Cambridge University Press & Assessment
978-1-107-00118-3 — Control Theory for Physicists
John Bechhoefer
Excerpt
More Information

PART I

CORE MATERIAL

Historical Introduction

1

> Any sufficiently advanced technology is indistinguishable from magic.
>
> *Arthur C. Clarke*[1]

This is a book about magic. Not the magic of wizardry and sorcerers, nor the magic of fairy tales or fables, but *real* magic: the magic that powers planes and runs computers, that keeps our investments high and our blood pressure low, our beer cold and our bodies warm. It is the magic behind physics experiments of extraordinary precision, from gravity-wave detectors that probe the cosmos to scanning probe microscopes that image atoms. It is the magic that regulates biological processes from the pupil size in our eyes to the gene expression in our cells. It is the magic made possible by control theory.

The study of control theory can lead to something of a culture shock for physicists. Of course, jargon, technical methods, and applications may all be new. But something more fundamental is at play: As physicists, we study the world as it is. We look for the fundamental laws that govern time and energy, fields and forces, matter and motion, at the level of individual particles and collective phenomena. We do this in settings that range from the very large scales of the cosmos to the very small scales of fundamental particles to the very complex systems that rule the human scale. But we do all of this on Nature's terms, content to describe the actual dynamics of real "physical" systems.

Control theorists ask, instead, what might be. They seek to alter the states and dynamics of a system to make it *better*. The word "better" already implies a human element, or at least an active agent that can influence its environment. The Ancient Greeks coined the notion of *teleology* to denote the purpose or end (*telos*) of an object. While science has moved away from endowing objects in themselves with purpose, engineers design machines or systems to accomplish predefined tasks. Control theory tells, in a precise way, how to accomplish these tasks and indicates what is possible or not. *Uncertainty* – about initial conditions, external disturbances, dynamical rules, etc. – can limit possibilities.

Since all systems are physical ones, ruled by the laws of physics, physics will play a role in our story. But in many ways, it will have a supporting role, as we seek to create "augmented" systems that perform in ways that seemingly ignore the laws of physics. Of course they do not. Even so, we will see that a larger, open, physical "supersystem" can give a subsystem effective dynamics with new laws and properties.

---

[1] *Profiles of the Future: An Inquiry Into the Limits of the Possible*, New York, Harper and Row, Rev. ed., 1973.

In this book, we will take a broad look at control, from both the fundamental point of view that seeks to understand what it can accomplish and what not, and how control in general meshes with other topics in physics such as thermodynamics and statistical physics. At the same time, we will also be interested in control for its practical applications. Just as control is fundamental to the technological devices of modern life, so too does it play a key role in the techniques an experimental physicist should know.

Sometimes called the "hidden technology," control is often invisible, despite its omnipresence in modern technology. We do not notice it until something fails. Planes are very safe, but occasionally they fall from the sky. Our bodies also depend on many control loops. To name one: to survive, we must maintain a core temperature within 27–44 °C, implying the need to keep maximum deviations to $< \pm 3\%$ and to regulate typical fluctuations to be $< \pm 0.3\%$. Again, we pay little attention to our body's temperature – except when it begins to deviate when we get sick or cold or hot. Our ability to ignore control under normal circumstances is a testament to its *robustness* to specific types of situations; our need to confront the often-drastic consequences of its failure is a consequence of its *fragility* to unforeseen circumstances. As we will see, the two aspects are linked.

In this introductory chapter, we present briefly the historical development of control and its theory, which gives some insight as to what "better" dynamics might mean. We then list some of these *goals* for control. Then we introduce, in an intuitive way, some of the principal *methods* of control, notably *feedback* and *feedforward*. We conclude with a discussion of the *types* of control systems.

1.1 Historical Overview

We can divide the development of control techniques and theory into five periods:

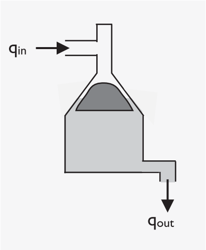
- Early control (before 1900)
- Preclassical period (1900–1940)
- Classical period (1930–1960)
- Modern control (1945–2000)
- Contemporary control (after 2000)

The overlaps are deliberate, as actual developments are not as well ordered chronologically as the classifications would imply. Although it seems logical to "begin at the beginning," this summary may be easier to follow after you have learned some of the material from later chapters. Partly for this reason, the discussion is relatively brief, with some aspects deferred to the relevant later chapters. Of course, a short exposition inevitably simplifies a complex story. The notes and references give pointers to more extensive presentations.

Cambridge University Press & Assessment
978-1-107-00118-3 — Control Theory for Physicists
John Bechhoefer
Excerpt
More Information

1.1.1 Early Control (before 1900)

The word *feedback* is of relatively recent origin, with the Oxford English Dictionary reporting its first use in 1920, in connection with an electrical circuit.[2] However, uses of feedback and the broader notion of control are far more ancient. Ktesibios (285–222 BC), a Greek working in Alexandria, Egypt, used feedback to improve the stability of water clocks, vessels that measure time by the outflow of water. However, as the fluid level in a vessel decreases, so too will its outflow rate. Keeping the level constant, or *regulating* it, stabilizes the rate of outflow. There are no original records of the device, but reconstructions based on Vitruvius's *De architectura* (~ 30–15 BC) and later Arab water clocks indicate that the mechanism was the same as that used in the modern flush toilet: a ball floating in the tank follows the water level. When the level is low, a float lets in more water, raising the level and increasing q_{out}; when high, the float shuts off the valve, decreasing q_{out} (see right).



In the Middle Ages, *mechanical clocks* powered by falling weights or springs were developed, with various ratchets ("escapements") that translate oscillating into rotational motion. These clocks also have feedback mechanisms to ensure constant rotation rates.

Because fluid density depends on temperature, the level of a fluid can be used to regulate temperature. René-Antoine Ferchault de Réamur (1683–1757) invented such a device, based on the temperature sensor of Cornelius Drebbel (1572–1663), a Dutch engineer working in England. In France, Jean-Simon Bonnemain (1743–1830) patented in 1783 an improved temperature controller based on a bimetallic rod that flexed when the temperature changed. He used it to make practical hot-water central heating for buildings.

The beginning of the Industrial Revolution, centered on England in the second half of the eighteenth century, led to the first important applications of feedback. The most prominent was the *governor*, which was developed to keep windmills turning at a constant rate and then adapted to the steam engine for more general purposes by James Watt in the late 1780s.[3] The issue was that variable loads would alter the rotation rate of the engine. To keep it constant, Watt and his partner Matthew Boulton adapted a *flyball* sensor for rotation rates that had been patented by Thomas Mead in 1787. As illustrated in Figure 1.1, the sensor has two heavy balls that rotate with the engine shaft. If the engine rotates too quickly, centrifugal force pushes the balls out, pulling down a lever and shutting off the throttle valve that lets steam in, thus slowing the motor. If it rotates too slowly, the balls fall in, pushing up the lever, opening the value, letting more steam in, and speeding up the motor. If all goes well, the steam-engine rotation rate settles at a desired value.

The nineteenth century saw a steady improvement in the technology of governors. The 1868 paper *On governors* by James Clerk Maxwell gave the first theoretical analysis. A flaw of governors was their tendency to make the engine "hunt" for the right

---

[2] The related term *feedforward* was first used even more recently, in 1952 (also according to the OED).

[3] By the 1670s, Christiaan Huygens had invented a governor to regulate pendulum clocks (Bateman, 1945).

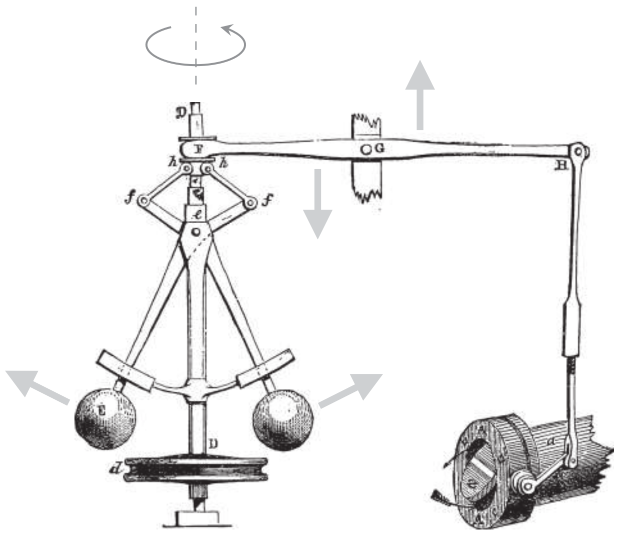


Fig. 1.1 Flyball governor and throttle valve, with rotation around the indicated axis. Flyballs move out, lever at F is pulled down, pivots about G, and pushes up at H, closing the throttle valve at lower right. Adapted from Routledge (1900).

rotation speed. In more modern language, there could be long-lived oscillatory transients before settling to a steady state. Even worse, the engine could become unstable and show erratic motion. Maxwell analyzed the conditions for stability of regulation against small perturbations using *linear stability analysis*. His stability conditions were generalized by Edward J. Routh and Adolf Hurwitz later in the nineteenth century. Although these early analyses of control systems eventually became part of the techniques of control theory in the mid twentieth century, they had little immediate impact on practical realizations, which was driven by the innovations of "tinkerers." Another emerging class of control applications concerned the position of a moving object. Thus, ships needed steering and missiles guiding to their target. In England, J. McFarlane Gray patented in 1866 a steering engine using feedback. In France, Jean Joseph Farcot introduced a range of position-control devices that he called *servomotors*. More generally, servomechanisms were used to *track* desired time-dependent trajectories, a generalization of the simpler goal of *regulation*, where the desired trajectory is simply a constant.

1.1.2 Preclassical Period (1900–1940)

Pre-1900 regulators were all *direct acting*: the elements that measured the quantity being regulated also had to change the system. The lever in a fluid-level regulator that moves in response to a change in level also opens the value that lets in more water. Of course, there is a "power source" (a high-pressure supply of water) that makes the response possible, but one "gadget" must still carry out two actions. Around 1900 a long process of abstraction began that led to distinct notions of *sensors*, *controllers*, and *actuators*. The sensor *measures* a quantity of concern, the controller *decides* how to respond, and the actuator *executes* the response. Each element can have its own, independent source of power. Such ideas, however, took several decades to become clear.

Meanwhile, the first decades of the twentieth century saw the beginnings of industrial process control. Applications included boiler control for steam generation, electric motor speed regulation, steering for ships and airplanes, temperature and pressure control, and more. A key development was of stand-alone controllers that could be added on to existing equipment. For example, around 1910, Elmer Sperry greatly improved the gyrocompass and designed a gyroscope autopilot to steer ships. The Sperry Gyroscope Company supplied the US Navy with navigational aids, as well as bomb sights and fire-control systems.

In 1922, Nicholas Minorsky gave a detailed analysis of such mechanisms, introducing the notion of three types of control. The first is *proportional* to the error between *set point* and actual signal, the second to the *integral* of that error, and the third to its *derivative*. Together, they form the three-term regulator, or proportional-integral-derivative (PID) control, which is discussed in Chapter 3. Although these ideas now seem very general, they were at first encountered separately in each domain of application. Thus, Minorsky's analyses were little known in the broader technical community for a number of years.

Another important development was the first airplane flight by Orville and Wilbur Wright. Others had built (and sometimes died testing) unsuccessful flying machines. The Wright brothers' success was based on their mastery of control, using flaps to alter yaw, pitch, and roll (three axes). Moreover, they recognized the advantages of an inherently *unstable* design stabilized by control (e.g., a human pilot). Unstable systems are more maneuverable than stable ones. They need active feedback to produce stable motion but can respond to disturbances (gusts of wind, abrupt change in terrain, etc.) much more quickly. The concept should be familiar: when we stand upright, we are unstable and must use (unconscious) small muscle movements to prevent ourselves from falling over. Indeed, the ability to walk on two legs is what distinguished the first hominids from other apes.[4]

Along with developments in mechanical control systems came parallel ones in electrical circuits. By the end of the nineteenth century, there was already a division between the power and signal applications of electricity. In both, the *amplifier* was a key element, allowing separation of the functions of sensor and actuator. Early high-power amplifiers took the form of relays and spring-based solenoids, which became the basis of many kinds of actuators.

For low-power electrical signals and their circuits, a key development was Lee de Forest's 1906 *grid audion*, a vacuum-tube amplifier that could boost the voltage level of a weak signal, compensating for signal losses in transmission and making possible

---

[4] What were the evolutionary advantages of walking upright? Darwin thought that it improved our ability to fight. But walking is also more efficient for traveling large distances on the ground (e.g., over grasslands). As with many evolutionary developments, the "why" is elusive (Wayman, 2012).

long-distance telephone networks. But the amplifiers had serious flaws: the signal gain was both nonlinear and prone to drifts, which led to distortion and volume variations.

Finally, there was a transformation in our view of living beings. Life in the nineteenth century was fixed on Newtonian, mechanical motion. Things alive *moved*, powered perhaps by the electric spark that jolted Frankenstein's monster to life or by some other unknown vital force. In the 1920s, the physiologist Walter Cannon introduced the term *homeostasis*, the ability to maintain conditions in the face of external perturbations. These conditions include the core temperature of the body and the concentrations of glucose, iron, oxygen, calcium, sodium, potassium, and other chemicals or ions. All these quantities are closely regulated, even when external conditions change dramatically: through hot or cold, our core temperatures are close to 37°C, our sodium levels stay between 135 and 145 milliequivalents per liter, and so on. The ability to regulate so many quantities in the body using multiple, hierarchical systems is one of the defining features of the modern view of life. Conversely, death is associated with a *failure cascade* that shuts down the essential functions of the body with its nested control loops, often one after the other. Understanding homeostasis was a goal of Wiener's influential book *Cybernetics*, a founding text of control theory, discussed below.

1.1.3 Classical Period (1930–1960)

At Bell Telephone Laboratories, a group of engineers was set up to address quality problems in the growing telephone network. Initial progress was slow, but on Tuesday morning, August 2, 1927, Harold Black had an epiphany while riding the Lackawanna Ferry across the Hudson to Manhattan to get to work. His idea, sketched out on a blank page of the *New York Times*, was that by taking a portion of the amplifier output signal and subtracting it from the input, one could reduce distortion, at the cost of a reduced gain. Thus was born the *negative-feedback amplifier*, which had the immediate effect of improving long-distance telephone calls and was a key development in the history of control. Its descendant, the *operational amplifier*, is described in Chapter 3. More broadly, efforts to understand what Black had created led to a "classical" formulation of control theory.

In the 1930s, Black's colleagues at Bell Labs, Harold Nyquist and Hendrik Bode, contributed theoretical analyses that put negative feedback and other ideas of *classical control* on a firmer footing. Work by Harold Hazen and Gordon Brown at MIT also was influential. In contrast to earlier studies based on solving ordinary differential equations in the time domain, they used frequency-domain methods based on the Laplace transform to derive a set of heuristic rules (often expressed in graphical form) for controllers of reasonable performance that work well for a relatively large class of systems. Bode's 1945 book *Network Analysis and Feedback Amplifier Design*, delayed because of the war, is perhaps the apotheosis of classical control. It considered *robustness* in depth, pointing out the fundamental compromises inherent in control: feedback that suppresses the response to disturbances at some frequencies will inevitably boost that response at other frequencies.

The next great impetus to the fledgling field of control engineering (and its counterpart, control theory) came with World War II. Engineers worked on a variety of control problems, notably the aiming of antiaircraft guns and automatic radar tracking. The *Radiation Lab* at MIT was a particularly important center for such research. To the scientists and engineers working at such centers, the war made particularly clear the need for unified, abstracted treatments of control based on concepts that were independent of specific applications. The classified results released en masse at the end of the war spurred rapid progress afterwards.

1.1.4 Modern Control (1945–2000)

After the end of World War II, control emerged as a distinct technical discipline. Engineering societies such as the American Society for Mechanical Engineers (AMSE), the Instrument Society of America (ISA), and the Institute of Radio Engineers (IRE, later the IEEE) all launched subgroups, and new professional societies such as the American Automatic Control Council (AACC) and International Federation of Automatic Control (IFAC) were created. Where MIT had stood almost alone as an academic center, many universities around the world added groups focusing on automatic control. The military-industrial complex took shape: think tanks such as the RAND corporation in Santa Monica, California and the Research Institute for Advanced Study in Baltimore and companies blurred military and industrial roles on scales larger than had been known before the War.[5] Prominent companies included IBM, General Electric, Hughes Aircraft, Bell Labs, Honeywell, Westinghouse, Leeds and Northrup in the United States, and Siemens (Germany), Schneider (France), ASEA (Sweden), and Yokogawa and Mitsubishi (Japan). Regular national and international conferences began: The first IFAC World Conference, in 1960 in Moscow at the height of the Cold War, marked the emergence of *modern control*.[6]

Modern control introduced state-space methods that marked a return to analysis in the time domain, in contrast to the frequency-domain methods characterizing classical control. The latter is fine for time-invariant, linear systems but cannot describe easily time-varying, nonlinear dynamics, which is omnipresent in applications. Although "modern," the state-space approach reaches back to the late nineteenth and early twentieth centuries. and includes figures such as Aleksandr Lyapunov in Russia and Henri Poincaré in France. A key insight was that knowing the system dynamics could improve performance spectacularly relative to the classical methods, which were developed assuming much less about the system under control. The resulting *optimal control* gave a systematic way to generate "the best" controller for a given task. With key contributions from Richard Bellman and Rudolph Kalman in the US and Lev Pontryagin in the Soviet Union, optimal control had spectacularly successful applications in the space program, particularly the Apollo moon-landing project.

---

[5] The RIAS was absorbed into the Martin Marietta Corporation, which survives as Lockheed Martin.

[6] Obviously, this use of "modern" is dated, as is "modern physics" (relativity and quantum mechanics), "modern art" (Impressionism, Dada, etc.), and "modern architecture" (Bauhaus, International Style, etc.).

The *digital computer* had a long gestation that was greatly advanced by war efforts – e.g., to formulate tables to aid in fire control. The history of computers is a separate story; in control, there was a gradual shift from *analog controllers* to *digital controllers*. The former had been implemented by external electrical, hydraulic, or mechanical circuits. Then came a long evolution to digital mainframes, minicomputers, microcomputers, laptops, and microcontrollers. In parallel came a shift from analog control methods for continuous-time dynamical systems to digital control methods for discrete dynamical systems.

At MIT in World War II, Norbert Wiener, introduced the stochastic analysis of control problems at roughly the same time and independently of efforts in the Soviet Union led by Andrei Kolmogorov. Wiener's primary technical publication, *The Extrapolation, Interpolation, and Smoothing of Time Series with Engineering Applications*, was circulated as a classified report in 1942 and eventually published in 1949.[7] His famous 1948 contribution, *Cybernetics: or Control and Communication in the Animal and the Machine*, showed that control theory applied not only to engineering systems but also to human, biological, and social systems. The book was inspired by the notion of homeostasis in organisms and by similar issues in controlling complex systems. Coined by Wiener from the Greek word for "governance," the word "cybernetics" was a tribute to Maxwell's 1868 governor paper.

The successes of optimal- and stochastic-control methods during the 1960s soon led to overconfidence, as it was forgotten how much the optimized performance depends on knowing system dynamics. In the 1970s and 1980s, control theory underwent something of an identity crisis. While state-space methods work well in the aerospace industry, where dynamics can be known accurately, they do poorly in industrial settings where the dynamics are more complex and harder to characterize (paper mills, chemical plants, etc.). This disenchantment led many practical engineers (and physicists) to avoid advanced techniques in favor of the tried-and-true PID controller. Indeed, academic research on control theory from 1960 through at least the 1970s had "negligible" impact on industrial applications. In response came a new subfield, *robust control*, to optimize the performance of systems whose underlying dynamics had at least moderate uncertainty. Its goal was to merge the robustness of classical control methods with the performance of modern control. In parallel came the subfields of *system identification* and *adaptive control*, where the goal was to *learn* better the system dynamics, either through independent or online measurements. Common applications of adaptive algorithms include noise-cancelling headsets, automobile cruise control, and thermostats.

1.1.5 Contemporary Control (after 2000)

Beginning around 2000, control theorists began to tackle increasingly complex systems. One notable example is the attempt to understand biological systems from

---

[7] Its nickname "Yellow Peril" came from the color of its cover, difficulty of its contents, and racism of its times.

an engineering perspective emphasizing control especially. The resulting field of *systems biology* contrasts, occasionally sharply, with a parallel effort in physics known as *biological physics*. Perhaps the most striking conclusion concerning control is strong empirical evidence that organisms have *internal models* of the world that allow them to anticipate and plan ahead. Recognizing the role of such planning and anticipation has been key to understanding how humans move and act in the world. See Example 3.4 for an application in connection with human balance – how we manage to stand upright without falling down. Later, in Chapter 10 we introduce *reinforcement learning*, a technique for learning how to plan and anticipate from repeated supervised trials.

Another example is the development of *autonomous*, self-driving vehicles. Indeed, the development of the automobile recapitulates the entire history of control. The early twentieth-century automobile was a mechanical device, like the governor and steam engine. In the 1970s, a variety of electrical control systems appeared, many based on microcontrollers that implemented feedback loops. By 2007, the typical automobile had 20–80 microprocessors, dealing with powertrain control to reduce emissions (e.g., by controlling the air-fuel ratio), performance optimization (e.g., variable cam timing), and driver assistance (e.g., cruise control and antilock brakes), and more. And, while the driver – the “human in the loop” – remains the ultimate controller for an automobile, many responsibilities are off-loaded (e.g., GPS and its associated navigational aids). At present, many companies seek to eliminate the driver from these control loops, a goal that must integrate many subproblems and use techniques from fields such as machine learning, big data, and wireless communications. Finally, control is expanding beyond the scale of single vehicles. Highways and smart phones already give real-time information on traffic for more efficient routing. In the future, platoons of trucks may travel in closely spaced groups that reduce traffic and increase fuel efficiency by controlling the collective air flow around the group (*drafting*).

Another application is to *climate science* and models of climate change, where it is crucial to understand feedbacks on both fast and slow timescales. On the one hand, water vapor is an effective greenhouse gas that is a “fast feedback” because the amount of water vapor in the air adjusts within days to changes in temperature. On the other hand, the area of land covered by glaciers and ice sheets adjusts much more slowly. (Glaciers melt, exposing darker surfaces, which absorb more sunlight.) Such *positive feedback* can lead to instability that will drive a dynamical system to another attractor (a new steady state or, sometimes, an oscillatory one). Negative feedbacks occur in climate models, too. As warmer temperatures lead to greater cloud cover, more light will be reflected away by the clouds, lessening absorption. Unfortunately, positive feedbacks seem likely to dominate.

The desire to understand complex systems has led to a discipline of *network science*. In the context of control, the goal is to understand collective network dynamics rather than individual dynamical systems. One focus has been to understand how the structure of a network affects one’s ability to control it. Applications are widespread, as networks are everywhere, from the world-wide web to the proteins that control cellular