
Chapter One

Introduction

Feedback is a central feature of life. The process of feedback governs how we grow, respond to stress and challenge, and regulate factors such as body temperature, blood pressure and cholesterol level. The mechanisms operate at every level, from the interaction of proteins in cells to the interaction of organisms in complex ecologies.

M. B. Hoagland and B. Dodson, *The Way Life Works*, 1995 [HD95].

In this chapter we provide an introduction to the basic concept of *feedback* and the related engineering discipline of *control*. We focus on both historical and current examples, with the intention of providing the context for current tools in feedback and control. Much of the material in this chapter is adapted from [Mur03], and the authors gratefully acknowledge the contributions of Roger Brockett and Gunter Stein to portions of this chapter.

1.1 What Is Feedback?

A *dynamical system* is a system whose behavior changes over time, often in response to external stimulation or forcing. The term *feedback* refers to a situation in which two (or more) dynamical systems are connected together such that each system influences the other and their dynamics are thus strongly coupled. Simple causal reasoning about a feedback system is difficult because the first system influences the second and the second system influences the first, leading to a circular argument. This makes reasoning based on cause and effect tricky, and it is necessary to analyze the system as a whole. A consequence of this is that the behavior of feedback systems is often counterintuitive, and it is therefore necessary to resort to formal methods to understand them.

Figure 1.1 illustrates in block diagram form the idea of feedback. We often use

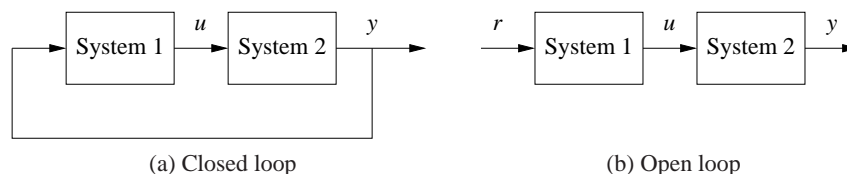


Figure 1.1: Open and closed loop systems. (a) The output of system 1 is used as the input of system 2, and the output of system 2 becomes the input of system 1, creating a closed loop system. (b) The interconnection between system 2 and system 1 is removed, and the system is said to be open loop.

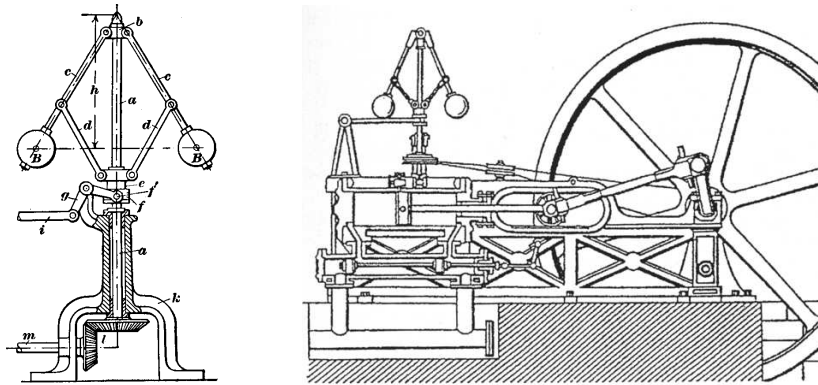


Figure 1.2: The centrifugal governor and the steam engine. The centrifugal governor on the left consists of a set of flyballs that spread apart as the speed of the engine increases. The steam engine on the right uses a centrifugal governor (above and to the left of the flywheel) to regulate its speed. (Credit: Machine a Vapeur Horizontale de Philip Taylor [1828].)

the terms *open loop* and *closed loop* when referring to such systems. A system is said to be a closed loop system if the systems are interconnected in a cycle, as shown in Figure 1.1a. If we break the interconnection, we refer to the configuration as an open loop system, as shown in Figure 1.1b.

As the quote at the beginning of this chapter illustrates, a major source of examples of feedback systems is biology. Biological systems make use of feedback in an extraordinary number of ways, on scales ranging from molecules to cells to organisms to ecosystems. One example is the regulation of glucose in the bloodstream through the production of insulin and glucagon by the pancreas. The body attempts to maintain a constant concentration of glucose, which is used by the body's cells to produce energy. When glucose levels rise (after eating a meal, for example), the hormone insulin is released and causes the body to store excess glucose in the liver. When glucose levels are low, the pancreas secretes the hormone glucagon, which has the opposite effect. Referring to Figure 1.1, we can view the liver as system 1 and the pancreas as system 2. The output from the liver is the glucose concentration in the blood, and the output from the pancreas is the amount of insulin or glucagon produced. The interplay between insulin and glucagon secretions throughout the day helps to keep the blood-glucose concentration constant, at about 90 mg per 100 mL of blood.

An early engineering example of a feedback system is a centrifugal governor, in which the shaft of a steam engine is connected to a flyball mechanism that is itself connected to the throttle of the steam engine, as illustrated in Figure 1.2. The system is designed so that as the speed of the engine increases (perhaps because of a lessening of the load on the engine), the flyballs spread apart and a linkage causes the throttle on the steam engine to be closed. This in turn slows down the engine, which causes the flyballs to come back together. We can model this system as a closed loop system by taking system 1 as the steam engine and system 2 as the governor.

When properly designed, the flyball governor maintains a constant speed of the engine, roughly independent of the loading conditions. The centrifugal governor was an enabler of the successful Watt steam engine, which fueled the industrial revolution.

Feedback has many interesting properties that can be exploited in designing systems. As in the case of glucose regulation or the flyball governor, feedback can make a system resilient toward external influences. It can also be used to create linear behavior out of nonlinear components, a common approach in electronics. More generally, feedback allows a system to be insensitive both to external disturbances and to variations in its individual elements.

Feedback has potential disadvantages as well. It can create dynamic instabilities in a system, causing oscillations or even runaway behavior. Another drawback, especially in engineering systems, is that feedback can introduce unwanted sensor noise into the system, requiring careful filtering of signals. It is for these reasons that a substantial portion of the study of feedback systems is devoted to developing an understanding of dynamics and a mastery of techniques in dynamical systems.

Feedback systems are ubiquitous in both natural and engineered systems. Control systems maintain the environment, lighting and power in our buildings and factories; they regulate the operation of our cars, consumer electronics and manufacturing processes; they enable our transportation and communications systems; and they are critical elements in our military and space systems. For the most part they are hidden from view, buried within the code of embedded microprocessors, executing their functions accurately and reliably. Feedback has also made it possible to increase dramatically the precision of instruments such as atomic force microscopes (AFMs) and telescopes.

In nature, homeostasis in biological systems maintains thermal, chemical and biological conditions through feedback. At the other end of the size scale, global climate dynamics depend on the feedback interactions between the atmosphere, the oceans, the land and the sun. Ecosystems are filled with examples of feedback due to the complex interactions between animal and plant life. Even the dynamics of economies are based on the feedback between individuals and corporations through markets and the exchange of goods and services.

1.2 What Is Control?

The term *control* has many meanings and often varies between communities. In this book, we define control to be the use of algorithms and feedback in engineered systems. Thus, control includes such examples as feedback loops in electronic amplifiers, setpoint controllers in chemical and materials processing, “fly-by-wire” systems on aircraft and even router protocols that control traffic flow on the Internet. Emerging applications include high-confidence software systems, autonomous vehicles and robots, real-time resource management systems and biologically engineered systems. At its core, control is an *information science* and includes the use of information in both analog and digital representations.

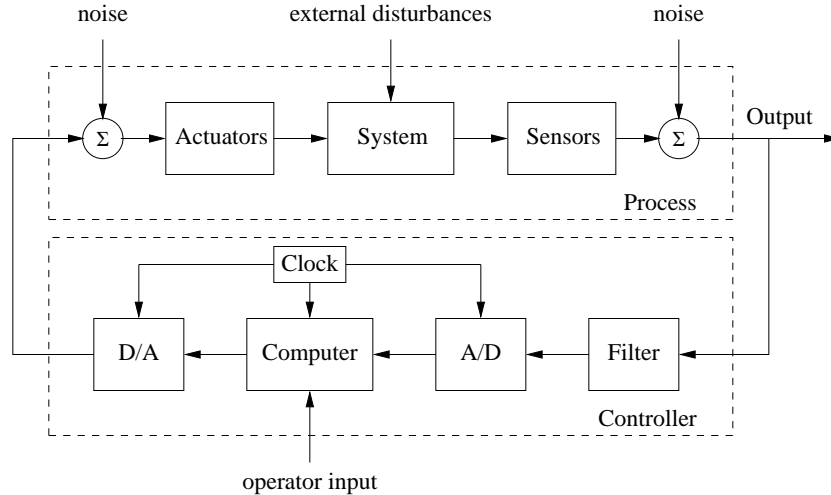


Figure 1.3: Components of a computer-controlled system. The upper dashed box represents the process dynamics, which include the sensors and actuators in addition to the dynamical system being controlled. Noise and external disturbances can perturb the dynamics of the process. The controller is shown in the lower dashed box. It consists of a filter and analog-to-digital (A/D) and digital-to-analog (D/A) converters, as well as a computer that implements the control algorithm. A system clock controls the operation of the controller, synchronizing the A/D, D/A and computing processes. The operator input is also fed to the computer as an external input.

A modern controller senses the operation of a system, compares it against the desired behavior, computes corrective actions based on a model of the system's response to external inputs and actuates the system to effect the desired change. This basic *feedback loop* of sensing, computation and actuation is the central concept in control. The key issues in designing control logic are ensuring that the dynamics of the closed loop system are stable (bounded disturbances give bounded errors) and that they have additional desired behavior (good disturbance attenuation, fast responsiveness to changes in operating point, etc). These properties are established using a variety of modeling and analysis techniques that capture the essential dynamics of the system and permit the exploration of possible behaviors in the presence of uncertainty, noise and component failure.

A typical example of a control system is shown in Figure 1.3. The basic elements of sensing, computation and actuation are clearly seen. In modern control systems, computation is typically implemented on a digital computer, requiring the use of analog-to-digital (A/D) and digital-to-analog (D/A) converters. Uncertainty enters the system through noise in sensing and actuation subsystems, external disturbances that affect the underlying system operation and uncertain dynamics in the system (parameter errors, unmodeled effects, etc). The algorithm that computes the control action as a function of the sensor values is often called a *control law*. The system can be influenced externally by an operator who introduces *command signals* to the system.

Control engineering relies on and shares tools from physics (dynamics and modeling), computer science (information and software) and operations research (optimization, probability theory and game theory), but it is also different from these subjects in both insights and approach.

Perhaps the strongest area of overlap between control and other disciplines is in the modeling of physical systems, which is common across all areas of engineering and science. One of the fundamental differences between control-oriented modeling and modeling in other disciplines is the way in which interactions between subsystems are represented. Control relies on a type of input/output modeling that allows many new insights into the behavior of systems, such as disturbance attenuation and stable interconnection. Model reduction, where a simpler (lower-fidelity) description of the dynamics is derived from a high-fidelity model, is also naturally described in an input/output framework. Perhaps most importantly, modeling in a control context allows the design of *robust* interconnections between subsystems, a feature that is crucial in the operation of all large engineered systems.

Control is also closely associated with computer science since virtually all modern control algorithms for engineering systems are implemented in software. However, control algorithms and software can be very different from traditional computer software because of the central role of the dynamics of the system and the real-time nature of the implementation.

1.3 Feedback Examples

Feedback has many interesting and useful properties. It makes it possible to design precise systems from imprecise components and to make relevant quantities in a system change in a prescribed fashion. An unstable system can be stabilized using feedback, and the effects of external disturbances can be reduced. Feedback also offers new degrees of freedom to a designer by exploiting sensing, actuation and computation. In this section we survey some of the important applications and trends for feedback in the world around us.

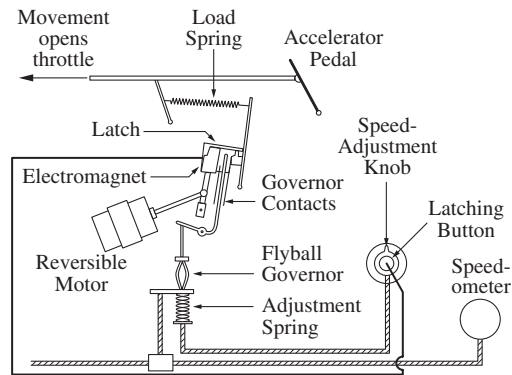
Early Technological Examples

The proliferation of control in engineered systems occurred primarily in the latter half of the 20th century. There are some important exceptions, such as the centrifugal governor described earlier and the thermostat (Figure 1.4a), designed at the turn of the century to regulate the temperature of buildings.

The thermostat, in particular, is a simple example of feedback control that everyone is familiar with. The device measures the temperature in a building, compares that temperature to a desired setpoint and uses the *feedback error* between the two to operate the heating plant, e.g., to turn heat on when the temperature is too low and to turn it off when the temperature is too high. This explanation captures the essence of feedback, but it is a bit too simple even for a basic device such as the thermostat. Because lags and delays exist in the heating plant and sensor, a good



(a) Honeywell thermostat, 1953



(b) Chrysler cruise control, 1958

Figure 1.4: Early control devices. (a) Honeywell T87 thermostat originally introduced in 1953. The thermostat controls whether a heater is turned on by comparing the current temperature in a room to a desired value that is set using a dial. (b) Chrysler cruise control system introduced in the 1958 Chrysler Imperial [Row58]. A centrifugal governor is used to detect the speed of the vehicle and actuate the throttle. The reference speed is specified through an adjustment spring. (Left figure courtesy of Honeywell International, Inc.)

thermostat does a bit of anticipation, turning the heater off before the error actually changes sign. This avoids excessive temperature swings and cycling of the heating plant. This interplay between the dynamics of the process and the operation of the controller is a key element in modern control systems design.

There are many other control system examples that have developed over the years with progressively increasing levels of sophistication. An early system with broad public exposure was the *cruise control* option introduced on automobiles in 1958 (see Figure 1.4b). Cruise control illustrates the dynamic behavior of closed loop feedback systems in action—the slowdown error as the system climbs a grade, the gradual reduction of that error due to integral action in the controller, the small overshoot at the top of the climb, etc. Later control systems on automobiles such as emission controls and fuel-metering systems have achieved major reductions of pollutants and increases in fuel economy.

Power Generation and Transmission

Access to electrical power has been one of the major drivers of technological progress in modern society. Much of the early development of control was driven by the generation and distribution of electrical power. Control is mission critical for power systems, and there are many control loops in individual power stations. Control is also important for the operation of the whole power network since it is difficult to store energy and it is thus necessary to match production to consumption. Power management is a straightforward regulation problem for a system with one generator and one power consumer, but it is more difficult in a highly distributed system with many generators and long distances between consumption and generation. Power demand can change rapidly in an unpredictable manner and

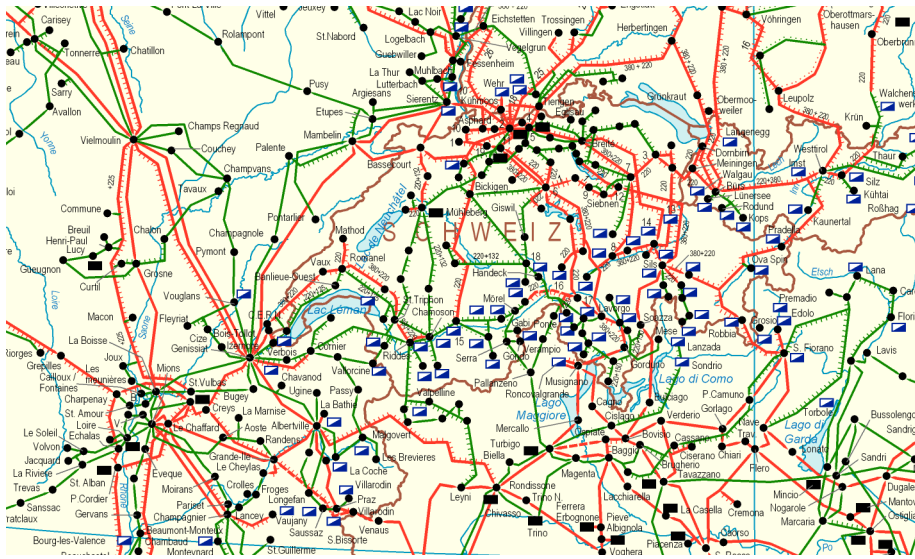


Figure 1.5: A small portion of the European power network. By 2008 European power suppliers will operate a single interconnected network covering a region from the Arctic to the Mediterranean and from the Atlantic to the Urals. In 2004 the installed power was more than 700 GW (7×10^{11} W). (Source: UCTE [www.ucte.org])

combining generators and consumers into large networks makes it possible to share loads among many suppliers and to average consumption among many customers. Large transcontinental and transnational power systems have therefore been built, such as the one shown in Figure 1.5.

Most electricity is distributed by alternating current (AC) because the transmission voltage can be changed with small power losses using transformers. Alternating current generators can deliver power only if the generators are synchronized to the voltage variations in the network. This means that the rotors of all generators in a network must be synchronized. To achieve this with local decentralized controllers and a small amount of interaction is a challenging problem. Sporadic low-frequency oscillations between distant regions have been observed when regional power grids have been interconnected [KW05].

Safety and reliability are major concerns in power systems. There may be disturbances due to trees falling down on power lines, lightning or equipment failures. There are sophisticated control systems that attempt to keep the system operating even when there are large disturbances. The control actions can be to reduce voltage, to break up the net into subnets or to switch off lines and power users. These safety systems are an essential element of power distribution systems, but in spite of all precautions there are occasionally failures in large power systems. The power system is thus a nice example of a complicated distributed system where control is executed on many levels and in many different ways.



(a) F/A-18 “Hornet”



(b) X-45 UCAV

Figure 1.6: Military aerospace systems. (a) The F/A-18 aircraft is one of the first production military fighters to use “fly-by-wire” technology. (b) The X-45 (UCAV) unmanned aerial vehicle is capable of autonomous flight, using inertial measurement sensors and the global positioning system (GPS) to monitor its position relative to a desired trajectory. (Photographs courtesy of NASA Dryden Flight Research Center.)

Aerospace and Transportation

In aerospace, control has been a key technological capability tracing back to the beginning of the 20th century. Indeed, the Wright brothers are correctly famous not for demonstrating simply powered flight but *controlled* powered flight. Their early Wright Flyer incorporated moving control surfaces (vertical fins and canards) and warpable wings that allowed the pilot to regulate the aircraft’s flight. In fact, the aircraft itself was not stable, so continuous pilot corrections were mandatory. This early example of controlled flight was followed by a fascinating success story of continuous improvements in flight control technology, culminating in the high-performance, highly reliable automatic flight control systems we see in modern commercial and military aircraft today (Figure 1.6).

Similar success stories for control technology have occurred in many other application areas. Early World War II bombsights and fire control servo systems have evolved into today’s highly accurate radar-guided guns and precision-guided weapons. Early failure-prone space missions have evolved into routine launch operations, manned landings on the moon, permanently manned space stations, robotic vehicles roving Mars, orbiting vehicles at the outer planets and a host of commercial and military satellites serving various surveillance, communication, navigation and earth observation needs. Cars have advanced from manually tuned mechanical/pneumatic technology to computer-controlled operation of all major functions, including fuel injection, emission control, cruise control, braking and cabin comfort.

Current research in aerospace and transportation systems is investigating the application of feedback to higher levels of decision making, including logical regulation of operating modes, vehicle configurations, payload configurations and health status. These have historically been performed by human operators, but today that

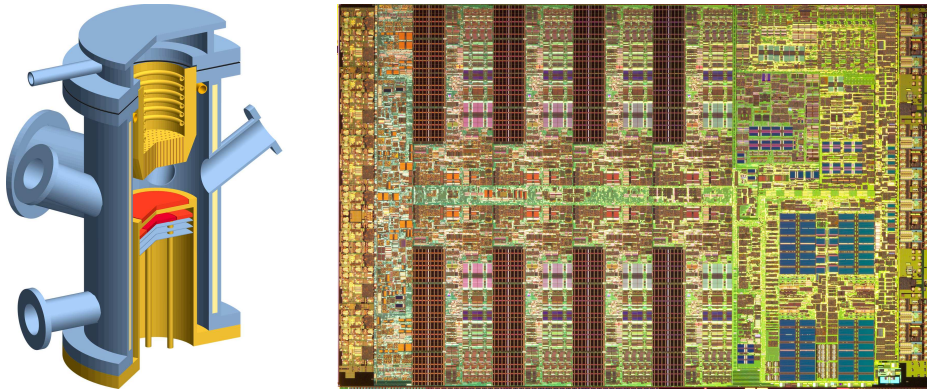


Figure 1.7: Materials processing. Modern materials are processed under carefully controlled conditions, using reactors such as the metal organic chemical vapor deposition (MOCVD) reactor shown on the left, which was for manufacturing superconducting thin films. Using lithography, chemical etching, vapor deposition and other techniques, complex devices can be built, such as the IBM cell processor shown on the right. (MOCVD image courtesy of Bob Kee. IBM cell processor photograph courtesy Tom Way, IBM Corporation; unauthorized use not permitted.)

boundary is moving and control systems are increasingly taking on these functions. Another dramatic trend on the horizon is the use of large collections of distributed entities with local computation, global communication connections, little regularity imposed by the laws of physics and no possibility of imposing centralized control actions. Examples of this trend include the national airspace management problem, automated highway and traffic management and command and control for future battlefields.

Materials and Processing

The chemical industry is responsible for the remarkable progress in developing new materials that are key to our modern society. In addition to the continuing need to improve product quality, several other factors in the process control industry are drivers for the use of control. Environmental statutes continue to place stricter limitations on the production of pollutants, forcing the use of sophisticated pollution control devices. Environmental safety considerations have led to the design of smaller storage capacities to diminish the risk of major chemical leakage, requiring tighter control on upstream processes and, in some cases, supply chains. And large increases in energy costs have encouraged engineers to design plants that are highly integrated, coupling many processes that used to operate independently. All of these trends increase the complexity of these processes and the performance requirements for the control systems, making control system design increasingly challenging. Some examples of materials-processing technology are shown in Figure 1.7.

As in many other application areas, new sensor technology is creating new opportunities for control. Online sensors—including laser backscattering, video microscopy and ultraviolet, infrared and Raman spectroscopy—are becoming more

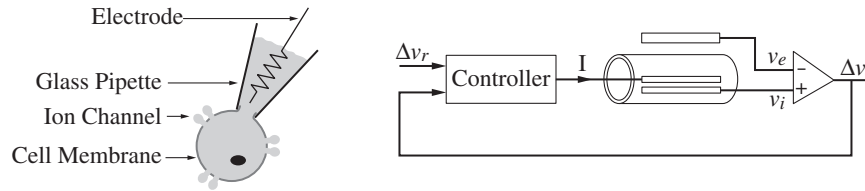


Figure 1.8: The voltage clamp method for measuring ion currents in cells using feedback. A pipet is used to place an electrode in a cell (left and middle) and maintain the potential of the cell at a fixed level. The internal voltage in the cell is v_i , and the voltage of the external fluid is v_e . The feedback system (right) controls the current I into the cell so that the voltage drop across the cell membrane $\Delta v = v_i - v_e$ is equal to its reference value Δv_r . The current I is then equal to the ion current.

robust and less expensive and are appearing in more manufacturing processes. Many of these sensors are already being used by current process control systems, but more sophisticated signal-processing and control techniques are needed to use more effectively the real-time information provided by these sensors. Control engineers also contribute to the design of even better sensors, which are still needed, for example, in the microelectronics industry. As elsewhere, the challenge is making use of the large amounts of data provided by these new sensors in an effective manner. In addition, a control-oriented approach to modeling the essential physics of the underlying processes is required to understand the fundamental limits on observability of the internal state through sensor data.

Instrumentation

The measurement of physical variables is of prime interest in science and engineering. Consider, for example, an accelerometer, where early instruments consisted of a mass suspended on a spring with a deflection sensor. The precision of such an instrument depends critically on accurate calibration of the spring and the sensor. There is also a design compromise because a weak spring gives high sensitivity but low bandwidth.

A different way of measuring acceleration is to use *force feedback*. The spring is replaced by a voice coil that is controlled so that the mass remains at a constant position. The acceleration is proportional to the current through the voice coil. In such an instrument, the precision depends entirely on the calibration of the voice coil and does not depend on the sensor, which is used only as the feedback signal. The sensitivity/bandwidth compromise is also avoided. This way of using feedback has been applied to many different engineering fields and has resulted in instruments with dramatically improved performance. Force feedback is also used in haptic devices for manual control.

Another important application of feedback is in instrumentation for biological systems. Feedback is widely used to measure ion currents in cells using a device called a *voltage clamp*, which is illustrated in Figure 1.8. Hodgkin and Huxley used the voltage clamp to investigate propagation of action potentials in the axon of the

giant squid. In 1963 they shared the Nobel Prize in Medicine with Eccles for “their discoveries concerning the ionic mechanisms involved in excitation and inhibition in the peripheral and central portions of the nerve cell membrane.” A refinement of the voltage clamp called a *patch clamp* made it possible to measure exactly when a single ion channel is opened or closed. This was developed by Neher and Sakmann, who received the 1991 Nobel Prize in Medicine “for their discoveries concerning the function of single ion channels in cells.”

There are many other interesting and useful applications of feedback in scientific instruments. The development of the mass spectrometer is an early example. In a 1935 paper, Nier observed that the deflection of ions depends on both the magnetic and the electric fields [Nie35]. Instead of keeping both fields constant, Nier let the magnetic field fluctuate and the electric field was controlled to keep the ratio between the fields constant. Feedback was implemented using vacuum tube amplifiers. This scheme was crucial for the development of mass spectroscopy.

The Dutch engineer van der Meer invented a clever way to use feedback to maintain a good-quality high-density beam in a particle accelerator [MPTvdM80]. The idea is to sense particle displacement at one point in the accelerator and apply a correcting signal at another point. This scheme, called *stochastic cooling*, was awarded the Nobel Prize in Physics in 1984. The method was essential for the successful experiments at CERN where the existence of the particles W and Z associated with the weak force was first demonstrated.

The 1986 Nobel Prize in Physics—awarded to Binnig and Rohrer for their design of the scanning tunneling microscope—is another example of an innovative use of feedback. The key idea is to move a narrow tip on a cantilever beam across a surface and to register the forces on the tip [BR86]. The deflection of the tip is measured using tunneling. The tunneling current is used by a feedback system to control the position of the cantilever base so that the tunneling current is constant, an example of force feedback. The accuracy is so high that individual atoms can be registered. A map of the atoms is obtained by moving the base of the cantilever horizontally. The performance of the control system is directly reflected in the image quality and scanning speed. This example is described in additional detail in Chapter 3.

Robotics and Intelligent Machines

The goal of cybernetic engineering, already articulated in the 1940s and even before, has been to implement systems capable of exhibiting highly flexible or “intelligent” responses to changing circumstances. In 1948 the MIT mathematician Norbert Wiener gave a widely read account of cybernetics [Wie48]. A more mathematical treatment of the elements of engineering cybernetics was presented by H. S. Tsien in 1954, driven by problems related to the control of missiles [Tsi54]. Together, these works and others of that time form much of the intellectual basis for modern work in robotics and control.

Two accomplishments that demonstrate the successes of the field are the Mars Exploratory Rovers and entertainment robots such as the Sony AIBO, shown in Figure 1.9. The two Mars Exploratory Rovers, launched by the Jet Propulsion



Figure 1.9: Robotic systems. (a) Spirit, one of the two Mars Exploratory Rovers that landed on Mars in January 2004. (b) The Sony AIBO Entertainment Robot, one of the first entertainment robots to be mass-marketed. Both robots make use of feedback between sensors, actuators and computation to function in unknown environments. (Photographs courtesy of Jet Propulsion Laboratory and Sony Electronics, Inc.)

Laboratory (JPL), maneuvered on the surface of Mars for more than 4 years starting in January 2004 and sent back pictures and measurements of their environment. The Sony AIBO robot debuted in June 1999 and was the first “entertainment” robot to be mass-marketed by a major international corporation. It was particularly noteworthy because of its use of artificial intelligence (AI) technologies that allowed it to act in response to external stimulation and its own judgment. This higher level of feedback is a key element in robotics, where issues such as obstacle avoidance, goal seeking, learning and autonomy are prevalent.

Despite the enormous progress in robotics over the last half-century, in many ways the field is still in its infancy. Today’s robots still exhibit simple behaviors compared with humans, and their ability to locomote, interpret complex sensory inputs, perform higher-level reasoning and cooperate together in teams is limited. Indeed, much of Wiener’s vision for robotics and intelligent machines remains unrealized. While advances are needed in many fields to achieve this vision—including advances in sensing, actuation and energy storage—the opportunity to combine the advances of the AI community in planning, adaptation and learning with the techniques in the control community for modeling, analysis and design of feedback systems presents a renewed path for progress.

Networks and Computing Systems

Control of networks is a large research area spanning many topics, including congestion control, routing, data caching and power management. Several features of these control problems make them very challenging. The dominant feature is the extremely large scale of the system; the Internet is probably the largest feedback control system humans have ever built. Another is the decentralized nature of the control problem: decisions must be made quickly and based only on local informa-

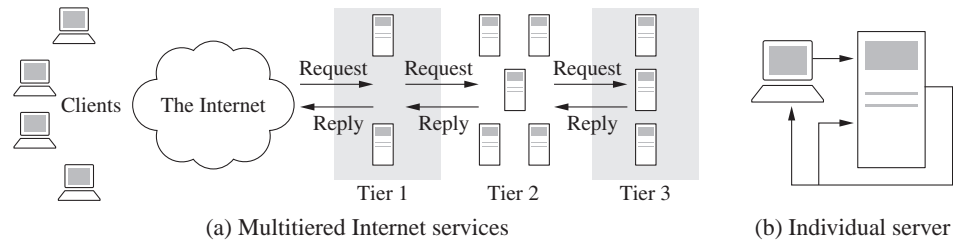


Figure 1.10: A multitier system for services on the Internet. In the complete system shown schematically in (a), users request information from a set of computers (tier 1), which in turn collect information from other computers (tiers 2 and 3). The individual server shown in (b) has a set of reference parameters set by a (human) system operator, with feedback used to maintain the operation of the system in the presence of uncertainty. (Based on Hellerstein et al. [HDPT04].)

tion. Stability is complicated by the presence of varying time lags, as information about the network state can be observed or relayed to controllers only after a delay, and the effect of a local control action can be felt throughout the network only after substantial delay. Uncertainty and variation in the network, through network topology, transmission channel characteristics, traffic demand and available resources, may change constantly and unpredictably. Other complicating issues are the diverse traffic characteristics—in terms of arrival statistics at both the packet and flow time scales—and the different requirements for quality of service that the network must support.

Related to the control of networks is control of the servers that sit on these networks. Computers are key components of the systems of routers, web servers and database servers used for communication, electronic commerce, advertising and information storage. While hardware costs for computing have decreased dramatically, the cost of operating these systems has increased because of the difficulty in managing and maintaining these complex interconnected systems. The situation is similar to the early phases of process control when feedback was first introduced to control industrial processes. As in process control, there are interesting possibilities for increasing performance and decreasing costs by applying feedback. Several promising uses of feedback in the operation of computer systems are described in the book by Hellerstein et al. [HDPT04].

A typical example of a multilayer system for e-commerce is shown in Figure 1.10a. The system has several tiers of servers. The edge server accepts incoming requests and routes them to the HTTP server tier where they are parsed and distributed to the application servers. The processing for different requests can vary widely, and the application servers may also access external servers managed by other organizations.

Control of an individual server in a layer is illustrated in Figure 1.10b. A quantity representing the quality of service or cost of operation—such as response time, throughput, service rate or memory usage—is measured in the computer. The control variables might represent incoming messages accepted, priorities in the oper-

ating system or memory allocation. The feedback loop then attempts to maintain quality-of-service variables within a target range of values.

Economics

The economy is a large, dynamical system with many actors: governments, organizations, companies and individuals. Governments control the economy through laws and taxes, the central banks by setting interest rates and companies by setting prices and making investments. Individuals control the economy through purchases, savings and investments. Many efforts have been made to model the system both at the macro level and at the micro level, but this modeling is difficult because the system is strongly influenced by the behaviors of the different actors in the system.

Keynes [Key36] developed a simple model to understand relations among gross national product, investment, consumption and government spending. One of Keynes' observations was that under certain conditions, e.g., during the 1930s depression, an increase in the investment of government spending could lead to a larger increase in the gross national product. This idea was used by several governments to try to alleviate the depression. Keynes' ideas can be captured by a simple model that is discussed in Exercise 2.4.

A perspective on the modeling and control of economic systems can be obtained from the work of some economists who have received the Sveriges Riksbank Prize in Economics in Memory of Alfred Nobel, popularly called the Nobel Prize in Economics. Paul A. Samuelson received the prize in 1970 for "the scientific work through which he has developed static and dynamic economic theory and actively contributed to raising the level of analysis in economic science." Lawrence Klein received the prize in 1980 for the development of large dynamical models with many parameters that were fitted to historical data [KG55], e.g., a model of the U.S. economy in the period 1929–1952. Other researchers have modeled other countries and other periods. In 1997 Myron Scholes shared the prize with Robert Merton for a new method to determine the value of derivatives. A key ingredient was a dynamic model of the variation of stock prices that is widely used by banks and investment companies. In 2004 Finn E. Kydland and Edward C. Prescott shared the economics prize "for their contributions to dynamic macroeconomics: the time consistency of economic policy and the driving forces behind business cycles," a topic that is clearly related to dynamics and control.

One of the reasons why it is difficult to model economic systems is that there are no conservation laws. A typical example is that the value of a company as expressed by its stock can change rapidly and erratically. There are, however, some areas with conservation laws that permit accurate modeling. One example is the flow of products from a manufacturer to a retailer as illustrated in Figure 1.11. The products are physical quantities that obey a conservation law, and the system can be modeled by accounting for the number of products in the different inventories. There are considerable economic benefits in controlling supply chains so that products are available to customers while minimizing products that are in storage. The real problems are more complicated than indicated in the figure because there may be

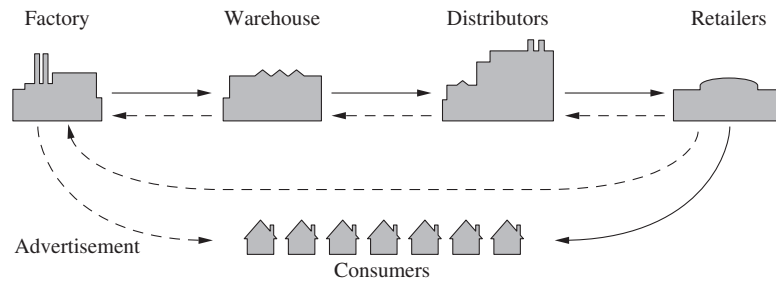


Figure 1.11: Supply chain dynamics (after Forrester [For61]). Products flow from the producer to the customer through distributors and retailers as indicated by the solid lines. There are typically many factories and warehouses and even more distributors and retailers. Multiple feedback loops are present as each agent tries to maintain the proper inventory level.

many different products, there may be different factories that are geographically distributed and the factories may require raw material or subassemblies.

Control of supply chains was proposed by Forrester in 1961 [For61] and is now growing in importance. Considerable economic benefits can be obtained by using models to minimize inventories. Their use accelerated dramatically when information technology was applied to predict sales, keep track of products and enable just-in-time manufacturing. Supply chain management has contributed significantly to the growing success of global distributors.

Advertising on the Internet is an emerging application of control. With network-based advertising it is easy to measure the effect of different marketing strategies quickly. The response of customers can then be modeled, and feedback strategies can be developed.

Feedback in Nature

Many problems in the natural sciences involve understanding aggregate behavior in complex large-scale systems. This behavior emerges from the interaction of a multitude of simpler systems with intricate patterns of information flow. Representative examples can be found in fields ranging from embryology to seismology. Researchers who specialize in the study of specific complex systems often develop an intuitive emphasis on analyzing the role of feedback (or interconnection) in facilitating and stabilizing aggregate behavior.

While sophisticated theories have been developed by domain experts for the analysis of various complex systems, the development of a rigorous methodology that can discover and exploit common features and essential mathematical structure is just beginning to emerge. Advances in science and technology are creating a new understanding of the underlying dynamics and the importance of feedback in a wide variety of natural and technological systems. We briefly highlight three application areas here.

Biological Systems. A major theme currently of interest to the biology commu-

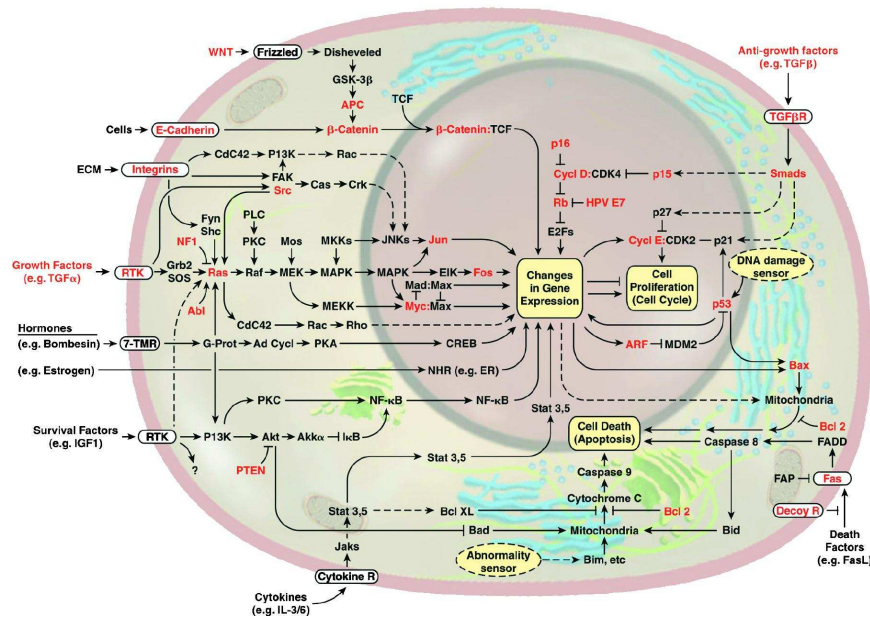


Figure 1.12: The wiring diagram of the growth-signaling circuitry of the mammalian cell [HW00]. The major pathways that are thought to play a role in cancer are indicated in the diagram. Lines represent interactions between genes and proteins in the cell. Lines ending in arrowheads indicate activation of the given gene or pathway; lines ending in a T-shaped head indicate repression. (Used with permission of Elsevier Ltd. and the authors.)

nity is the science of reverse (and eventually forward) engineering of biological control networks such as the one shown in Figure 1.12. There are a wide variety of biological phenomena that provide a rich source of examples of control, including gene regulation and signal transduction; hormonal, immunological and cardiovascular feedback mechanisms; muscular control and locomotion; active sensing, vision and proprioception; attention and consciousness; and population dynamics and epidemics. Each of these (and many more) provide opportunities to figure out what works, how it works, and what we can do to affect it.

One interesting feature of biological systems is the frequent use of positive feedback to shape the dynamics of the system. Positive feedback can be used to create switchlike behavior through autoregulation of a gene, and to create oscillations such as those present in the cell cycle, central pattern generators or circadian rhythm.

Ecosystems. In contrast to individual cells and organisms, emergent properties of aggregations and ecosystems inherently reflect selection mechanisms that act on multiple levels, and primarily on scales well below that of the system as a whole. Because ecosystems are complex, multiscale dynamical systems, they provide a broad range of new challenges for the modeling and analysis of feedback systems. Recent experience in applying tools from control and dynamical systems to bacterial networks suggests that much of the complexity of these networks is due to the presence of multiple layers of feedback loops that provide robust functionality

to the individual cell. Yet in other instances, events at the cell level benefit the colony at the expense of the individual. Systems level analysis can be applied to ecosystems with the goal of understanding the robustness of such systems and the extent to which decisions and events affecting individual species contribute to the robustness and/or fragility of the ecosystem as a whole.

Environmental Science. It is now indisputable that human activities have altered the environment on a global scale. Problems of enormous complexity challenge researchers in this area, and first among these is to understand the feedback systems that operate on the global scale. One of the challenges in developing such an understanding is the multiscale nature of the problem, with detailed understanding of the dynamics of microscale phenomena such as microbiological organisms being a necessary component of understanding global phenomena, such as the carbon cycle.

1.4 Feedback Properties

Feedback is a powerful idea which, as we have seen, is used extensively in natural and technological systems. The principle of feedback is simple: base correcting actions on the difference between desired and actual performance. In engineering, feedback has been rediscovered and patented many times in many different contexts. The use of feedback has often resulted in vast improvements in system capability, and these improvements have sometimes been revolutionary, as discussed above. The reason for this is that feedback has some truly remarkable properties. In this section we will discuss some of the properties of feedback that can be understood intuitively. This intuition will be formalized in subsequent chapters.

Robustness to Uncertainty

One of the key uses of feedback is to provide robustness to uncertainty. By measuring the difference between the sensed value of a regulated signal and its desired value, we can supply a corrective action. If the system undergoes some change that affects the regulated signal, then we sense this change and try to force the system back to the desired operating point. This is precisely the effect that Watt exploited in his use of the centrifugal governor on steam engines.

As an example of this principle, consider the simple feedback system shown in Figure 1.13. In this system, the speed of a vehicle is controlled by adjusting the amount of gas flowing to the engine. Simple *proportional-integral* (PI) feedback is used to make the amount of gas depend on both the error between the current and the desired speed and the integral of that error. The plot on the right shows the results of this feedback for a step change in the desired speed and a variety of different masses for the car, which might result from having a different number of passengers or towing a trailer. Notice that independent of the mass (which varies by a factor of 3!), the steady-state speed of the vehicle always approaches the desired speed and achieves that speed within approximately 5 s. Thus the performance of

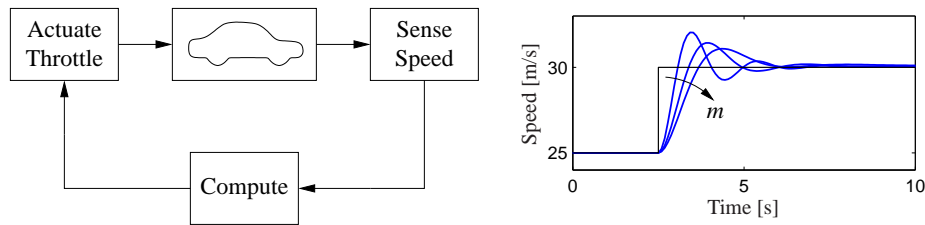


Figure 1.13: A feedback system for controlling the speed of a vehicle. In the block diagram on the left, the speed of the vehicle is measured and compared to the desired speed within the “Compute” block. Based on the difference in the actual and desired speeds, the throttle (or brake) is used to modify the force applied to the vehicle by the engine, drivetrain and wheels. The figure on the right shows the response of the control system to a commanded change in speed from 25 m/s to 30 m/s. The three different curves correspond to differing masses of the vehicle, between 1000 and 3000 kg, demonstrating the robustness of the closed loop system to a very large change in the vehicle characteristics.

the system is robust with respect to this uncertainty.

Another early example of the use of feedback to provide robustness is the negative feedback amplifier. When telephone communications were developed, amplifiers were used to compensate for signal attenuation in long lines. A vacuum tube was a component that could be used to build amplifiers. Distortion caused by the nonlinear characteristics of the tube amplifier together with amplifier drift were obstacles that prevented the development of line amplifiers for a long time. A major breakthrough was the invention of the feedback amplifier in 1927 by Harold S. Black, an electrical engineer at Bell Telephone Laboratories. Black used *negative feedback*, which reduces the gain but makes the amplifier insensitive to variations in tube characteristics. This invention made it possible to build stable amplifiers with linear characteristics despite the nonlinearities of the vacuum tube amplifier.

Design of Dynamics

Another use of feedback is to change the dynamics of a system. Through feedback, we can alter the behavior of a system to meet the needs of an application: systems that are unstable can be stabilized, systems that are sluggish can be made responsive and systems that have drifting operating points can be held constant. Control theory provides a rich collection of techniques to analyze the stability and dynamic response of complex systems and to place bounds on the behavior of such systems by analyzing the gains of linear and nonlinear operators that describe their components.

An example of the use of control in the design of dynamics comes from the area of flight control. The following quote, from a lecture presented by Wilbur Wright to the Western Society of Engineers in 1901 [McF53], illustrates the role of control in the development of the airplane:

Men already know how to construct wings or airplanes, which when driven through the air at sufficient speed, will not only sustain the

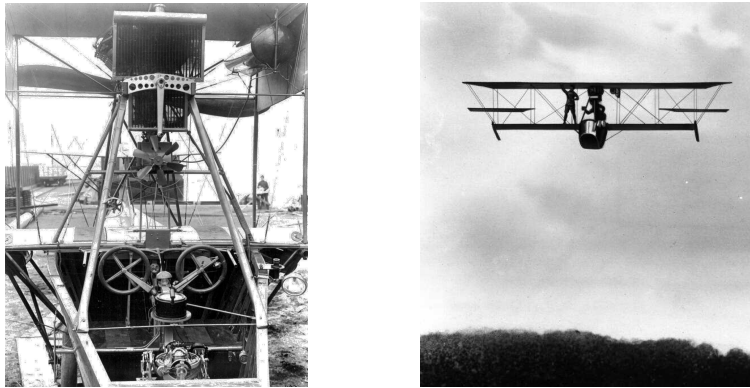


Figure 1.14: Aircraft autopilot system. The Sperry autopilot (left) contained a set of four gyros coupled to a set of air valves that controlled the wing surfaces. The 1912 Curtiss used an autopilot to stabilize the roll, pitch and yaw of the aircraft and was able to maintain level flight as a mechanic walked on the wing (right) [Hug93].

weight of the wings themselves, but also that of the engine, and of the engineer as well. Men also know how to build engines and screws of sufficient lightness and power to drive these planes at sustaining speed ... Inability to balance and steer still confronts students of the flying problem ... When this one feature has been worked out, the age of flying will have arrived, for all other difficulties are of minor importance.

The Wright brothers thus realized that control was a key issue to enable flight. They resolved the compromise between stability and maneuverability by building an airplane, the Wright Flyer, that was unstable but maneuverable. The Flyer had a rudder in the front of the airplane, which made the plane very maneuverable. A disadvantage was the necessity for the pilot to keep adjusting the rudder to fly the plane: if the pilot let go of the stick, the plane would crash. Other early aviators tried to build stable airplanes. These would have been easier to fly, but because of their poor maneuverability they could not be brought up into the air. By using their insight and skillful experiments the Wright brothers made the first successful flight at Kitty Hawk in 1903.

Since it was quite tiresome to fly an unstable aircraft, there was strong motivation to find a mechanism that would stabilize an aircraft. Such a device, invented by Sperry, was based on the concept of feedback. Sperry used a gyro-stabilized pendulum to provide an indication of the vertical. He then arranged a feedback mechanism that would pull the stick to make the plane go up if it was pointing down, and vice versa. The Sperry autopilot was the first use of feedback in aeronautical engineering, and Sperry won a prize in a competition for the safest airplane in Paris in 1914. Figure 1.14 shows the Curtiss seaplane and the Sperry autopilot. The autopilot is a good example of how feedback can be used to stabilize an unstable system and hence “design the dynamics” of the aircraft.

One of the other advantages of designing the dynamics of a device is that it allows for increased modularity in the overall system design. By using feedback to create a system whose response matches a desired profile, we can hide the complexity and variability that may be present inside a subsystem. This allows us to create more complex systems by not having to simultaneously tune the responses of a large number of interacting components. This was one of the advantages of Black's use of negative feedback in vacuum tube amplifiers: the resulting device had a well-defined linear input/output response that did not depend on the individual characteristics of the vacuum tubes being used.

Higher Levels of Automation

A major trend in the use of feedback is its application to higher levels of situational awareness and decision making. This includes not only traditional logical branching based on system conditions but also optimization, adaptation, learning and even higher levels of abstract reasoning. These problems are in the domain of the artificial intelligence community, with an increasing role of dynamics, robustness and interconnection in many applications.

One of the interesting areas of research in higher levels of decision is autonomous control of cars. Early experiments with autonomous driving were performed by Ernst Dickmanns, who in the 1980s equipped cars with cameras and other sensors [Dic07]. In 1994 his group demonstrated autonomous driving with human supervision on a highway near Paris and in 1995 one of his cars drove autonomously (with human supervision) from Munich to Copenhagen at speeds of up to 175 km/hour. The car was able to overtake other vehicles and change lanes automatically.

This application area has been recently explored through the DARPA Grand Challenge, a series of competitions sponsored by the U.S. government to build vehicles that can autonomously drive themselves in desert and urban environments. Caltech competed in the 2005 and 2007 Grand Challenges using a modified Ford E-350 offroad van nicknamed "Alice." It was fully automated, including electronically controlled steering, throttle, brakes, transmission and ignition. Its sensing systems included multiple video cameras scanning at 10–30 Hz, several laser ranging units scanning at 10 Hz and an inertial navigation package capable of providing position and orientation estimates at 5 ms temporal resolution. Computational resources included 12 high-speed servers connected together through a 1-Gb/s Ethernet switch. The vehicle is shown in Figure 1.15, along with a block diagram of its control architecture.

The software and hardware infrastructure that was developed enabled the vehicle to traverse long distances at substantial speeds. In testing, Alice drove itself more than 500 km in the Mojave Desert of California, with the ability to follow dirt roads and trails (if present) and avoid obstacles along the path. Speeds of more than 50 km/h were obtained in the fully autonomous mode. Substantial tuning of the algorithms was done during desert testing, in part because of the lack of systems-level design tools for systems of this level of complexity. Other competitors in the race

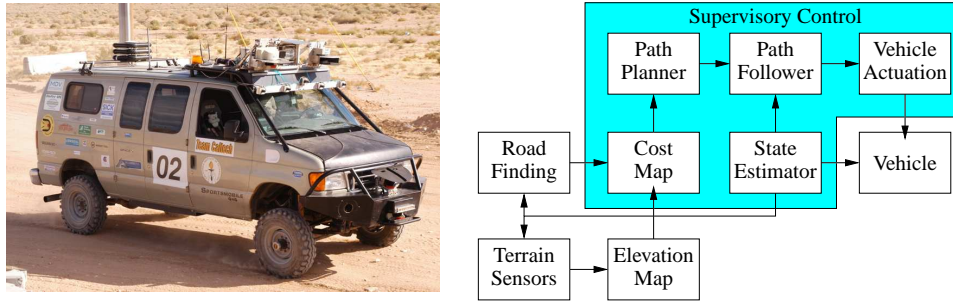


Figure 1.15: DARPA Grand Challenge. “Alice,” Team Caltech’s entry in the 2005 and 2007 competitions and its networked control architecture [CFG+06]. The feedback system fuses data from terrain sensors (cameras and laser range finders) to determine a digital elevation map. This map is used to compute the vehicle’s potential speed over the terrain, and an optimization-based path planner then commands a trajectory for the vehicle to follow. A supervisory control module performs higher-level tasks such as handling sensor and actuator failures.

(including Stanford, which won the 2005 competition) used algorithms for adaptive control and learning, increasing the capabilities of their systems in unknown environments. Together, the competitors in the Grand Challenge demonstrated some of the capabilities of the next generation of control systems and highlighted many research directions in control at higher levels of decision making.

Drawbacks of Feedback

While feedback has many advantages, it also has some drawbacks. Chief among these is the possibility of instability if the system is not designed properly. We are all familiar with the effects of *positive feedback* when the amplification on a microphone is turned up too high in a room. This is an example of feedback instability, something that we obviously want to avoid. This is tricky because we must design the system not only to be stable under nominal conditions but also to remain stable under all possible perturbations of the dynamics.

In addition to the potential for instability, feedback inherently couples different parts of a system. One common problem is that feedback often injects measurement noise into the system. Measurements must be carefully filtered so that the actuation and process dynamics do not respond to them, while at the same time ensuring that the measurement signal from the sensor is properly coupled into the closed loop dynamics (so that the proper levels of performance are achieved).

Another potential drawback of control is the complexity of embedding a control system in a product. While the cost of sensing, computation and actuation has decreased dramatically in the past few decades, the fact remains that control systems are often complicated, and hence one must carefully balance the costs and benefits. An early engineering example of this is the use of microprocessor-based feedback systems in automobiles. The use of microprocessors in automotive applications began in the early 1970s and was driven by increasingly strict emissions standards,

which could be met only through electronic controls. Early systems were expensive and failed more often than desired, leading to frequent customer dissatisfaction. It was only through aggressive improvements in technology that the performance, reliability and cost of these systems allowed them to be used in a transparent fashion. Even today, the complexity of these systems is such that it is difficult for an individual car owner to fix problems.

Feedforward

Feedback is reactive: there must be an error before corrective actions are taken. However, in some circumstances it is possible to measure a disturbance before it enters the system, and this information can then be used to take corrective action before the disturbance has influenced the system. The effect of the disturbance is thus reduced by measuring it and generating a control signal that counteracts it. This way of controlling a system is called *feedforward*. Feedforward is particularly useful in shaping the response to command signals because command signals are always available. Since feedforward attempts to match two signals, it requires good process models; otherwise the corrections may have the wrong size or may be badly timed.

The ideas of feedback and feedforward are very general and appear in many different fields. In economics, feedback and feedforward are analogous to a market-based economy versus a planned economy. In business, a feedforward strategy corresponds to running a company based on extensive strategic planning, while a feedback strategy corresponds to a reactive approach. In biology, feedforward has been suggested as an essential element for motion control in humans that is tuned during training. Experience indicates that it is often advantageous to combine feedback and feedforward, and the correct balance requires insight and understanding of their respective properties.

Positive Feedback

In most of this text, we will consider the role of *negative feedback*, in which we attempt to regulate the system by reacting to disturbances in a way that decreases the effect of those disturbances. In some systems, particularly biological systems, *positive feedback* can play an important role. In a system with positive feedback, the increase in some variable or signal leads to a situation in which that quantity is further increased through its dynamics. This has a destabilizing effect and is usually accompanied by a saturation that limits the growth of the quantity. Although often considered undesirable, this behavior is used in biological (and engineering) systems to obtain a very fast response to a condition or signal.

One example of the use of positive feedback is to create switching behavior, in which a system maintains a given state until some input crosses a threshold. Hysteresis is often present so that noisy inputs near the threshold do not cause the system to jitter. This type of behavior is called *bistability* and is often associated with memory devices.

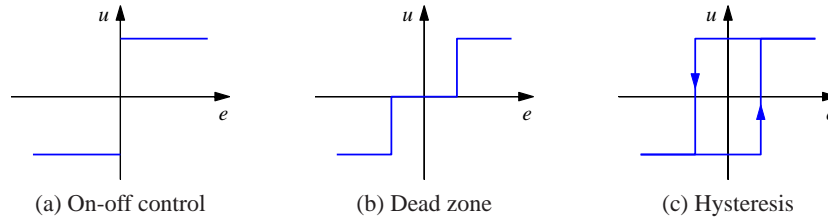


Figure 1.16: Input/output characteristics of on-off controllers. Each plot shows the input on the horizontal axis and the corresponding output on the vertical axis. Ideal on-off control is shown in (a), with modifications for a dead zone (b) or hysteresis (c). Note that for on-off control with hysteresis, the output depends on the value of past inputs.

1.5 Simple Forms of Feedback

The idea of feedback to make corrective actions based on the difference between the desired and the actual values of a quantity can be implemented in many different ways. The benefits of feedback can be obtained by very simple feedback laws such as on-off control, proportional control and proportional-integral-derivative control. In this section we provide a brief preview of some of the topics that will be studied more formally in the remainder of the text.

On-Off Control

A simple feedback mechanism can be described as follows:

$$u = \begin{cases} u_{\max} & \text{if } e > 0 \\ u_{\min} & \text{if } e < 0, \end{cases} \quad (1.1)$$

where the *control error* $e = r - y$ is the difference between the reference signal (or command signal) r and the output of the system y and u is the actuation command. Figure 1.16a shows the relation between error and control. This control law implies that maximum corrective action is always used.

The feedback in equation (1.1) is called *on-off control*. One of its chief advantages is that it is simple and there are no parameters to choose. On-off control often succeeds in keeping the process variable close to the reference, such as the use of a simple thermostat to maintain the temperature of a room. It typically results in a system where the controlled variables oscillate, which is often acceptable if the oscillation is sufficiently small.

Notice that in equation (1.1) the control variable is not defined when the error is zero. It is common to make modifications by introducing either a dead zone or hysteresis (see Figure 1.16b and 1.16c).

PID Control

The reason why on-off control often gives rise to oscillations is that the system overreacts since a small change in the error makes the actuated variable change over the full range. This effect is avoided in *proportional control*, where the characteristic

of the controller is proportional to the control error for small errors. This can be achieved with the control law

$$u = \begin{cases} u_{\max} & \text{if } e \geq e_{\max} \\ k_p e & \text{if } e_{\min} < e < e_{\max} \\ u_{\min} & \text{if } e \leq e_{\min}, \end{cases} \quad (1.2)$$

where k_p is the controller gain, $e_{\min} = u_{\min}/k_p$ and $e_{\max} = u_{\max}/k_p$. The interval (e_{\min}, e_{\max}) is called the *proportional band* because the behavior of the controller is linear when the error is in this interval:

$$u = k_p(r - y) = k_p e \quad \text{if } e_{\min} \leq e \leq e_{\max}. \quad (1.3)$$

While a vast improvement over on-off control, proportional control has the drawback that the process variable often deviates from its reference value. In particular, if some level of control signal is required for the system to maintain a desired value, then we must have $e \neq 0$ in order to generate the requisite input.

This can be avoided by making the control action proportional to the integral of the error:

$$u(t) = k_i \int_0^t e(\tau) d\tau. \quad (1.4)$$

This control form is called *integral control*, and k_i is the integral gain. It can be shown through simple arguments that a controller with integral action has zero steady-state error (Exercise 1.5). The catch is that there may not always be a steady state because the system may be oscillating.

An additional refinement is to provide the controller with an anticipative ability by using a prediction of the error. A simple prediction is given by the linear extrapolation

$$e(t + T_d) \approx e(t) + T_d \frac{de(t)}{dt},$$

which predicts the error T_d time units ahead. Combining proportional, integral and derivative control, we obtain a controller that can be expressed mathematically as

$$u(t) = k_p e(t) + k_i \int_0^t e(\tau) d\tau + k_d \frac{de(t)}{dt}. \quad (1.5)$$

The control action is thus a sum of three terms: the past as represented by the integral of the error, the present as represented by the proportional term and the future as represented by a linear extrapolation of the error (the derivative term). This form of feedback is called a *proportional-integral-derivative (PID) controller* and its action is illustrated in Figure 1.17.

A PID controller is very useful and is capable of solving a wide range of control problems. More than 95% of all industrial control problems are solved by PID control, although many of these controllers are actually *proportional-integral (PI) controllers* because derivative action is often not included [DM02]. There are also more advanced controllers, which differ from PID controllers by using more sophisticated methods for prediction.

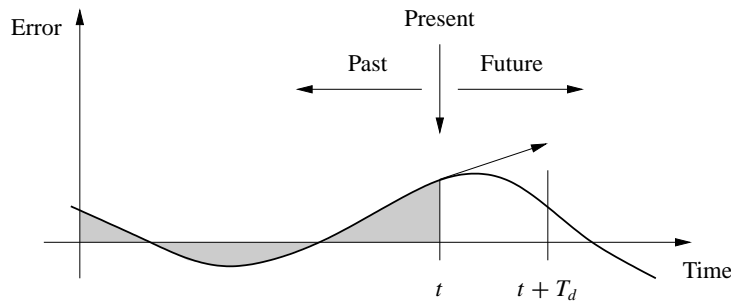


Figure 1.17: Action of a PID controller. At time t , the proportional term depends on the instantaneous value of the error. The integral portion of the feedback is based on the integral of the error up to time t (shaded portion). The derivative term provides an estimate of the growth or decay of the error over time by looking at the rate of change of the error. T_d represents the approximate amount of time in which the error is projected forward (see text).

1.6 Further Reading

The material in this section draws heavily from the report of the Panel on Future Directions on Control, Dynamics and Systems [Mur03]. Several additional papers and reports have highlighted the successes of control [NS99] and new vistas in control [Bro00, Kum01, Wis07]. The early development of control is described by Mayr [May70] and in the books by Bennett [Ben79, Ben93], which cover the period 1800–1955. A fascinating examination of some of the early history of control in the United States has been written by Mindell [Min02]. A popular book that describes many control concepts across a wide range of disciplines is *Out of Control* by Kelly [Kel94]. There are many textbooks available that describe control systems in the context of specific disciplines. For engineers, the textbooks by Franklin, Powell and Emami-Naeini [FPEN05], Dorf and Bishop [DB04], Kuo and Golnaraghi [KG02] and Seborg, Edgar and Mellichamp [SEM04] are widely used. More mathematically oriented treatments of control theory include Sontag [Son98] and Lewis [Lew03]. The book by Hellerstein et al. [HDPT04] provides a description of the use of feedback control in computing systems. A number of books look at the role of dynamics and feedback in biological systems, including Milhorn [Mil66] (now out of print), J. D. Murray [Mur04] and Ellner and Guckenheimer [EG05]. The book by Fradkov [Fra07] and the tutorial article by Bechhoefer [Bec05] cover many specific topics of interest to the physics community.

Exercises

1.1 (Eye motion) Perform the following experiment and explain your results: Holding your head still, move one of your hands left and right in front of your face, following it with your eyes. Record how quickly you can move your hand before you begin to lose track of it. Now hold your hand still and shake your head left to right, once again recording how quickly you can move before losing track of your