with different rates for heating and cooling. The ratio of the rate of heating and cooling processes at any stage depends only on the value of the setpoint.

In one situation, the heating rate and the cooling rate are equal: at the temperature that requires application of exactly 50% of the maximum applicable power to the heater. What is that "balanced" temperature? Theoretically, from **Figure 2**'s thermal model, the dependence between power, W, and temperature, T_b, should be linear. That is, the stabilized temperature corresponding to 50% applied power should be exactly in the middle between the minimum and maximum temperature (**Figure 3**). The assumption here is that the maximum reachable temperature requires 100% power and the minimum—that is, ambient—temperature requires 0% power, both to be applied for an unlimited time.

Practical measurements, however, show that the temperature-power relationship is not perfectly linear. The reason is that the thermal resistance, R_b, is not constant; it changes with respect to the temperature gradient (the difference between T_b and T_a). Heat transfer between the barrel and the surrounding environment consists of radiation and convection (conduction by air) components. It is probably the convection-cooling process that changes its characteristics with the temperature gradient. The direct result of the nonlinear power-temperature relationship is that the temperature corresponding to the 50% output power is higher than the midpoint between minimal (ambient) and the maximal temperatures (**Figure 4**).

Figure 2



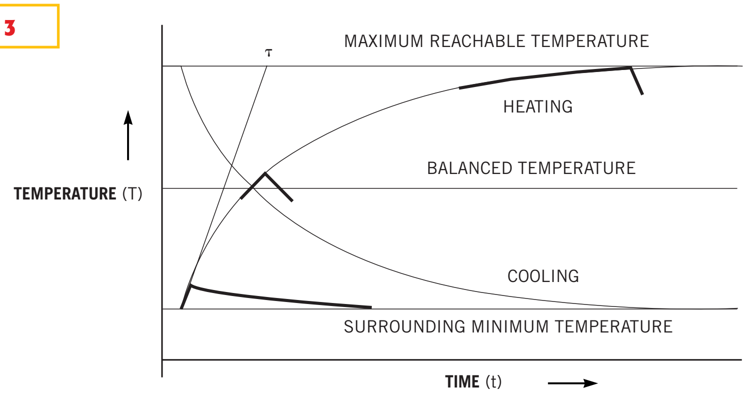
A simple heating-system model has thermal resistances and capacitances but ignores heat-transport delays.

EFFECT ON PID CONSTANTS

Different rates of the heating and cooling processes affect the required values of the PID constants, which basically depend upon the gain and the time constant of the controlled system. At least two main autotuning methods measure these two parameters to compute PID constants. Autotuning methods generally provide only one set of PID parameters, which are suitable for the setpoints close to the balanced temperature. These parameters require no modifications, because, as noted, the heating and cooling rates are equal near the balanced temperature.

A different situation exists when the temperature setpoint is different from the balanced temperature. The further away the setpoint is from the balanced temperature, the further the PID con-

Figure 3



Heating and cooling rates vary with temperature. They are equal at the balanced temperature, which is theoretically midway between the ambient and maximum temperatures.

stants are from their optimal values. Just how far away they are depends on how far the setpoint is from the balanced temperature and whether the current temperature is below or above the setpoint.

One of the two most common autotuning methods can calculate the system gain and the time constant; the other cannot. Instead, the other method directly calculates the rate of temperature change, which is basically the same information. However, both autotuning methods calculate the K_P, K_I, and K_D PID constants as inversely proportional to the rate of temperature change. The rate of temperature change is not a linear function of temperature because the derivative of the exponential is also exponential. However, a straight line is a reasonable approximation for the rate of change (**Figure 5**).

The rate of heating at the balanced temperature is $1/2\tau$, where τ is a time constant of the exponential (**Figure 3**). At the ambient temperature, $1/2\tau$ is the rate of heating. The rate of cooling is infinitesimally small at the ambient temperature. The opposite situation exists at the other end of the exponential curve, close to the maximum temperature. There, the rate of heating is infinitesimally small, and the rate of cooling is high.

To compensate for varying heating and cooling rates, the K_P, K_I, and K_D PID constants must vary, so the basic PID controller requires modification. You can add an adapter and heater model (**Figure 6**).

The heater model must initially determine a value of the balanced temperature that corresponds to 50% application of power to the heater. Autotuning can assist in this determination. Next, the heater model must find the position of the current setpoint temperature relative to the balanced temperature and then calculate the values of two coefficients you need for modification of the PID constants. The adapter performs the actual modification with two linear equations. During heating, the **equation** is:

$$K_{XM} = K_X \bullet T_{MAX} / 2(T_{MAX} - T_{SP}). \quad (4)$$

And, during cooling, it is:

$$K_{XM} = K_X \bullet T_{MAX} / 2T_{SP}, \quad (5)$$

where K_{xm} is a modified K_x PID constant, T_{MAX} is the maximum reachable temperature of the system, and T_{SP} is the current setpoint temperature.

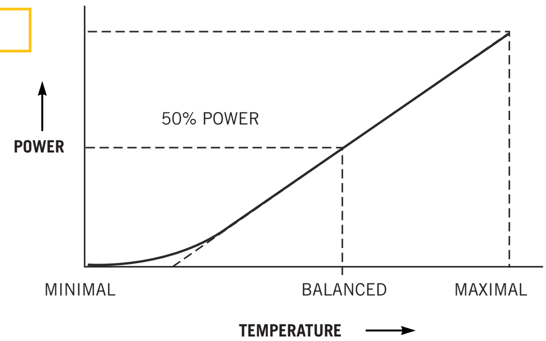
This modification of the PID constants brings them much closer but not equal to their optimal values. The autotuning process can calculate PID constants that provide optimal behavior during a warm-up phase but not for maintaining a stabilized temperature. Likewise, it can calculate constants that are optimal for maintaining a stabilized temperature but not for a warm-up phase. Higher than required PID constants are especially harmful during a warm-up phase, because they usually cause big overshoots before the temperature stabilizes. Addressing this problem requires the introduction of some nonlinearity.

INTRODUCING NONLINEARITY

With the introduction of nonlinearity, the entire range of temperature control splits into three ranges, with a different control mechanism acting in each one; hence the term "nonlinearity." **Figure 7** illustrates a process, for example, that incorporates a variation of "bang-bang" temperature control in one range and linear control in a linear-control zone. In a third range, where temperatures are higher than in the linear-control zone, the controller outputs 0% power.

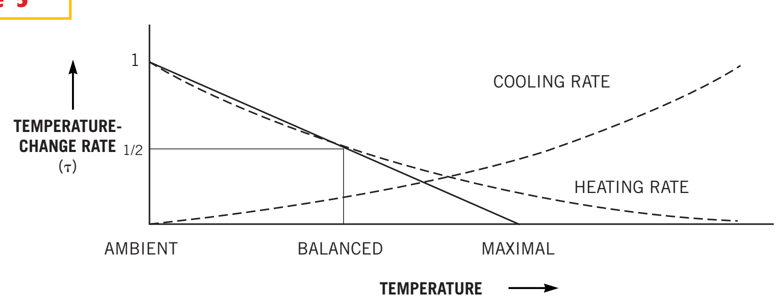
In traditional bang-bang control, the heating process initially applies maximum power to the heating element. If the rising temperature reaches a certain value (the so-called switching value) before it gets to the setpoint, the heating process ends, and the cooling process starts. In the case of unforced cooling, power applied to the heating element then drops to 0%. For an optimally selected switching temperature, the system eventually reaches the setpoint temperature without any overshoot. After that, the heating element receives only the power necessary to maintain the setpoint temperature.

Figure 4



The relationship between applied power and temperature is not completely linear, because thermal resistance changes with temperature. Thus, the balanced temperature deviates from the midpoint between minimum and maximum temperature.

Figure 5



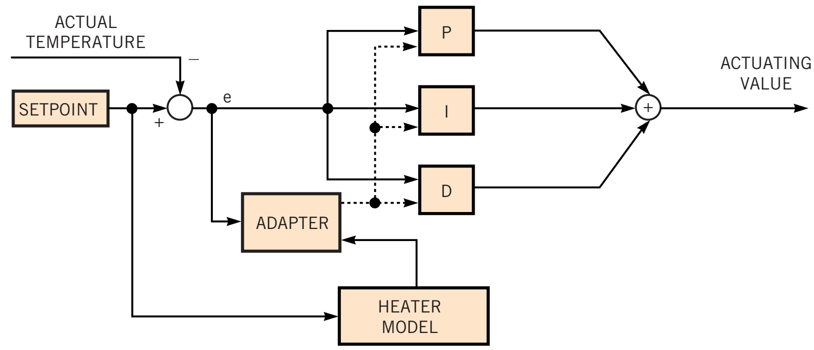
The rate of temperature change is nonlinear, but a straight line is a reasonable approximation for it.

Bang-bang temperature control sounds simple, but applying it is not an easy task. The linear-control-zone method is simple to apply, however, and it provides results that are almost as good as pure bang-bang control. The linear-control-zone approach differs from the bang-bang method in that, to avoid overshoots, the linear-control-zone band must begin at a lower temperature than the bang-bang switching temperature, because a nonzero power always applies to the heater inside the linear-control-zone band. Also, you need not precisely define the linear-control-zone, because the PID control inside the band handles the tasks of reaching and maintaining the setpoint temperature.

At Point 1 in **Figure 7**, a system is initially stable at a starting temperature with a required starting power. The selection of a new setpoint temperature results in application of full power to the heater (points 1 to 2), because the current temperature is below the beginning of the linear-control-zone band. The temperature rises during application of 100% power (points 2 to 3). When the temperature reaches the beginning of the linear-control-zone band (Point 3), the PID compensator takes over (points 3 to 4), and the power level becomes less than 100%. Line P in the **figure** shows the contribution of the PID compensator's proportional member; Curve I is the contribution of the integral member. The **figure** omits the contribution of the derivative member for clarity. The actual PID output depends upon the values of the PID constants and may have various shapes, such as Curve C.

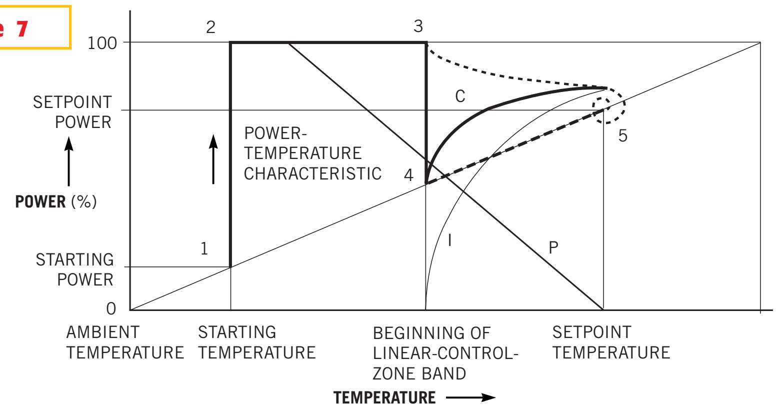
In an optimally tuned system, **Figure 7**'s phase diagram would follow the bold dashed line until it reached the setpoint temperature and power (Point 5). In real systems, however, the phase diagram follows the dotted spiral, which shows oscillation (overshoots and undershoots) around the setpoint. The implementation of an additional nonlinearity, the regulation error limit, can reduce these overshoots. The regulation error limit clamps the otherwise-varying value of the regulation error to a small constant value throughout almost the entire linear-control-zone band. The regulation is unclamped only when the temperature is close to the setpoint value. Limiting the regulation error creates the equivalent of a tracking-control system, which tracks a moving setpoint value. Because the setpoint value is only marginally ahead of the actual temperature, even rather high values of PID constants provide little danger of high overshoots.

Figure 6



The addition of a heater model and an adapter to a PID controller provides adaptive modification of PID constants, which is useful when the setpoint temperature is not near the balanced temperature.

Figure 7



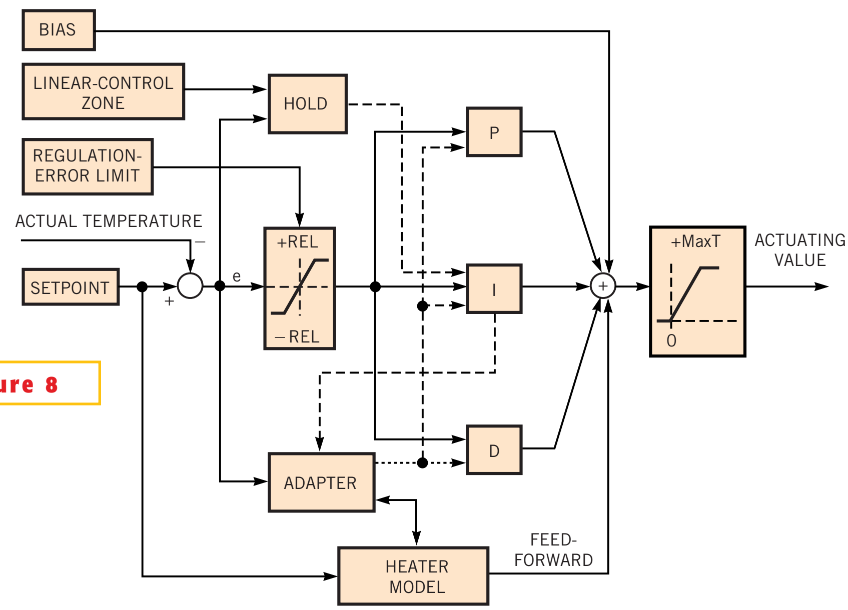
A band of linear-control-zone nonlinearity, as this phase diagram of a heating process illustrates, helps avoid overshoots.

INTRODUCING A FEED-FORWARD MEMBER

You can, however, add a feed-forward member to the control system to make the power level more closely follow the optimal curve (**Figure 7**'s bold dashed line). If you know what power level is necessary for a required setpoint temperature, you can simply replace the contribution of the controller's integral member with the output of a suitable feed-forward member. The system's heater model does know the power level, however, because it helps the adapter block to modify the PID constants, and, to do this, it must know the relationship of power and temperature in the system. Because the heater model knows the current value of the setpoint, it can easily determine the corresponding power value. This value can then become the contribution of a feed-forward member, and the power level more closely follows the optimal curve. The feed-forward member replaces the integral member and, in some cases (when overshoots are still unacceptably high), the proportional member, both of which the control system should disable at this point. As the temperature approaches the setpoint, however, reactivating the integral member is desirable. Otherwise, the system exhibits a steady-state error, because the heater model might not model the system as precisely as necessary.

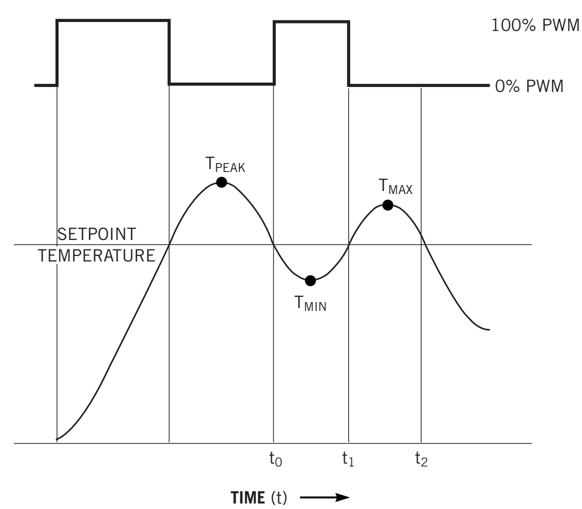
The adapter gets a new task with this approach. Once the temperature stabilizes at the setpoint value, the required power-output value is easy to calculate, because it is simply a sum of the integral member and the feed-forward member from the heater model. (Because the temperature is stabilized at the setpoint, the regulation error is zero, so the contributions of the proportional and derivative members are zero.) The adapter can now move the content of the integrator to the feed-forward block and clear the integrator. The overall output of the control system does not change. After autotuning predicts the power-temperature characteristic, the adapter must perform the last step of modifying this characteristic to account for the latest knowledge of power-output values. Usually, simply changing the slope, or gain, of the characteristic is adequate. Now, the heater model knows precisely what power is necessary to maintain the required setpoint temperature; when you change the setpoint, it can more precisely predict and provide the required output power.

Figure 8



An improved temperature controller incorporates PID control, adaptive PID constants, an approximation of bang-bang control during warm-up, and a learning feed-forward member that can predict and generate required power output.

Figure 9



The relay-feedback method creates power pulses that cause a process variable, such as temperature, to oscillate around a setpoint value.

FINAL DESIGN OF TEMPERATURE CONTROLLER

The temperature controller now has all the possible bells and whistles. It uses an improved PID compensation to reduce the integration errors and the noise that the derivative member generates. It also dynamically adapts the PID constants according to the value of the setpoint temperature and according to the current temperature, and it approximates the optimal (bang-bang) control during the warm-up phase by using the linear-control-zone and regulation-error-limit nonlinearities. The controller also uses a learning feed-forward member that can predict and generate the power output required to maintain any setpoint temperature (**Figure 8**).

Figure 8 shows a bias parameter, which sums together with all the other blocks outputs to create the actuating value. Under normal conditions, the bias parameter's contribution is zero, but, in certain applications, nonzero bias may be useful. For example, a high-level control system can use it to prevent undershoots and overshoots if heat disturbances are predictable.

Figure 8 also shows the output limiter. Its purpose is to guard the actuating value, which is now the sum of five output values. You usually calculate the actuating value in terms of temperature, because the controlled and measured process variable is temperature. It can acquire any value from zero to a maximum temperature, MaxT. Because no forced cooling occurs, negative values have no meaning.

The output value requires certain ad-

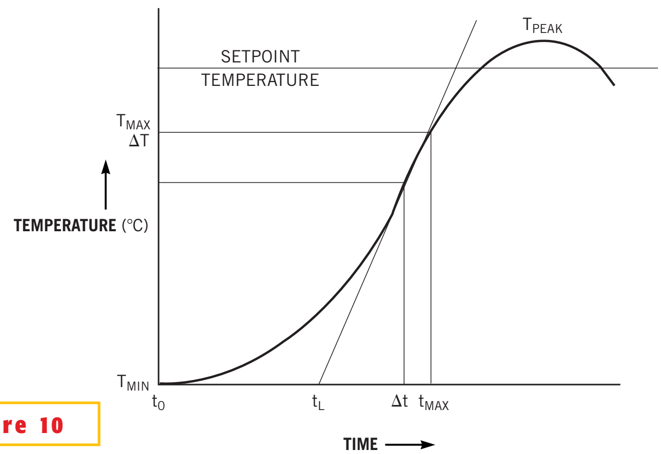


Figure 10

Step-response autotuning derives from the step-response of an open-loop system.

ditional conversions. An analog voltage or current corresponding to the actuating variable in degrees Celsius or Fahrenheit only rarely drives the heater. Most applications use PWM-based methods and ac-current-powered heaters. The PWM period should be identical to the temperature sampling and processing period. For fast processes, this period equals the 50- or 60-Hz ac cycle, and the application of a phase-angle firing technique is necessary. For slower processes, the PWM period can be much longer. In any case, avoid phase-angle firing, because it is usually a source of intense system noise. From a noise standpoint, the optimal switching method is zero-crossing firing. A heater turns on or off only when the ac voltage crosses a zero value.

A PWM-based method requires converting the actuating value from terms of temperature to a time value of a duty cycle of the PWM period. For zero-crossing firing, the duty-cycle value should be an integer multiple of the ac period. Applying a digital differential analysis can ensure that any value of the duty cycle gets converted into a correct integer multiple of the ac period, thus avoiding deterioration of long-term control precision. For example, if the actuating value is 40% and the PWM period is 100 msec, the heater should turn on for 40 msec every 100 msec. However, 40 is not an integer multiple of 16.66, which is the 60-Hz ac period in msec. To overcome this difficulty, the digital-differential-analysis block can drive the heater for two ac cycles (33.33 msec) and add the "unused" part of the 40-msec period (6.67 msec) to the duty cycle of the next period. In this next period, the heater turns on for two ac cycles and passes an unused value of 13.34 msec to the third duty cycle. Eventually, a duty cycle of 53.34 msec results in the heater being on for three ac cycles. The unused 4.36 msec of that duty cycle passes to the next cycle and so on.

IMPLEMENTATION OF AUTOTUNING

Autotuning is useful for providing initial PID constants for a control system. Manually tuning a PID controller is a painful process, as anyone who has tried it well knows. Following a recommended tuning procedure is helpful. If the procedure is built into a control system as autotuning, it's even more helpful.

Two common tuning methods for heating processes with slight modifications can provide reasonable initial values not only for PID coefficients, but also for the additional control blocks in **Figure 8**'s enhanced controller.

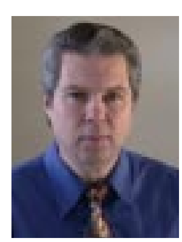
The first autotuning method derives from the relay-feedback method for identification of the dynamic behavior of a controlled process. This method applies a sequence of power pulses (**Figure 9**) to the heater, which is running in an open loop. The series of pulses forces the process variable (temperature) to oscillate around the setpoint value. This value is a switching temperature; whenever the actual temperature crosses it, the output (actuating value) changes from 0 to 100% or vice versa. At the end of process, the measured values of t_0, t_1, t_2, T_{MIN}, T_{MAX}, T_{PEAK} go into calculating all the PID parameters, which include K_P, K_I, K_D, PWMPeriod, regulation-error limit, linear-control zone, and FF_G (a gain used by the feed-forward block).

The second autotuning method, the step-response method, comes from the step response of an open-loop system (**Figure 10**). At the end of the step-response autotuning process, the measured values of t_0, T_{MIN}, Δt, ΔT, t_{MAX}, T_{MAX}, and T_{PEAK} go into calculating the same set of PID constants that the relay-feedback method calculates.

Which autotuning method is better? Both get wide use, so each method must have certain pros and cons. Generally, the relay-feedback method should provide more reliable results, because it more thoroughly "exercises" with the tuned system. In addition, it can find the gain of the system, which is important for the heater model (feed-forward) block. The heater model can simulate the power-temperature characteristic (**Figure 4**); however, without knowledge of the gain value, this characteristic cannot properly scale. If a control system uses the step-response autotuning method, then it must use some default value (usually one) of scaling. However, once the system reaches a stable temperature, the adapter block finds the proper scaling value.

On the other hand, you cannot use relay-feedback autotuning in multiple-zone systems, such as in a barrel in an injection-molding machine, which typically has eight adjacent heating zones. The zones affect each other to such an extent that temperature oscillations become completely distorted. In a situation such as this, only the step-response method can provide usable results.□

AUTHOR'S BIOGRAPHY



Peter Galan is a senior software engineer at OCM Technology Inc (Ottawa, ON, Canada), where he participates in research, development, and design of industrial-control modules for motion, temperature, and power control. He holds a PhD in automated control systems from Technical University (Kosice, Czechoslovakia) and a master of engineering degree in applied cybernetics from Czech Technical University (Prague). Galan's spare-time interests include photography and video.