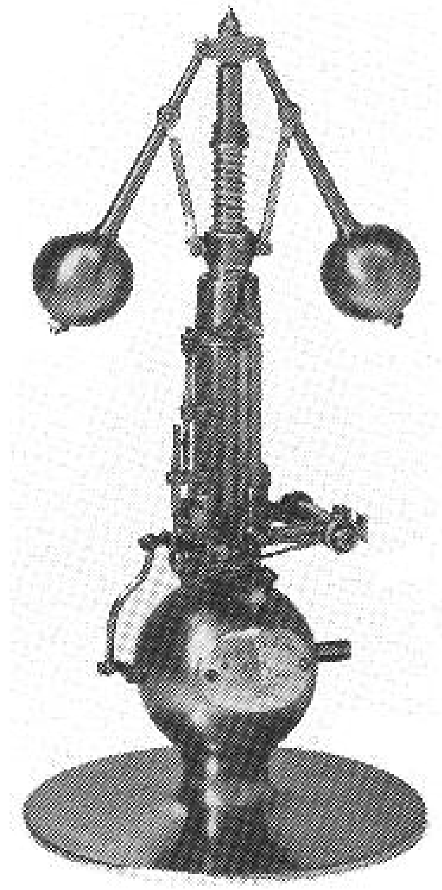
Feedback Control

Joseph Durham
Carleton College,

Feedback Control



Joseph Durham

Department of Physics and Astronomy
Carleton College
Advisor: Bruce Thomas

A feedback controlled system is a system that compares its own output to a desired value and automatically takes corrective action. This paper will examine linear feedback systems, including the basics of automobile cruise control, operational amplifiers, and PID controllers. System performance can be improved by the use of feedback, especially when faced with external disturbances. To realize these gains, however, the feedback system must be properly designed. Design tools such as Nyquist stability criteria, Bode plots, and pole-zero placement will be examined.

Table of Contents:

Title page image: Otto Mayr describes the as: “Patent model of a steam engine governor by Oliver A. Kelley and Estus Lamb, 1864.”[1]

I. INTRODUCTION

Over the last hundred years, the uses of feedback control have grown at an astounding rate. A feedback controlled system is a system that compares its output to a desired value, and then automatically takes "corrective action." Many modern conveniences, including automobile cruise control systems and thermostats, rely heavily on feedback. While its uses continue to grow, the utility of feedback control was first shown more than two thousand years ago. In ancient Alexandria, feedback was used to help increase the accuracy of water clocks. Despite its success, feedback was relegated to a handful of specific applications until the growth of steam and combustion engines. Centrifugal governors on steam engines first shed the spotlight on feedback control, but it remained an engineering art. The demand for long distance telephone communication spurred development of the mathematical formalism of feedback, and since then the field has exploded. It is now a crucial part of the design of electronics and amplifiers. More exotic applications include stabilizing aircraft, keeping satellites in orbit, and designing robots. Feedback has even crept into other sciences: psychologists have examined the role of feedback in human behavior and learning, and biologists are studying chemical feedback pathways in the human body. Another form of feedback is also used by corporations looking to maximize customer satisfaction. The tremendous growth of the field of feedback control can be attributed to the important gains it brings to a vast array of applications.

In examining the gains and pitfalls of feedback systems, I'll first examine a basic cruise control system to demonstrate the power of feedback. From this example, I'll build up the general formalism of linear feedback control and flesh out some of the specific benefits. After clarifying the basic idea of feedback, the history of its use by humans will be examined in more depth. This will segue into the development of the operational amplifier and the formalism of the concept of stability. Here the control system engineering tools developed by Harry Nyquist and Hendrik Bode will be discussed. Next, I'll demonstrate the power of transfer functions and pole-zero placement in system design and analysis. I'll conclude with a look at proportional-integral-derivative controllers and a summary of feedback control methods.

II. SIMPLIFIED CRUISE CONTROL APPLICATION

One example of linear feedback control is a simplified model of a car cruise control system. The purpose of this example is to show the improvements that feedback can bring to a system, even in the presence of external factors. Before I get into the specifics of cruise control, there is an important control system design tool that helps visualize system dynamics, the block diagram. A *block diagram* consists of several blocks and arrows connecting them: the blocks represent a process or action and the arrows represent information or energy flow. In control system applications, the process over which control is desired in known as the *plant*. A collection of other processes are then connected to the plant in such a way that the output is adjusted towards a specific value. I'll call the plant in this case *P*, and it will take in an input*u* and produce an output *y*. The diagram for process *P* is shown in Fig. (1).

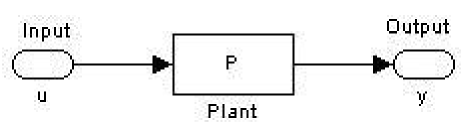


FIG. 1. Process P takes an input u and produces an output y.

Mathematically, this process can be represented by the equation:

$$y = Pu \,. \qquad (1)$$

In a car, there are several interconnected components that determine system behavior. The first is the angle of the accelerator pedal. This sends a signal to a small motor that controls the fuel intake of the engine. The motor, often called the throttle, adjusts the opening of a valve in the engine that restricts fuel intake. The opening is described by the throttle angle; the larger the angle the more fuel is available in the engine. So, the output of the throttle changes the amount of fuel flowing into the engine. A block diagram of this part of the car is shown in Fig. (2).

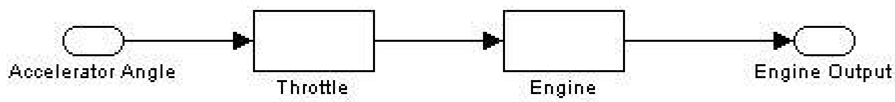


FIG. 2. Input to the system is the accelerator angle. The throttle takes in this signal and adjusts the throttle angle accordingly. The changing throttle angle changes the output of the engine.

The output of the engine is then transferred to the roadway by the wheels. To determine the actual speed of the car, a couple of other factors must then be examined. Newton's Laws form the basis for any such development, most importantly the second law:

$$\vec{F}_{net} = m\vec{a} \,. \tag{2}$$

For a simplified model of a car on level ground, the engine is producing a force in the forward direction while the motion of the car produces some drag in the opposite direction. Because of this, I only need to consider the vector components in the direction of travel. The force of the engine will be F_e, while the drag F_d will be treated as linearly dependent on the speed of the car:

$$F_d = bv \tag{3}$$

$$F_{net} = F_e - bv = ma \tag{4}$$

where b is the drag coefficient and m is the mass of the car. When these forces balance out and there is no net acceleration, the car is said to be in equilibrium. The flat ground equilibrium velocity r is then given by:

$$r = F_e / b \quad (5)$$

If the car begins ascending a hill, without a change in engine output its speed will drop. Some of the engine's work now has to go into increasing potential energy, while the drag forces will initially dissipate energy at the same rate. The speed will drop until a new equilibrium is reached. The dynamics of how this occurs are not important to this development; I'll only look at the steady state velocity. Assuming that the incline is a relatively small angle, the new equilibrium will be governed by:

$$F_{net} = F_e - bv - mg\theta = ma = 0 \quad (6)$$

$$v = \frac{F_e - mg\theta}{b}. \quad (7)$$

Eq. (7) forms the model of the equilibrium velocity of the car based on engine output. For any road grade, the model shows how the car will react. Based on this model, the behavior of the car on road can be represented by Fig. (3).

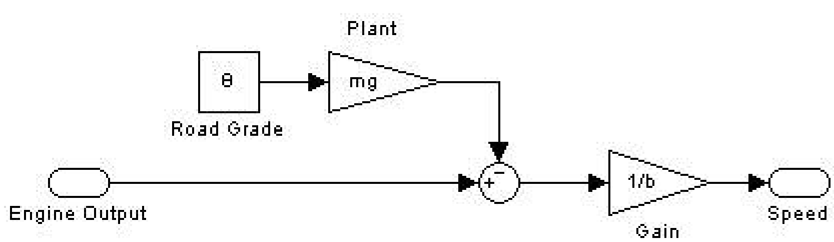


FIG. 3. The force put out by the engine, F_e, is modified by the effects of gravity when traveling on an incline θ. This force is then divided by the coefficient of drag, b, to get the equilibrium velocity of the system. See Appendix A for block meaning.

Putting Fig. (3) together with Fig. (2), the total behavior of the system could be seen. However, the actual cruise control portion of the system has not yet been introduced.

When traveling on a flat stretch of road, if the driver presses the "set" button on the cruise control, the current speed is saved into an onboard computer. This speed, *r*, given by Equation (5), is the desired speed of the system. The onboard computer then compares the desired speed and the actual speed, and adjusts the throttle accordingly. This takes the place of the accelerator pedal input into the system, the driver's foot is freed of its responsibility and the computer now has control. To simplify the diagram, the computer and throttle blocks will be combined and labeled controller. Fig. (4) shows the diagram with a cruise control system included.

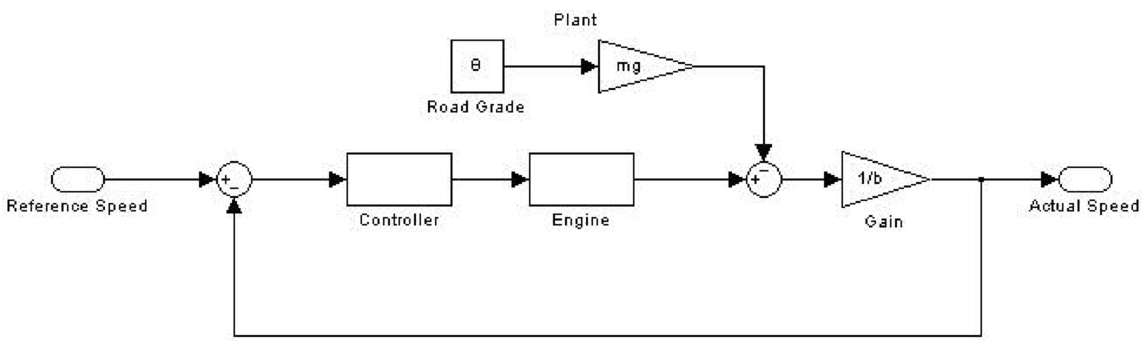


FIG. 4. Block diagram for a car with cruise control engaged traveling on a road with grade *θ*. See Appendix A for block meaning.

In this example, the relationship between throttle angle γ and speed will be treated as linear in the region of operation. Using an engine constant *E* to relate throttle angle and output force yields $F_e = E\gamma$. The feedback loop can now be constructed, where the controller will produce a signal $\gamma = C(r - v)$, where *C* is the controller constant. The equilibrium velocity of the system, *v*, is now given by Eq. (8).

$$v = \frac{EC(r - v) - mg\theta}{b} \qquad (8)$$

Now solving for *v*, the magic of feedback appears:

$$\left(1+\frac{EC}{b}\right)v=\frac{EC}{b}r-\frac{mg}{b}\theta \qquad (9)$$

$$v=\frac{EC}{b+EC}r-\frac{mg}{b+EC}\theta. \qquad (10)$$

Comparing this equation to the result with out feedback, the benefits will be plain. If the driver maintains a constant accelerator angle, the engine force will be constant. Therefore, as shown below, Eq. (5) and (7) can be combined to yield Eq. (11).

$$r={F_e}/{b} \qquad (5)$$

$$v=\frac{F_e-mg\theta}{b} \qquad (7)$$

$$v=r-\frac{mg}{b}\theta \qquad (11)$$

In this example, the reference speed will be 20 m/s (about 45 mph). For simplicity, I'll arrange the constants such thata 1 degree slope would reduce speed by 1 m/s and a 2 degree slope by 2 m/s. Using *b* = 10 Ns/m, the approximate value at 20 m/s for a Porsche 911 (see Appendix B), the gain on the incline would then be $570\frac{N}{\deg}$. For open loop control, each degree of incline adds 5% to the error from constant speed. With a simple feedback circuit, the speed will hold much more constant. Choosing a value of 1000 for *EC*, yields the following result. In the no cruise control case, Eq. (11) becomes Eq. (12).

$$v=r-1\tfrac{m}{s\cdot\deg}\theta \qquad (12)$$

For the cruise control case, Eq. (10) becomes Eq (13).

$$v=\frac{1000}{1010}r-\frac{10\,\tfrac{m}{s\cdot\deg}}{1010}\theta=0.99r-0.0099\tfrac{m}{s\cdot\deg}\theta \qquad (13)$$

The cruise control system has reduced the effect of road grade on speed by a factor of 100! This is a tremendous increase in stability of *v* for changing *θ*. If the feedback gain was increased, then the independence of v from θ would increase further. However, it does come at a cost. Notice that now for no incline, the equilibrium speed has been reduced to 0.99r, an error of 1%. This is referred to as the *steady state error* of the feedback system, and must be addressed in almost all feedback applications. Since we set the controller output directly proportional to the difference between v and r, the engine would shut off if $v = r$. Fig. (5) illustrates both the desensitivity and steady state error of this cruise control system on a sinusoidally oscillating roadway. It compares the speed of a car with its accelerated pedal held constant and one with cruise control engaged (all other things being equal). One method for avoiding steady state error will be discussed in section VII on PID controllers. Despite this drawback, the cruise control example has proven that feedback control systems produce important benefits in real world applications. The next section is a general treatment of linear feedback systems, and will shed more light on the benefits and drawbacks of using feedback.

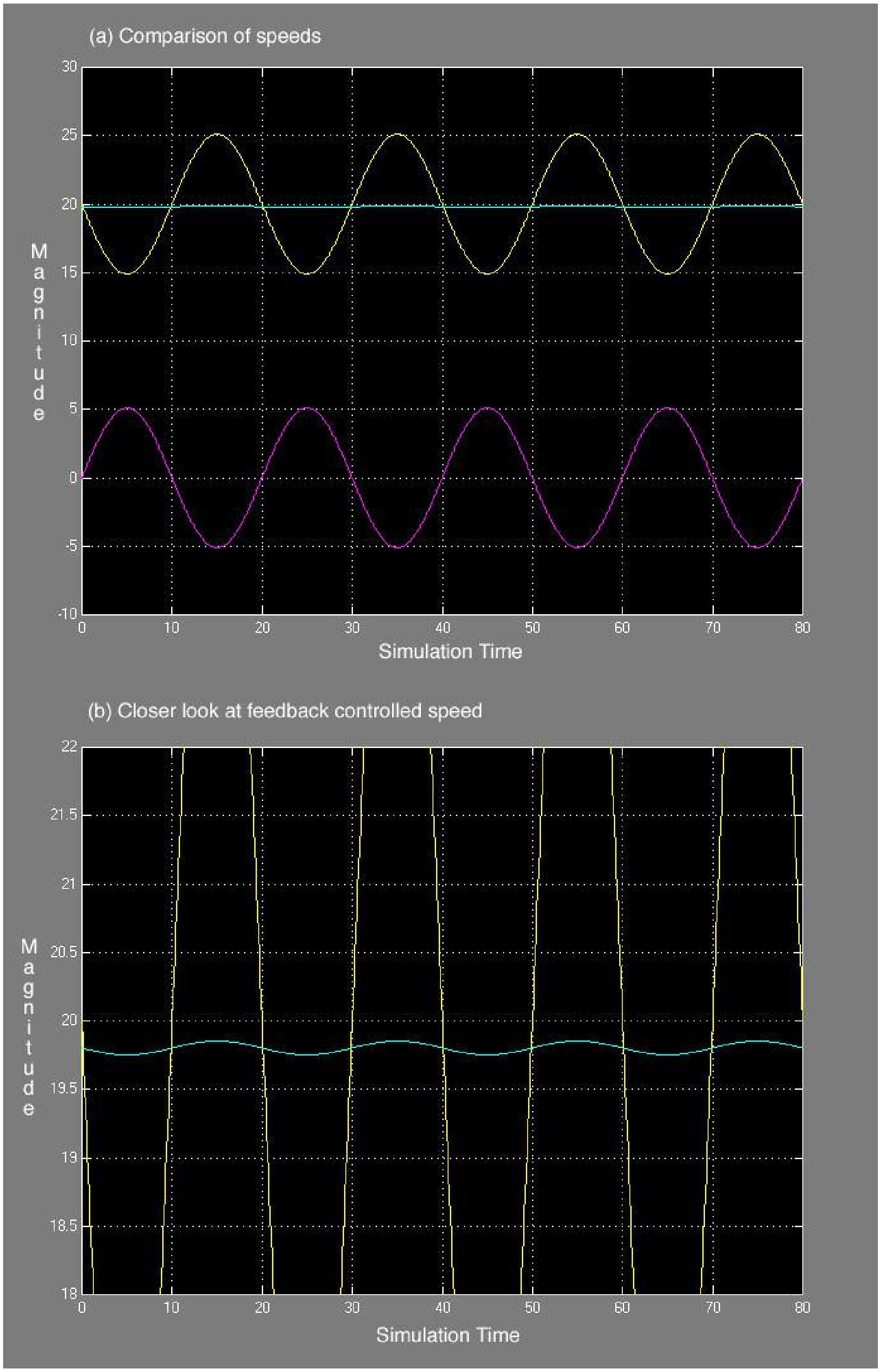


FIG. 5. For these two graphs, two cars are traveling on the same road. The road grade is sinusoidal, shown in purple. (a) shows the speeds for the car with its accelerator pedal held at a constant angle (yellow), and one with cruise control engaged (teal). (b) is a zoomed in subsection of (a), showing the desensitivity and steady state error in the feedback controlled speed.

III. LINEAR FEEDBACK FORMALISM

As was mentioned at the start of Section II, in any feedback control application, there is some process, *P*, over which control is desired. In the lingo of control system engineering, *P* is referred to as the *plant*. Again, this process takes some input *u* and produces an output *y*, represented by Eq. (1) (reproduced below).

$$y = Pu \tag{1}$$

This is what is known as an *open-loop* system. No information about the output of the system is looped back to affect its behavior. It contains no feedback, and if the input changes by Δu, then the output changes by $P\Delta u$. A feedback controlled system, on the other hand, is error driven: the input to the system is modified based on a comparison of the actual and desired outputs of the process. If we want the system to produce an output *r*, then the feedback setup would look like Fig. (6).

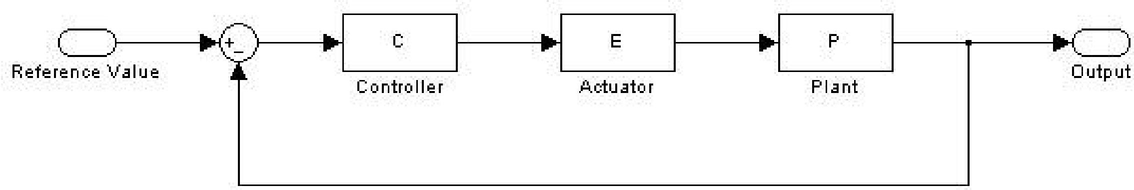


FIG. 6. General feedback control setup of plant, *P*, with controller and actuator.

Output *y* is compared to the reference value *r*. The *controller* then acts on the difference between *y* and *r*, feeding an appropriate input into the *actuator*. The actuator then changes the system response based on this signal. In the cruise control example just discussed, the onboard computer (controller) changes the engine throttle angle. This change in throttle angle then corrects the output of the engine (actuator), adjusting the behavior of the car on the road (plant). For this general treatment, all these different components will produce outputs linearly proportional to their inputs. As a result, this development applies to *proportional* feedback devices. A proportional controller sends a signal to the actuator proportional to the difference between y and r. Other kinds of controllers will be treated at the end of the paper, in section VII. Since both actuator and controller are linear, the controller and actuator can be treated as a single block that takes in the difference of y and r and produces a signal modified by C. I will combine them for this theoretical development, but in the construction of a feedback control mechanism the controller and actuator are often distinct parts. The output y is then given by:

$$u = C(r - y) \tag{14}$$

$$y = Pu \tag{15}$$

$$y = PC(r - y)\,. \tag{16}$$

Notice that y is subtracted from r; this is what is known as negative feedback. As y gets closer to the desired value, the controller and actuator will presumably need to exert less control on the process. If instead the sum of y and r was taken, this would be an example of positive feedback. Although not good for precise control of a dynamic system, positive feedback is used to create oscillators.

The result in Eq. (16) is the general form of the cruise control feedback system seen in Section II. While this works well for systems whose output is supposed to be driven to the reference value, other applications require amplification of the reference signal. Adding a linear gain to the feedback path, F, can produce this amplification. As will be shown shortly, if F is less than one the system output will be a multiple of r. Fig. (7) shows the block diagram corresponding to this setup.

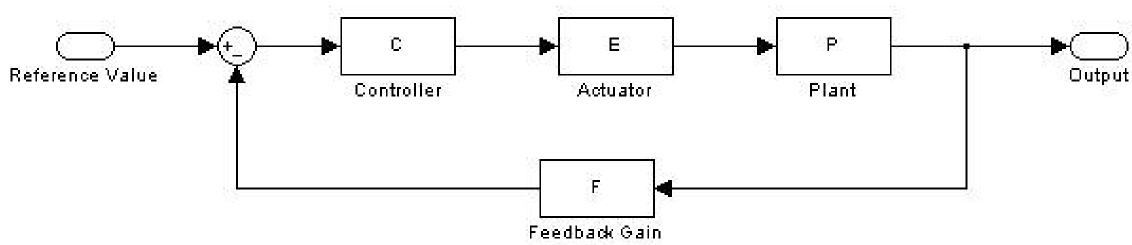


FIG. 7. General feedback system setup with feedback gain *F* added.

In this case, Eq. (16) will be modified into the following form:

$$y = PC(r - Fy)\ . \tag{17}$$

Eq. (17) represents a general negative feedback configuration. Solving for the output y yields:

$$y = \frac{PC}{1 + PCF} r\ . \tag{18}$$

In Eq. (18), *PCF* is referred to as the *gain* of the feedback system. In a standard feedback application, *PC* will be made as large as possible. Transistor based operational amplifiers can have PC values of 10^6 or greater.[2] At such large values, Eq. (18) can be well approximated by:

$$y \approx \frac{1}{F} r\ . \tag{19}$$

The feedback gain *F* then controls the effective amplification of the system. This approximation also sheds light on the basis for the steady state error uncovered in the cruise control example. For *F* = 1 and a large value of *PC*, the effect of the one in the denominator is small, but it is not always negligible. The one is a result of the use of proportional control, and also the cause of steady state error. Fortunately, there are methods to eliminate steady state error by using more than just proportional control, as will be covered in Section VII on PID controllers.

Another of the important tradeoffs of a feedback control system also stems from Eq. (18). Without the feedback system in place, the output would be $y = PCr$. If the value of *PC* is huge, the open-loop output would be much larger than the output with feedback included. For an amplifier, this large gain is important. In fact, when feedback amplifiers were originally being conceived, many people rejected the idea because the effective gain of the amplifier is reduced by including feedback. In essence, a feedback control system gives away large amounts of amplification in order to improve other properties of the system. Feedback enthusiasts eventually won out, because the lost gain can be recovered by using multiple amplifiers. On the other hand, the major improvements in system response can be vital. Two of these improvements will be examined here. First, if some disturbance, *d*, enters the plant as shown in Fig. (8), the feedback system will mitigate its affect.

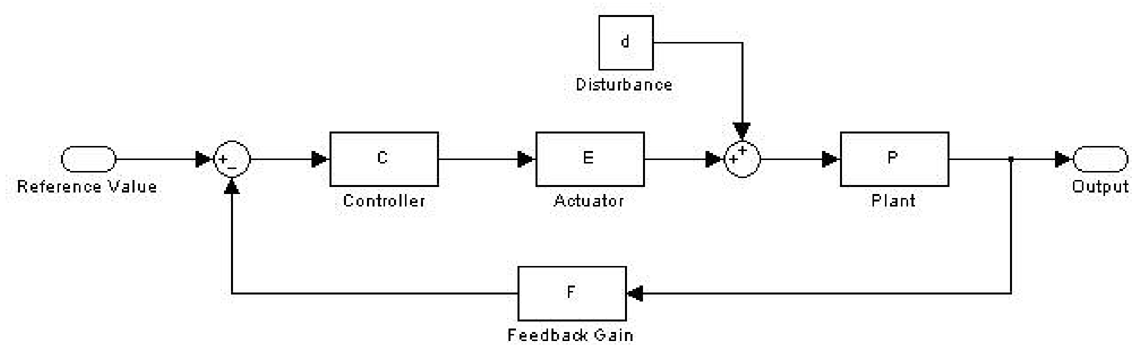


FIG. 8. General feedback system diagram with disturbance *d* influencing the plant.

Now the output y is given by:

$$y = PC(r - Fy) + Pd \tag{20}$$

$$y = \frac{P}{1 + PCF}\left(Cr + d\right). \tag{21}$$

Under the open-loop condition, the disturbance would have been magnified by *P* and added to the output. With feedback, the system automatically resists the disturbance. This is clearly demonstrated by comparing change in y due to the disturbance for both the open-loop case (Δy_{OL}) and the feedback case (Δy_{CL}), shown in Eq. (24).

$$\Delta y_{OL} = Pd \tag{22}$$

$$\Delta y_{CL} = \frac{Pd}{1 + PCF} \tag{23}$$

$$\frac{\Delta y_{CL}}{\Delta y_{OL}} = \frac{1}{1 + PCF} \tag{24}$$

The closed-loop system suffers a smaller change in output by a factor of $1 + PCF$, which is referred to as the *desensitivity* of the system.[3] Since *PC* is usually a very large value, the system has been desensitized to disturbances, like road grade in the cruise control example. The disturbance need not be an external influence: electronic amplifiers suffer gain fluctuations due to changing temperature. A feedback enhanced process is less susceptible to changes in its own properties!

The second major benefit is that process *P* can often be made to imitate some other process *Q*.[4] If we wanted the process *P* to behave like *Q*, the controller C would need to be setup likeEq. (27).

$$y = Qr = \frac{PC}{1 + PCF} r \tag{25}$$

$$Q = \frac{PC}{1 + PCF} \tag{26}$$

$$C = \frac{Q}{P(1 - FQ)} \tag{27}$$

By adding a controller setup like Eq. (27), the properties of process P would be changed to imitate process Q. In theory, a controller can be found for most applications to satisfy the above condition. The mathematics for uncovering the behavior of a controller given desired and actual responses will be discussed in Section VI on transfer functions and pole-zero placement. However, humans had been making use of feedback for nearly two thousand years before the development of control system theory. The next section looks at the history of human use of feedback, from antiquity through locomotion.

IV. HISTORY OF FEEDBACK CONTROL

There are numerous opinions on what constituted the first human use of a mechanical feedback loop. Hand-eye coordination is certainly a great example of a feedback circuit (try writing while watching your actions in a mirror for an experience of positive feedback). Human attendants have made corrections to systems to achieve a desired result at least since the first use of fire. However, the invention of a mechanical system that automatically adjusts its own behavior to reach a desired end clearly represents an important step in the sophistication of humanity. According to Otto Mayr, the first use of an automated, external feedback loop was the design for a water clock from around 250 BCE.[5] Although limited evidence remains, Mayr contends that Ktesibios of Alexandria invented the first feedback controlled device. Only the water clock has the distinction of being the first feedback system developed, however all three developments outlined below were invented quite independently. The self-regulating oven and the centrifugal governor were also ingenious engineering breakthroughs which occurred before the formalism of feedback theory.[6,7]

A. Water Clock

To use water to keep time, a constant flow of water is needed. A vessel of water with a spout at the bottom will only produce a constant flow if the level of the water is constant (See Fig. (9)). The pressure, *P*, at the bottom of the vessel is determined by the weight of the volume of water above the opening.[8] In Eq. (28), the area of the spout is denoted *A*, the density of water is denoted ρ, the depth of the water *h*, and gravitational acceleration *g*.

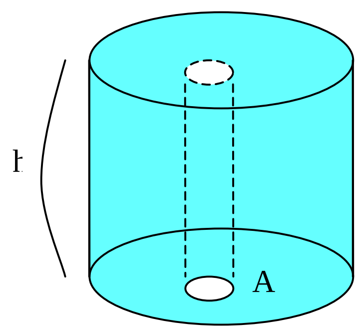


FIG. 9. Vessel of water with a spout at the bottom. The pressure over the area of the spout is dependent on the depth of the water.

$$P = \frac{F}{A} = \frac{mg}{A} = \frac{(hA\rho)g}{A} = \rho gh \qquad (28)$$

The flow produced by this pressure depends on the nature of the spout, but is always dependent on the pressure on the spout opening.[9] The level of the vessel would then drop because of this flow, causing a drop in pressure. To keep the level constant, the vessel would have to be replenished by a constant flow equal to the outflow, seemingly just moving the problem up a level. Level dependent pressure is one reason that water was not used in hour glasses, the level would drop much faster at the beginning of the hour than at the end. As discussed by Clark Ritz in his comps paper, the granular nature of sand means that for a column of sand, pressure is independent of height of the column.[10] The flow of sand is then independent of height making it a simpler measure of the passage of time. Ktesibios devised a system that kept the level of one vessel constant, even when feeding it from a non-constant source.[11] The basic idea of his design is shown in Fig. (10).

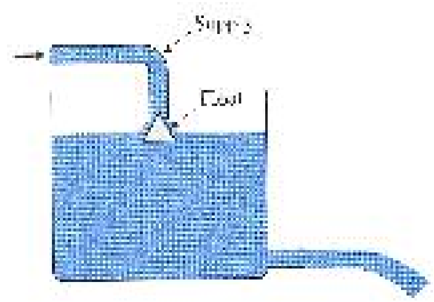


FIG. 10. Float controlled fluid level, as first used in the water-clock of Ktesibios.[12]

The major breakthrough of Ktesibios' design is the use of a float to control the rate of input. In the equilibrium case, the float in the constant level vessel will restrict the water input to the rate of output. If the water level should rise, the float would further plug the input pipe, limiting or blocking flow into the vessel. Notice that for this configuration, the float is both the controller and the actuator. Other examples of feedback control from antiquity exist, mainly for level control of fluids, some of which separated the sensor and actuator.[13] It took nearly two-thousand years before another truly unique example of feedback control surfaced.

B. Self-regulating Oven

In the early 1600's, alchemists were still searching for a way to turn ordinary metals into gold. One hypothesis was that transmutation into gold might occur if the base metal was held at a constant temperature for a long time: an application ripe for the application of feedback.[14] Enter Cornelis Drebbel, a Dutch chemist who served the kings of England. He developed an oven that maintained a constant temperature, the idea behind which is pictured in Fig. (11). The heat produced by a fire is dependent on the availability of oxygen, so Drebbel used the temperature of the oven to control the air intake. The thermal expansion of the fluid in the thermometer adjusts the opening of the damper. Not surprisingly, the alchemy application did not pay the bills, but the oven was soon modified to serve as an incubator for chicken eggs.[15] It was only another hundred years before the next novel application of feedback control, which again held promise for the automation of food production.

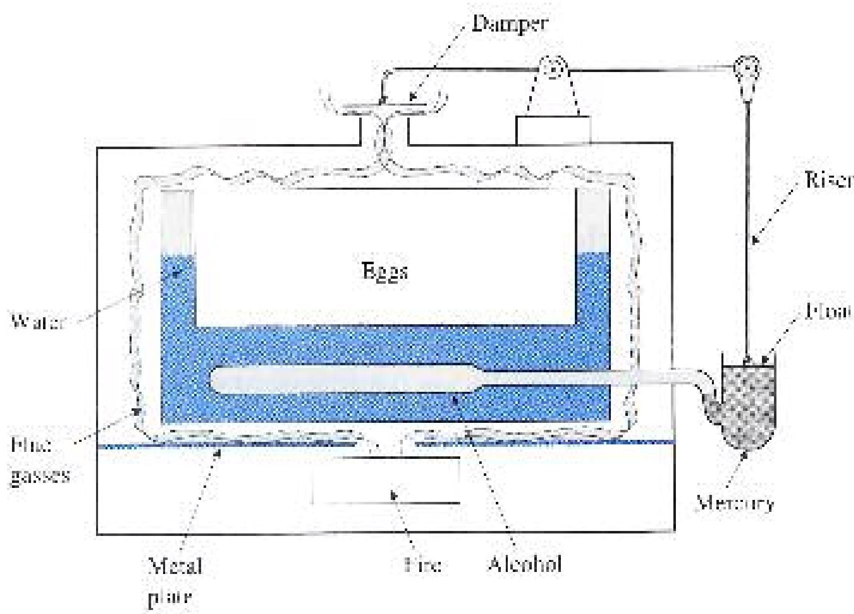


FIG. 11. Self-regulating oven design, first conceived by Cornelis Drebbel in the 1600's.[16]

C. Centrifugal Governor

Windmills harness the power of the wind to grind grain between two large stones, producing flour. While certainly easier than grinding all the grain by hand, the inherent gustiness of wind makes the process less than ideal. Given the immense friction between the two millstones, if the wind died just for a moment, the stones might grind to a halt.

Considerable work would be needed to get the large masses rotating again. The use of feedback can drastically reduce fluctuations in an output even if the input varies widely. During the latter half of the 1700's, millwrights invented several systems for keeping the millstones rotating on gusty days.[17] Unfortunately limited information on the inventors or the thought behind the inventions is left, mostly just patent applications remain. The most important invention, in terms of its effect on the future of feedback control, was Thomas Mead's patent of 1787.[18] In his design, two pendulums were made to rotate at the speed of the millstones; the centrifugal motion of the pendulums was connected to height of the rotating millstone. At high wind speeds, the stone was allowed to drop and grind away at the grain. If the wind suddenly waned and the speed of the stones started to drop, the pendulums would swing in, the stone would be picked up and the friction would be reduced. While the governing of speed was a very important invention, it took an application of the concept in another arena to make feedback famous.

The big break that brought feedback out of the shadows came with the development of steam and internal combustion engines.[19] When the load on an engine is suddenly increased, something needs to be adjusted on the engine to increase its power output. For steam engines, a throttle on the line between the boiler and the pistons controls how much steam is used to power the engine. When the throttle opens wider, more steam is let through and the engine output is increased. Without an opening of the throttle, when the load on an engine increases, the speed of the engine will drop and if the load is big enough, the engine will stall. Engines were built with manual throttle controls, but clearly an automated system is superior. Removal of the human element and response time of an attendant would increase the reliability of the engine. All of these arguments very closely parallel those that brought about the use of feedback to maintain mill speed. In 1788, James Watt modified the centrifugal mechanism developed for mills to fit the steam engines his firm was developing.[20]

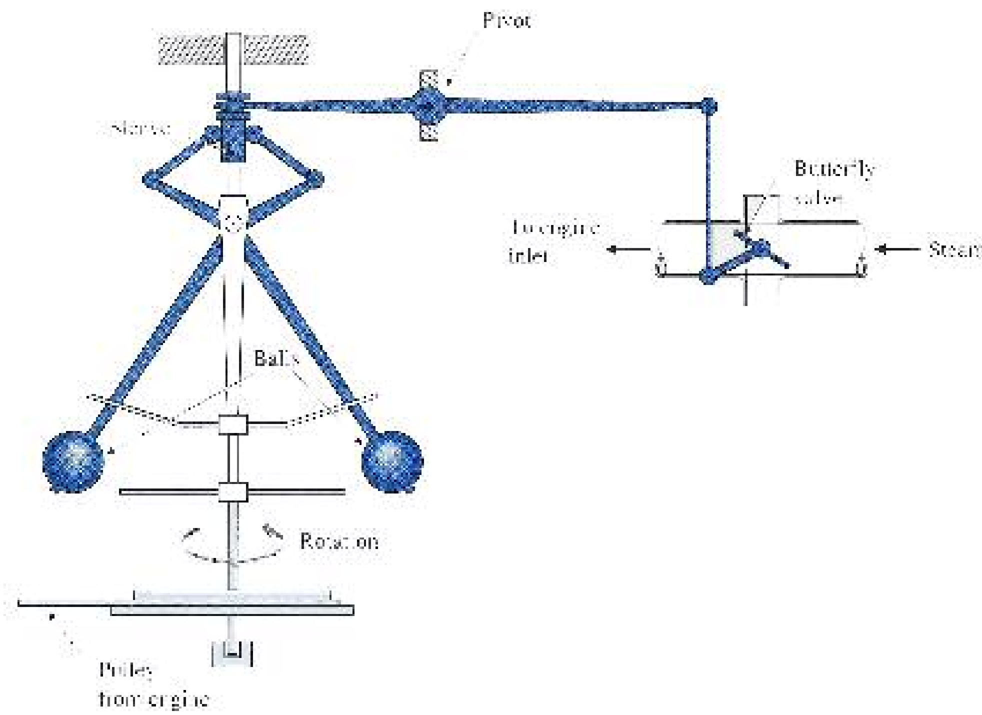


FIG. 12. Centrifugal governor for use in steam engines.[21]

Watt and his colleagues attempted to keep their governor technology secret, since they could not patent a borrowed concept. While the technology itselfspread rather slowly, the concept of the centrifugal governor for steam engines spread rapidly across England. In fact, in a rather humorous twist, in 1804 somebody in England attempted to file a patent for:

> … the centrifugal pendulum as applied to the speed regulation of mills, instead of giving a detailed explanation, he simply referred to it as "a pair of centrifugal balls – like the governor of the steam engine …"[22]

Originally developed to control the speed of mills, the centrifugal governor gained its notoriety from success on steam engines. The same concept was adapted to internal combustion engines, and the concept is still use in cars to this day. While Watt was too "practical" to study the theoretical side of the governor, several mathematicians developed concepts surrounding feedback control in the 19th century.[23]

D. Feedback Theory

One of the concepts fundamentally important to feedback control is the idea of *stability*. Anyone who has ever experienced the potentially ear piercing whine of a public address system, has experienced a feedback system becoming unstable. In this example, the microphone picks up some of the output of the speakers. If the amplification of the microphone is high enough, the small input will produce an increase in the volume of the output signal. This can quickly spiral out of control; hence the sound control board at most electrically enhanced outdoor functions. The first mathematician to discuss instability and treat it using differential equations was G. B. Airy, who was also an astronomer.[24] Airy adapted the centrifugal governor method to rotate telescopes counter to Earth's rotation. He noticed in 1840 that for certain setups "the machine (if I may so express myself) became perfectly wild."[25]

In 1868, James Clerk Maxwell studied the differential equations of the governor and determined that they depended on the roots of certain characteristic equations (see Section VI on transfer functions).[26] He successfully deduced the stability criteria for 2nd and 3rd order polynomials. E. J. Routh generalized Maxwell's method to polynomials of any order, to win the 1877 Adams Prize.[27,28] While still useful today for lower-order systems, Maxwell and Routh's method is too cumbersome for higher-orders. Also notable was A. M. Lyapunov's study (1893) of the stability of non-linear differential equations of motion.[29] The mathematical formalism of feedback engineering received a huge boost in the 1930's, when a non-mechanical need for feedback control surfaced.

V. OPERATIONAL AMPLIFIER

In 1915 a transcontinental telephone line opened, connecting the United States as never before.[30] This was quite a technical feat, as any transmission line will attenuate the signal that passes through it. As the signal produced by a microphone in New York travels across the country, its amplitude drops with the attenuation of the transmission line. To arrive in San Francisco differentiable from the background noise, the signal had to be amplified. Bell Telephone Laboratories was founded in 1925 as a cooperative venture between AT&T and Western Electric, partly to address this issue.[31] The lab conducted research into all areas associated with telephone and telegraph communications, from speakers and microphones to transmission. Several important breakthroughs surrounding feedback amplifiers occurred at Bell Labs during the 1930's.

Vacuum tubes were the best suited amplifiers of the day, but unfortunately they were not perfect. After going through a series of these vacuum tubes, the output signal was distorted because of their inherent nonlinearities. If too many amplifiers were placed in series, the voice would become unintelligible. While the 1915 line worked well enough, economic pressures were pushing telephone companies towards transmission lines with smaller wires and thus higher attenuation.[32] In fact, by the time coaxial cable and its Megahertz carrier frequencies connected the country, nearly 100 times as many amplifiers would be necessary for transcontinental communication.[33] Bell Labs began working on a linear amplifier to meet these needs.

The major break came from Harold Black, who realized that feedback was a solution that would solve issues with both the linearity and gain fluctuations of amplifiers. After graduating from Worcester Polytechnic Institute in 1921 with a degree in electrical engineering, Black went to work for Western Electric.[34] Eventually he ended up at Bell Labs, working on linearizing the latest breed of vacuum tube amplifier. While development of the feedback amplifier is what brought him fame, he first realized that feedforward could solve the linearity problems. In Black's mind, the output of an amplifier consisted of two overlapped parts, the linearly amplified signal and some non-linear distortion.[35] This concept is shown in Fig. (13), where the output signal (purple) is the sum of the linear signal (blue) and the distortion (red). The feedforward setup was then a natural solution. The output of the amplifier was fed into another electronic circuit. This circuit would calculate the distortion signal and then subtract that distortion from the amplifier output. A block diagram representing this setup is shown in Fig. (14).

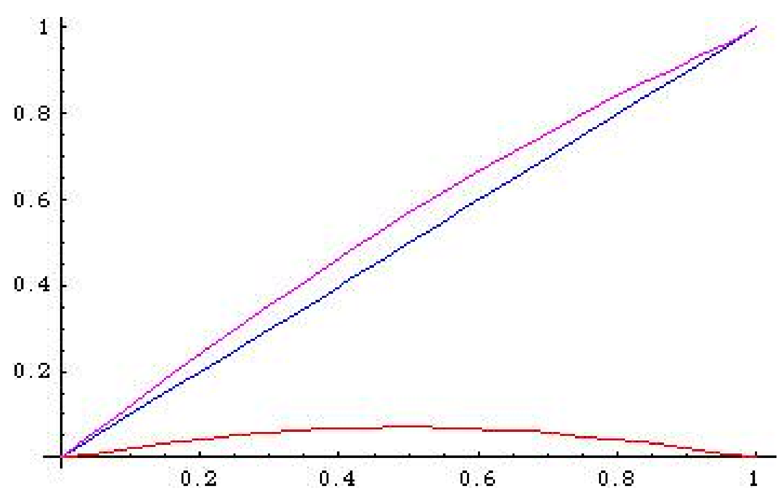


FIG. 13. Harry Black's conception of amplifier output (purple) was the overlapping of a linearly amplified signal (blue) and some distortion (red).

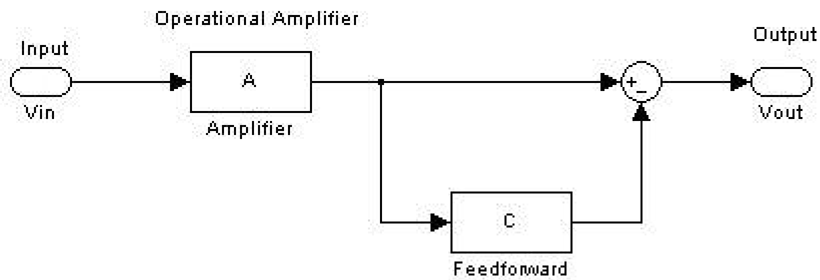


FIG. 14. Block diagram of one of Harry Black's first attempts to linearize vacuum tube amplifiers, in this case using feedforward, *C*.

Unfortunately, the feedforward term *C* requires precise setup for the amplifier gain *A*. If the slope of the linearly amplified signal changes, then *C* has to be recalibrated to be able to pick out the distortion.[36] Since amplifier gains fluctuate with temperature, humidity, and many other factors, recalibration would have to be done almost continuously. While possible in the lab, this cannot be done in the field, especially not when a more elegant solution exists. Since *C* requires knowledge of *A*, why not just feed the output back through *A*? When Black reached this epiphany, the operational amplifier was born.

An operational amplifier (op-amp) is an electronic amplifier with a huge open-loop gain *A*, which can be many orders of magnitude.[37] After feedback is included, the gain of the op-amp is determined by feedback gain *B*, just like was seen in Section II and Eq. (19). These days they are built from transistors instead of the vacuum tubes that Black was accustomed to, but the principle remains the same. A block diagram representation of an op-amp is shown in Fig. (15), while an electrical schematic is shown in Fig. (16).

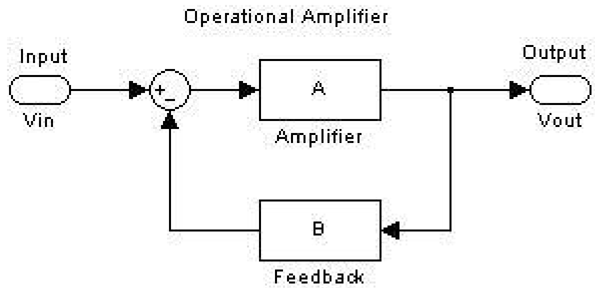


FIG. 15. Block diagram of an operational amplifier with amplification *A* and feedback gain *B*.

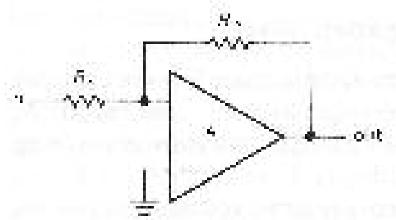


FIG. 16. Electrical schematic of an operational amplifier or op-amp. The amplifier itself is the triangular block labeled A, but the amplification of the circuit also depends on the two resistors, R_1 and R_2. Together, these resistors form the feedback gain *B* in Eq. (29).[38]

The voltage output of an op-amp is given by Eq. (29).[39]

$$V_{out} = \frac{A}{1+AB} V_{in} \tag{29}$$

Again, I'll show that changes or nonlinearities in the amplifier, in this case Δ*A*, are reduced by the desensitivity of the system. If the overall gain of the amplifier is denoted *G*, then Eq. (32) shows the relative deviation magnitude caused by Δ*A*.[40]

$$G = \frac{A}{1+AB} \tag{30}$$

$$\frac{\Delta G}{\Delta A} = \frac{1}{1+AB} - \frac{A}{(1+AB)^2} B = \frac{1+AB-AB}{(1+AB)^2} = \frac{1}{(1+AB)^2} \tag{31}$$

$$\frac{\Delta G}{G} = \frac{1}{1+AB} \frac{\Delta A}{A} \tag{32}$$

This was the improvement that Black was looking for: increase in linearity and decrease in sensitivity to amplifier fluctuations. Black, however, met with some rather obstinate resistance from both higher-ups at Bell Labs and the US Patent Office. To get vacuum tubes to have high gains was a real chore, so giving away the gain seemed a steep price to pay.[41] In addition, feedback was relatively well known to electrical engineers working with vacuum tubes. Their opinion of it really put the negative in negative feedback. Electrical engineers often worked very hard to eliminate feedback in their amplifier circuits because of the instability it produced. Just like a 'singing' public address system, amplifiers were prone to the development of oscillations that would overpower any useful information. The basis for this is that *B* has to be treated as impedance, not just a resistance. Even if no actual capacitor is included in the circuit, there will be a small capacitance inherent in the wires and solder joints of the circuit. The impedances for a capacitance *C* and resistance *R* are as follows:[42]

$$Z_R = R \tag{33}$$

$$Z_C = \frac{1}{i\omega C} \tag{34}$$

where *i* represents an imaginary number and ω is the frequency of oscillation. Including a frequency dependent imaginary term in the gain is inviting instability. Plotting *AB* in the complex plane, the magnitude of this vector is known as the *loop gain* of the system. Properties of the loop gain determine whether or not the system will be stable. The angle between the positive real axis and the loop gain is then the phase shift caused by amplification.[43] If the phase shift ever reaches 180° for some frequency, output signals at this frequency are essentially the inverse of the input scaled by the magnitude of *AB*. When this is then subtracted from the input, it is inverted again such that it constructively overlaps with the input! In Fig. (17) the input signal is the larger, green signal, while the blue signal is about to be fed back into the input. The blue signal has had its sign invert to subtract it from the green signal in standard negative feedback fashion. However, if a

180° phase shift is placed in the feedback path, the blue signal will slide half a wavelength with respect to the green. Suddenly negative feedback has become positive! If the magnitude of $AB \geq 1$ in these circumstances, then the feedback loop will oscillate. The man who developed this understanding of when instability strikes was a researcher with AT&T, Harry Nyquist.

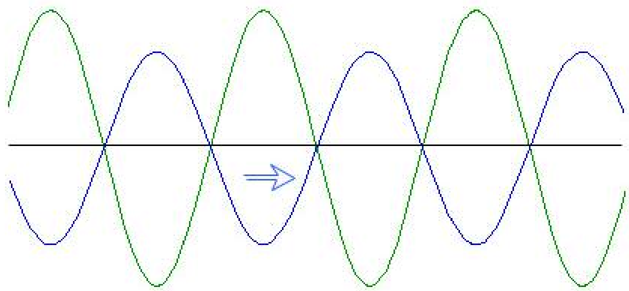


FIG 17. Input signal is green, negative feedback signal is blue. If a 180° phase shift is induced in the feedback path, negative feedback will become positive.

A. Nyquist Stability Criteria

Nyquist was a Swedish immigrant who received a Ph.D. in physics from Yale.[44] In 1928, Black joined Nyquist in conducting a trial of the new feedback amplifiers. These studies formed the basis of the famous *Nyquist stability criterion,* outlined in his 1932 paper on "Regeneration Theory."[45] In his paper, Nyquist defined a *stable* circuit as one where "all disturbances impressed upon the circuit died out in a finite time." In an *unstable* circuit, "disturbances went on indefinitely", perhaps even growing in amplitude. From this definition, and the insight on phase changes and gain, Nyquist developed his criteria and a simple, empirical way to test for instability. Capacitance and inductance have frequency dependent, imaginary impedances. Therefore it is necessary to look at the whole frequency range of the system to determine if instability will strike. The loop gain of a feedback circuit (*AB* for the feedback amplifier above) is plotted for the range of frequencies until it is clear that higher frequencies will not be amplified. If this line contains or encircles the point (-1,0), where the amplifier has unity gain and 180 degree phase shift, then the circuit is unstable. An example of a Nyquist plot, as these plots are known, is shown in Fig. (18).

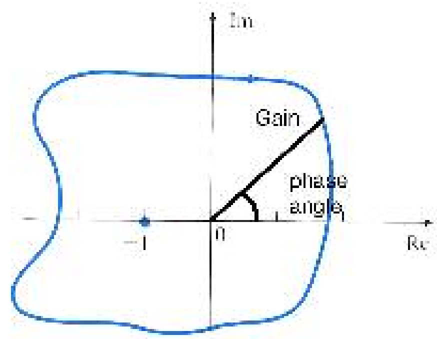


FIG. 18. Example of a Nyquist plot. Gain and phase a plotted over a range of frequencies. Since the point (-1,0) is encircled, the system depicted in this plot would be unstable.[46]

To test for instability, an engineer no longer had to connect the feedback path and see what happened. The feedback path could be left disconnected and the input signal compared to the feedback signal. Plotting the amplitude gain and phase shift for a range of frequencies, it was then graphically apparent whether oscillations would strike.

B. Bode Plot

While Nyquist's method for graphical analysis of stability is a powerful tool, it has one main drawback: amplifier parameters fluctuate. An amplifier could be stable on a cold morning, but unstable when the temperature rose in the afternoon.[47] Some kind of margin of error needed to be built into the stability analysis so that the amplifier could be declared stable for a range of operating conditions. This was the reasoning thatoccurred to Hendrik W. Bode, another Bell Labs researcher, as he was attempting to use feedback to linearize an existing amplifier.[48] The task drove him so batty that he eventually scrapped the whole effort and built a new feedback amplifier that met his original needs. He also left behind several colorful quotes from the effort, including:

> The engineer who embarks upon the design of a feedback amplifier must be a creature of mixed emotions. On the one hand, he can rejoice in the improvements in the characteristics of the structure which feedback promises to secure him. On the other hand, he knows that unless he can finally adjust the phase and attenuation characteristics around the feedback loop so the amplifier will not spontaneously burst into uncontrollable singing, none of these advantages can be actually realized.[49]

Bode's method for dealing with this issue was again a graphical analysis tool. In this case, gain and phase shift are plotted on a logarithmic plot versus frequency; an example of a *Bode plot* is shown in Fig. (19).

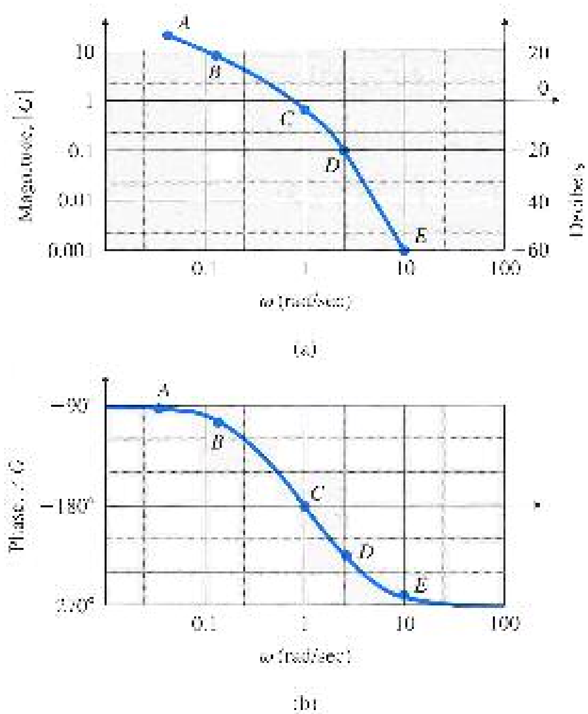


FIG. 19. Example of a Bode plot, showing how close system is to instability.[50]

This plot reveals two important quantities that describe the margin before instability strikes: the *phase margin* and the *gain margin*. For a particular setup, the phase margin is the number of degrees by which the phase change is less than 180° when the gain amplitude equals 1.[51] In Fig. (19), the frequency at which this occurs is just less than 1, so just to the left of point C. The phase margin in this case is around 10 degrees. The gain margin is the factor by which the gain is less than 1 when the phase change equals 180°.[52] For Fig. (19), the gain margin is about 0.7 and can be measured at point C. In general, amplifiers with insufficient gain and phase margins could become unstable if the conditions changed. As a result, amplifiers are now designed to be stable to much higher frequencies than they are actually used, so that they are free of instability concerns. In the case of the amplifier Bode was working with, while it was designed to work at 1 MHz, the amplifier was stable all the way to 30 MHz.[53] Bode and Nyquist plots are both powerful, graphical analysis methods, but a more equation driven method for the analysis of feedback system response exists.

VI. TRANSFER FUNCTIONS

A. Laplace Transforms

Laplace transforms are a fundamental tool for any control engineer. A Laplace transform takes a signal from the time domain into the frequency domain. The definition for the Laplace transform $Y(s)$ for some function $y(t)$, where s is the complex frequency given by $s = \sigma + i\omega$, is:

$$Y(s) = L\{y(t)\} = \int_0^\infty e^{-st} y(t) dt \,. \qquad (35)$$

This technique is useful for controls because of the concept of *impulse response* $h(t)$. The impulse response of a system is the response to an impulse of minimal duration. Taking the Laplace transform of $h(t)$ produces the *transfer function* of the system $H(s)$. The transfer function of a system is a rational polynomial of s and completely describes the response of a system to any input. Readers interested in the proof of this utility should see Appendix C. While the Laplace transform is an important tool, the transfer function of a system is more easily determined from the differential equations governing system behavior. For the rest of this section, a damped harmonic oscillator will be used to demonstrate the power of transfer functions. It may seem surprising that a damped harmonic oscillator can be treated with feedback techniques, but feedback is present. For any object, acceleration determines changes in velocity and position. The feedback appears for a harmonic oscillator because its acceleration is dependent on position and velocity.

B. Transfer Function from Differential Equation

The one-dimensional damped oscillator is a true physics classic, with position dependent restoring force with constant *k* and viscous friction with coefficient *b*. The differential equation is given in Eq. (37):

$$F_{net} = -kx - b\dot{x} = m\ddot{x} \tag{36}$$

$$\ddot{x} + \frac{b}{m}\dot{x} + \frac{k}{m}x = 0. \tag{37}$$

The following change of variables is convenient, and specifying initial conditions the system response is now fully determined:

$$\ddot{x} + 2\gamma\dot{x} + \omega^2 x = 0 \tag{38}$$

$$\dot{x}(0) = 0, x(0) = \alpha. \tag{39}$$

To find the transfer function X(s) for this equation, the following trait of Laplace transforms is crucial (see Appendix C):

$$L\{\dot{x}\} = -x(-0) + sL\{x\}. \tag{40}$$

Using Eq. (40), the transforms to $\dot{x}$ and $\ddot{x}$ become:

$$L\{\dot{x}\} = -\alpha + sX(s) \tag{41}$$

$$L\{\ddot{x}\} = -0 + sL\{\dot{x}\} = -\alpha s + s^2 X(s)\,. \tag{42}$$

Applying the Laplace transform to the entire equation then yields:

$$-\alpha s + s^2 X(s) + 2\gamma\left(-\alpha + sX(s)\right) + \omega^2 X(s) = 0\,. \tag{43}$$

So the transfer function is then:

$$X(s) = \frac{\alpha s + 2\gamma\alpha}{s^2 + 2\gamma s + \omega^2}\,. \tag{44}$$

C. Pole-Zero Placement

The transfer functions seen in control system applications are rational polynomials in the complex variable *s*. This is a result of their derivation from differential equations and the associated sinusoids and exponentials. See Appendix C for more examples and information on transfer functions. In the lingo of controls, when the numerator has a root, it is referred to as a *zero*: the transfer function is 0 at that point.[54] Roots in the denominator are referred to as *poles*: the transfer function spikes at these locations and unstable behavior may occur. As will be shown, the locations of these poles and zeros go a long way to describing the behavior of the system. The poles and zeros for the harmonic oscillator are given by Eq (44), reproduced below.

$$X(s) = \frac{\alpha s + 2\gamma\alpha}{s^2 + 2\gamma s + \omega^2} \tag{44}$$

This function has a zero when α is 0. Clearly, the system has no response if it begins in the stable equilibrium and the transfer function demonstrates this. The poles of the system are more interesting, given by the roots of the denominator. With $\upsilon = \sqrt{\gamma^2 - \omega^2}$, the poles are located at:

$$s = -\gamma \pm \upsilon . \tag{45}$$

To solve for the time domain response of the system, I'll put the transfer function into a form appearing in Table C1 and then correlate the response. The first step is to separate the numerator:

$$X(s) = \alpha\left[\frac{s}{(s+\gamma+\upsilon)(s+\gamma-\upsilon)} + \frac{2\gamma}{(s+\gamma+\upsilon)(s+\gamma-\upsilon)}\right]. \tag{46}$$

Then, making use of the fact that $(-\gamma+\upsilon)-(-\gamma-\upsilon) = 2\upsilon$, and multiplying by $\frac{2\upsilon}{2\upsilon}$, the transfer function can be expressed as:

$$X(s) = \frac{\alpha}{2\upsilon}\left[\frac{2\upsilon s}{(s+\gamma+\upsilon)(s+\gamma-\upsilon)} + 2\gamma\frac{2\upsilon}{(s+\gamma+\upsilon)(s+\gamma-\upsilon)}\right]. \tag{47}$$

Both of these two forms appear in Table C1, so the time-domain response is then:

$$x(t) = \frac{\alpha}{2\upsilon}\left[\left((\gamma+\upsilon)e^{-(\gamma+\upsilon)t} - (\gamma-\upsilon)e^{-(\gamma-\upsilon)t}\right) - 2\gamma\left(e^{-(\gamma+\upsilon)t} - e^{-(\gamma-\upsilon)t}\right)\right]. \tag{48}$$

Pulling out common terms and simplifying, this becomes:

$$x(t) = \frac{\alpha}{2\upsilon}e^{-\gamma t}\left[(\upsilon+\gamma)e^{\upsilon t} + (\upsilon-\gamma)e^{-\upsilon t}\right]. \tag{49}$$

First, you'll notice that $x(0) = \alpha$, so the initial condition has been preserved. To bring the role of poles to light, we'll look at several features of Eq. (49). Fig. (20) displays each of these cases graphically. When $\gamma = 0$, the poles of the system are imaginary. Based on the harmonic oscillator equations, the undamped system should oscillate at $\omega = \sqrt{k/m}$. The transfer function has yielded this solution as well. In general, purely imaginary poles will generate oscillatory responses, as shown in purple in Fig. (20).[55] If instead $\gamma < \omega$, the system has complex poles. As we would expect, the response of this underdamped system is to oscillate inside of a decaying exponential envelope. This response is shown in blue. When $\gamma > \omega$, the poles of the system are real and are shown in red in Fig. (20). Again, the transfer function result agrees with standard harmonic oscillator theory: an overdamped oscillator decays exponentially. The critical damping case of $\gamma = \omega$ is shown in yellow in Fig. (20), but requires a different equation than Eq. (49). If $\upsilon = 0$ in Eq. (46), then using Table C1 the system response is:

$$x(t) = \alpha\left[(1-\gamma t)e^{-\gamma t} + 2\gamma(te^{-\gamma t})\right] = \alpha(1+\gamma t)e^{-\gamma t} . \qquad (50)$$

This is the same solution as the critically damped oscillator. The transfer function yielded the proper result in each case! In dealing with a system with multiple sets of poles, the zeros of a system play a larger role.

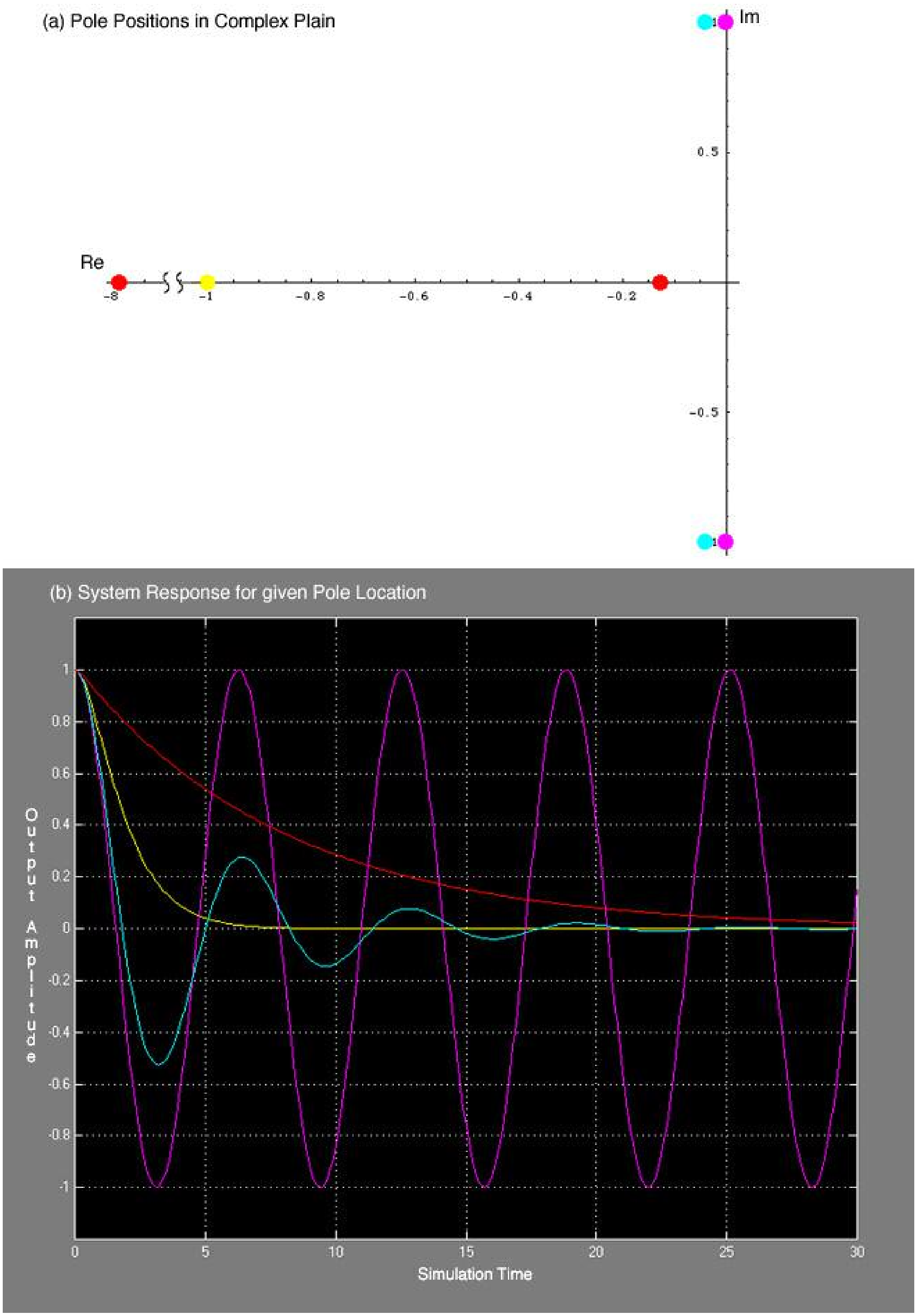


FIG. 20. Graph (a) shows the pole positions corresponding to responses seen in graph (b) for a damped harmonic oscillator.

In a system with multiple sets of poles, each set will produce an exponential that will enter into the total response of the system.[56] The effect of zeros is slightly more complex. The proximity of zeros to the poles of a system controls the relative initial amplitude of the poles.[57] For example, if a zero is placed directly on top of a pole, the affect of that pole is negated; both pole and zero disappear from the transfer function and no longer affect the system response. In practice pole-zero cancellation is trickier than this makes it sound. Fluctuations in system parameters will move either the pole or the zero so they no longer totally overlap. Pole-zero placement, however, is a very important tool for controls and can customize the response of a feedback system. The plant will have some built in poles and zeros, but the controller can be configured to yield any result desired. One way to accomplish this is to use the controller and actuator to adjust the parameters of the plant to maneuver the poles and zeros of the plant. Another way is to setup the controller so that its zeros minimize the poles of the plant and it introduces its own poles. An example of this technique is to use a configurable PID controller.

VII. PID CONTROLLER

As was seen in the cruise control example, steady state error plagues purely proportional control systems. However, if an additional term is added based on the time integral of error, problems with steady state error can be avoided. When the proportional term settles in on a steady state solution, the integral term starts to accumulate the error. This term then modifies the controller output to push the error to zero. At an error of zero, the proportional term no longer has an affect on system response and the error integral term is no longer changing. Control comes from the history of the system stored in the integral, and with no change in system dynamics will maintain zero error.[58] Combining the use of proportional and integral control law is often referred to as PI control. The controller equation, Eq. (14), becomes:

$$u = C_P(r - y) + \frac{1}{C_I}\int_{t_o}^{t}(r - y)dt\,. \qquad (51)$$

Where C_P is the coefficient for the proportional term (previously just C) and C_I is the coefficient for the integral term. While PI control is good for eliminating steady state error, the integral term takes time to change. This can significantly affect the response to an impulse or sudden change in input, slowing the system's response to changes and inducing overshoot. To regain the lost response time, it is necessary to add a term that can predict changes in the error.

Including a derivative of the error in the control law allows the rate of change of error to factor into the equation. For a PD controller, the response time and settling rate will be vastly improved.[59] The controller equation for a PD controller is shown below:

$$u = C_P(r - y) + C_D(\dot{r} - \dot{y})\,. \qquad (52)$$

If the system receives a sudden shock, the error will change quickly. In a proportional control system, little control would be exercised until the error grew in magnitude. With the derivative term, the control system will notice the changing error and immediately take corrective action. Additionally, as the system approaches the steady state solution, the derivative term will act as a sort of damper to potentially eliminate overshoot. As the system output approaches the desired value, the proportional term will continue to drive the system towards equilibrium. However, the derivative term will notice the decreasing error and, if the constants are chosen properly, cause the system to act like a critically damped oscillator. The drawback is that for just PD control the system will not settle exactly on the desired value, still suffering from steady state error.

To get the best of both worlds, controllers incorporating all three terms have been developed, referred to as PID controllers.[60] The PID has constants associated with each term in the control equation, which have to be tuned to meet the application. Fig. (21) shows the block diagram of a standard PID controller. The transfer function for a PID controller is derived in Appendix D and shown in Eq. (53).

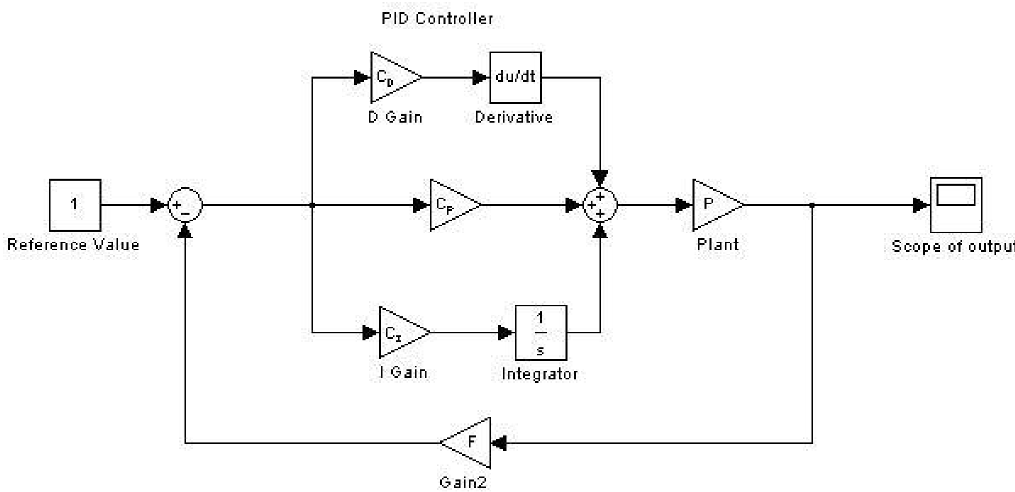


FIG. 21. Block diagram for a PID controller. See Appendix A for explanation of block meaning.

$$Y = \frac{C_I C_D \alpha s^2 + (C_P C_I + \alpha)s + 1}{s\left[C_I C_D s^2 + C_I\left(C_P + \frac{1}{P}\right)s + 1\right]} \tag{53}$$

The poles and zeros of this transfer function are given by Eq. (D11) and (D12). In terms of pole-zero placement, it's clear from the transfer function that the poles and zeros of a PID controlled system can be tailored to produce any response. Particularly, in Fig. (22), the only parameter changed was C_D, and the system response covers the board of possibilities. This feature is what gives the PID controller its utility: for any linear plant, the response of the system is completely customizable.

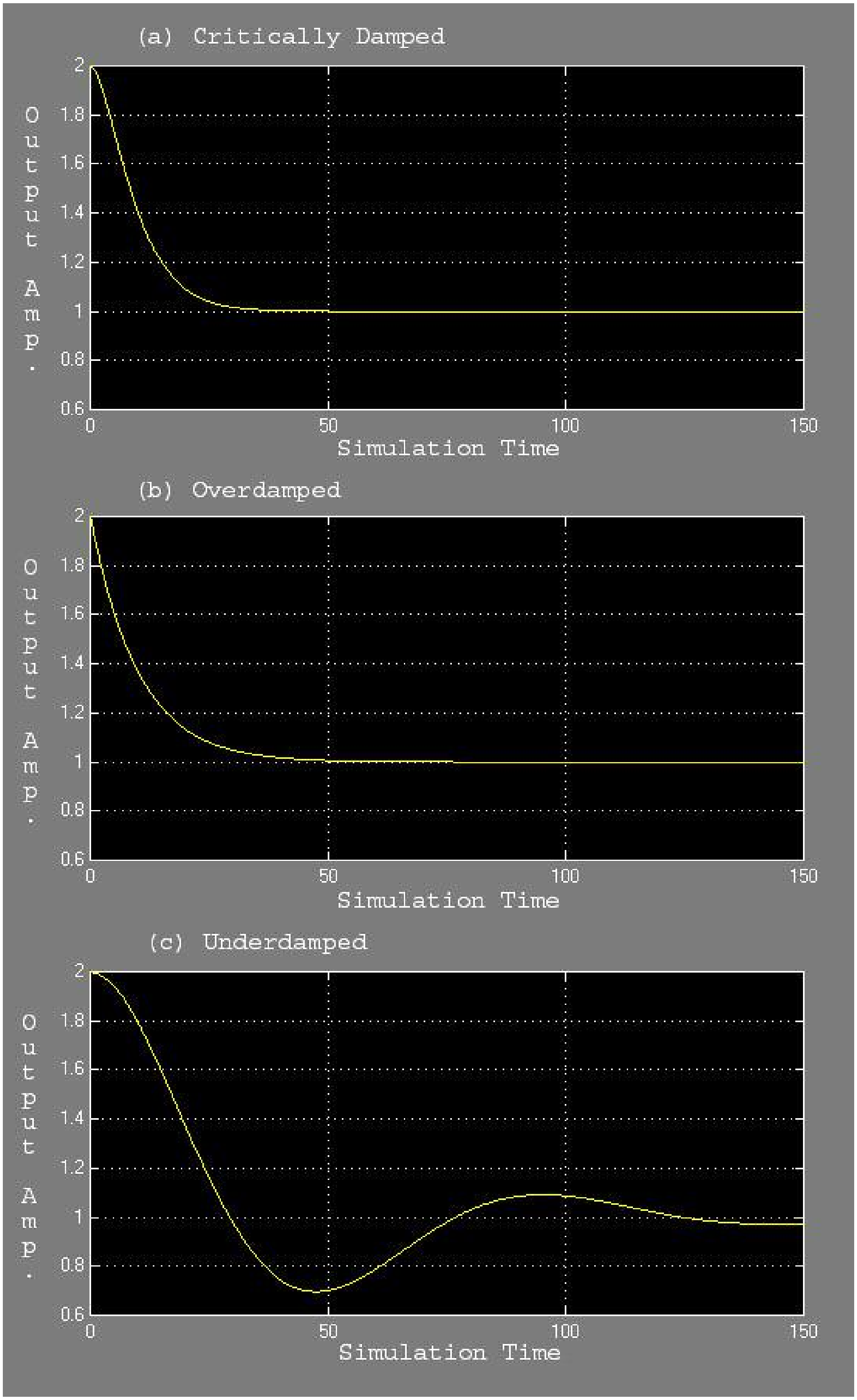


FIG. 22. Comparison of PID controller output for changing C_D.

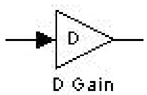
VIII. SUMMARY OF CLASSICAL CONTROLS

The linear control system design techniques discussed in this paper are referred to as classical controls. Automatic feedback control of processes began in antiquity with water clocks. Fluid level control was the only really example of pure feedback control until the development of controlled temperature ovens in the 1600s. Feedback control then experienced a couple big breakthroughs. First, centrifugal governors on steam engines in the early 1800s brought feedback notoriety. The development of the op-amp at Bell Labs in the 1930s brought the mathematical formalism of feedback into focus. Several important tools emerged from the minds of Bell Labs engineers: Harry Nyquist's stability criterion and Hendrik Bode's concepts of gain and phase margin. Completing the classical controls toolkit is the idea of the transfer function. This lends itself to control system analysis and design based on the locations of the poles and zeroes of the system. The culmination of linear feedback control is the PID controller, whose response can be tuned to optimally control any process.

APPENDIX A: BLOCK DIAGRAM SYMBOLS

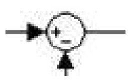
Block diagrams are an important graphical tool for control system engineering. The diagram consists of blocks and arrows connecting the different blocks. Each block represents some process or action in the system, while the arrows indicate the flow of energy or information through the system. All block diagrams were created using Simulink©, a graphical programming add-on to the mathematical simulation package MATLAB©. Pieces of Fig. 23 are reproduced below and each of its blocks explained:

Gain block:



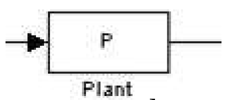
Always triangular. Multiplies input by the coefficient displayed.

Summer:



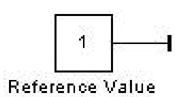
Always circular. Adds inputs with sign shown.

Function:



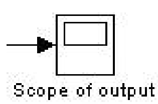
Always rectangular. Used in this paper to represent general processes. Since all processes in this paper are linear, effectively these blocks multiply by coefficient displayed. (Side note: In actual usage, to multiply by P the function would have to be written P*u where u indicates block input. This has been left out for clarity.)

Constant:



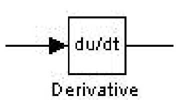
Always square. Constant value input to simulation.

Scope:



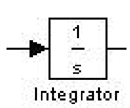
Graphically displays input versus simulation time.

Derivative:



Takes continuous time derivative of input.

Integral:



Takes continuous time integral of input.

APPENDIX B: DERIVATION OF DRAG COEFFICIENT

In the cruise control example, a drag force proportional to speed was used. Viscous drag is a complicated phenomenon; this approximation was chosen for its simplicity since viscous drag is not the focus of the paper. While this is a common approximation in physics, automobile manufacturers release drag information based on a different approximation:

$$F_d = \tfrac{1}{2}\rho A C_D v^2 . \qquad \text{(B1)}$$

In Eq. (B1) ρ is the density of air at STP ($1.3\,\tfrac{kg}{m^3}$),[61] v is the speed of the car, A is the cross-sectional area of the car, and C_D is a unit-less coefficient based on the shaped of the car. This coefficient has a value of 0.34 for a 2004 Porsche 911 GT2.[62] The exact cross-sectional area of the GT2 is not given, so I'll approximate as a rectangle with the given height 1.275 m and width of 1.830 m. At 20 m/s (around 45 mph), the drag force experienced by the GT2 is given by Eq. (B2).

$$F_d = \tfrac{1}{2}(1.3\tfrac{kg}{m^3})(1.275m)(1.830m)(0.34)(20\tfrac{m}{s})^2 = 200N \qquad \text{(B2)}$$

To get the same drag for from my approximation, the value of b is given in Eq. (B3).

$$b = \frac{F_d}{v} = 10\tfrac{Ns}{m} \qquad \text{(B3)}$$

APPENDIX C: UTILITY OF THE TRANSFER FUNCTION

The transfer function of a system is defined as the Laplace transform of the system's impulse response.[63] The impulse response of a system is the behavior the system exhibits after an impulse of minimal duration. For an impulse *p* applied at time σ to system *y* with impulse response *h*, this can be represented by Eq. (C1).

$$y(t) = p \cdot h(t - \sigma) \tag{C1}$$

By Eq. (35), the transfer function $H(s)$ for a system with impulse response $h(t)$ is given by:

$$H(s) = L\{h(t)\} = \int_0^\infty e^{-st} h(t)dt \,. \tag{C2}$$

The goal of this appendix is to show that the transfer function can be used to describe the response of a system to any input. Systems in this paper, and to which the following derivation applies, are both linear and time-invariant. They therefore follow two important rules:[64]

- The *principle of superposition*: "if the system has an input that can be expressed as a sum of signals, then the response of the system can be expressed as the same sum of the individual responses to the respective signals." This allows each input signal to be treated individually, then summed together to get the total system response.
- Second, the response of the system can be described by the convolution of the input and the unit impulse response of the system. Here the Laplace transform comes to the rescue, as convolution in the time domain is just multiplication in the frequency domain.

Based on Fourier series, we know that any input can be expressed using linear combinations of complex exponentials, like e^{st} where $s = \sigma + i\omega$.

$$u(t) = \sum \left(e^{s_1 t} + e^{s_2 t} + \ldots \right) \tag{C3}$$

Applying rule 1, we can then treat each $e^{s_k t}$ separately. Not only that, but each $u(t)$ can be represented by a series of impulses. So, to find the response of a system at time t, we need only sum up the response from this series of impulses. Eq. (C1) then becomes:

$$y(t) = \int_{-\infty}^{t} u(\sigma) h(t-\sigma) d\sigma \qquad \text{(C4)}$$

This is the convolution of impulses and impulse responses, just as rule 2 predicted. As noted, the Laplace transform will allow us to get away from the convolution. Eq. (C4) can be morphed into a Laplace transform by the following tricks. First, change integration variables to $\tau = t - \sigma$.

$$y(t) = \int_{0}^{\infty} u(t-\tau) h(\tau) d\tau \qquad \text{(C5)}$$

Next, applying Eq. (C3) to this, we get:

$$y(t) = \int_{0}^{\infty} h(\tau) e^{s_1(t-\tau)} d\tau + \int_{0}^{\infty} h(\tau) e^{s_2(t-\tau)} d\tau + \ldots \; . \qquad \text{(C6)}$$

Pulling out the e^{st}:

$$y(t) = e^{s_1 t} \int_{0}^{\infty} h(\tau) e^{-s_1 \tau} d\tau + e^{s_2 t} \int_{0}^{\infty} h(\tau) e^{-s_2 \tau} d\tau + \ldots \; . \qquad \text{(C7)}$$

This is just a series of Laplace transformations of $h(\tau)$, so we arrive at:

$$y(t) = H(s_1) e^{s_1 t} + H(s_2) e^{s_2 t} + \ldots \; . \qquad \text{(C8)}$$

The response of the system is given by the sum of complex exponentials weighted by the transfer function of the system. Not only is the transfer function useful in this way, it is possible to determine the system response without even working through integrals. The transfer function of a system can be found just from the differential equations governing the system. This is shown in the paper, and requires the following fact. If we let *X* be the transfer function of a system *x*, then the transform of $\dot{x}$ is given by:

$$L\{\dot{x}\} = \int_0^\infty e^{-st}\dot{x}(t)dt \, . \tag{C9}$$

This looks like a great opportunity to use integration by parts.

$$L\{\dot{x}\} = e^{-st}x(t)\Big|_0^\infty - \int_0^\infty \frac{d}{dt}\left(e^{-st}\right)x(t)dt \tag{C10}$$

Applying the limits and using the definition of a Laplace transform, we get:

$$L\{\dot{x}\} = -x(0) + sX \, . \tag{C11}$$

Eq. (C11) is a very useful result. Finally, Table C1 shows the Laplace transform of a few different functions of time used in this paper. A more complete listing can be found inside the front cover of Franklin's text.[65]

$f(t)$	$F(s)$
1	$1/s$
t	$1/s^2$
e^{-at}	$1/(s+a)$
$(1-at)e^{-at}$	$s/(s+a)^2$
$e^{-at} - e^{-bt}$	$\dfrac{(b-a)}{(s+a)(s+b)}$
$be^{-bt} - ae^{-at}$	$\dfrac{(b-a)s}{(s+a)(s+b)}$

Table C1. Laplace transforms for selected functions.

APPENDIX D: PID CONTROLLER TRANSFER FUNCTION

To develop the transfer function of any system, the easiest place to start is the differential equation governing system dynamics. The equation for the output of a PID controller linear plant *P* is:

$$y = P\left[C_P\left(r - Fy\right) + C_D\left(\dot{r} - F\dot{y}\right) + \frac{1}{C_I}\int\left(r - Fy\right)dt \right]. \qquad \text{(D1)}$$

For simplicity, I'll consider only constant reference values, so $\dot{r} = 0$. Integrating *r* over some time *t* then yields:

$$\frac{y}{P} = C_P\left(r - Fy\right) + C_D\left(F\dot{y}\right) + \frac{1}{C_I}\left(rt - \int Fydt\right). \qquad \text{(D2)}$$

Collecting all *y* terms, this becomes:

$$C_D F\dot{y} + \frac{y}{P} + C_P Fy + \frac{F}{C_I}\int ydt = C_P r + \frac{rt}{C_I}. \qquad \text{(D3)}$$

Applying the Laplace transform to both sides (see Appendix C, Table C1), it begins to take transfer function form:

$$C_D\left(-\alpha + sY\right) + \left(C_P + \frac{1}{PF}\right)Y + \frac{F}{C_I}\left(\frac{Y - \alpha}{s}\right) = \frac{C_P r}{Fs} + \frac{r}{FC_I s^2}. \qquad \text{(D4)}$$

Isolating *Y*, we'll eventually by left with the transfer function.

$$\frac{C_I C_D\left(-\alpha s + s^2 Y\right) + C_I\left(C_P + \frac{1}{PF}\right)sY + FY - F\alpha}{C_I s} = \frac{r\left(C_P C_I s + 1\right)}{FC_I s^2} \qquad \text{(D5)}$$

$$C_I C_D\left(-\alpha s + s^2 Y\right) + C_I\left(C_P + \frac{1}{PF}\right)sY + FY - F\alpha = \frac{r(C_P C_I s + 1)}{Fs} \qquad \text{(D6)}$$

$$Y\left[C_I C_D s^2 + C_I\left(C_P + \frac{1}{PF}\right)s + F\right] = \frac{r(C_P C_I s + 1)}{Fs} + C_I C_D \alpha s + F\alpha \qquad \text{(D7)}$$

$$Y\left[C_I C_D s^2 + C_I\left(C_P + \frac{1}{PF}\right)s + F\right] = \frac{C_I C_D \alpha s^2 + \left(rC_P C_I + F^2\alpha\right)s + r}{Fs} \quad \text{(D8)}$$

Finally, we arrive at the transfer function:

$$Y = \frac{C_I C_D \alpha s^2 + \left(rC_P C_I + F^2\alpha\right)s + r}{Fs\left[C_I C_D s^2 + C_I\left(C_P + \frac{1}{PF}\right)s + F\right]}. \quad \text{(D9)}$$

Let's say that $F = 1$, so y is supposed to be equal to r. For simplicity, I'll also set r = 1.

We now have:

$$Y = \frac{C_I C_D \alpha s^2 + \left(C_P C_I + \alpha\right)s + 1}{s\left[C_I C_D s^2 + C_I\left(C_P + \frac{1}{P}\right)s + 1\right]}. \quad \text{(D10)}$$

This function has the following poles and zeros:

Zeros:

$$s = \frac{-C_P C_I - \alpha \pm \sqrt{\left(C_P C_I + \alpha\right)^2 - 4C_I C_D \alpha}}{2C_I C_D \alpha} \quad \text{(D11)}$$

Poles:

$$s = 0, \frac{-C_I\left(C_P + \frac{1}{P}\right) \pm \sqrt{\left(C_I\left(C_P + \frac{1}{P}\right)\right)^2 - 4C_I C_D}}{2C_I C_D}. \quad \text{(D12)}$$

ANNOTATED BIBLIOGRAPHY:

1) "911 GT2", available online at *http://www.porsche.be/911/gt2/default.htm*. (accessed April 4 2004).

Technical data sheet for the 2004 Porsche 911 GT2, which I used to calculate the coefficient of viscous friction for the simplified cruise control example.

2) H. W. Bode, *Network analysis and feedback amplifier design*. (D. Van Nostrand company, inc., New York, 1945).

The original text on feedback control for electronic applications, it was written from handouts created by Bode for "afterhours classes" on feedback control at Bell Labs. This book was also the first exposition on frequency response methods for feedback control.

3) M. E. El-Hawary, *Control System Engineering*. (Reston Publishing, Reston, VA, 1984).

A very similar book to Franklin's Feedback Control of Dynamic Systems*, this takes an engineering textbook approach to classical controls. It also contains more of the mathematical background needed for dealing with systems as matrices.*

4) G. F. Franklin, J. D. Powell and A. Emami-Naeini, *Feedback Control of Dynamic Systems*, Third ed. (Addison-Wesley Publishing, Reading, Mass., 1994).

A classic textbook on classical feedback control methods. This book was the main source for my understanding of Laplace transforms and transfer functions. Also used for it's nice and clear figures.

5) G. F. Franklin, J. D. Powell and M. L. Workman, *Digital Control of Dynamic Systems*, Third ed. (Addison-Wesley Publishing, Reading, Mass., 1997).

Feedback control is now often done using digital technology and computers. This text updates many of the methods from Feedback Control of Dynamic Systems *for digital applications.*

6) P. Horowitz and W. Hill, *The Art of Electronics, Second Edition*, 2nd ed. (Cambridge University Press, Cambridge, 1989).

This book is a tremendous resource for all things electronic. For this paper, it was the source for much of the technical detail of transistor operational amplifiers.

7) "Bell Labs History", available online at http://www.bell-labs.com/about/history/. (accessed March 15 2004).
The Bell Labs history website contains some history of the lab, but surprisingly little on operational amplifiers

8) E. B. Lee and L. Markus, *Foundations of Optimal Control Theory*. (Robert E. Krieger Publishing, Malabar, FL, 1986).

While not actually discussed in this paper, this book looks at how to apply the calculus of variations to controls. These methods allow an engineer to design a control system with, for example, an optimized settling time.

9) J. R. Leigh, *Control theory: a guided tour*. (P. Peregrinus LTD, London, 1992).

A clear introduction into the theory behind feedback control, this was the first book I turned to. It contains great explanations of the ideas behind feedback control without completely shying away from equations. It's a great book for anyone looking to get started with feedback control.

10) O. Mayr, *The origins of feedback control*. (M.I.T. Press, Cambridge, 1970).

This book was my main source for the history of feedback control. Originally written in German, the author also translated his book into English.

11) O. Mayr, *Feedback mechanisms in the historical collections of the National Museum of History and Technology*. (Smithsonian Institution Press, Washington, 1971).

While it contains plenty of descriptions, I mostly used mainly for the pictures. It contains lots of pictures of governors, including the one of the front cover of this paper.

12) D. A. Mindell, *Between Human and Machine: Feedback, Control, and Computing before Cybernetics*. (Johns Hopkins University Press, Baltimore, 2002).

I found this book to be a great source for the story surrounding the development of the operational amplifier. It also contains some information not presented in this paper on mechanical feedback systems from between the steam engine governor and WWII.

13) R. Murphy, *Introduction to AI Robotics.* (MIT Press, Cambridge, Mass., 2000).

Although not specifically addressed in this paper, this is a far-out and extremely cool application for control system ideas.

14) R. L. Reese, *University Physics*. (Brooks/Cole Publishing, 2000).

A classic introduction to physics, I used Reese for some of the basic physical principles behind feedback control applications.

15) C. E. Rohrs, J. L. Melsa and D. G. Schultz, *Linear Control Systems*. (McGraw-Hill, Inc., New York, 1993).

Another classical controls textbook, this one has particular emphasis on system modeling.

---

[1] Image from O. Mayr, *Feedback mechanisms in the historical collections of the National Museum of History and Technology*. (Smithsonian Institution Press, Washington, 1971), pp. 20.
[2] P. Horowitz and W. Hill, *The Art of Electronics, Second Edition*, 2nd ed. (Cambridge University Press, Cambridge, 1989), pp. 196-208.
[3] P. Horowitz and W. Hill, *The Art of Electronics, Second Edition*, 2nd ed. (Cambridge University Press, Cambridge, 1989), pp. 233.
[4] J. R. Leigh, *Control theory: a guided tour*. (P. Peregrinus LTD, London, 1992), pp. 19.
[5] O. Mayr, *The origins of feedback control*. (M.I.T. Press, Cambridge, 1970), pp. 11-16.
[6] O. Mayr, *The origins of feedback control*. (M.I.T. Press, Cambridge, 1970), pp. 56-65.
[7] O. Mayr, *The origins of feedback control*. (M.I.T. Press, Cambridge, 1970), pp. 90-108.
[8] R. L. Reese, *University Physics*. (Brooks/Cole Publishing, 2000), pp. 496-501.
[9] R. L. Reese, *University Physics*. (Brooks/Cole Publishing, 2000), pp. 511-516.
[10] C. Ritz, “Booming Sands”, senior comprehensive exercise paper, (Carleton College, 2004). Included in this volume.
[11] O. Mayr, *The origins of feedback control*. (M.I.T. Press, Cambridge, 1970), pp. 11-16.
[12] Figure from G. F. Franklin, J. D. Powell and A. Emami-Naeini, Feedback Control of Dynamic Systems, Third ed. (Addison-Wesley Publishing, Reading, Mass., 1994), pp. 8.
[13] O. Mayr, *The origins of feedback control*. (M.I.T. Press, Cambridge, 1970), pp. 16-40.
[14] O. Mayr, *The origins of feedback control*. (M.I.T. Press, Cambridge, 1970), pp. 64.
[15] O. Mayr, *The origins of feedback control*. (M.I.T. Press, Cambridge, 1970), pp. 59.
[16] Figure from G. F. Franklin, J. D. Powell and A. Emami-Naeini, Feedback Control of Dynamic Systems, Third ed. (Addison-Wesley Publishing, Reading, Mass., 1994), pp. 8.
[17] O. Mayr, *The origins of feedback control*. (M.I.T. Press, Cambridge, 1970), pp. 90-108.
[18] O. Mayr, *The origins of feedback control*. (M.I.T. Press, Cambridge, 1970), pp. 100-107.
[19] G. F. Franklin, J. D. Powell and A. Emami-Naeini, *Feedback Control of Dynamic Systems*, Third ed. (Addison-Wesley Publishing, Reading, Mass., 1994), pp. 9.
[20] O. Mayr, *The origins of feedback control*. (M.I.T. Press, Cambridge, 1970), pp. 109.
[21] Figure from G. F. Franklin, J. D. Powell and A. Emami-Naeini, *Feedback Control of Dynamic Systems*, Third ed. (Addison-Wesley Publishing, Reading, Mass., 1994), pp. 10.
[22] O. Mayr, *The origins of feedback control*. (M.I.T. Press, Cambridge, 1970), pp. 113.
[23] G. F. Franklin, J. D. Powell and A. Emami-Naeini, *Feedback Control of Dynamic Systems*, Third ed. (Addison-Wesley Publishing, Reading, Mass., 1994), pp. 10-11.
[24] G. F. Franklin, J. D. Powell and A. Emami-Naeini, *Feedback Control of Dynamic Systems*, Third ed. (Addison-Wesley Publishing, Reading, Mass., 1994), pp. 11.

[25] G. F. Franklin, J. D. Powell and A. Emami-Naeini, *Feedback Control of Dynamic Systems*, Third ed. (Addison-Wesley Publishing, Reading, Mass., 1994), pp. 11.
[26] G. F. Franklin, J. D. Powell and A. Emami-Naeini, *Feedback Control of Dynamic Systems*, Third ed. (Addison-Wesley Publishing, Reading, Mass., 1994), pp. 11.
[27] G. F. Franklin, J. D. Powell and A. Emami-Naeini, *Feedback Control of Dynamic Systems*, Third ed. (Addison-Wesley Publishing, Reading, Mass., 1994), pp. 11.
[28] G. F. Franklin, J. D. Powell and A. Emami-Naeini, *Feedback Control of Dynamic Systems*, Third ed. (Addison-Wesley Publishing, Reading, Mass., 1994), pp. 215-223.
[29] G. F. Franklin, J. D. Powell and A. Emami-Naeini, *Feedback Control of Dynamic Systems*, Third ed. (Addison-Wesley Publishing, Reading, Mass., 1994), pp. 11.
[30] D. A. Mindell, *Between Human and Machine: Feedback, Control, and Computing before Cybernetics*. (Johns Hopkins University Press, Baltimore, 2002), pp. 112.
[31] D. A. Mindell, *Between Human and Machine: Feedback, Control, and Computing before Cybernetics*. (Johns Hopkins University Press, Baltimore, 2002), pp. 113
[32] D. A. Mindell, *Between Human and Machine: Feedback, Control, and Computing before Cybernetics*. (Johns Hopkins University Press, Baltimore, 2002). 114-116.
[33] D. A. Mindell, *Between Human and Machine: Feedback, Control, and Computing before Cybernetics*. (Johns Hopkins University Press, Baltimore, 2002), pp. 128.
[34] D. A. Mindell, *Between Human and Machine: Feedback, Control, and Computing before Cybernetics*. (Johns Hopkins University Press, Baltimore, 2002), pp. 117.
[35] D. A. Mindell, *Between Human and Machine: Feedback, Control, and Computing before Cybernetics*. (Johns Hopkins University Press, Baltimore, 2002), pp. 118.
[36] D. A. Mindell, *Between Human and Machine: Feedback, Control, and Computing before Cybernetics*. (Johns Hopkins University Press, Baltimore, 2002), pp. 118.
[37] P. Horowitz and W. Hill, *The Art of Electronics, Second Edition*, 2nd ed. (Cambridge University Press, Cambridge, 1989), pp. 196-208.
[38] Figure from P. Horowitz and W. Hill, *The Art of Electronics, Second Edition*, 2nd ed. (Cambridge University Press, Cambridge, 1989), pp. 193.
[39] P. Horowitz and W. Hill, *The Art of Electronics, Second Edition*, 2nd ed. (Cambridge University Press, Cambridge, 1989), pp. 233.
[40] P. Horowitz and W. Hill, *The Art of Electronics, Second Edition*, 2nd ed. (Cambridge University Press, Cambridge, 1989), pp. 234.
[41] D. A. Mindell, *Between Human and Machine: Feedback, Control, and Computing before Cybernetics*. (Johns Hopkins University Press, Baltimore, 2002), pp. 118-119.
[42] P. Horowitz and W. Hill, *The Art of Electronics, Second Edition*, 2nd ed. (Cambridge University Press, Cambridge, 1989), pp. 32.
[43] G. F. Franklin, J. D. Powell and A. Emami-Naeini, *Feedback Control of Dynamic Systems*, Third ed. (Addison-Wesley Publishing, Reading, Mass., 1994), pp. 340.
[44] D. A. Mindell, *Between Human and Machine: Feedback, Control, and Computing before Cybernetics*. (Johns Hopkins University Press, Baltimore, 2002), pp. 125.
[45] D. A. Mindell, *Between Human and Machine: Feedback, Control, and Computing before Cybernetics*. (Johns Hopkins University Press, Baltimore, 2002), pp. 126.
[46] Figure adapted from G. F. Franklin, J. D. Powell and A. Emami-Naeini, *Feedback Control of Dynamic Systems*, Third ed. (Addison-Wesley Publishing, Reading, Mass., 1994), pp. 364.
[47] D. A. Mindell, *Between Human and Machine: Feedback, Control, and Computing before Cybernetics*. (Johns Hopkins University Press, Baltimore, 2002), pp. 121.
[48] D. A. Mindell, *Between Human and Machine: Feedback, Control, and Computing before Cybernetics*. (Johns Hopkins University Press, Baltimore, 2002), pp. 127-129.
[49] D. A. Mindell, *Between Human and Machine: Feedback, Control, and Computing before Cybernetics*. (Johns Hopkins University Press, Baltimore, 2002), pp. 129.
[50] Figure from G. F. Franklin, J. D. Powell and A. Emami-Naeini, *Feedback Control of Dynamic Systems*, Third ed. (Addison-Wesley Publishing, Reading, Mass., 1994), pp. 369.
[51] G. F. Franklin, J. D. Powell and A. Emami-Naeini, *Feedback Control of Dynamic Systems*, Third ed. (Addison-Wesley Publishing, Reading, Mass., 1994), pp. 375.

[52] G. F. Franklin, J. D. Powell and A. Emami-Naeini, *Feedback Control of Dynamic Systems*, Third ed. (Addison-Wesley Publishing, Reading, Mass., 1994), pp. 375.
[53] D. A. Mindell, *Between Human and Machine: Feedback, Control, and Computing before Cybernetics*. (Johns Hopkins University Press, Baltimore, 2002), pp. 130.
[54] G. F. Franklin, J. D. Powell and A. Emami-Naeini, *Feedback Control of Dynamic Systems*, Third ed. (Addison-Wesley Publishing, Reading, Mass., 1994), pp. 118-120.
[55] G. F. Franklin, J. D. Powell and A. Emami-Naeini, *Feedback Control of Dynamic Systems*, Third ed. (Addison-Wesley Publishing, Reading, Mass., 1994), pp. 118-126.
[56] G. F. Franklin, J. D. Powell and A. Emami-Naeini, *Feedback Control of Dynamic Systems*, Third ed. (Addison-Wesley Publishing, Reading, Mass., 1994), pp. 131-138.
[57] G. F. Franklin, J. D. Powell and A. Emami-Naeini, *Feedback Control of Dynamic Systems*, Third ed. (Addison-Wesley Publishing, Reading, Mass., 1994), pp. 131-138.
[58] G. F. Franklin, J. D. Powell and A. Emami-Naeini, *Feedback Control of Dynamic Systems*, Third ed. (Addison-Wesley Publishing, Reading, Mass., 1994), pp. 181-183.
[59] G. F. Franklin, J. D. Powell and A. Emami-Naeini, *Feedback Control of Dynamic Systems*, Third ed. (Addison-Wesley Publishing, Reading, Mass., 1994), pp. 183-185.
[60] G. F. Franklin, J. D. Powell and A. Emami-Naeini, *Feedback Control of Dynamic Systems*, Third ed. (Addison-Wesley Publishing, Reading, Mass., 1994), pp. 185-187.
[61] R. L. Reese, *University Physics*. (Brooks/Cole Publishing, 2000), pp. 12.
[62] "911 GT2", *available online at http://www.porsche.be/911/gt2/default.htm*.
[63] G. F. Franklin, J. D. Powell and A. Emami-Naeini, *Feedback Control of Dynamic Systems*, Third ed. (Addison-Wesley Publishing, Reading, Mass., 1994), pp. 89-92.
[64] G. F. Franklin, J. D. Powell and A. Emami-Naeini, *Feedback Control of Dynamic Systems*, Third ed. (Addison-Wesley Publishing, Reading, Mass., 1994), pp. 86.
[65] G. F. Franklin, J. D. Powell and A. Emami-Naeini, *Feedback Control of Dynamic Systems*, Third ed. (Addison-Wesley Publishing, Reading, Mass., 1994), inside front cover.