FISHER

FISHER-ROSEMOUNT™ Managing The Process Better.™

technical monograph 40

Fundamentals of Dynamic Valve Performance

Floyd D. Jury
Technical Performance Consultant

This paper provides a basic understanding of dynamic valve performance and how it relates to process variability. The concept of Internal Model Control (IMC) is also introduced. Emphasis is upon understanding the fundamentals rather than specific hardware. Although the reader will encounter some mathematics in this paper, the level is pretty much confined to basic high school algebra.

Fundamentals of Dynamic Valve Performance

THE CONTROL OBJECTIVE

Companies are in business to make a profit through the production of a quality product. A quality product is one that conforms to specifications. Any deviation from the established specifications means lost profit due to wasted product, reprocessing costs, or excessive product requirements. In order to ensure that they meet the specifications and produce a quality product at the lowest possible cost, the goal is to "do it right the first time" with the least amount of cost.

Dr. W. Edwards Deming, the father of the modern quality movement, points out that every process has variation associated with it.

The causes of process variation, according to Deming, come in two forms;

- SPECIAL CAUSES
- COMMON CAUSES

SPECIAL CAUSES of process variation tend to be infrequent, relatively large disturbances which are identifiable and assignable to a special situation. Examples of special causes of process variation are such things as;

- Equipment malfunction
- Operator errors
- Gross upsets in steam pressure
- Fouling
- Catalyst decay
- Grossly defective batches of raw materials
- Etc.

A quality expert would describe a process with Special Cause variation as "being out of control." One can attempt to compensate for these special cause variations through the use of process control technology, but the efforts are usually expensive if not outright counterproductive. Often it simply moves the disturbance to another area of the plant.

The best solution for improving quality with Special Cause variation present is to identify the source of those causes and eliminate them.

COMMON CAUSES of process variation are frequent, short-term, random disturbances that are inherent in every process.

ELIMINATING PROCESS VARIABILITY

Eliminating the common causes of process variability and manufacturing a uniform product in the face of such things as nonuniform, randomly varying raw materials and other process disturbances is the prime function of control valves and other instrumentation.

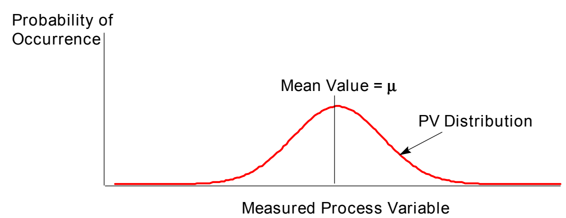
Since it is not possible to make every machine and material perfect, we must accept the fact that process parameters and raw material compositions will have minor variations in their characteristics and performance, but we will expect them all to be within their own tolerance limits.

As each machine, process, and raw material varies randomly within its own tolerance band, the net effect on the overall process will be a band of process performance values which form a distribution about some average value.

UNDERSTANDING THE STATISTICS OF VARIATION

When a process is in "statistical process control" it is stable and predictable because it is undisturbed by extraneous or special causes. When this is the case, the distribution of data typically follows a "normal" or "Gaussian" frequency distribution which is the familiar bell-shaped curve shown below.

The process variable data values are plotted along the horizontal axis, and the probability of occurrence of each particular value is plotted in the vertical direction.



This distribution curve implies that when we make a sample measurement of the process variable there is a high probability that it will be relatively near to the mean value. By the same token, there is significantly lower probability that the measured value will be very far above or below the mean value.

If there was no process deviation, then every time we sampled the process variable it would be exactly on the mean, or set point value.

In real-life situations, however, we know that we will always see some variation in the measured process variable. Sometimes it will be above the set point and sometimes it will be below the set point. On the average, we would expect to see about the same number of deviations above the set point as we would see below the set point thus forming a curve which is symmetrical about the mean value (μ).

The closer the data points are to the set point (mean value) the more frequently they will tend to occur, while those data points which represent larger deviations from the set point will tend to occur with less frequency.

The spread of the data is the amount of variation about the mean value; i.e., the variation from set point.

Mathematicians have three basic measurements or metrics which they use to describe the spread of the data. These metrics are range, variance, and standard deviation.

RANGE is simply the magnitude of the difference between the highest and lowest values of data in the set. Range is a rather coarse measurement of variability spread and is of little use to us.

A better measurement is VARIANCE which is also called the "mean square deviation from the average" as illustrated by the formula below.

$$\text{Variance} = \sigma^2 = \frac{\Sigma(x_i - \mu)^2}{n-1}$$

where
X_i = an individual data point
n = the number of data points
μ = the average of all the data points

NOTE: The formula given above is really the variance of a finite sample rather than the variance of the total population. The variance of the total population is referred to as (σ), while the variance of a sample is typically referred to as (S). For a large sample size ($n \geq 30$), there is typically little difference between these two statistics. Even though we are actually talking about finite sample measurements, we will continue to use sigma (σ) throughout this text since this is often done in common usage and we also have a special role assigned to (S) in later discussions.

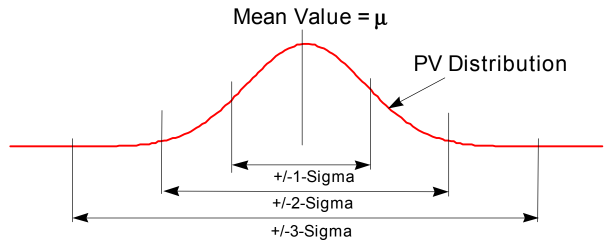
In the variance equation, the term (X_i - μ) is simply the amount that the individual data point deviates from the mean value. Since we don't usually care whether that deviation is above or below the mean, that term is squared to eliminate any opportunity for the algebraic signs to cause terms to cancel each other.

The "squared deviation" terms for each of the individual data points are then averaged for all the points. The term (n - 1) rather than n is used in the averaging equation as an adjustment factor to account for a finite sample size.

STANDARD DEVIATION (σ) is simply the square root of the variance; i.e.,

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\Sigma(x_i - \mu)^2}{n-1}}$$

Standard Deviation (σ) is often simply referred to as "sigma." In layman's terms, Standard Deviation can be visualized simply as the average deviation from the set point (mean).



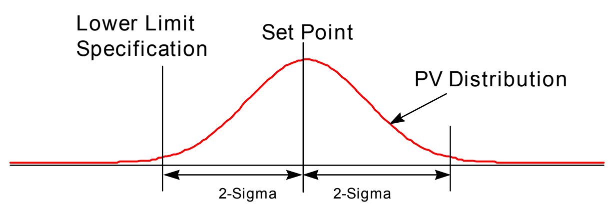
Thus, sigma (σ) is a statistically derived parameter which describes the "spread" of the data about the mean value. The larger the value of σ, the greater the spread.

Sigma is also a parameter which tells us how much of the total population is contained in a given region centered about the mean value; i.e.

- $\pm$ 1 Sigma contains 68.26% of the total population
- $\pm$ 2 Sigma contains 95.45% of the total population
- $\pm$ 3 Sigma contains 99.73% of the total population

In performance testing, 2-sigma (2σ) is the metric most often used to indicate valve performance and to compare one valve against another in terms of reducing process variability.

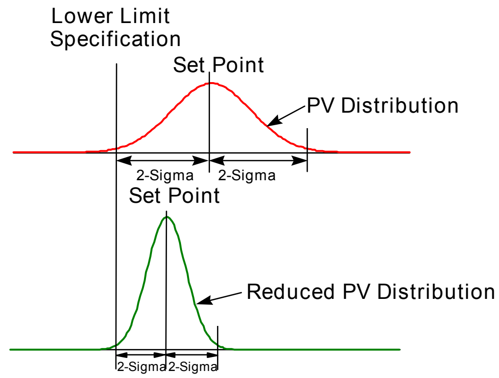
Thus, assuming a manufacturer wishes to ensure that at least 95% of his manufactured product meets the minimum specification, he must adjust the set point of the process at least a 2-sigma distance above the lower specification limit.



This adjustment of set point will ensure that only 2.275% of the finished product will fail to meet the specifications on the first try. (Don't forget that the tail of products above the 2-sigma limit on the upper end also meet the specification.)

While this setting will produce acceptable results, it also means that 97.725% of the finished product is being produced at quality levels which considerably exceed the requirement. This may be costing significant dollars in terms of excessive catalysts, raw materials, etc.

While the manufacturer cannot reduce the percentage which must exceed the lower specification limit, he can reduce the degree to which these products must exceed the specification limit by decreasing the "spread" of the distribution; i.e., going to a lower value of sigma as shown in the following diagram.



This is referred to as a reduction in “variance” or “process variability.”

This reduction in Process Variability allows the manufacturer to move the process set point closer to the lower specification limit which;

- Produces a more uniform product
- Improves manufacturing efficiencies
- Improves profitability

Thus, reducing process variability is really the manufacturer’s goal.

Manufacturers tend to understand well that spending large sums of money on elaborate systems of instrumentation are often justified in order to achieve this type of reduction in process variability.

What is not as well understood is that significant reductions in process variability can also be achieved simply by choosing the right control valve for the application and ensuring that the control loop is tuned for the most effective reduction of process variability.

Surveys have shown that as many as 80% of the control loops in plants are not achieving the level of product uniformity of which they are capable.

The reason for this is twofold;

- The right process control equipment is not being utilized.
- The process control equipment which is installed is not being operated at optimum performance.

Research data show that process variability can be reduced by as much as 50% by the straightforward application of

- Good engineering design practices
- Intelligent selection of valves and instruments
- Careful system installation
- Good instrument maintenance
- Proper loop tuning methods

In general, process variance can be reduced by

- Maintaining stable loops
- Minimizing the effects of load disturbances
- Reducing the unwanted changes in loop gain
- Reducing the effects of nonlinearities in the loop

Selection of the control valve can have a significant impact on the variability of the process.

In addition to hardware selection, tuning of the loop can also impact the process variability.

TUNING THE LOOP USING THE IMC APPROACH

There are many different approaches that have been developed over the years for tuning a control loop, but the technique that is used by Fisher in its dynamic performance testing program is known as “IMC Tuning.” This technique is used because it provides a reliable way to ensure consistent tuning from one test setup to another.

IMC stands for “Internal Model Control,” and is so named because the actual, experimentally determined system characteristics are internally modeled into the controller algorithm so as to produce a desired closed-loop system response. The IMC tuning approach was first introduced in the 1980’s and has achieved great popularity in the industry for several reasons.

The first reason for IMC’s popularity is its relative simplicity. It is based on common time domain step response testing and it only involves one tuning parameter.

Secondly, the one tuning parameter involved relates directly to the closed-loop time constant and to the robustness of the control loop.

NOTE: The term “robustness” as used here should be interpreted as “conservativeness”; i.e., a robust loop is one that is conservatively tuned.

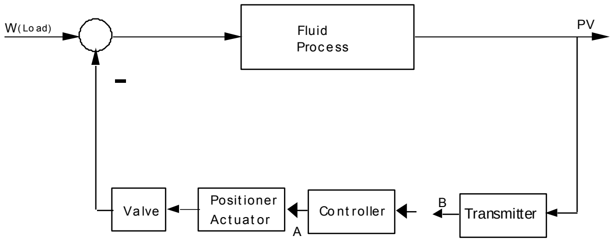
Another reason for IMC’s popularity is that when the loop is tuned using the IMC rules, the closed-loop load response will exhibit no oscillation or overshoot. Experience indicates that this tends to minimize interactions between control loops and enhances overall process disturbance rejection.

The first step in the IMC tuning approach is to experimentally determine the dynamic characteristics of the process. This is typically done using a step response test on the actual system.

USING STEP RESPONSE TO DETERMINE THE PROCESS TRANSFER FUNCTION

Mathematically deriving a transfer function for all the elements in a process loop is not always feasible. It is not only difficult to do, but may not accurately represent the actual element.

Typically, the combined loop transfer function is developed using a step response test of the open loop as explained below.



The controller is placed in manual operation and a small (2%-3%) step change is introduced at the output of the controller (point A). This disturbance will travel around the entire loop and show up back at point B where it will be measured.

The response at point B is the response of all the elements in the loop EXCEPT THE CONTROLLER. This is called the response of the "Process" (P).

It is typical to find that most of the higher order dynamics of the process will be dominated by one or two elements which will result in a process response that appears to be either a first-order response or a heavily damped second-order which responds similarly to a first-order.
In many cases, it is sufficient for purposes of process control analysis to approximate this loop response with a first-order lag whose time constant is τ (tau); i.e.,

$$P = \frac{K_P}{\tau S + 1}$$

It should be clearly understood that this transfer function represents the OPEN-LOOP characteristics of everything in the loop except the controller. This function represents those elements in the loop which the controller is going to attempt to control.

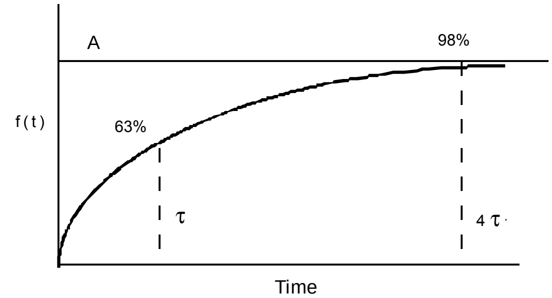
Notice that we have named this loop function "P" implying that this is the "process" transfer function. This brings up an important issue of terminology.

Many control engineers tend to think of the process as being the level tank, the pressure vessel, the mixing vat, etc. as being "the process;" however, this fluid process is only one of the many elements in the loop.

On the other hand, most plant operators tend to think of the "process" as being the thing that they need to control with their controller. Thus, their concept of "the process" includes everything in the loop, except the controller; i.e., the control valve, the fluid process, the transmitter, etc.. This is precisely what the transfer function (P) from the above loop step response test represents.

From this point on, when we speak of "the process," we will usually be referring to the (P) function as defined above.

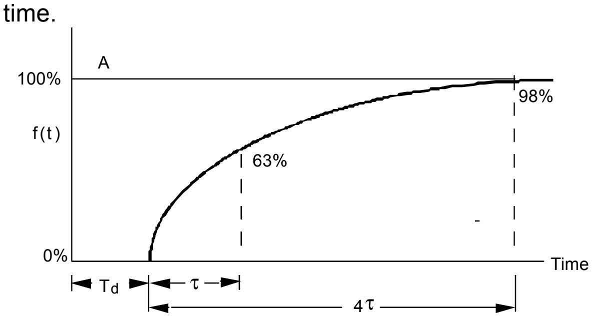
The time constant (τ) for this process function is usually determined by measuring the time it took the loop response at point B to achieve 63% of the final value when subjected to the step input at point A.



So far in these discussions we have assumed linear elements. The Laplace transform method of analysis is only good for linear analysis. Unfortunately most real life systems have many nonlinearities.

One of the most common nonlinearities in the loop is dead time. Dead time, as might be supposed, is time when nothing appears to be happening due to transport lag times, dead zones in a relay, long digital sampling times, etc.

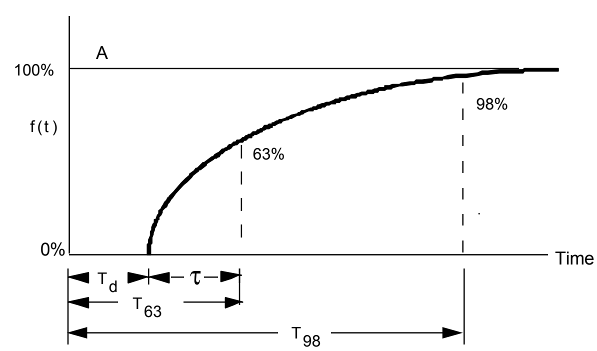
These types of dead time will exhibit themselves on the measured response as follows, where T_d is the dead time.



The above diagram provides an opportunity to discuss two terms which frequently cause some confusion. These two terms are Tee-63 (T_{63}) and Tee-98 (T_{98}).

Tee-63 represents the time in which the response function reaches 63% of the final response, and Tee-98 represents the time in which the response function reaches 98% of the final response.

The following diagram makes it clear that Tee-63 and τ are not the same unless there is no dead time since Tee-63 includes the dead time. By the same token, Tee-98 also includes the dead time.



The presence of dead time requires a couple of precautions that need to be observed in order to ensure the most accurate representation of the process.

The first precaution involves the actual conduct of the test. Due to any dead band nonlinearity present, the system is likely to respond differently to a decreasing step input than it would to an increasing step input.

Basically, a two-step change procedure is used to deal with the dead band. One step is designed to position the valve at the dead band and the second step is designed to take the valve through the dead band.

Thus, in order to ensure that the system is properly positioned to begin an INCREASING step input test, the system should first be preconditioned by subjecting it to an UP-DOWN-DOWN-UP cycle of steps before conducting the first actual UP-step measurement.

To ensure further accuracy, this entire sequence should be repeated at least twice more to obtain a total of three up-step measurements.

Three down-step measurements are also obtained during this same sequence.

Final values for the process dead time (T_d) and the process time constant (τ) are then obtained by averaging the six measurements (three up and three down) of each of these two parameters.

The second precaution revolves around the decision of how the dead time nonlinearity should be treated in a linear analysis situation.

Since dead time is a nonlinearity that cannot be easily handled in the type of linear analysis we do, it is common practice to make some type of "adjustment" in the process parameters in order to develop a pseudo-linear system that we can deal with.

The method that Fisher Controls uses to account for dead time will be discussed later after some new information is introduced.

DERIVING THE IMC TUNING RULES

For many years, control practitioners have been taught the Ziegler-Nichols method or some other version of the quarter-decay damping ratio method of tuning. These methods were an attempt to develop a systematic approach for tuning control loops.

These techniques depend upon tuning the system so that the response to a step input is an oscillation that decays rapidly; i.e., the second overshoot is a quarter of the amplitude of the first overshoot, etc. Hence the name of the method.

With these types of tuning, the loop will tend to resonate and amplify any disturbances that are present near its natural frequency. In any manufacturing environment there is always going to be broad-band noise present in the system. Hence the loop will always find disturbance energy to amplify at its natural frequency.

Because of these problems, as well as others, many users have tended to resort to a more pragmatic, trial-and-error approach to tuning.

It is not unusual for loop gains to change by a factor of four or more. Thus, the Ziegler-Nichols method of tuning, as well as many of the trial-and-error methods of tuning that are based on a specific gain margin of two subject the loop to the possibility of instability when the gain changes by a factor of two or more.

Many practitioners who are aware of these problems tend to compensate by detuning the loop to provide even larger gain and phase margins. As a result of this, and the oscillatory nature of system response, process variability tends to suffer.

In recent years, a newer method of tuning, called the IMC tuning method, has gained popularity because it helps to avoid these problems. IMC stands for "Internal Model Control." Later we'll see the reason for this name.

The primary goals of a control system are to
- Obtain satisfactory response speeds to changes in the process set point,
- Minimize or eliminate overshoot,
- Hold deviations from set point to a minimum.

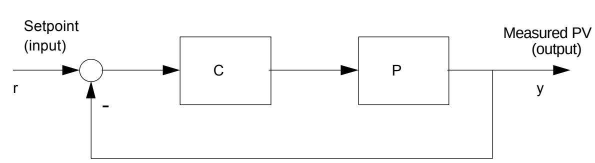
As we will see, the IMC method of tuning is designed to address all of these goals.

It should be pointed out from the start that the IMC tuning rules are based upon analyzing the closed-loop response to a set point change rather than a load disturbance.

Part of the reason for this is due to the simplicity of the step response test which is part of the procedure. It is much easier to test the system with a set point change than to produce a load change.

While the IMC tuning rules are derived based upon a response to set point, the tuning itself is applied to closed-loop load disturbances. As we will demonstrate later, this tuning still produces a load response with no overshoot. This is one of the advantages of the IMC tuning process.

The first step in deriving the IMC tuning rules is to draw the block diagram showing the controller (C) function for which we are to derive the algorithm and the process (P) function which represents everything else in the loop. This diagram is shown below.



Using the conventional loop reduction rule, we can then write the closed-loop set point transfer function for this system.

$$\frac{y}{r} = \frac{PC}{1+PC}$$

Typically, the desired set point response would be a first-order lag with unity gain; i.e., the transfer function we want for this system is

$$\frac{y}{r} = \frac{1}{\lambda S + 1}$$

Since these two expressions represent the same thing, we can set them equal to each other and solve for the controller algorithm (C) which will give us the desired first-order system response.

$$\frac{y}{r} = \frac{PC}{1+PC} = \frac{1}{\lambda S + 1}$$

Cross multiplying and solving for C gives

$$PC(\lambda S + 1) = 1 + PC$$

$$PC(\lambda S) = 1$$

$$PC = \left(\frac{1}{\lambda S}\right)$$

$$C = \frac{1}{P}\left(\frac{1}{\lambda S}\right)$$

Thus, we see that the required controller algorithm is a function of the open-loop process (P) characteristics and the desired closed-loop time constant (λ).

This algorithm clearly illustrates why this technique is called "Internal Model Control (IMC)"; i.e., because the controller algorithm internally contains a model of the process (P).

Earlier we showed how a step response test can be used to develop a realistic representation of the process.

For example, we often find that the first-order transfer function below is adequate to represent the process.

$$P = \frac{K_P}{\tau S + 1}$$

Substituting this experimentally determined process function into the controller algorithm gives

$$C = \left(\frac{\tau S + 1}{K_P}\right)\left(\frac{1}{\lambda S}\right)$$

We can algebraically rearrange this equation into the following form.

$$C = \frac{1}{K_P}\left(\frac{\tau}{\lambda} + \frac{1}{\lambda S}\right)$$

$$= \left(\frac{1}{K_P}\right)\left(\frac{\tau}{\lambda}\right)\left(1 + \frac{1}{\tau S}\right)$$

The basis for the IMC tuning rule is to let

$$T_R = \tau$$

$$K_C = \left(\frac{1}{K_P}\right)\left(\frac{\tau}{\lambda}\right)$$

where

T_R = Controller reset time
K_C = Controller gain

Thus, the required controller algorithm becomes

$$C = K_C\left(1 + \frac{1}{T_R S}\right)$$

We recognize this as a PI (Proportional-plus-Integral) controller.

Thus, we conclude that when the process can be adequately represented by a first-order transfer function, IMC tuning will produce a first-order set point response; i.e., the IMC tuning has transformed an open-loop system with time constant τ to a closed-loop system with time constant λ.

Rearranging the IMC controller gain equation gives

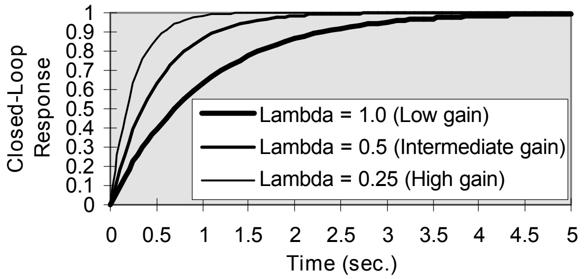
$$\lambda = \frac{\tau}{K_P K_C}$$

Once the Reset Time (T_R) has been set equal to the process time constant (τ), the only adjustable parameter in this equation is K_C; i.e.,

$$\lambda = \left(\frac{1}{K_C}\right)\left(\frac{T_R}{K_P}\right)$$

Thus, we see that for a given plant application the closed-loop response (λ) of the system is determined solely by the proportional gain setting (K_C). As the controller gain increases, Lambda becomes smaller and the system response is speeded up as shown in the illustration below.

EFFECT OF GAIN ON SYSTEM RESPONSE



The above diagram represents the closed-loop response to a set point change. Note that the system response is truly first-order with no overshoot.

This diagram might tempt us to believe that we could always continue to improve the system response by simply continuing to increase the gain. This might be true if we were truly dealing with a linear world as we have assumed so far.

Unfortunately, the real world is full of nonlinearities that we must deal with. These nonlinearities are hidden by the first-order process model we have used.

The greater the mismatch between the model and the real process, the more the loop needs to be de-tuned in order to achieve acceptable performance. This is one of the major reasons why some process control engineers recommend relatively conservative tuning rules.

We'll talk about these nonlinearities later after we have addressed the issue of processes which cannot be adequately represented by a first-order transfer function.

OTHER PROCESSES

IMC tuning does not require the use of a first-order process model. Any reasonable process model could be used; however, one is not always guaranteed of obtaining a controller algorithm that can be implemented.

It should be noted that many people call this technique "Lambda tuning" and equate the technique with the choice of a first-order process model.

For this reason, some individuals may believe that "Lambda tuning" is not appropriate for many processes such as temperature where a second-order process model would be more appropriate.

The following discussion shows that the IMC tuning concept works with any process since the controller algorithm (C) actually contains the process function (P).

The first point to emphasize is that the IMC tuning method is always the same regardless of the process used. The only thing that varies is what we choose to represent the process. Thus, as was developed earlier, the controller equation shown below is always our starting point when considering other processes.

$$C = \frac{1}{P}\left(\frac{1}{\lambda S}\right)$$

As before, we can use a step response test to develop a realistic estimate of the process; however, we need to use good engineering judgment when it comes to making a choice of the transfer function that we will use to represent the test results from the process.

For example, when we are dealing with a temperature process, we know that a second-order process model would be more logical than a first-order model. If the decision is made to use a second-order model, the transfer function representing the process would be as follows.

$$P = \frac{K_P}{(\tau_1 S + 1)(\tau_2 S + 1)}$$

The technique for defining τ_1 and τ_2 from the step response test are not as well defined as for the first-order lag; however, there are techniques recommended in various control handbooks. One such technique recommended in the TAPPI literature is presented here without proof.

τ_1 = The time from the step response test where the response has reached 61% of the final value.

τ_2 = The time from the step response test where the response has reached 10% of the final value.

As before, we substitute the experimentally determined process function into the controller algorithm which gives

$$C = \left(\frac{(\tau_1 S + 1)(\tau_2 S + 1)}{K_P}\right)\left(\frac{1}{\lambda S}\right)$$

We can algebraically rearrange this equation into the following form.

$$C = \left(\frac{\tau_1}{K_P \lambda}\right)\left[\left(1 + \frac{1}{\tau_1 S}\right)(\tau_2 S + 1)\right]$$

The basis for the IMC tuning rule is to let

$$T_R = t_1$$
$$T_D = t_2$$
$$K_C = \frac{T_R}{K_P\, l}$$

where

T_R = Controller reset time
T_D = Controller derivative time
K_C = Controller gain

We, of course, recognize this as a three-mode, PID (Proportional-plus-Integral-plus Derivative) controller.

Thus, we conclude that when the process can be adequately represented by a second-order transfer function, IMC tuning will still produce a first-order response.

The above example clearly illustrates the general technique that can be followed to determine the controller function for any of the common process functions that we might elect to use to most closely match the actual process conditions.

The table below shows the IMC controller function which results from several common process functions.

PROCESS TYPE	PROCESS TF	DESIRED CONTROLLER	NET OPEN LOOP RESPONSE
Flow	Static Gain (i.e., a very fast first order lag)	Integrator (or PI with very broad PB)	Integrator
Level	Integrator	P	Integrator
Pressure	First order	PI	Integrator
Temperature	Second order (i.e. two first order lags)	PID	Integrator

Note that in each case we end up with an integrator in the open loop; i.e.,

$$PC = \left(\frac{1}{\lambda S}\right)$$

Using the loop reduction rule, one can easily show that closing the loop around an integrator will always result in a first order lag for the closed-loop set point response.

LOAD RESPONSE WITH IMC TUNING

Recall that the IMC tuning rules were developed by analyzing the system response to a set point change; however, in actual operation we are more concerned about system response to other disturbances in the process variable.

These disturbances actually come in two forms; i.e., random load disturbances which we can control with our instrumentation and noise introduced into the system by the instrumentation itself.

The IMC tuning procedures we have discussed will do nothing to eliminate the instrumentation noise, but with proper filtering of the measured process variable signal, we can reduce the tendency of the controller to respond to this instrumentation noise.

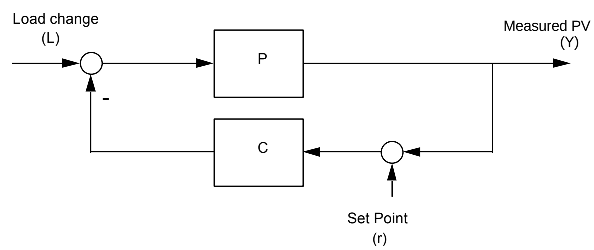
One important advantage of IMC tuning is the ability to minimize interaction between loops. Since the closed-loop time constant of each loop is determined simply by adjusting the controller gain of that loop, one can easily manipulate the loop time constants of several related loops so as to minimize interaction between these loops.

Another important advantage of the IMC tuning technique is its ability to achieve optimum load disturbance rejection.

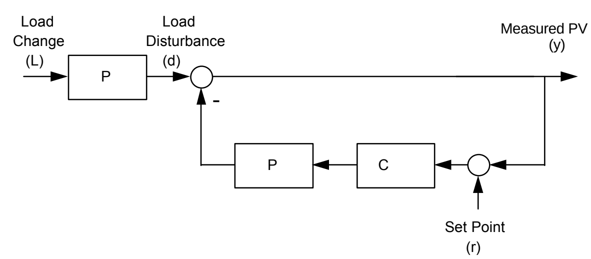
Regardless of where these disturbances are introduced in the loop they show up as deviations in the process variable.

We can see that more clearly by taking the general system diagram shown below and rearranging it to illustrate this different perspective.

In this diagram, a load change is shown as an input to which the process reacts to produce some disturbance in the measured process variable.



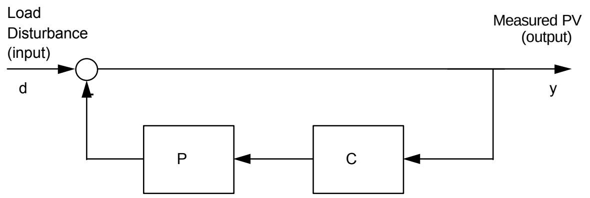
In order to change this diagram into one with a more useful perspective, we need to recognize that we can reconfigure any block diagram such as this in any way that makes useful sense to us, provided that the net effect of all paths through the system are unaffected. Thus, we can rearrange the above diagram as follows.



Here we have simply multiplied each of the inputs to the summing junction by the process transfer function (P) rather than adding these inputs first and then multiplying the sum by the process transfer function. The net result is the same, but it gives us another perspective on the system.

Now, instead of thinking of an actual load change (L) as the input to the system, we can ignore this part of the diagram, and simply consider the effect that this load change has on the process (d) as the input disturbance. This has real advantages to us, because it is typically this load disturbance effect (d) which we measure in our system rather than the actual load change itself.

If we consider these disturbances as “inputs” to the process variable that we want the system to “reject,” our perspective of the loop becomes as shown below.



This new perspective of the loop places both the process and the controller in the feedback path. Applying the loop reduction rule results in the following closed-loop transfer function.

$$\frac{y}{d} = \frac{1}{1+PC}$$

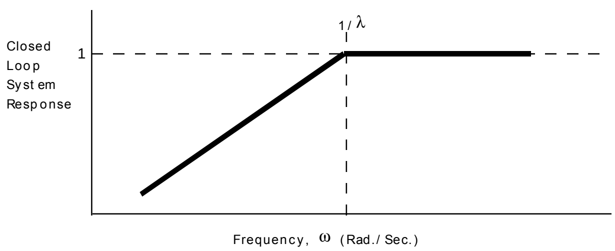
From the previous IMC derivation, we recall that for an IMC tuned system

$$PC = \frac{1}{\lambda S}$$

Therefore, the closed-loop transfer function for the system becomes

$$\frac{y}{d} = \frac{1}{1+PC} = \frac{1}{1+\frac{1}{\lambda S}} = \frac{\lambda S}{\lambda S + 1}$$

We should recognize this closed-loop transfer function as a combination of a differentiator and a first-order lag. Represented on a Bode gain diagram, this transfer function would appear as follows.



Note that this transfer function ensures that there will be no overshoot response to any input load disturbance; i.e., there is no peak in the system response function.

The typical load disturbance will be an on-going disturbance comprised of a random, multifrequency input. In order to best understand how the system will respond to this type of input, we can use this Bode diagram which depicts the system load response to any input.

This Bode diagram shows the amount of amplification or attenuation that occurs at every frequency of input disturbance. This diagram shows that there will be attenuation (or “rejection”) of the low frequency input load disturbances below the bandwidth frequency (1/λ), but there will be neither amplification or attenuation at frequencies above the bandwidth frequency.

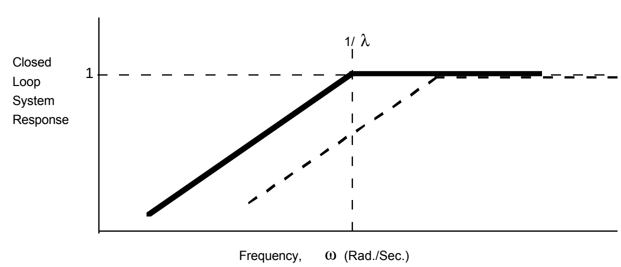
Note from this Bode diagram that the bandwidth of the closed loop is basically determined by (1/λ) which is a function of the proportional gain of the controller. Thus, as we continue to increase the controller gain, we not only improve the response time of the system, but we increase the potential for improving the load rejection ability of the system. The limiting factor, of course, will be the amount of gain increase the system can tolerate before the effects of nonlinearities and higher order dynamics begin to reduce these benefits.

This closed-loop load response function acts like a “high-pass filter.” In other words, the system would attenuate process disturbances at frequencies below (1/λ), but any disturbances at higher frequencies would simply be passed through with no attenuation. This means we need to tune the system such that the break frequency (1/λ) is a higher frequency than any expected load disturbances.

The high frequency “pass” behavior is why IMC tuning does nothing to alleviate the high-frequency instrumentation noise from the system.

On the other hand, note that the presence of the differentiator ensures lots of attenuation of the low frequency disturbances. This low-frequency range (typically 1 Hz or less) is where we would expect to see most of the process load disturbances occurring. For example, a relatively fast process loop with a closed-loop time constant of λ = 0.2 seconds would translate to a break frequency (f) of only 0.8 Hertz; i.e., f = 1/2πλ. Thus, the main benefit of aggressively tuning the system is the increased load rejection capability derived from the differentiator response at low frequencies.

Since we can logically view the Bode gain plot as the amount of “amplification” the system applies to any load disturbance, it makes senses to believe that the lower we can drive this gain curve the more we “reject” (or attenuate) these disturbances and the better we make the process variability. This is exactly what happens when we speed up the system response (smaller Lambda) by increasing the controller gain. This is illustrated by the dashed curve below.



Note that every frequency below the system bandwidth (1/λ) now has more attenuation or load rejection capability than before.

We can extend this line of thinking even further to state that the area under this Bode gain curve represents the variance of the system. This can be shown mathematically, but the proof is beyond the scope of this text.

Thus, when we increase the controller gain and move the response function further to the right, we are in effect reducing the area under this curve and reducing the variance of the process.

NONLINEARITIES

With the linear systems we have studied so far, the IMC tuning method produces an inherently stable system such that we could, theoretically, continue to increase the proportional gain indefinitely with resulting improvements in process variability.

The only limitation would be when we speed the system response up so much that we begin to encroach upon any of the higher order dynamics that may have been previously ignored.

Unfortunately, this is not a linear world. In real life systems we often see discontinuous nonlinearities such as

- Friction
- Backlash
- Relay dead zone
- Component saturation
- Sample time in transmitters and controllers
- Etc.

In addition, we have other forms of continuous nonlinearities such as:

- Shaft windup
- Filters
- Processes which change gain with change in throughput
- Valve characteristics that change valve gain with throughput
- Etc.

Many of these nonlinearities are extremely complicated to handle or explain mathematically, but the general observation can be made that all of these nonlinearities introduce undesirable phase shift and gain effects which place limitations on our ability to improve the performance of the system.

System stability is always a concern, of course, and nonlinearities can definitely influence stability; however, a more insidious effect of nonlinearity is the undesirable influence it has on process variability.

As a result, these nonlinearities cause the need to be more conservative in tuning the proportional gain, which means that we will have a limit as to how much we can reduce the process variability.

This is why it is so important to review the control hardware components of the loop to determine what nonlinearities are present and try to reduce or eliminate them through more judicious component selection.

The effects of nonlinearity can play havoc with an idealistic view of the system. For example, let's look at what happens when just one simple nonlinearity such as dead time is introduced into the system. In this case, the process transfer function becomes

$$P = \frac{K_P e^{-T_d S}}{\tau S + 1}$$

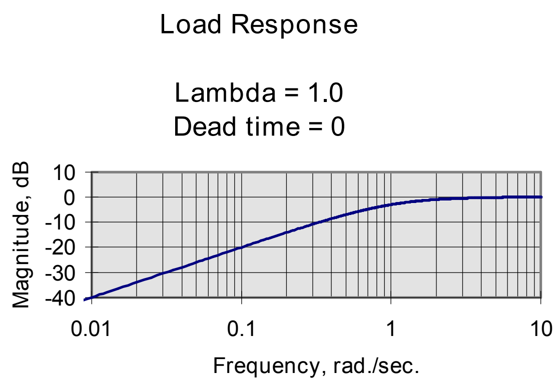
Following the same IMC derivation techniques as before, we arrive at the following closed-loop load response transfer function.

$$\frac{y}{d} = \frac{\lambda S}{\lambda S + e^{-T_d S}}$$

This transfer function is more complex than it appears on the surface. The dead time function is actually oscillatory in the frequency domain.

It is common in process control to simplify this equation through various approximations for the exponential function. To do so, however, masks the true nature of the dead time effect.

It is instructive to plot this transfer function, without simplification, for three different dead time situations. The following plot of this transfer function is with a Lambda of 1.0 second and zero dead time.

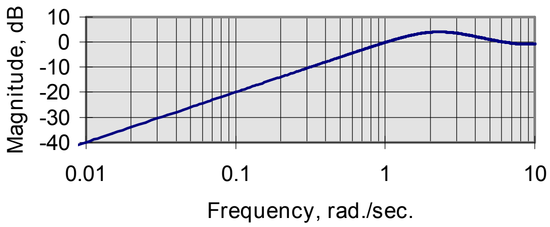


Here we see the simple high-pass filter function that we discussed earlier for purely linear systems.

If we keep Lambda at the same value (1.0 second) and increase the dead time to 0.5 second, we get the following.

Load Response

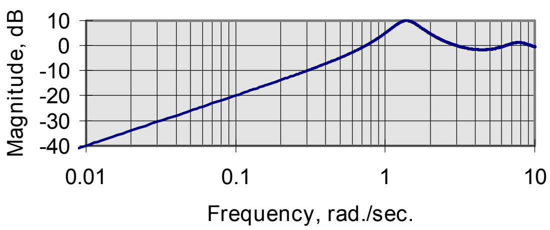
Lambda = 1.0
Dead time = 0.5



Note that the amplitude is greater than zero dB in the range from 1-6 rad./sec.. This actually represents amplification of the disturbance. Thus, the nonlinear effect of dead time has actually increased the system variance; i.e., the area under the Bode curve. As the dead time increases to equal the system time constant (1.0 second) we get even more disastrous results as shown below.

Load Response

Lambda = 1.0
Dead time = 1.0



This allows us to see rather graphically how important it is to reduce or eliminate nonlinearities such as dead time from the system.

All three of the above situations assume the same system tuning; i.e., a Lambda of 1.0 second which is equal to the process time constant (i.e., $\lambda/\tau = 1$).

Now let's keep the open-loop process time constant (τ) the same ($\tau = 1.0$ second) and see what happens when we tune the closed-loop response (λ) to different relative conditions.

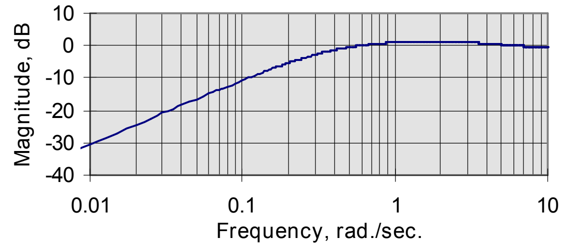
In a purely idealistic, linear system, increasing the controller gain (decreasing λ) would simply improve the load rejection capability of the system as we saw earlier. In order to provide a more realistic situation, we will assume a dead time equal to half the process time constant (i.e., $T_d = 0.5$ second).

The dead time will stay the same in each case as we explore three relative closed-loop responses (i.e., three Lambda-over-Tau ratios).

In the first case, we will use the conservative tuning ratio ($\lambda/\tau = 3$) recommended by some process control engineers.

Load Response

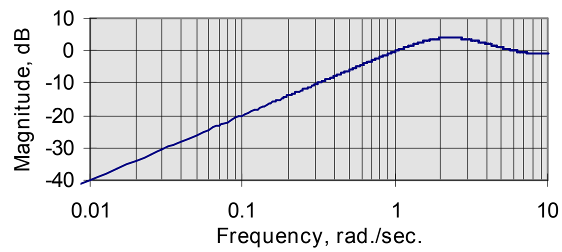
Lambda/Tau = 3.0
Dead time = 0.5



While the dead time still has an effect on increasing the process variance, the effect is rather minimal as we see here. If we get more aggressive with the tuning and increase the ratio to $\lambda/\tau = 1$, we get the following.

Load Response

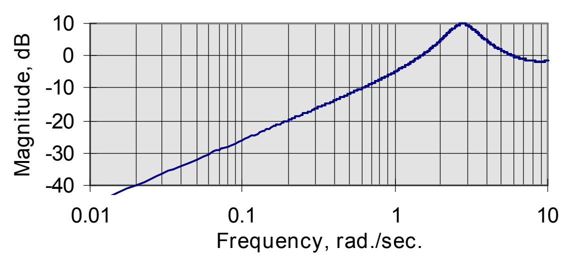
Lambda/Tau = 1.0
Dead time = 0.5



Note that there is a significant increase in process variance at this more aggressive tuning. If we were to go further and increase the controller gain for a ratio of $\lambda/\tau = 0.5$ which is recommended by some process engineers, we would get the following.

Load Response

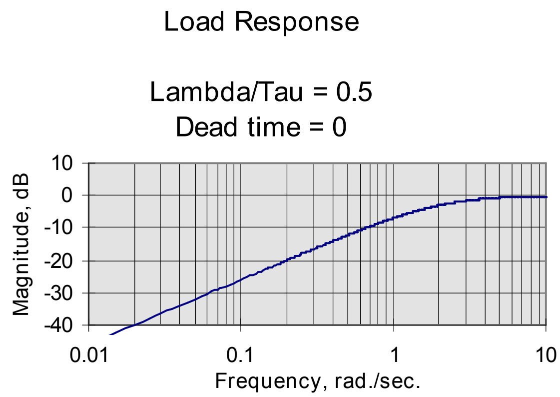
Lambda/Tau = 0.5
Dead time = 0.5



Not only has this more aggressive tuning increased the process variance due to more area under the curve in the range between 1.5 rad./sec. and 6 rad./sec., but the plot shows that the closed-loop load response will tend to be oscillatory and will amplify load disturbances with frequencies in this same range.

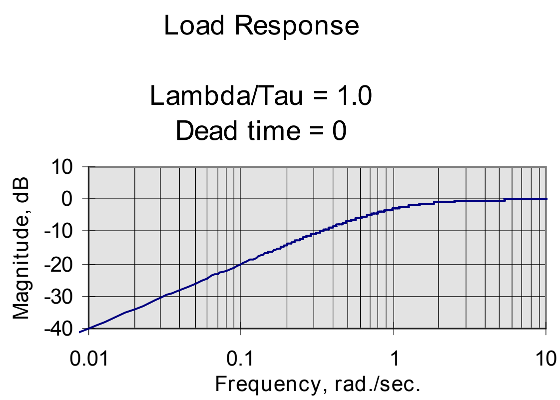
We might be tempted to conclude from this that we should always tune the system for the more conservative

$\lambda/\tau = 3$ ratio; however, we must recognize that if we do this our system response time will suffer. We saw this effect occur earlier when we showed the time response curves, but we can also see it here on these same Bode diagrams if we let the dead time equal zero.

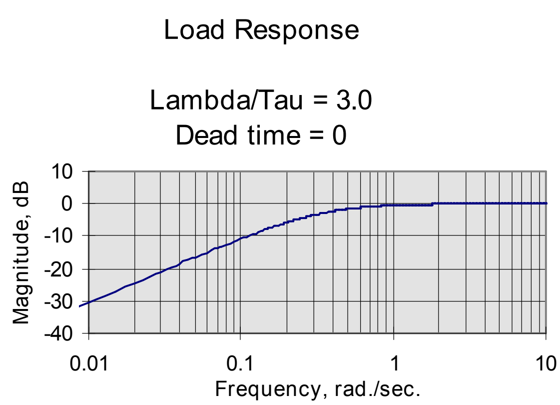


The above curve for zero dead time and a $\lambda/\tau = 0.5$ ratio shows a closed-loop bandwidth response (i.e., the -3 dB frequency) of 2 rad./sec..

If we were to decrease the tuning to a more conservative $\lambda/\tau = 1$ ratio, we see below that the closed-loop bandwidth response has decreased to 1 rad./sec..



The even more conservative tuning ratio of $\lambda/\tau = 3$ has reduced the closed-loop system response to 0.3 rad./sec. as illustrated below.



Notice that as we decreased the gain, the Bode load response function moved further and further to the left to lower and lower frequencies. As it does so, we begin to pick up more and more area under the curve, which means that our process variance is increasing.

We need to be very careful about the conclusions we draw from these exercises. The answer to reducing process variance is not always so simple when there are a variety of nonlinearities involved.

When we increase the gain of the system, we pick up additional process variance due to the dead time "peaking" effect of the system response function which introduces additional area under the curve, but at the same time, this increased gain moves the response function to a higher frequency which reduces the area under the curve.

The question then becomes, what is the trade off; i.e., which effect benefits us the most in terms of reducing the process variability. The answer is that we can't always tell for certain until we actually test the system for process variance in an actual closed-loop performance test.

Based on the IMC tuning rules derivation we see there are only two steps in the IMC tuning process. The first step is to adjust the Integral or Reset Time to equal the process time constant as measured from the step response test; i.e.,

$$T_R = \tau$$

The second step is to adjust the Proportional Gain (K_C) of the controller which raises or lowers the loop gain and establishes the desired closed-loop response (λ).

Despite the simplicity of this technique, the nagging question remains as to how aggressively one should tune the loop gain in view of all the uncertainties involved when nonlinearities are present.

Process control engineers often have strong opinions about how aggressively to tune the gain. It is useful to recognize that many of these opinions are based on past experience in dealing with unstable loops.

If our goal was simply a stable loop, tuning would be a lot easier; i.e., we would simply raise the Proportional Gain until the loop began to cycle and then we would lower the gain to provide some comfortable margin.

On the other hand, if our goal was to minimize the visual oscillatory effects of nonlinearities on the system response, we would tune the system according to more conservative guidelines.

But, because of all the uncertainties involved with nonlinearities, this approach to loop tuning does not guarantee the best process variability performance.

FISHER DEAD TIME PROCEDURE

Although it is theoretically possible to analytically account for dead time in the process model as was done in the previous example, when we substitute such a process model into the IMC tuning equation, we obtain an expression for the controller algorithm which can be difficult to implement with conventional controllers.

A common approach to resolving this dilemma is to proceed as though there is no dead time and then modify the final controller gain setting to account for the dead time actually measured during the step-response test.

There have been numerous techniques proposed over the years for making this dead time modification, but one that is currently being used by Fisher Controls is discussed below.

This technique, which has been recommended by numerous process control researchers, simply modifies the (λ/τ) ratio used in the IMC tuning rule. As developed previously, the tuning rule for controller gain (without dead time involved) was as shown below.

$$K_C = \left(\frac{1}{K_P}\right)\left(\frac{\tau}{\lambda}\right)$$

This can be rearranged into the following form which involves the (λ/τ) ratio.

$$K_C = \frac{1}{K_P\left(\frac{\lambda}{\tau}\right)}$$

The dead time compensation rule modifies the (λ/τ) ratio in this expression as follows.

$$K_C = \frac{1}{K_P\left(\frac{\lambda + T_d}{\tau}\right)}$$

We can rearrange this equation as follows.

$$K_C = \frac{1}{K_P\left(\frac{\lambda}{\tau} + \frac{T_d}{\tau}\right)}$$

Thus, we see that the dead time adjustment factor simply increases the (λ/τ) ratio by an amount equal to the ratio of the process dead time (T_d) to the process time constant (τ).

In essence, this dead time adjustment factor decreases the controller gain thus detuning the system slightly to compensate for the dead time nonlinearity in the process.

When nonlinearities are involved (which is all the time), the only real way to know for certain the best gain setting to reduce process variability is to develop some kind of systematic closed-loop comparison test.

That is why Fisher has developed the performance testing loops in which various valves can be compared using a standard input disturbance and systematic tuning procedures. These closed-loop dynamic performance tests under actual loaded conditions provides Fisher with a wealth of information that aids in the design and development of Fisher valves.

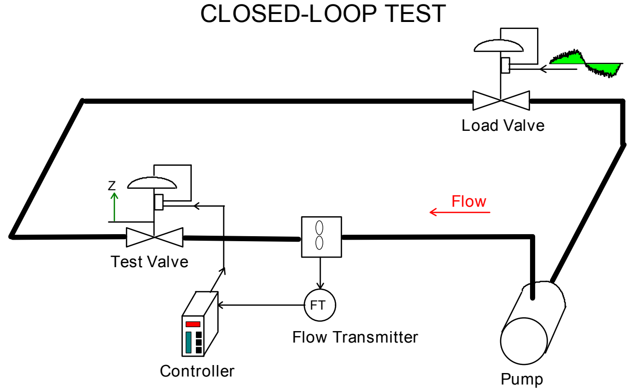
While these closed-loop tests also provide valuable information regarding the application of Fisher valves and their performance compared to other competitive equipment, it is important to recognize that the end user of these valves also needs to measure the process variability of each loop in order to ensure precise knowledge of the valve performance in the specific, as-installed application.

Furthermore, by continuously monitoring the process variability performance of the valve, the user can ensure that the desired performance is being maintained over time. Process control equipment is available which will allow on-line process variability measurements with relative simplicity.

CLOSED-LOOP PERFORMANCE TESTING

The diagram below illustrates a typical setup that Fisher uses to evaluate dynamic performance of valves under a wide range of operating conditions.

CLOSED-LOOP TEST



The valve to be tested is placed in the flow loop as shown.

The first step in conducting a test is to determine what mean flow rate is going to be used. Each valve is typically tested at a minimum of two different flow levels since the effects of system nonlinearities tend to be different at different flow levels.

Once the mean flow rate set point for the test is established, the controller is taken out of the loop (set on manual), and the pre-recorded series of load valve movements is input to the load valve to produce an open-loop disturbance in the process flow.

The process variable (flow rate) is continually measured and recorded during this manual test run. The resulting data points are then statistically analyzed to determine the standard deviation (σ) of the flow disturbance thereby quantifying how much process variability is inherent in the process when there is no attempt to control it; i.e., when the loop is in manual operation.

This Standard Deviation (σ) is then used as an objective measure of how much disturbance exists in the process and for which the test valve will be expected to compensate when the loop is closed with the controller in place and properly tuned.

The controller is then placed on automatic and the loop is placed in operation. The loop is tuned using the systematic IMC tuning procedure previously described, and the same pre-recorded load disturbance is played into the system.

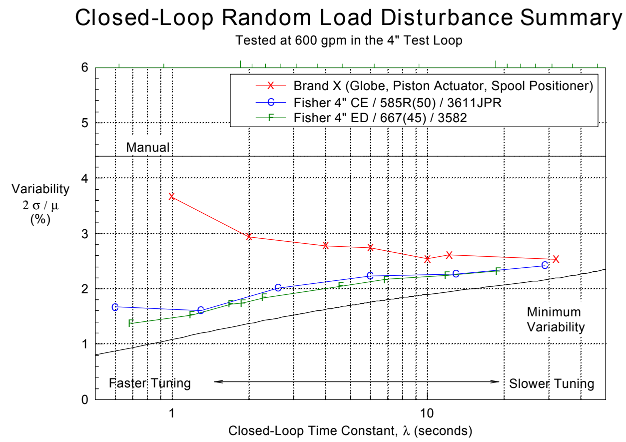
Again, the process variable (flow rate) is measured and recorded during this test run. The resulting data points are then statistically analyzed to determine how well the control loop using this test valve was capable of holding the set point flow in the face of the given random process disturbance.

This test is then repeated at several different controller tuning settings and at different flow conditions and the results are then plotted on a graph for comparison with other valves.
The same load disturbance sequence of random motion is played back to the control loop for each valve under test and for each test condition. This standard disturbance provides a basis for distinguishing between the performances of different valves at different loop tuning conditions in terms of their measured ability to reduce process variability under identical load conditions.

Plots are then made which compare the performance of the various valves being tested.

In order to provide a consistent standard for comparison from valve to valve, the same IMC tuning rules are applied to each test valve.

The result of these systematic Performance Tests is a chart such as that shown below which compares the ability of several different valves to reduce process variability for different tuning conditions; i.e., at selected values of closed-loop time constant (λ).



Since both the mean (μ) and the standard deviation (σ) represent measurements of the process variable, they will both have the same units of measurement. In the flow test loop previously described, these units will be gallons per minute (gpm).

The mean (μ) represents the set point flow condition, and (2σ) represents the magnitude of flow variation which will account for 95% of the measured data points.

We can use the ratio ($2\sigma/\mu$) as a convenient way to express the magnitude of the process variation as a percentage of the set point condition. This will provide an easy way to gain some perspective about the relative significance of the variation.

When comparing two or more valves through performance testing it is important to make those comparisons on an equitable and consistent basis.

As the loop is changed from one test valve to another the process transfer function (P) also changes since the valve is part of the process. This means that both K_P and τ change.

Since each valve tested has a different valve gain which contributes differently to the process gain (K_P), it follows that the actual controller gain settings for each valve will also have to be different in order to achieve the same closed-loop response (λ).

The most meaningful way to display these valve comparisons is to plot the process variability as a function of the closed-loop response (λ). Remember, when two valves have the same Lambda value, this means that they have the same system response time. This does not mean, however, that both valves will have the same capability to reduce process variability as we can see clearly from the previous process variability diagram.

Comparing the process variability performance of several valves when plotted as a function of λ will more clearly differentiate the good valves from those that don't perform as well.

This test data clearly illustrates that not all valves provide the same dynamic performance despite the fact that they all theoretically meet the conventional purchase specifications.

Note that both the Fisher CE and ED valves do a good job of following the minimum variability line over a wide range of controller tunings, which indicates excellent dynamic performance with minimum variability. In contrast, another popular valve (brand X) fared considerably less well and actually increases in variability as the system is tuned more aggressively for decreasing closed-loop time constants.

THE PAYOFF

Obviously all three valves are capable of controlling the process and reducing the variability, but some do it better. For example, consider what would happen if you replaced the brand X valve with the Fisher CE valve, and tuned the system to a 2.0 second closed-loop time constant.

The test data shows that this would result in a 1.1% improvement in process variability. This may not seem like much, but the results over a period of time can be impressive.

A valve which can provide this much improvement every minute of every day can rack up significant dollar savings over a single year.

By maintaining closer adherence to the set point, you can achieve a reduction in throughput by moving the set point closer to the lower specification limit. The 1.1% improvement in this example converts to a throughput savings of 9504 gallons per day. Using a conservative material cost of 30 cents per gallon the Fisher valve would contribute $2851 per day directly to your bottom line profits. This adds up to an impressive $1,040,615 per year.

The excellent performance of the Fisher valve in this example provides strong evidence that a superior control valve can have a profound economic impact on your bottom line.

IMC TUNING SUMMARY

The ability of the control valve to reduce process variance depends upon many factors. We cannot just look at any one thing in isolation as we did for the dead time situation earlier.

The interaction between all of these nonlinear factors becomes so complex that one cannot analytically predict with certainty how any valve will perform when it comes to reducing process variability. That is why it is so important to conduct the kind of process variability comparison tests that Fisher has instituted.

It is probably safe to assume, as a generality, that a conservative approach with lower rather than higher controller gain will produce more docile behavior of the loop, however, that may not be the best situation from the standpoint of reduced process variability.

Some process engineers often recommend a very conservative tuning with $\lambda/\tau = 3.0$. A plant with loops tuned to this conservative extent will tend to be rather well behaved which usually appeals to the person who likes to see things running smoothly.

While Fisher can still show improvement in process variability at such conservative settings, the test results usually demonstrate that the difference between well designed valves and average, run-of-the-mill valves tends to increase dramatically as the tuning is allowed to become more aggressive. This can be clearly seen on the process variability curves shown above where the difference in process variability gets better and better as the controller gain is increased (i.e., faster tuning with decreased λ values).

Thus, Fisher generally feels that a more aggressive tuning is usually possible and desirable; however, this will obviously depend upon the comfort level of the individual doing the tuning, or upon other operating conditions which have nothing to do with reduced process variability.

Fisher highly recommends that process variability be continuously measured as a way to quantify the performance of each control loop. This will not only provide you with the data you need to make the most intelligent control valve selections, but it will also ensure that the desired performance is being maintained over time. Process control equipment is available which will allow on-line process variability measurements with relative simplicity.

- END -

FISHER® Fisher Controls

For information, contact Fisher Controls:
Marshalltown, Iowa 50158 USA
Cernay 68700 France
Sao Paulo 05424 Brazil
Singapore 0512

Printed in USA