Control: A Perspective

K. J. Åström and P. R. Kumar

Abstract

Feedback is an ancient idea, but feedback control is a young field. Nature long ago discovered feedback since it is essential for homeostasis and life. It was the key for harnessing power in the industrial revolution and is today found everywhere around us. Its development as a field involved contributions from engineers, mathematicians, economists and physicists. It is the first systems discipline; it represented a paradigm shift because it cut across the traditional engineering disciplines of aeronautical, chemical, civil, electrical and mechanical engineering, as well as economics and operations research. The scope of control makes it the quintessential multidisciplinary field. Its complex story of evolution is fascinating, and a perspective on its growth is presented in this paper. The interplay of industry, applications, technology, theory and research is discussed.

Key words: Feedback, control, computing, communication, theory, applications.

1 Introduction

Nature discovered feedback long ago. It created mechanisms for and exploits feedback at all levels, which is central to homeostasis and life. As a technology, control dates back at least two millennia. There are many examples of control from ancient times [460]. Ktesibios (285–222 BC) developed a feedback mechanism to regulate flow to improve the accuracy of water clocks. In the modern era, James Watts’s use of the centrifugal governor for the steam engine was fundamental to the industrial revolution. Since then, automatic control has emerged as a key enabler for the engineered systems of the 19th and 20th centuries: generation and distribution of electricity, telecommunication, process control, steering of ships, control of vehicles and airplanes, operation of production and inventory systems, and regulation of packet flows in the Internet. It is routinely employed with individual components like sensors and actuators in large systems. Today control is ubiquitous in our homes, cars and infrastructure. In the modern post-genomic era, a key goal of researchers in systems biology is to understand how to disrupt the feedback of harmful biological pathways that cause disease. Theory and applications of control are growing rapidly in all areas.

The evolution of control from an ancient technology to a modern field is a fascinating microcosm of the growth of the modern technological society. In addition to being of intrinsic interest, its study also provides insight into the nuances of how theories, technologies and applications can interact in the development of a discipline. This paper provides a perspective on the development of control, how it emerged and developed. It is by no means encyclopedic. To describe the field, we have, somewhat arbitrarily, chosen the years 1940, 1960 and 2000 as separators of four periods, which are covered in sections with the titles; Tasting the Power of Feedback Control: before 1940, The Field Emerges: 1940–1960, The Golden Age: 1960–2000, and Systems of the Future: after 2000. We provide a reflection on the complexity of the interplay of theory and applications in a subsequent section.

It was only in the mid 20th century that automatic control emerged as a separate, though multidisciplinary, discipline. The International Federation of Automatic Control (IFAC) was formed in 1956 [513,443,354], the first IFAC World Congress was held in Moscow in 1960,
and the journal Automatica appeared in 1962 [42,149]. By 2000 IFAC had grown to 66 Technical Committees. As a key enabler of several technological fields, control is quintessentially multidisciplinary. This is clearly reflected in the diverse organizations, AIAA, AIChE, ASCE, ASME, IEEE, ISA, SCS and SIAM that are included in the American Automatic Control Council (AACC) and IFAC.

There is yet another sense in which control has been multidisciplinary — in its search for theories and principles physicists, engineers, mathematicians, economists, and others have all contributed to its development. The physicist Maxwell laid the theoretical foundation for governors [456]. Later, one of the first books [343] was written by a physicist, a mathematician and an engineer. The mathematicians Richard Bellman [59], Solomon Lefschetz [269], and L. S. Pontryagin [534] contributed to the early development of modern control theory. Indeed, respect for mathematical rigor has been a hallmark of control systems research, perhaps an inheritance from circuit theory [272,95].

Control theory, like many other branches of engineering science, has developed in the same pattern as natural sciences. Although there are strong similarities between natural and engineering science, there are however also some fundamental differences. The goal of natural science is to understand phenomena in nature. A central goal has been to find natural laws, success being rewarded by fame and Nobel prizes. There has been a strong emphasis on reductionism, requiring isolation of specific phenomena, an extreme case being particle physics. The goal of engineering science, on the other hand, is to understand, invent, design and maintain man-made engineered systems. A primary challenge is to find system principles that make it possible to effectively understand and design complex physical systems. Feedback, which is at the core of control, is such a principle. While pure reductionism has been tremendously successful in natural science, control is more complex since interaction is a key element of engineered systems.

Many overviews of control have been presented in connection with various anniversaries. IFAC held a workshop in Heidelberg in September 2006 to celebrate its 50th anniversary [324]. Automatica celebrates its 50th anniversary in 2014 [149]. A comprehensive overview of sensors and industrial controllers was published on the 50th anniversary of the International Society of Automation (ISA) [632]. The American Society of Mechanical Engineers (ASME) published a series of papers on the history of control in connection with the 50th anniversary of the Journal Dynamic Systems, Measurement, and Control in 1993 [552]. The IEEE Control Systems Society sponsored the reprint of 25 seminal papers on control theory, selected by an editorial board [50]. The European Journal of Control published a special issue: On the Dawn and Development of Control Science in the XX-th Century in January 2007, in which researchers reflected on their view of its development [82]. A special issue on the history of control systems engineering [43] was published in 1984 at the centennial of IEEE. The IEEE Control Systems Society organized a workshop on the Impact of Control: Past, Present and Future in Berchtesgaden, Germany, in 2009. Material from the workshop was combined with an extensive collection of success stories and grand challenges in a comprehensive report [576]. The National Academy of Engineering published two studies about the future of engineering at the turn of the century [488,489]. They point out the growing importance of systems and the role of modeling and simulation for computer based design and engineering. The US Air Force Office of Scientific Research (AFOSR) sponsored a panel to study future directions in control, dynamics and systems, which resulted in a comprehensive report [485], summarized in [486].

The field of control is even attracting the attention of historians, perhaps an indication that it has had a complex development process that needs to be brought to light. There are books on the history of control [65,66,80], on individual researchers [318], and on organizations and projects [448,472,471]. There are sessions on the history of the field at many control conferences.

Paradoxically, in spite of its widespread use, control is not very much talked about outside a group of specialists; in fact it is sometimes called the “hidden technology” [30]. One reason could be its very success which makes it invisible so that all the attention is riveted to the end product device. It is more difficult to talk about ideas like feedback than to talk about devices. Another reason is that control scientists have not paid enough attention to popular writing; a notable exception is the 1952 issue of Scientific American which was devoted to Automatic Control [657,122].

By 1940 control was used extensively for electrical systems, process control, telecommunication and ship steering. Thousands of governors, controllers for process control, gyro-compasses and gyro-pilots were manufactured. Controllers were implemented as special purpose analog devices based on mechanical, hydraulic, pneumatic and electric technology. Feedback was used extensively to obtain robust linear behavior from nonlinear components. Electronic analog computing was emerging; it had originally been invented to simulate control systems [307]. Communication was driven by the need for centralized control rooms in process control and fire control systems. The benefits derived from the power of control were the driving force.

Although the principles were very similar in the diverse industries, the commonality of the systems was not widely understood. A striking illustration is that features like integral and derivative action were reinvented and patented many times in different application fields.
The theoretical bases were linearized models and the Routh-Hurwitz stability criterion. A few textbooks were available [644,351]. Research and development were primarily conducted in industry.

Control was an established field by 1960 because of its development during the Second World War. Servomechanism theory was the theoretical foundation. Tools for modeling from data, using frequency response, together with methods for analysis and synthesis, were available. Analog computing was used both as a technology for implementation of controllers and as a tool for simulation. Much of the development had been driven by requirements from applications and practical needs. After a long and complex evolution there finally had emerged a holistic view of theory and applications, along with many applications in diverse fields. Control systems were mass produced, large companies had control departments, and there were companies which specialized in control. An international organization IFAC had been created, its first World Congress was held in Moscow in 1960. Most of the research and development had been done in research institutes, and industries with collaborations with a few universities. By 1960 more than 60 books on control had been published.

However, many changes began occurring around 1960; the digital computer, dynamic programming [59], the state space approach to control [360], and the linear quadratic regulator [358] had appeared, with the Kalman filter just around the corner [359]. There commenced a very dynamic development of control, which we have dubbed the Golden Age. There were challenges from the space race and from introduction of computer control in the process industry as well as in many other applications such as automobiles and cellular telephones. There was a rapid growth of applications and a very dynamic development of theory, and many subspecialties were developed. University education expanded rapidly both at the undergraduate and the graduate levels. One consequence was that the parity that had been achieved between theory and practice after many decades was once again breached, this time in the reverse direction. Pure theory seized attention to a significant extent and there emerged a perception among some that there was a “gap” [41], and that the holistic view had been lost [73].

It is of course difficult to have a good perspective on recent events but our opinion is that there are indications that yet another major development and spurt is now in progress. By around 2000, there had occurred a phase transition in technology, due to the emergence and proliferation of wireline and wireless networking, and the development of sensors, powerful computers, and complex software. At the turn of the century there were therefore new challenges; control of networks and control over networks, design of provably safe embedded systems, and autonomy and model based design of complex systems. The dramatic growth in technological capabilities thus provided many opportunities but also presented many challenges that require a tight integration of control with computer science and communication. This recognition led to the creation of many major research programs such as ARTIST2 [20] and ArtistDesign [21] focused on embedded systems in EU, and Cyber-Physical Systems in the US [44].

Closer interaction with physics, biology and medicine is also occurring. Control is a key ingredient in devices such as adaptive optics and atomic force microscopes. Control of quantum and molecular systems is being explored. The need and interest in using ideas from systems and control to obtain deeper insight into biological systems has increased. The field of systems biology has emerged and groups with control scientists and biologists have been created; noteworthy are the departments of bioengineering in engineering schools.

2 Tasting the Power of Feedback Control

In order for the industrial revolution to occur, it required power, and control was essential to harness steam power. Therefore a major development of control coincided with the industrial revolution. Feedback control was a powerful tool. It made it possible to reduce the effect of disturbances and process variations, to make good systems from bad components, and to stabilize unstable systems. The major drawback was that feedback could cause instabilities. Recognition and solution of these problems led to major advances in control. As the industrial revolution progressed, the emergence of new technical innovations and industries made control an essential and central part of the electrical, chemical, telephone and other industries. The evolution of control and industry have been strongly connected ever since.

2.1 The Centrifugal Governor

The need for control devices appeared already in the operation of windmills. The centrifugal governor, which dates back to 1745, was invented to keep windmills running at constant speed [460]. Similar requirements appeared when steam power was used in the textile industry to keep looms and other machines running at constant speed. James Watt successfully adapted the centrifugal governor to fit the steam engine and patented it in 1788. The centrifugal governor combines sensing, actuation and control.

Designing a governor was a compromise; heavy balls are needed to create strong actuation forces but they also result in sluggish response. Other practical difficulties were created by friction and backlash in the mechanical devices. The basic governor yields proportional action because the change in the angle is proportional to
the change in velocity. Such a governor gives a steady state error. A controller with integral action has the remarkable property that it always approaches the correct steady state if the closed loop system is stable. Integral action was introduced around 1790 in a governor designed by the Périer brothers. They used a hydraulic device where the inflow to a vessel was proportional to the velocity and the steam valve was driven by the level [460, p 110–113]. In 1845 Werner and William Siemens introduced integral action by using differential gears [65, p 21-22]. The Siemens brothers also introduced derivative action based on an inertia wheel. The governor became an integral part of all steam engines. The governor was further developed over a 200 year period stretching from late 1700, as is well described in [65].

Theoretical investigation of governors started with the paper by Maxwell [456]. He analyzed linearized models and demonstrated the benefits of integral action. He also found that the stability of the closed loop system could be determined by analyzing the roots of an algebraic equation. Maxwell derived a stability criterion for third order systems and turned to his colleague Routh who solved the general problem [569]. Vyshnegradskii analyzed a steam engine with a governor independently of Maxwell [670], and also developed a stability criterion for third order systems. Vyshnegradskii’s results were engineering oriented and strongly coupled to the design of governors. He had been trained as a mathematician, and was director of St. Petersburg’s Technological Institute, where he pioneered courses in machine-building with a strong science base. He ended his career as Minister of Finance of the Russian empire [14].

Vyshnegradskii’s results were used by Stodola [626] to design water turbine governors. He used more complicated models and turned to his colleague Hurwitz at Eidgenössische Technische Hochschule (ETH), Zurich for help with stability analysis. Hurwitz developed a general stability criterion using techniques other than used by Routh [320]. Today we know the result as the Routh-Hurwitz criterion. Stodola also introduced dimension free variables and time constants. Interesting perspectives on the work of Maxwell and Vyshnegradskii are given in [548,14,65,77]. There was little interaction between the scientists [239, p 172–173]. Routh and Hurwitz were not aware of each other’s contributions and they used different mathematical techniques [445]. Stodola only mentioned Routh in his later papers [14].

At the beginning of the 19th century there was a firmly established engineering base for controlling machines with governors. Many companies invented and manufactured governors. According to Bennett [65] (page 74), there were more than 75000 governors installed in England in 1868. Proportional, integral and derivative actions were understood and implemented by mechanical or hydraulic devices. The theoretical foundation was based on work by Maxwell and Vyshnegradskii and the Routh-Hurwitz criterion. Education in control started at a few universities. Tolle compiled the results in a textbook [644] “Der Regelung der Kraftmaschinen (Control of Power Machines)” in 1905. Analysis and design were based on linearization and examination of the roots of the characteristic equation. The aerodynamicist Joukowski at Moscow University published the first Russian book [351] on control, “The Theory of Regulating the Motion of Machines,” in 1909.

2.2 Generation and Transmission of Electricity

The electric power industry emerged in the late 19th century and grew rapidly at the beginning of the 20th century. Electricity was generated by water turbines or by boiler-turbine units, and was originally distributed locally with DC networks. Control of turbines and boilers were major application areas as discussed in Section 2.1. While the early development of governors was largely empirical, the demands from the electric industry required a more systematic approach, and theory started to be developed and applied. Vyshnegradskii’s paper [670] had a strong influence on engineering practice and was widely used in control systems for turbine control [14]. Tolle’s book [644] was reprinted in 1909 and 1929, and remained a standard work on control of electrical machines for a long time.

New control problems emerged when the distance between generation and consumption increased. Many generators were connected in large networks to supply sufficient power and to increase reliability. Challenges arose when the electrical networks expanded. The generators had to all be synchronized when the transmission switched from DC to AC. Stability problems were encountered in the control of frequency and voltage. For safe operation it was necessary to understand the response of generators to disturbances such as load changes, faults, lightning strikes, etc. Charles Steinmetz had developed the foundations of alternating current analysis, with his introduction of “complex imaginary quantities” and phasors in the late eighteen hundreds [624]. This work addressed only steady-state behavior and could not deal with dynamics. Motivated by this, Harold Hazen, working under Vannevar Bush at Massachusetts Institute of Technology (MIT), built a “network analyzer” in the late 1920s. The analyzer was a laboratory model of a power network, built using phase-shifting transformers and transmission lines, and was reconfigurable using a plug board from a telephone exchange. The system was replicated at General Electric and other power companies.

The emergence of the electrical industry was a game changer because it was developed by large industries in collaboration with public and state utilities which were large actors. Due to the requirement of operating large networks safely, utilities and electric compa-
cies built groups for research and development to understand, design and operate them. Research and development teams were created in companies like General Electric, Westinghouse, ASEA, BBC, Alstom, and in many public and state utilities like ENEL, EDF and the Swedish State Power Board in Europe. One example is the General Electric Research Laboratory that was created by Thomas Edison, Willis R. Whitney, and Charles Steinmetz in 1900. It was the first industrial research laboratory in the US.

### 2.3 Industrial Process Control

Automation of process and manufacturing industries evolved from the late 19th century and accelerated at the beginning of the 20th century. The production processes in the chemical, petroleum, pulp and paper, and pharmaceutical industries required accurate control of pressure, temperature and flow to achieve good product quality. Boilers, reactors, distillation columns, mixers and blenders were the typical processes. A wide variety of sensors for different physical variables was developed. The actuators were typically valves and pumps. Pneumatics became the common technology to implement the sensing, actuation and control functions. Sensors and actuators had to be located at the process. Originally the controllers were attached to the process equipment; they communicated with sensors and actuators via pressure signals. The connectors, the pressure tubes and the signal levels (3 to 15 psi) were standardized, permitting equipment from different vendors to combined. The controllers were later moved to be combined a central control room where recorders for signals were also provided, greatly simplifying the work and the working environment of the operator.

Development of the controllers was driven by engineering insight rather than theory. The effects of integral and derivative action were rediscovered by tinkering. An interview [91] with John Ziegler, from the Taylor Instrument Company, provides a perspective:

*Someone in the research department was tinkering with the Fulscope (a pneumatic PI controller) and somehow had got a restriction in the feedback line to the capsule that made the follow-up in the bellows. He noted that this gave a strange kicking action to the output. They tried it on the rayon shredders and it gave perfect control of the temperature.*

The controller components were also used as pneumatic analog controllers to simulate processes. Since the simulator used pneumatic signals it could easily be connected to a pneumatic controller. Feedback was used extensively in the sensors and actuators, and the controllers themselves. The key idea was to create good linear behavior by combining passive components in the form of volumes with restrictions with pneumatic amplifiers that had high gain, very similar to the feedback amplifiers discussed later in Section 2.6.

The controllers became standardized general purpose devices, not built for a specific process like the governor, and they were equipped with dials that permitted adjustment of the parameters of the PID controller. The first general purpose PID controller was the Stabilog developed by Foxboro; the gain could be adjusted between 0.7 and 100. It appeared in 1931, soon after other manufacturers developed similar products. Since there could be many controllers in a process, there was a need for methods for finding good values of the controllers for different processes. Ziegler and Nichols [703] developed tuning rules where the controller parameters could be determined from simple experiments on the process.

The emergence of sensors, instruments and controllers led to the creation of new companies. The industry was highly diversified, and by mid 1930 there were more than 600 control companies, with Bailey, Brown, Fisher & Porter, Foxboro, Honeywell, Kent, Leeds & Northrup, Siemens, Taylor Instruments, and Yokogawa, among the leading ones [632]. Bennett ([66], p 28) estimates that about 75000 controllers were sold in the US in the period 1925-35.

The industrial structure for process control differed from that in the communications and power industries. Ideas were not disseminated but protected as proprietary secrets. In process control there were a large number of companies. Concentrated resources that were available in the communications and power industries were lacking as was a theoretical foundation.

### 2.4 Ship Steering

There were many interesting developments in ship steering. Actuation was a major issue because of the large forces required to turn the rudder of a large ship. The word “servo motor” was coined by the French engineer Farciot who developed hydraulic steering engines [65]. These devices, which provided actuation, were important ingredients in the automation of ship steering. Control also benefited from advances in sensors.

Major advances in ship steering were inspired by exploitation of gyroscopic action. The collection of ideas and devices based on gyroscopic action had a major impact, and has been labeled “the gyro culture” [448,472,471].

The first gyro compass was developed by Anschütz-Kaempfe who started the company Anschütz in 1905. The company collaborated with Max Schuler who was head of the Institute of Applied Mechanics at the University of Göttingen. Schuler invented a clever technique to make the gyro compass insensitive to the motion of...
the ship (Schuler tuning) [586]. Schuler also taught a control course at the university [585,450].

In 1910 Sperry started the Sperry Gyroscope Company to develop a gyro compass and many other devices based on gyroscopes. The company Brown also developed a gyro compass a few years later, and there were court battles with Anschütz about intellectual property rights [448]. Sperry combined the gyro compass with an electric motor connected to the steering wheel to obtain a gyro-pilot. By observing experienced pilots Sperry had found that:

An experienced helmsman should ‘meet’ the helm, that is, back off the helm and put it over the other way to prevent the angular momentum of the ship carrying it past its desired heading.

Sperry tried to create an electro-mechanical device with this behavior. The design, which is well documented in [65,318,472], is a typical PID controller; the function of meeting the helm is obtained by derivative action. Integral action is obtained by the motor which drives the steering wheel. Amplification was often based on on-off devices and feedback is exploited to obtain linear behavior. Sperry’s gyro-pilot relieved the helmsman of the tedious job of adjusting the rudder to keep the course. The gyro-pilot had adjustments to set the desired course, and to change the controller parameters. There was also a lever to connect and disconnect it. Sperry’s gyro-pilot, which was nicknamed the “Metal-Mike,” was very successful. Sperry also provided recorders so that the steering errors of automatic and manual control could be compared. In 1932 there were more than 400 systems installed [318].

There were interesting theoretical developments in ship steering by Minorsky [473] who was educated at St. Petersburg Imperial Technical Institute. He presented a taxonomy of controllers and recommended the use of PID control for ship steering. His design method based on a simplified linear model is what is today called pole placement. Minorsky built an autopilot which was tested, but it did not lead to a product, and he sold his patents to Bendix [66]. Later Minorsky became a professor at Stanford University and wrote a book on nonlinear oscillations [474]. In Bennett’s book [65, p 147-148] and in [68], there are interesting discussions of the contributions of Sperry, Minorsky and Anschütz, and their impact on actual auto-pilot design.

New control problems in ship steering appeared in the First World War in connection with the intensive program for modernization of the navies [52]:

Touched off by the gyro-compass and its repeaters of data transmitters, the possibilities of transmitting target bearings, turret angles, true azimuth, and ship’s heading automatically from topside to plotting rooms to guns opened a vast new field.

The orientation and distance to the target were measured by optical devices, typically observers aft and forward on the ship. The future position of the target was computed and the large gun turrets was oriented by servos. Self-synchronous motors (synchros) transmitted the information from the optical devices to the computer, and from the analog computer to the servos. The computers were analog electro-mechanical devices using wheel and disk integrators [472]; they were manufactured by the Ford Instrument Company, General Electric, and Sperry.

2.5 Flight Control

There were many experiments with manned flight in the 18th century. One reason why the Wright brothers succeeded was that they understood the relations between dynamics and control. Wilbur Wright expressed it in the following way when lecturing to the Western Society of Engineers in 1901 [461]:

Men already know how to construct wings ... Men also know how to build engines ... 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.

By combining their insight with skilled experiments, the Wright brothers made the first successful flight in 1905. An interesting perspective on their success is given in the 43rd Wilbur Wright Memorial Lecture delivered by Charles Stark Draper at the Royal Aeronautical Society on May 19, 1955 [175]:

The Wright Brothers rejected the principle that aircraft should be made inherently so stable that the human pilot would only have to steer the vehicle, playing no part in stabilization. Instead they deliberately made their airplane with negative stability and depended on the human pilot to operate the movable surface controls so that the flying system - pilot and machine - would be stable. This resulted in an increase in maneuverability and controllability.

The fact that the Wright Flyer was unstable stimulated development of autopilots [318]. Sperry used his understanding of gyroscopes and autopilots for ships to design an autopilot for airplanes. The deviations in orientation were sensed by gyroscopes, and the rudder and ailerons were actuated pneumatically. There was a spectacular demonstration of the autopilot in a competition in Paris in 1912. Sperry’s son Lawrence flew close to the ground with his hands in the air while his mechanic walked on
the wing to demonstrate that the autopilot could cope with disturbances.

The success of the Wright brothers is an early example of what we today call integrated process and control design. The key idea is that automatic control gives the designer extra degrees of freedom. The Wright Brothers made a maneuverable airplane and relied on the pilot to stabilize it. Mjorsky was well aware of these issues. He captured it in the phrase [473]: *It is an old adage that a stable ship is difficult to steer.* It is interesting to observe that modern high performance fighters are designed to be unstable; they rely on a control system for stabilization.

There was also a strong culture in Germany associated with development of autopilots [516]. Lufthansa had international long distance flights in the 1920s. There was a demand for directional control to fly safely in all weather conditions. The German companies Askania, Siemens and Möller-Patin developed autopilots that competed with Sperry’s equipment.

Sperry continued to develop autopilots; a refined model A-2 used air-driven gyroscopes and pneumatic-hydraulic actuators. A spectacular demonstration of the benefits of autopilots was when the Sperry A-2 autopilot was used in Wiley Post’s solo flight around the World in 1933. Airlines started to introduce autopilots in the early 1930s and companies like Bendix and Honeywell started to make autopilots.

The autopilots made extensive use of feedback both in the individual components and in the systems. Although there was a good theoretical understanding of flight dynamics based on linearized equations and analysis of the characteristic equation as early as 1911, the theoretical work did not have any impact on practical autopilot design until the mid-1950s [464]. One reason was lack of computational tools. As in the case of ship steering, engineering ability was more important than theory.

### 2.6 Long Distance Telephony

Graham Bell patented the telephone in 1876. Originally the phones were connected with wires to a central location with a switchboard. The number of phones grew rapidly. Many phone calls were transmitted over the same wire using frequency separation. The telephone industry was highly centralized, more so than the electric industries because it was driven by private or state monopolies and large industries.

One driver of communications, and indirectly but profoundly of control, was the growth of transcontinental telephony in the USA [472]. Around 1887, Oliver Heaviside showed that adding inductance to lines could be used to reduce distortion. In 1899, Mihajlo Pupin of Columbia University patented the loading coil [550], while, at about the same time, George Campbell of AT&T, developed it and implemented it on a telephone cable in 1900. This was subsequently used in long distance lines and cables. Transmission of signals over long distances was however passive, and the loading coil technique reached its limits in terms of allowable distortion and attenuation around 1911 with its implementation in the New York – Denver line. In 1913, AT&T bought the rights to the triode which Lee de Forest [421] had invented in 1907, and had it further studied and developed by Harold Arnold. It used eight repeaters (amplifiers) to connect New York and San Francisco, extending the line from Denver to California. The number of repeaters increased as more cities were interconnected, but distortion then became a major problem, as noted by Bode [95]:

> Most of you with hi-fi systems are no doubt proud of your audio amplifiers, but I doubt whether many of you would care to listen to the sound after the signal had gone in succession through several dozen or several hundred even of your fine amplifiers.

Consequently, there was great impetus to increase the capacity of telephone lines by using carrier multiplexing, which together with the employment of cable, greatly increased the number of repeaters needed. This required high quality amplifiers with low distortion. The electronic tube was the prime device for amplification at the time, but it had severe drawbacks such as a nonlinear characteristic that changed with time. There were many efforts but no real progress was made until Harold Black of Bell Labs developed the negative feedback amplifier in 1927 [83]. The critical idea was to provide an amplifier with feedback via passive linear elements to reduce the distortion in amplifiers. We quote from Bode [95]:

> The causes of distortion were of various sorts. They included power supply noises, variations in gain and so on, the dominant problem, however, was the intermodulation due to the slight nonlinearity in the characteristics of the last tube. Various efforts were made to improve this situation, by the selection of tubes, by careful biasing, by the use of matched tubes in push-pull to provide compensating characteristics, and so on. Until Black’s invention, however, nothing made a radical improvement of the situation.

It should be noted that Black used the word “stable” to describe constancy of the amplifier gain in spite of temperature changes, rain, weather, component aging, etc., but not its immunity to “singing” or oscillation [83]. Feedback was an enabler which made it possible to make a good amplifier even while employing components with many undesirable features. A perspective on the invention is given in Black’s paper [84], which was written 50 years after the invention:
I suddenly realized that if I fed the amplifier output back to the input, in reverse phase, and kept the device from oscillating (singing, as we called it then), I would have exactly what I wanted: a means of canceling out the distortion in the output. ... By building an amplifier whose gain is deliberately made, say 40 decibels higher than necessary and then feeding the output back on the input in such a way as to throw away the excess gain, it had been found possible to effect extraordinary improvement in constancy of amplification and freedom from non-linearity.

It took nine years for Black’s patent to be granted, partially because the patent officers refused to believe that the amplifier would work. They did not believe that it was possible to have a stable feedback loop with a loop gain of several hundred [84].

Instability or “singing” was frequently encountered when experimenting with feedback amplifiers. Thus the technological challenge of long distance telephonic communication led to the issue of stability of the feedback loop. Harry Nyquist encountered the problem in 1932, when he participated in a joint project with Black to test the negative feedback amplifiers in a new carrier system. To address this, Nyquist used ideas that were very different from the stability results of Maxwell and Vyshnegradskii. Instead of analyzing the characteristic equation, he explored how sinusoidal signals propagated around the control loop, resulting in the “Nyquist criterion” [507]. Stability of electronic amplifiers was independently investigated by Kupfmüller [402]. He introduced signal-flow diagrams and analyzed the circuits using integral equations [517].

The performance requirements of communication required further advances in the design of feedback loops. While working on the design of an equalizer network in 1934, Hendrik Bode developed a deep insight into feedback amplifiers. He investigated the relationship between attenuation and phase and introduced the concepts of gain and phase margin and the notion of minimum phase [93]. He also proved that there are fundamental limitations to control system design. In particular he showed that the integral of the logarithm of the magnitude of the sensitivity function is constant, which means that control is inherently a compromise; making the sensitivity smaller for one frequency increases it at other frequencies. He also showed that there were even more stringent limitations if systems are not minimum phase. Bode also developed tools to design feedback amplifiers based on graphical methods (Bode plots) that we today call loop shaping. A particular difficulty was to deal with the large variations in the gain of the triode. He showed that a constant phase margin can be maintained for very large gain variations by shaping the loop transfer function so that its Nyquist curve is close to a straight line through the origin, which he called the “ideal cut-off characteristic.”

Bode’s design method was the first example of robust control. His results were based on the theory of complex variables and are summarized in the seminal book [94].

The AT&T Company started an industrial research laboratory as part of its strategy of controlling American telecommunications. To implement the strategy the company wanted to control the rate and direction of technical change by obtaining, or preventing others from obtaining, key patents. The research laboratories played a major part in ensuring that AT&T kept control of the technology and the patent rights [66, p 70-71]. The environment at Bell Labs, which had a mix of scientists like Bode, Shannon and Nyquist and engineers like Black, was a very fertile ground for technology development and basic research. The lab has had 13 Nobel Laureates. Insight into the personalities and the research environment is presented in Mindell’s book [472].

A major difference between the telecommunications industry and the other industries where control was used was that the industry was supported by a research laboratory with many qualified researchers. Theory was interleaved with the practical development, and repeaters for land lines and underwater cables were mass produced.

2.7 Early Electro-Mechanical Computers

It was recognized early on that computers could be used to simulate and thereby understand the behavior of dynamic systems in the absence of a mathematical solution. In fact, mechanical devices for integrating differential equations had been designed already in 1876 by William Thomson (Lord Kelvin) [639,640], who used a ball-and-disk integrator to perform integration. Motivated by the problems of simulating power system networks, Vannevar Bush improved the mechanical design significantly, and also designed a torque amplifier to avoid loading [524]. Bush’s first mechanical differential analyzer had six integrators [125]. The differential analyzer at MIT was used for a variety of applications beyond power systems.

A first step was the “product integraph,” a device for integrating the product of two functions [126], which was an important element in network analysis. This required human tracking of each of the input waveforms of the functions that then generated an electrical signal fed to a watt-hour meter whose output was a turning wheel. If the output of this calculation was to be used as the input to a next stage, then to avoid loading the wheel, a servomotor was used to replicate the movement. It served as the mechanical analog of the amplifier repeater in the telephone network. The next stage of evolution in 1931 was to feed the output signals of the integrators after the servos back to the inputs, which provided the capability to solve differential equations. Servomechanisms played
the crucial role in connecting the stages of computation. Thus control played a central role in the construction of this early electro-mechanical analog computer.

In turn, the development of this “computer” stimulated Hazen to pursue further work on servo-mechanisms [290]. However this work did not explicitly make the connection with the earlier work of Nyquist and Bode. It did however, cite the earlier work of Minorsky who had introduced the PID Controller in connection with steering of US Navy ships [473]. Early work on servomechanisms were also done at Bell Labs [100,446].

The next generation of the differential analyzer was the “Rockefeller Differential Analyzer,” which transmitted data electronically, and thus allowed reconfiguration of the system by resetting telephone switches rather than by the more time-consuming process of mechanically rotating shafts. This project was funded at MIT by Warren Weaver of the Rockefeller Foundation, a partnership which played a very important role in the subsequent development of anti-aircraft fire control. Punched paper tape could be used to “program” this computer, making it a “hybrid” digital/analog system. Motivated by this, Claude Shannon examined in his MIT Master’s Thesis the problem of switching circuits and showed how Boolean algebra can be used for design [597]. Subsequently, George Sibitz built on this work in making progress towards the digital computer. Shannon later investigated the class of problems that could be solved by the differential analyzer [598].

Copies of the differential analyzers were built by several universities and research organizations. Nichols used the differential analyzer at MIT when he developed the tuning rules for PID control [91]. The analog computer at the University of Manchester was used to analyze control of systems with time delays [130].

In 1938 George Philbrick of Foxboro invented an electronic analog computer called Polyphemus for simulation of process control systems [307]. This system was used extensively at Foxboro for training and demonstration. Analog computing would later have a major impact on control.

3 The Field Emerges

Control emerged as a discipline after the Second World War. Prior to the war it was realized that science could have a dramatic impact on the outcome of the war. Fire-control systems, gun-sights, autopilots for ships, airplanes, and torpedoes were developed. Significant progress was also made in process control. There was close collaboration between military agencies, industry, research labs, and university [517,472]. Engineers and researchers with experiences of control systems from different specialties were brought together in cross disciplinary teams. It was recognized that there was a common foundation for all control problems, even if the application areas were very diverse.

Fire control was one of the major challenges. Cities, factories, and ships needed guns to protect them from attacking aircraft. Radar was used as a sensor, electric or hydraulic motors were used to direct the guns. Communication was required, because the radar and the guns were separated physically. An additional difficulty was that the radar signal was noisy. Fire control for ships also had to deal with the motion of the ships. Early fire control systems used manual control which became infeasible when the speed of aircraft increased. Automated aiming was implemented using mechanical analog computers. Feedback was used extensively both at the system level and at the component level.

Germany had a strong tradition in control; Tolle’s textbook [644] appeared already in 1905. By 1940 here was control expertise at many companies Askania, Kreiselgeräte, Siemens, and at some universities. The VDI (Verein Deutscher Ingenieure, Association of German Engineers) had recognized the importance of control and they had organized a committee on control engineering in 1939, with Hermann Schmidt as a chairman and Gerhard Ruppel as a secretary.

Germany had severe restrictions on military activities in the Versailles treaty; for example, it was not allowed to have an airforce. In spite of this there were many secret projects. Auto-pilots for aircraft and missiles were developed [61,517]. The navy established a secret company “Kreiselgeräte (Gyro devices)” in 1926. The company played a central role in navigation and guidance throughout the Second World War [252,448]. Several companies manufactured autopilots in 1940. According to Oppelt [516], thousands of autopilots were produced every month. Siemens alone had manufactured 35,000 systems by the end of the war. The autopilots were based on gyroscopes and analog computing using pneumatic, hydraulic, and electro-mechanical technologies.

The German army secretly created a Ballistics Council to develop military rockets. The program, which was led by Walter Dornberger, started in 1930 and it was transferred to Peenemünde in 1937. At that time the group had about 90 people [385,61]. Guidance and control were critical elements. Several missiles were developed among them the cruise missile V-1 and the ballistic missile V-2. Much research and development was required for the guidance systems. Askania developed and produced the autopilot for V-1. The V-2 missile and its guidance system were developed by a team led by Wernher von Braun [61]. More than 8,000 V-1’s and 3000 V-2’s were launched during the war. The German rocket scientists subsequently went to the USA and the USSR.
after the war and played leading roles in the development of missile technology. The USSR launched the first artificial Earth satellite, Sputnik, in 1957, triggering the Space Race. The first rocket to reach the Moon was the Soviet Union’s Luna 2 mission in 1959.

Research in the USSR was highly centralized [403,79]. The Academy of Sciences directed the research and there were large engineering institutes for applications: Electrotechnical, Boiler and Turbine, Power Engineering, Naval and Aviation. The USSR had a long tradition in automatic control going back to Vyshegradskii, Joukowski, and Lyapunov, recall that Joukowski’s book was published in 1909. Control also benefited from a strong tradition in nonlinear dynamics with schools in Moscow, led by Mandelstam and Andronov [15], and in Kiev, led by Krylov and Bogolyubov [397]. A technical Commission on remote control and automation was created in 1934, with A. A. Chernyshov as chairman. An All-Union Conference on automatic control and dispatch design was organized in 1936 with about 600 participants [403]. The Institute of Automation and Remote Control was founded in Moscow in 1939 [16].

It became a power house for control systems research with many prominent researchers, including A. A. Andronov, M. A. Aizerman, A. A. Butkovsky, A. A. Feldbaum, N. N. Krasovskii, B. Ya. Kogan, A. Ya. Lerner, B. N. Petrov, V. V. Solodovnikov, Ya. Z. Tsypkin, and S. V. Yemelyanov. The institute published the journal Automatika i Telemekhanika, which was translated to English and widely read in the west. In 1944 Andronov organized a research seminar at the Institute of Automation and Remote Control with a group of very talented researchers [652]. He correctly predicted a grand era of automation [422]. Mathematicians like Pontryagin and Gamkrelidze from the Steklov Institute of Mathematics gave significant contributions (the Maximum Principle). There were also institutes in many other cities, for example Leningrad, Sverdlovsk, and Kiev.

In the US a group of scientists, including Karl T. Compton (president of MIT), James B. Conant (president of Harvard), and Frank Jewett (director of Bell Labs), led by Vannevar Bush, petitioned President Roosevelt to form an organization that could exploit scientific and technical expertise for the war effort [684, p 182–184]. The result was the formation of the National Defense Research Committee (NDRC) in 1940, with Bush as its chair. Within a year the NDRC became a part of the Office of Scientific Research and Development (OSRD). Its director Bush reported directly to the President. NDRC built on laboratories around MIT and Bell Labs. The Instrumentation Laboratory at had been created in 1930 [166,17] by Charles Stark Draper with the mission of making precise measurements of velocity and angular rate. Pioneering work on servomechanisms has been done by Harold Hazen in the 1930s [290,289]. In 1939 the US Navy requested a special course on servomechanism and fire control. The course was given by Gordon Brown who shortly after created the Servomechanisms laboratory [684, p 212–217]. NDRC also created the Radiation Laboratory at MIT, which at one time had about 4000 researchers. The laboratories had an interdisciplinary staff with a wide range of academic and industrial backgrounds [472]. There were fertile interactions between engineers and scientists in the different groups and engineers in industry [472,448].

The Bureau of Ordnance of the US Navy funded joint projects between the Servomechanism and Instrumentation Laboratories at MIT. Gordon Brown developed improved hydraulic systems to turn the turrets and Draper designed the Mark 14 gun-sight based on gyro. The gunsight was manufactured by Sperry, and more than 85000 systems were produced by the end of the war [472].

The Instrumentation Laboratory was renamed the Draper Laboratory in 1970 and became a not-for-profit research organization in 1973 the Servomechanism Laboratory remained as part of MIT and is now the Laboratory for Information and Decision Systems [477].

The Radiation Laboratory was dissolved after the war but it was decided to publish the research in a series of 28 volumes. We quote from the foreword to the series:

_The tremendous research and development effort that went into the development of radar and related techniques during World War II resulted not only in hundreds of radar sets for military use but also in a great body of information and techniques... Because this basic material may be of great value to science and engineering, it seemed most important to publish it as soon as security permitted. The Radiation Laboratory of MIT, ..., undertook the great task of preparing these volumes. The work described herein, however, is the collective result of work done at many laboratories, Army, Navy, university, and industrial, both in this country and in England, Canada, and other Dominions. ... The entire staff agreed to remain at work at MIT for six months or more after the work of the Radiation Laboratory was complete._

Most of the volumes deal with radar and microwaves
but at least two of them are highly relevant to control; Volume 27 Computing Mechanisms and Linkages and particularly Volume 25 Theory of Servomechanisms. Although there are earlier books on servomechanisms [100,283,279,446,94], Volume 25 [343] is unquestionably a landmark. The multidisciplinary nature of control is illustrated by the fact that the prime authors include Hubert James, a physics professor of Purdue, Nathaniel Nichols, director of research at Taylor Instrument Company, and Ralph Phillips, a professor of mathematics at University of Southern California. The book was followed by others written by authors from the Servomechanism Laboratory [121].

Before the outbreak of the war, research and development in control in the UK was carried out by the Admiralty Research Laboratory, the Royal Aircraft Establishment, the Telecommunication Research Establishment, the National Physical Laboratory and in industries in shipbuilding, chemical, and electrical industries [539,67]. A committee, under the chairmanship of Sir Henry Tizard, Rector of Imperial College London, was created in 1935 to examine the problem of the defense of Britain from air attack. Many schemes were explored and it was decided to focus on the development of radar. Successful experiments were carried out in late 1935. Working ground stations were available by 1940 and airborne station in in the spring of 1941 [684, p 193-194]. When Churchill became prime minister he selected Professor Frederick Lindemann (Viscount Cherwell) as his scientific advisor, and there were frequent conflicts between Tizard and Lindemann [608,147]. There was an extensive exchange of ideas and hardware with the USA [459] [684, p 195].

The Admiralty explored use of radar for naval gunnery. The development was done at companies like Metropolitan-Vickers where Arnold Tustin was one of the leading researchers. The company had experience in servo systems and analog computing because they had built a mechanical differential analyzer in 1935. Tustin also chaired a group of companies working for the Admiralty [78]. A Servo Panel was formed in 1942 with Hartree as a chairman and Porter as a secretary [539]. The mission of the panel was to exchange experiences of servo systems; Tustin and Whiteley were among the members. The Servo Panel was followed by a more formal organization the Interdepartmental Committee on Servomechanisms and Related Devices (ICSR) established by the Ministry of Supply in 1944. The mission was to follow research in the field, to advise the Ministry and to act as an advisory body on servomechanism to any firm engaged in Government work.

There were also activities in many other countries France [223], Italy [271], Japan [384] and China [139], and even in small countries such as Sweden [32] where the Research Institute of National Defense (FOA) was created in 1945.

### 3.1 The Development of Servomechanism Theory

The early work on control had a rudimentary theory based on linear differential equations and Routh-Hurwitz stability criterion. The frequency response method developed by Bode [93] and Nyquist [507] was a paradigm shift. Bode [95] expressed the differences between process control and telecommunications as follows:

> The two fields are radically different in character and emphasis. ... The fields also differ radically in their mathematical flavor. The typical regulator system can frequently be described, in essentials, by differential equations by no more than perhaps the second, third or fourth order. On the other hand, the system is usually highly nonlinear, so that even at this level of complexity the difficulties of analysis may be very great.... As a matter of idle curiosity, I once counted to find out what the order of the set of equations in an amplifier I had just designed would have been, if I had worked with the differential equations directly. It turned out to be 55.

The fire control problems were particularly challenging because they involved radar and optical sensing, prediction and servoing. Servomechanisms had been investigated early at MIT by Hazen in connection with work on network analyzers and Bush’s differential analyzer [290], as described in Section 2.7. By combining it with the ideas of Bode and Nyquist it was refined to a coherent method to analyze and design control systems at the Radiation Laboratory. Many applications centered around servo problems; typical examples were gun-sights and radar. One of the pioneers, Hall of MIT, expresses it as follows [280]:

> Real progress results from strong stimulation. ... The war brought three problems to the controls engineer. The first was handling problems and systems of considerable dynamics complexity dictated by the requirements of more accurate and rapid fire-control systems. The second was that of designing systems that would cope with large amounts of noise, occasioned by the use of radar in fire control. The third problem, raised by the guided missile, was that of designing so accurately a dynamic system that it could be used successfully almost at once with negligible field trials.

The key elements of servomechanism theory are block diagrams, transfer functions, frequency response, analog computing, stochastic processes and sampling. The mathematical foundation was based on linear systems, complex variables, and Laplace transforms.

A block diagram is an abstraction for information hiding, where systems are represented by blocks with inputs and outputs. The internal behavior of the systems is hidden in the blocks. The behavior of the blocks was
described by ordinary differential equations, or transfer functions derived using Laplace Transforms. A central idea was that relations between signals in a block diagram could be found by algebra instead of manipulation of differential equations, an idea which goes back to Heaviside. Block diagrams and transfer functions allowed a compact representation of complex systems. An important consequence was that the similarity of many different control systems became apparent because their block diagrams revealed that they had the same structure.

An important factor that significantly contributed to the success of servomechanism theory was that the transfer function of a system could be determined experimentally by investigating the response to sinusoidal inputs. In this way it was possible to deal with systems whose physical modeling was difficult. Control engineers were fearless in finding models of technical systems by injecting sinusoidal perturbations and observing the responses. An example is given in [8,510], where in an attempt to determine the dynamics of the Swedish power network, the full output of a major power station was used to perturb the system. Special frequency analyzers were also developed to generate sinusoidal signals and to compute the transfer functions.

Graphical design methods for controller design were based on shaping the frequency response of the loop transfer function (loop shaping). The design method yielded controllers in the form of rational transfer functions; they were not restricted to PID controllers. Compensators were often obtained as combinations of lead and lag networks. The limitations caused by process dynamics that are not minimum phase were apparent in the design procedure. The graphical representations of Bode and Nichols charts were easy for engineers to use since they also provided significant physical insight, as is illustrated by the following quote from an engineer in ASEA’s research department [528,242]:

We had designed controllers by making simplified models, applying intuition and analyzing stability by solving the characteristic equation. At that time, around 1950, solving the characteristic equation with a mechanical calculator was itself an ordeal. If the system was unstable we were at a loss, we did not know how to modify the controller to make the system stable. The Nyquist theorem was a revolution for us. By drawing the Nyquist curve we got a very effective way to design the system because we know the frequency range which was critical and we got a feel for how the controller should be modified to make the system stable. We could either add a compensator or we could use extra sensor.

The design methods were originally developed for systems with one input and one output; they could be extended to systems with several inputs and outputs by combining the Nyquist plots for different loops [243]. Disturbances are a key ingredient in a control problem; without disturbances and process uncertainties there is no need for feedback. Modeling of disturbances is therefore important. In servomechanism theory, it was proposed to model disturbances as stochastic processes [343,614,655]. The book [343] has formulas for calculating the mean square error for linear systems driven by stochastic processes. A key problem in fire control was to predict the future motion of an aircraft. Solutions to this problem were given independently by Wiener [683] and Kolmogorov [389]. The work had no impact on the fire control systems during the war [472, p 280–283]. Newton, Gould and Kaiser [500] used Wiener's prediction theory to design control systems that minimize mean square fluctuation. An interesting feature of their approach is that they converted the feedback problem to an equivalent feedforward problem, which was much easier to solve. Today we call this approach Youla parameterization. Other books on control of systems with random processes are [674,615,412,163].

The Semi-Automatic Ground Environment (SAGE), a semi-automatic system for detecting missiles approaching North America, was developed in the late 1950s at the Lincoln Laboratory [561]. The system consisted of a network of radar, computers and command and control centers. The scanning radar stations provided periodic samples of missile position; this spawned much research in sampled data systems. Significant contributions were made by Franklin and Jury in the control group at Columbia University led by Ragazzini [553,352]. There was also significant research on sampled data systems by Tustin in the UK [655], and by Tsypkin at the Institute of Automation and Remote Control in the USSR [651]. Earlier Oldenburger and Sartorius [511] had worked on sampling motivated by chopper-bar systems used in process control.

Since fire and flight control systems involved a human in the loop it was natural to investigate the dynamic characteristics of a human in a feedback loop [656,518,90, Chapter 13]. Partially inspired by this, Norbert Wiener coined the term Cybernetics (Control and Communication in the Animal and the Machine) in the book [682] published in 1948. Wiener emphasized interdisciplinary research, convergence of control, communication, biology and system theory. Ross Ashby explored the origin of the adaptive ability of the nervous systems in the book [26], resonating with the idea of cybernetics [27]. An engineering view of cybernetics was however given in Tsien's book Engineering Cybernetics [646], which anticipated much of the development of control after 1954. Cybernetics caught the imagination of both professionals and the public in general but eventually it fell into disrepute, perhaps because of a lack of any significant research outcome, over-promising, and over-exploitation. The word survived in some institutions, Yakubovich founded the Department of Theoretical Cybernetics in Leningrad in 1970. The control department
at the Norwegian Institute of Technology was named Teknisk Kybernetikk.

Information about servomechanisms was spread in many conferences leading to the formation of the International Federation of Automatic Control in 1956. The Department of Scientific and Industrial research in the UK arranged a conference in Cranfield in July 1951. The proceedings [658] was edited by Tustin, a central person in control research in the UK. Another conference was arranged by ASME in December 1953. The conference proceedings was edited by Rufus Oldenburger, Director of Research of the Woodward Governor Company, and it was dedicated to Harry Nyquist [514]. (The highest ASME award in control systems is the Oldenburger Medal.) The Italian National research council arranged a series of meetings in Milan culminating in an International Congress on the Problems of Automation in April 1956 with more than 1000 attendees [150].

3.2 The Wide Applicability of Servomechanism Theory

Although servomechanism theory was developed primarily for the fire control problem it quickly became clear that the theory had wide applicability to practically any control problem. All fields where control had been used earlier were invigorated by influx of ideas from servomechanism theory. The associated systems engineering methodology, that had been developed to deal with complex systems, also had very wide applicability.

Pioneering work on numerically controlled machine tools was done at MIT’s Servomechanism Laboratory [684, p. 218-225]. A numerically controlled three-axis milling machine was demonstrated in 1952. The first version of a language APT for programming the machines was later developed. APT was widely used through the 1970s and is still an international standard.

Servomechanism theory had a strong impact on process control. Oldenburger and Sartorius of Siemens showed that concepts and methods from servomechanism theory were useful for process control [511]. Smith [607] and Eckman [184] made similar observations. Equipment for process control were also improved. Electronics replaced pneumatics, but valve actuation was still pneumatic because of the forces required. One consequence was that the delay in the pneumatic lines used for signal transmission was reduced significantly. The linearity and precision of sensors and actuators were improved significantly by using force feedback. Feedback was also used extensively to improve the electronic controllers.

Drive systems with electric motors were improved significantly when the thyristor became available in the mid 1950s. There were major developments in power systems as electric power networks increased in size and complexity. High voltage DC transmission systems were developed. They required sophisticated electronics and control systems for AC to DC and DC to AC conversions. The first system was a 20MW 100kV transmission from mainland Sweden to the island of Gotland in 1954 [410].

The systems engineering capability required to build complex systems became an important part of control during the war. A dramatic demonstration of the advances of control was made in September 1947 when the aircraft “Robert E. Lee” made a completely autonomous transatlantic flight [464]:

The aircraft had a Sperry A-12 autopilot with approach coupler and a Bendix automatic throttle control. ... It also had some special purpose IBM equipment that permitted commands to its automatic control to be stored on punch cards fed automatically. From the time that the brakes were released for takeoff from Stephenville, Newfoundland, until the landing was completed at Brize-Norton, England the next day, no human had touched the control. The selection of radio station, course, speed, flap setting, landing gear position, and the final application of wheel brakes were all accomplished from a program stored on punched cards. The complete automation of aircraft flight appeared to be at hand.

3.3 From Mechanical to Electronic Computers

Controllers developed before 1940 were special purpose analog computers based on mechanical, pneumatic or electrical technology. There was a breakthrough when mechanical technology was replaced by electronics. The invention of the operational amplifier [554,128,529,307] was the key. By providing the operational amplifiers with input and feedback impedances it was possible to create components that could add and integrate signals. Adding multipliers and function generators made it possible to develop powerful computing devices for implementation of control systems.

Electronic analog computing had significant advantages over electro-mechanical devices, particularly in airborne equipment where low weight was important. The operational amplifier was also used to build general purpose analog computers. They were fast because operation was parallel. It was even possible to run them in repetitive operation so that effects of parameter variations could be visualized instantaneously. The number of differential equations that could be solved was equal to the number of integrators; large installations had more than 100 integrators. The computers were programmed with a detachable patch panel. Problems of oscillations arose if there was an algebraic loop, i.e., a closed loop without an integrator.

The analog computer became a popular tool for research institutes, and electrical, aerospace and chemical companies. The computers were typically run by well staffed
computing centers that attended to the hardware and assisted with programming. There were also smaller analog computers that could be placed on a desk top. Several institutes built their own systems, and universities also acquired analog computers.

The analog computers made it possible to simulate large systems. For the first time it was possible to use mathematical models to explore the behavior of systems under a wide range of operating conditions. Analog computers could also be used for hardware-in-the-loop simulation where real components were combined with simulation. Analog computing became an academic subject [316,442].

Digital computing emerged with the ENIAC, developed in the mid 1940’s by Mauchly and Eckert of the University of Pennsylvania’s Moore School of Electrical Engineering. Mauchly and Eckert left the university and formed a company that became Univac. The first computer Univac 701 appeared in 1951. A year later IBM announced the IBM 701. Several companies entered the computer business but by 1960 IBM totally dominated the industry.

In 1944 the Servomechanism Laboratory at MIT got a contract from the US Navy to develop a general purpose simulator for training naval bombers. Originally it was attempted to base the simulator on analog computing, but the program shifted to digital computing inspired by the emerging new technology. The computer was called “Whirlwind” after the name of the project [560]. The project changed direction several times. In the beginning of the 1950s it was used in the SAGE program, where Whirlwind became an early example of real-time computing. It was connected to radar stations for feasibility studies in the SAGE program. Whirlwind was designed as a 16-bit machine with 2K of memory. When experimenting with memory, Forrester explored magnetic cores in 1949, and core memory was installed two years later [219]. Forrester and others patented the technology which became the standard random-access memory for a twenty year period. Ken Olsen worked on the core memory in the Whirlwind team as a student. Later he moved to Lincoln Labs to make TR-0, a transistorized version of the Whirlwind. In 1957 he founded Digital Equipment (DEC). DEC’s PDP1, which appeared in 1959, was the first of a long string of successful minicomputers [136].

3.4 Communication

There was a need for centralization of control rooms, both in fire-control systems and in process control. Precision of the synchros, that were used for fire control, was improved and standardized. There were significant advances in synchros for communication of angles in fire-control systems, and the synchros and associated equipment became standard commodities.

In process control the pneumatic tubes that were used for communication were replaced by electrical systems. Signal levels were standardized to 4–20 mA. The fact that the zero signal corresponds to a nonzero current was used for diagnostics. The electric systems reduced the time delay caused by the limited speed of sound in the pneumatic systems. Cabinets with maneuvering equipment, controllers and recorders improved significantly. They were also augmented by relay panels for automatic start-up and shutdown and for safety interlocks. Centralized control rooms became common in process control.

There was a seminal breakthrough in communication theory with the publication of Shannon’s paper on information theory in 1948 [599]. Shannon defined what is the “capacity” of a communication link, showed what are the appropriate tools to study it, and characterized it in terms of the mutual information. Shannon also studied whether feedback could be used to increase capacity, and showed that for a discrete memoryless channel it could not; however its implication for control is limited since it does not address delay or a finite horizon. These themes are being revisited currently, as detailed in Section 5.2.

3.5 The Growth of Institutions and Research Labs

Control was nurtured in large laboratories that were created during the Second World War, such as the laboratories around MIT and in Moscow. The Radiation Laboratory was closed after the war but some of the MIT labs such as the Draper Lab and the Instrumentation Lab continued to operate. Lincoln Lab at MIT was established in 1951 to build the air defense system SAGE, many of the engineers having previously worked at the Radiation Lab. There were also significant control groups at General Electric, Hughes Aircraft, Bell Labs, Minneapolis Honeywell, Westinghouse and Leeds and Northrup.

There was a strong control group at Columbia University under the leadership of John Ragazzini and Lofti Zadeh, created around 1950. Among the graduate students were future leaders like Rudolf Kalman, John Bertram, Gene Franklin, and Eliahu Jury. Seminal work on sampled data systems was conducted; there was a weekly seminar dominated by Kalman and Bertram [232]. The group at Columbia dissolved in the late 1950s when Jury and Zadeh moved to Berkeley, Franklin to Stanford, Kalman to RIAS, and Bertram to IBM.

The RAND corporation in the US was created as a think tank, operated by the Douglas Aircraft Company and financed by the Air Force. In the 1950’s it carried out significant research related to control. George Danzig developed linear programming [162]. Bellman, who had done his PhD under Solomon Lefschetz at Princeton, developed dynamic programming [56,58,59].
Solomon Lefschetz had established a center for research in nonlinear differential equations and dynamics at Princeton in the late 1940s. In 1955 the Glenn Martin Company created the Research Institute for Advanced Study (RIAS) in Baltimore with very close relations to the Princeton group. Lefschetz and many of his group members joined RIAS, among them were: Bellman, Bhatia, Hale, Kalman, Kushner, LaSalle, Lee, and Marcus, who all would make contributions to control. Lefschetz and many of his colleagues moved to Brown University in 1964 to form the Lefschetz Center for Dynamical Systems. Lawrence Marcus moved to the University of Minnesota to create the Center for Control Science and Dynamical Systems.

In the late 1950s IBM and other computer manufacturers saw the potential for using computers for process control. They started a research group in control in the Department of Mathematics at the T. J. Watson Research Center in Yorktown Heights, with Kalman as the first leader [185]. Kalman left after a short time and John Bertram, a classmate of Kalman at Columbia, took over as the leader. The group later moved to San Jose. IBM also started laboratories in Europe; the IBM Nordic Laboratory in Stockholm was devoted to process control.

In England several of the major researchers moved to universities. Tustin became head of Electrical Engineering at Imperial College in 1953 where Westcott already was a lecturer, and Coales moved to Cambridge in 1953 [67,459,677]. National Physical Laboratory in England started a research group in control.

China had a long tradition in science. Professor Hsuen-shen Tsien had worked with von Karman at Caltech and the Jet Propulsion laboratory on missile guidance. In 1954 he wrote the remarkable book “Engineering Cybernetics” [646]. Tsien returned to China in 1955, he gave lectures based on control on and proposed to establish research facilities for aeronautics and missile development - the Fifth Academy of the Ministry of National Defense [137]. The Chinese Academy of Sciences created The Institute of Automation in 1956. The mathematician Z. Z. Guan established a research laboratory in control theory at the Institute of Mathematics, Chinese Academy of Sciences in 1962. The Chinese Association of Automation (CAA) was founded in 1961 after substantial preparatory work [139].

There were similar activities in many other countries with growth of control research in industry and academia [82]. Research institutes were also created by the academies of science in Budapest and Prague. The Research Institute of National Defense in Stockholm had one group for analog computing and another for missile guidance and control theory. In 1955 Saab created a new division called R-System, patterned after RAND corporation and sponsored by the Swedish Air Force [32].

3.6 The Onset of Control Education

Most of the research in control had been done in industry and research institutes and at a few universities. When servomechanism theory emerged it was recognized as a useful technology that could be widely applied. Control courses were introduced at practically all engineering schools. Control groups were created in many companies, and new industrial enterprises specializing in control were established. Many textbooks were written. In addition to [607,446,184,343], other popular US books were [121,143,637,103,645]. Among the books from the USSR we can mention [616,669,5,391]. Books were also published in Germany [515,512,585], UK [538,449,676] and France [254]. A list of early textbooks on control was compiled in connection with the 50th anniversary of IFAC [250]. The list includes 33 books published in 1960 and earlier. The book by Truxal [645] is representative of the state of the art of control education in the mid 1950s. The topics covered included linear systems theory based on Laplace transforms, the root locus method, stochastic processes, sampled data systems, analysis of nonlinear systems based on phase-plane and describing function methods. The book summarized many of the results and presented a systematic method for controller design inspired by circuit theory [272,659].

3.7 The Emergence of Professional Control Societies

The American Society of Mechanical Engineers (ASME) created a division for instruments and regulators in 1943. The Instrument Society of America (ISA) was founded in 1946 by companies interested in industrial instruments. They published a journal in 1954 that was later called inTech.

Much of the early work in automatic control was classified because of its military connection. After the war there was a need for more open interaction. The IRE (now IEEE) formed a Professional Group on Automatic Control in 1954. A journal that was to become the IEEE Transactions on Automatic Control was started in 1954 with George Axelby as the editor.

There were also international activities. In 1956 there were plans for no less than eight national meetings on automatic control in Europe. Wise leadership resulted in the formation of IFAC, which became the international forum for the field of control [142,443,354]. Many organizational issues were settled in a meeting in Heidelberg in 1956 with participants from 19 countries. An organizational structure was set up with triennial World Congresses, symposia, and workshops. Harold Chestnut from the General Electric Research Laboratory was elected as the first president and it was decided to hold the first World Congress in Moscow in 1960. This conference had a great impact because it provided an opportunity for
researchers who had been working in isolation to meet with colleagues who had worked on similar problems.

4 The Golden Age

Any field would have been proud of the accomplishments that control had achieved by 1960, but more was to come. The space race and the use of digital computers to implement control systems triggered new developments. Servomechanism theory was not well suited for systems with many inputs and many outputs, performance had to be optimized, and computer control gave rise to new challenges. Modeling based on injection of sinusoidal signals was time consuming for process control. These challenges required new tools, and control scientists eagerly turned to mathematics for new ideas. Many subspecialties were explored, which required focused and deep dives into applied mathematics. In contrast to the previous era when theory lagged applications, in this era the theoretical investigation went ahead of practice. Many ideas were investigated in an open-loop manner, without the benefit of feedback from implementation. In some cases, the computational power was not yet powerful enough, or networking had yet to emerge, to permit testing of the ideas. Research and education expanded significantly and there was ample funding. The development was also heavily influenced by the advances in computing. In 1960 computers were slow, bulky, unreliable and expensive. In 2000 they were fast, small, reliable and cheap.

The appropriate theory was state-space based rather than frequency domain based [360]. The earlier work of Aleksandr Lyapunov in the USSR on stability of differential equations [445] was found to be very useful in addressing the problem of stability of systems described by differential equations [367,367]. In the USSR, the problem of optimal control of systems based on differential equations was investigated by Pontryagin and his coworkers [534,99], and by researchers at the Institute of Control Sciences. This was a generalization of the earlier work on calculus of variations [204,363]. Rudolf Kalman laid a broad foundation for linear systems [357,360,364,361,362]. The state-space theory found immediate application. Swerling [634], Kalman [368], and Kalman and Bucy [364] extended the filtering theory of Wiener so that it addressed transient behavior as well as time-varying systems. Richard Bellman developed dynamic programming for the optimization of both deterministic as well as stochastic systems, including a foundation for Bayesian adaptive control [58,59,55]. In the ensuing five decades, all these efforts were thoroughly investigated, and a grand edifice of “systems theory” was developed. The concepts of linear systems, optimal control, dynamic programming, partially observed systems, system identification, adaptive control, nonlinear estimation, robust control, nonlinear systems, distributed parameter systems, decentralized systems, discrete-event systems, etc., were all explored. What is very interesting is that many of the ideas were investigated at a time when the technology was not yet available for their implementation.

The aerospace industry has always been at the frontier of technology with extreme demands on safety and performance. During 1960–1980 process control was a strong driver for computer control, while around 1980 the automotive industry became a technology driver. Problems of manufacturing and queuing also drove the development of control with applications in operations research.

The golden age was a very prolific period; our treatment is by no means complete and we apologize for omissions. In particular we do not adequately cover mechatronics, robotics, distributed parameter control (PDEs), Hamiltonian control, to mention just a few of many such examples.

4.1 The Space Race

Space travel and ballistic missiles posed many challenges. There were guidance, control and estimation problems. How to make effective use of moderate sized rockets to put a satellite in orbit? How to find efficient trajectories for interplanetary travel? How to minimize heating at reentry into the earth’s atmosphere? How to control rockets during launch, coasting and reentry? How to determine position, velocity and orientation from accelerometers, gyroscopes and star sights?

The Soviet program was led by Sergei Korlev with German engineers and scientists from Peenemünde as consultants. The first rocket, R-7 Semyorka, was based on the V2 with a new control system. Semyorka was used to launch Sputnik in 1957. Four years later Yuri Gagarin became the first astronaut. Wernher von Braun with several coworkers joined the Army Ballistic Missile Agency in Huntsville Alabama. Sputnik caused much consternation in the US. A new agency, NASA, was created in 1958. In 1961 President Kennedy announced the goal of landing a man on the moon within 10 years. NASA received significant funding and quickly grew to 8000 persons. Much research and development was sub-contracted to industry and universities.

The new challenges in the aerospace industry could not be met by servomechanism theory, and many new avenues were explored. Large resources were focused, with highly diversified groups, to solve specific engineering problems. Control research benefited dramatically from a strong influx of ideas from applications, mathematics and computing.

Inertial navigation was an enabler for intercontinental missiles and space flight; it required significant development of gyroscopes, accelerometers, computers and
The emergence of the digital computer spawned speculations about its use for control; indeed the Whirlwind (See Section 3.3 computer was designed for that very purpose. Today it is hard to grasp the state of computers in the 1950s. We illustrate with the following quote from a 1958 paper of Kalman [357] where he described an attempt to implement an adaptive controller:

> In practical applications, however, a general-purpose digital computer is an expensive, bulky, extremely complex, and somewhat awkward piece of equipment. ... For these reasons, a small special-purpose computer was constructed.

A perspective on the tremendous impact of computing is illustrated by the following quote of Herman Goldstine, Head of the Mathematics Department at IBM Research in Yorktown Heights, delivered at a staff meeting in 1962:

> When things change by two orders of magnitude it is revolution not evolution.

Combining Goldstein’s statement with Moore’s Law it follows that from 1971 onwards computers have enjoyed a revolution every 10 years. There has been a tremendous impact on how control systems are designed and implemented.

The poor capability and the poor reliability of general purpose computers was the reason why the Polaris ICBM used a digital differential analyzer (DDA), an emulation of an analog computer [471, p 98], [448]. The computer was developed at the Instrumentation Laboratory. It was followed by the Apollo Guidance Computer which was implemented using integrated circuits with a conventional computer architecture [471, ch 6]. The first version of the computer, Block I, had a core memory of 1K 16 bit words and a read only memory of 24K words. The clock speed was 1 Mhz. Versions of the AGC were later used to show the feasibility of fly-by-wire. By the time Block I, the first version of AGC, flew in August 1966, computers had been controlling industrial processes for 7 years.

There were major developments in industrial process control. Even if the computers were slow, bulky and expensive, their capabilities matched the basic requirements of process control. Process companies saw potential for improved operation, and computer companies saw business opportunities. Control groups were formed in the process industries and feasibility studies were executed jointly with computer companies [286]. The first system in operation was a Ramo-Woolridge RW-300 computer at the Port Arthur refinery in Texas. The early installations used supervisory control where the computer provided set-points to PID controllers that handled the basic control loops.

When IBM entered the field they used a small transistorized, scientific computer, IBM 1620, as a base. An interesting aside is that Ted Hoff was inspired by the IBM 1620 when he developed the first microcomputer. The IBM 1720 was based on the 1620 [287]. It had variable word length, one hardware interrupt, and analog and digital inputs and outputs. An upgraded version was announced as the Process Control Computer System IBM 1710 in 1961. A typical configuration was a CPU with a core memory of 40K decimal digits (80 K-bytes), and a hard disk with 2M decimal digits (4M-bytes), 80 analog inputs, 22 pulse counter inputs, 100 digital outputs and 45 analog outputs. The computer had a clock rate of 100 kHz. Control was done in a supervisory mode. Typical installations performed supervisory control of many loops, production planning, quality control and production supervision [188]. In 1964 the IBM 1800 was announced. It was the first computer designed for real time process control applications. The machine was successful and several thousand machines were delivered [287]. Many computer companies entered the field.

When computers became more powerful and more reliable it was possible to let them control actuators directly. A systems architecture called Direct Digital Control (DDC) emerged in 1962 when Imperial Chemical Industries (ICI) in England used a Ferranti Argus computer to control a soda ash plant. Computer control was used for all control functions including the low level loops. There were sensors for 224 variables and the computer controlled 129 valves directly. Computer control permitted operator panels to be simplified, and the system could be reconfigured by programming instead of re-wiring.

Computerized process control developed rapidly as technology went through the phases of special purpose machines, mini-computers and microcomputers, and there were many actors. Computer companies started to withdraw from the field which was taken over by instrumentation companies. It was attractive to distribute computing. In 1975 Honeywell and Yokogawa introduced distributed control systems (DCS), the TDC 2000 and the CENTUM. The systems permit direct digital control in functionally and spatially distributed units. The systems have standardized units for interaction with the process, with analog and digital signals and human-machine interfaces. Several manufacturers followed, and DCS became the standard for process control systems.

Use of computer control in the process industry expanded rapidly as distributed control systems (DCS)
Traditionally process control systems had two types of equipment: a control panel with controllers, recorders and displays, and a relay cabinet for start and stop sequences and safety interlocks. When minicomputers emerged the control panel was replaced by a DCS system. There was a similar development of the relay systems that were also used for automation in the manufacturing industry. General Motors challenged the electronics industry with requirements for a standard machine controller that could replace the relays. Several companies responded to the challenge. A system from Digital Equipment based on a mini-computer was rejected. A successful demonstration of a special purpose system was made by Bedford Associates and Modicon in 1969. The unit was rugged with conductive cooling and no fans. In 1971 Allen Bradley developed a device called Programmable Logic Controller (PLC). The system architecture was based on round robin schedulers with different cycle rates. PLCs were originally programmed in a graphical language called ladder diagrams (LD), which emulated the ladder logic used to describe relay circuits. Later several different programming styles were standardized [322,425]; function block diagrams (FBD), sequential function charts (SFC) and structured text (ST). The PLCs developed rapidly and became a standard tool for automation.

Process control systems are typically widely distributed. Wires from sensors and actuators, typically 4–20 mA current loops, were brought to a central cabinet and distributed to the computer. These systems were expensive, difficult to maintain and upgrade; the systems had a lifetime of tens of years. When networks appeared in the 1970s it was natural to replace expensive wiring with networks and several different systems emerged. National standards were developed in Germany (PROFIBUS [547]) and in France (FIP [690]), and in the US the manufacturers formed the consortium FOUNDATION Fieldbus [205] which absorbed FIP. There were divergent opinions driven by commercial interests of the vendors [203]. After more than a decade the IEC in 2000 introduced a standard, IEC 61784, which included many of the different suppliers’ features. Similar standards appeared in the building industry. Some vendors used Ethernet instead.

### 4.3 Automotive Applications

The automotive area is an important application area for control. It is a strong technology driver because of its scale; about 40 million cars were manufactured in the year 2000. By providing a large market, the automotive industry contributed strongly to the development of the micro-controller, a microprocessor with integrated analog and digital inputs and outputs. The automotive industry also stimulated the development of inexpensive emission sensors, accelerometers and gyroscopes. Together with the aerospace industry it was an early adopter of model based design, and provided a fertile ground for research in modeling, integrated process and control design, and implementation of control systems [379,275]. The impact of the automotive industry on control became stronger towards the turn of the century and even stronger in the 21st century.

The excellent experience with these systems inspired car manufacturers to introduce more sophisticated systems such as collision avoidance, parking assist and autonomous driving [133].

Environmental concerns and recurring oil crises created demands for reduced emissions and reduced fuel consumption. In 1967 California established The Clean Air Resources Board, and requirements on automotive exhaust emissions became federal US laws in 1970. Feedback emission control made it possible to satisfy the new laws. The control system used a new oxygen sensor (lambda sensor), a catalytic converter, and a feedback system which kept oxygen levels at the converter very close to the stoichiometric condition. General Motors was one of the early adopters; we quote from John Cassidy who was head of the control group at GM:

> I recall a meeting with Ed Cole, an engineer by background, who was then president of GM. A workable closed loop system was possible using a fairly simple circuit based on an operation amplifier. Mr. Cole made the decision at that meeting that GM would take an advanced technical approach based on the newly emerging microprocessor technology. Others in the industry followed.

Systems went into production in the late 1970s. Once computer based feedback control was introduced in cars, its use expanded rapidly into many other functions. Anti-lock braking systems (ABS) were introduced to prevent the wheels from locking up. Electronic braking systems (EBS) and electronic stability control (ESC) controlled the brakes individually to improve stability and steering. These systems used accelerometers and gyroscopes as sensors. Automatic cruise control had been used earlier, but implementation by computer control was much more convenient. A consequence is that cruise control is now a standard feature. Adaptive cruise control, based on radar sensors, was introduced to maintain a constant distance to the car in front. The excellent experience with these systems inspired car manufacturers to introduce more sophisticated systems such as collision avoidance, parking assist and
autonomous driving [133].

In the beginning, control systems were typically add-on features. Over time there has been a move towards integrated design of mechanics and control. Control of turbochargers permits smaller engines. Hybrid and electrical vehicles are even more prominent examples of co-design of systems and control.

In 1986 Pravin Varaiya initiated an ambitious research project, Program for Advanced Technology for Highways (PATH), at the University of California, Berkeley, in collaboration with Caltrans [523]. Platooning of cars that were linked electronically was explored. In 1997, the program demonstrated platooning of cars traveling at 60mph separated by 21 feet on a San Diego freeway, and showed that capacity could be doubled. The PATH program still continues with much effort directed towards control of traffic flow. Platooning is particularly efficient for heavy duty vehicles [427,7].

4.4 Optimal Control

The early rockets did not have great thrust, and so a crucial problem was to launch the rocket most effectively. Attempts to solve problems of this type led to the development of optimal control theory. Major contributions were made by mathematicians and control engineers. There was a revitalization of the classical calculus of variations which has its origins in the Brachistochrone problem posed by Bernoulli in 1696 [246]. Pontryagin and his coworkers in Moscow followed the tradition of Euler and Lagrange and developed the maximum principle [534]. They were awarded the 1962 Lenin Prize for Science and Technology. In the United States, Bellman instead followed the ideas of Hamilton and Jacobi and developed dynamic programming [59,56].

The case of linear systems with quadratic criteria was solved by Bellman in special cases [56], and a complete solution was provided by Kalman [358]. The books by Athans and Falb [38] and Bryson and Ho [123] presented the results in a form that was easily accessible to engineers; they also dealt with computational issues. A spectacular demonstration of the power of optimal control was given by Bryson [124]. He calculated the optimal trajectory for flying an aircraft from sea level to 20 km, and found that it could be done in 332 seconds. The optimal trajectory was flight tested and the plane reached 20 km in 330 seconds. The traditional quasi-steady analysis predicted that the airplane could not get up to 20 km. Optimal control grew rapidly, many books were written and courses were introduced in control curricula [418,12,424].

Another computational approach to optimal control, model predictive control, which emerged from industry is now widely used [148,447,551,131,559,563]. The paper [457] was selected for the first High Impact Paper Award at the IFAC World Congress in Milan in 2011.

4.5 Dynamic Programming

Multi-stage decision making was a problem of interest to the RAND Corporation, supported by the U.S. Air Force, in the 1950s. Richard Bellman was attracted to this problem. He initiated the field of dynamic programming and developed the principle of optimality [59]. It is of particular interest in the case of stochastic systems since it provides optimal policies in state-feedback form. Howard developed the policy iteration algorithm [314] (see Section 4.14), which is a very efficient algorithm to determine optimal policies when the number of states and actions is finite. It has become very popular in operations research and industrial engineering; see Section 4.14. This was further sharpened by Blackwell [86]. He comprehensively showed the differences arising in the infinite horizon case from positive and negative cost functions as well as the case of discounted cost functions [87–89,629]. The continuous time version of the dynamic programming equation is the Hamilton-Jacobi-Bellman equation for the optimal cost-to-go.

Dynamic programming is, however, computationally complex; it suffers from the “curse of dimensionality.” With the advent of fast computers, methods to approximate the cost-to-go function by nonlinear functions, e.g., neural networks have received attention. In 1995, TD-Gammon, a temporal difference based learning scheme using a neural network trained by self-play [636] played at the level of a world class human player.

Dynamic programming has also become useful as a method to establish qualitative properties of the optimal solution. This has been found to be extremely useful in areas such as inventory control and production planning [662,75]; as described in Section 4.14. The teaching of Markov Decision Processes, which is dynamic programming for finite state stochastic systems, is a standard part of the curriculum of operations research and industrial engineering departments.

Dynamic programming has found wide applicability. In the Internet, the Distributed Bellman Ford algorithm for determining the shortest path between two nodes on a graph is a key element of distance-vector based routing algorithms such as RIP [292,452] and IGRP [328]. With increasing interest in fast computational methods for machine learning and artificial intelligence, the ideas of dynamic programming are becoming widely used.

4.6 Dynamic Games

Game theory was pioneered by John von Neumann in his attempt to develop a foundation for economic behavior [667]. He analyzed both static two person zero-sum
games where one agent’s cost is the negative of that of the other agent, as well as static teams, where all the agents have the same cost criterion that they are seeking to minimize. For two person zero-sum “matrix games” where each agent has only a finite number of choices, he showed that there is a saddle-point in randomized strategies [668]. Subsequently, Nash [495] showed a similar result for static nonzero-sum games.

At the same time that Bellman was developing dynamic programming at RAND, Rufus Isaacs was studying dynamic continuous time two-person zero-sum games. The “Isaacs equation” is a two-sided version of the Hamilton-Jacobi equation [337,338]. This differential game theory was applied to military problems such dog-fights and tank battles [694], [301], and later to robust control [48].

At around the same time, Shapley [600] and Everett [199] were also investigating discrete-time games. Zachrisson [694] provided a particularly cogent treatment of Markov games. Interest continued in the subsequent decades with the investigation of Nash equilibria, Pareto optimality, Stackelberg solutions and incentives in dynamic games [621,602,305,49].

4.7 Linear Systems

Linear approximations have been extremely useful for analysis and design of control systems. Differential equations were used in the early development, but there was a switch to frequency response when the servomechanism theory was introduced. In the 1960s there was a return to differential equations because frequency response was not well suited for numerical computations, and it was inconvenient for systems with many inputs and many outputs. The return to differential equations became known as “the state space approach,” because Newton’s notion of state played a central role. It was also called “modern control theory” to separate it from servomechanism theory. The mathematical sophistication of the research, and consequently also textbooks, increased. The books by Zadeh and Desoer [695], Brockwell [114] and Kailath [355] were popular.

The reformulation of the models naturally raised two questions: can all states be reached by appropriate choices of the control signal and can the state be reconstructed from measurements of the outputs. Kalman posed these questions and defined the notions of reachability and observability [360,366,362,253]. Kalman’s results also provided clear insight into the relationship between the linear differential equations and the associated transfer functions, which cleared up a classical question on the effect of cancellation of poles and zeros in a transfer function [92].

There was also work on the structure of linear feedback systems in a classic setting. Horowitz [308] introduced a controller architecture, with two degrees of freedom, that combined feedback and feedforward so that requirements on command signal following could be separated from requirements on robustness and disturbance attenuation. The servomechanism was analyzed in the state-space model [165].

The theory of linear systems drew heavily on linear algebra, matrix theory and polynomial matrices. Results from numerical linear algebra could also be exploited for computations [415]. The size of textbooks grew to 700 pages and more, when chapters on state-space theory were added to classic material on servomechanisms [229,401,508,169].

In standard state space theory, the state space is the Euclidean space and time is a real variable. Extensions to systems over rings was also established [365]. A uniform framework for linear systems, finite state machines and automata can be established. The introductory signals and systems book by Lee and Varayia [420] is written in this spirit. A theory of discrete event systems was initiated in [556] to address control theoretic notions of controllability, observability, aggregation, decentralized and hierarchical control for automata and formal language models [557,587,98]. Lately there has been significant interest in hybrid systems [117,451,257] which have a combination of continuous and discrete behavior.

Singular perturbation theory [388] and descriptor systems [177] were introduced to deal with systems having widely differing time scales. Differential algebraic systems were used to model large electrical circuits [244]. Inspired by circuit theory, Willems [533] introduced system models called behavioral systems, that de-emphasized the role of inputs and outputs, which also were described as differential algebraic systems. Differential algebra is a natural framework for modeling physical systems, and it is the mathematical framework behind the modeling language Modelica [643]. There is an extensive body of literature on infinite dimensional dynamical systems [429,156,69,47]; control of fluid flow is one application area [1].

The field of linear systems has been declared many times to be “mature” from a research point of view, but interest has repeatedly been renewed due to new viewpoints and introductions of new theories.

4.8 State Feedback Kalman Filtering and LQG

When state-space theory is used for design, it is natural to use state feedback because the state contains all relevant information about the past. A linear controller can then be represented by a matrix which maps state variables to control variables. Kalman formulated the design problem for state models as an optimization problem where the criterion to be minimized is a quadratic form
in states and control variables, the so-called LQ problem. He solved the problem elegantly and showed that the optimal feedback is given by a solution to a Riccati equation. To quote from [358]:

One may separate the problem of physical realization into two stages:
(A) Computation of the “best approximation” \( \hat{x}(t_1) \) of the state \( x(t_1) \) from knowledge of (the output) \( y(s) \) for \( t \leq t_1 \).
(B) Computing (the control) \( u(t_1) \) given \( x(t_1) \).

... Somewhat surprisingly, the theory of Problem (A), which includes as a special case Wiener’s theory of filtering and prediction of time series, turns out to be analogous to the theory of Problem (B) developed in this paper. This assertion follows from the duality theorem discovered by the author.

Kalman’s solution also applies to linear time-varying systems. The corresponding problem for difference equations is very similar, and led to a reformulation of the theory of sampled systems. The condition for a solution is that the system is reachable. A remarkable property of the solution is that it gives a closed loop system with infinite gain margin and a phase margin of 60°. Glad extended these results on robustness of the LQ controller to nonlinear systems [256]; which was further generalized in [590].

Kalman also showed that the optimal filter for a linear system with Gaussian noise is a process model driven by the measured observation with the gain specified by a Riccati equation. The condition for solvability is that the system is observable. The optimality of the controller consisting of state feedback and a Kalman filter, which is known as the LQG controller, was first proven in a special case by the economist Simon [603]. There are some subtleties about the separation that have only recently been sorted out [247].

The controllers obtained by servomechanism theory can be viewed as compensators that shape the frequency response of the loop transfer function. The LQG controllers have a very different interpretation. They have two elements, a state feedback and a Kalman filter or an observer. The dynamics of the controller comes from the observer which is a dynamic model of the process and its environment. This idea is captured by the internal model principle introduced by Francis and Wonham [228]. A reference signal generator can be added to the LQG controller to provide command signal following using an architecture with two degrees of freedom [558, Section 7.5]. The LQG controller is very well suited for systems with many inputs and many outputs. The computations required for design are based on solid algorithms from numerical linear algebra. The LQG controller does not automatically provide integral action, illustrating the fact that it is important to capture all aspects when formulating an optimization problem. Integral action can be provided by augmenting the process model with a model of the disturbances.

The LQG paradigm has proved to be a useful tool for iteratively designing linear control systems due to the explicit form of the solution, as well as the well developed asymptotic theory for the infinite horizon case. It is a standard tool for design of control system [424, 12].

The important issue of what information is available to a decision maker in a system was studied by Witsenhausen [686]. He showed that even in a linear Gaussian system with a quadratic cost, if there is no memory of what observation was made in a previous stage, then a linear control law is not optimal. He showed the several complexities that arise depending on the information available to agents in a distributed system at the time that they have to make a decision [688, 687]. Information structures that lead to tractable solutions were further investigated in [304].

4.9 Nonlinear Systems

Linear theory has, somewhat surprisingly, been extremely useful for analysis and synthesis of control systems even though most real systems are nonlinear. The necessity of considering nonlinear effects was part of classical control theory; to quote from Truxal [645, p. viii]:

Fifth, the designer must be acquainted with the basic techniques available for considering nonlinear systems. He must be able to analyze the effects of unwanted nonlinearities in the system and to synthesize nonlinearities into the system to improve dynamic performance.

Typical nonlinearities he mentions are friction, backlash, saturation, and hysteresis [514, 264, 39].

Approximate methods for analyzing nonlinearities were developed in nonlinear dynamics [397, 15, 474]. One method to explore limit cycles, called harmonic balance, consisted of investigating the propagation of the first harmonic, similar to Nyquist’s analysis of linear systems. A version of this method became known as the describing function method [654, 386]. On-off control was popular in the early days of control because it was possible to obtain high gain with simple devices; significant theory was also developed [647, 648, 217, 650].

Lyapunov stability theory was used extensively in the USSR [453]. Much research was stimulated in the West when it was popularized by Kalman and Bertram [367], who had picked up the ideas from Lefschetz at Princeton. Useful extensions were provided by Krasovskii and LaSalle [414, 392]. Willems showed that the notions of energy and dissipation are closely related to Lyapunov...
theory and developed a theory for dissipative systems [685]. Lyapunov theory is now commonly used both for analysis and design [230]. The notions of control Lyapunov functions and input-to-state stability introduced by Sontag [618] have proven useful. Khalil’s book [375] is a popular standard text.

The problem of the stability of a system obtained by feedback around a memory-less nonlinearity and a linear feedback system was proposed by Lurie and Postnikov [444]. Aizermann conjectured that the closed loop system would be stable if the nonlinearity was sector bounded and if the linear system is stable for any linear gain in the sector [4]. The conjecture was false but it stimulated much creative research. Originally the problem was approached by Lyapunov theory but a major breakthrough was made by Popov who provided a stability condition in terms of a restriction of the Nyquist plot of the linear part [535,536]. Yakubovich [692] showed that Popov’s results could be expressed and extended in terms of linear matrix inequalities (LMI’s).

Yet another approach to stability was presented by Sandberg [577] and Zames [697] at the National Electronics Conference in 1964. The presentations were later followed by detailed publications [578,579,698,699]. Zames focused on input-output properties and avoided the notion of state space. He had picked up functional analysis from Singer at MIT and he introduced the small gain theorem and the passivity theorem. These concepts generalize the notions of gain and phase for linear systems. The ideas garnered much following and they quickly became part of the core of control theory [167,663].

In the 1970s there was also an influx of ideas from differential geometry [101], leading to the development of geometric control theory. Brockett, Jurdjevic, Hermann, Krener, Lobry, and Sussman were key researchers who drove the research agenda. The notions of controllability and observability of nonlinear systems were investigated for systems which are affine in the control [299,288,115,633,393,437,438,116,298]; the criteria were based on Lie algebra. Feedback linearization was introduced as a technique for design of nonlinear systems [319]. Fliess used differential algebra to define the notion of differential flatness which became a powerful method to design feedforward and tracking [212–214]. Computer algebra was used to compute Lie brackets. Isidori and Byrnes introduced the notion of zero dynamics, an extension of the zeros of a linear system [340]. There are many interesting applications of geometrical control theory, e.g., attitude control of spacecraft [601], aircraft flying at high angles of attack [625, Section 7.4], backing of trailers [211], walking robots [678], and quantum systems [317,377]. Geometric control theory is part of the core of nonlinear control theory with several books [339,503].

4.10 Stochastic Systems

Dynamic programming can be used even when the state of the system is only noisily observed, by considering the conditional probability distribution of the state as the “hyperstate” [29]. The optimality of separated policies was thoroughly investigated by Striebel [631].

For linear Gaussian systems, by the separation theorem, the hyperstate is finite dimensional since the conditional probability distribution is Gaussian and thus described completely by the conditional mean and conditional covariance. As described in Section 4.8, when the cost function is further taken to be a quadratic function of the state and control one obtains the separation theorem with certainty equivalence [603,638,349,540]. The cost function consisting of the expected value of the exponential of a quadratic cost can also be solved explicitly since it is multiplicatively decomposable [342]. It can be used to model risk-averting or risk-seeking behavior, and also has connections to differential games and robust control.

Bellman also expounded on the fact that dynamic programming could be used to develop adaptive controllers for systems when the parameters are unknown, by viewing the conditional distribution of the unknown parameters as the hyperstate [55]. In this case control serves a dual purpose, as a tool for exciting the system and determining its characteristics, and also as a tool to move the state to a desirable region. This was dubbed “dual control” by Feldbaum [202].

Conceptually it is very attractive to formulate and solve the adaptive control problem using dynamic programming. There are, however, significant computational problems because of the large state space – the curse of dimensionality. For that reason an alternative non-Bayesian certainty equivalence approach was pursued, resulting in the self-tuning approach; see Section 4.12. An early Bayesian attempt was to approximate the loss function locally by a quadratic function [458]; another approach is to estimate the cost-to-go using Monte Carlo methods [74].

One special adaptive control problem, which captures the quintessential tradeoff implied by the dual roles of control, is the multi-armed bandit problem. In a more useful incarnation it models the problem of testing drugs whose efficacies are unknown. In the bandit version, it features several slot machines with unknown probabilities of rewards, with the probabilities themselves modeled as random variables with a prior probability distribution. A compulsive gambler has to play one arm each day, with the goal of maximizing the expected total reward obtained by playing the arms. This problem has the special structure that nothing is learned about an arm if it is not played on a given day; hence its hyperstate remains unchanged. For the case of discounted rewards,
this celebrated problem was shown to have a very appealing structure by Gittins and Jones [255]. Every arm has an index, defined by its hyperstate, and the optimal policy is to just play the arm with the highest index. The index of an arm is the maximal expected discounted reward up to a stopping time divided by the discounted time.

With the advent of powerful computation, the problem of “partially observed Markov decision processes,” (POMDPs) [606], has acquired great attention as a methodology for modeling and solving problems in machine learning and artificial intelligence [531,619,501,641,493,245,596].

Beginning in the late 1950s, there was great interest in developing optimal filters for nonlinear systems. In the discrete-time case, obtaining the conditional distribution of the state of the system given past noisy measurements amounts simply to an application of Bayes Rule. By allowing for unnormalized distributions where the denominator in Bayes Rule is ignored, one can obtain linear recursive equations for the conditional distribution [400]. In the continuous time case featuring nonlinear stochastic differential equations, the optimal filtering equations are also nonlinear [628,404,405,235]. However, by propagating the unnormalized probability distribution, the resulting equations are linear [178,181,696,483]. The central difficulty is that except in special cases [63] the filters are generally not finite-dimensional. As in the case of dynamic programming, with the availability of increasingly fast computers, one can judiciously exploit the capably to perform simulations to approximate unknown distributions; an example in this vein is particle filtering [282,261] which is useful for nonlinear non-Gaussian systems.

Early in the 1960s there was already interest in developing stochastic control theory for continuous time systems [215,209]. There was a great effort in the 1960s and 1970s in developing a theory of optimal control of continuous nonlinear stochastic systems described by stochastic differential equations for partially observed systems. This work has found application principally in mathematical finance [469], as noted in [478]. There were deep mathematical challenges, and several control researchers delved into the field and conducted frontline mathematical research into stochastic differential equations and martingale theory. Issues related to the nature of solution of stochastic differential equations, existence of optimal solutions, representation of the optimal solution in the case of partial (i.e., noisy) observations, etc., were investigated [216,62,179,180,146,164,208]. A good account is available in [102]. The problem of existence of solutions to the Hamilton-Jacobi-Bellman equations was addressed by the viscosity approach [153,431,430,432].

Motivated originally by problems in biology, a filtering theory for counting processes was developed by Snyder [609]. The problem of interest was to estimate the underlying intensity of a process given measurement of “ticks.” This spurred much mathematical work in stochastic processes [109,661,97]. It has found application in queuing systems [108]. As one example, it has been used to analyze flows of customers in queuing networks [671]. The stochastic control of point processes was also investigated [96].

### 4.11 Identification

One factor that contributed to the success of servomechanism theory was that the transfer function of a process could be obtained empirically by frequency response. Frequency response was, however, not suitable for process control because the processes were typically slow and it took a very long time to perform the experiments. It was also desirable to obtain models that additionally captured noise characteristics, for example to apply LQG controllers.

For computer control it was natural to use discrete time models. Much inspiration came from time series analysis where Box and Jenkins [104] had developed methods of estimating parameters in time series. Three popular models are auto-regressive (AR), moving average (MA) and auto-regressive moving average (ARMA) models. These models are difference equations driven by discrete time white noise. The models do not have inputs, and for control applications it was necessary to extend the models by adding controlled inputs. The presence of inputs also raised interesting problems of finding input signals that provide a sufficiently rich excitation. By combining ideas from probability theory, statistics and time series analysis, it was possible to obtain powerful methods with good statistical properties. An early application was to determine paper machine dynamics and to design control laws that minimized fluctuations in quality variables [34,31]. Research in this area, which became known as system identification, started in the 1960s. Identification brings control engineers, probabilists, statisticians and econometricians together. Typical issues are not only statistical issues such as consistency and efficiency but also control inspired problems such as input selection and experiments in open and closed loop [251]. Several books were written as the research progressed [400,506,435,611]. The Matlab toolbox developed by Ljung has led to system identification techniques being widely used in industry and academia. The IFAC symposia series on System Identification which started in Prague in 1967 is still continuing.

### 4.12 Adaptive Control

Adaptive control emerged in the 1950s in flight and process control [268,226,357]. Supersonic flight and ballistic missiles posed new challenges because the dynamic behavior of air vehicles changes drastically with altitude
and Mach number. Autopilots based on constant-gain, linear feedback can be designed to work well in one flight condition but not for the whole flight envelope. Many adaptive flight control systems were proposed and flight tested [268,475]. Interest in adaptive flight control diminished toward the end of the 1960s. One reason was the crash of a rocket powered X15 with an adaptive controller [182]. Another was the success of gain scheduling based on air-data sensor [622]. There was also early interest in adaptation in process control [226,357].

Research in the 1950s and early 1960s contributed to a conceptual understanding of Bayesian adaptive control, as described in Section 4.10. However, as noted there, due to its complexity, an alternative non-Bayesian certainty equivalence approach was pursued, resulting in the self-tuning approach.

Draper and Li investigated on-line optimization of aircraft engines and developed a self-optimizing controller that would drive the system towards optimal operation. The system was successfully flight tested [85,174] and initiated the field of extremal control. Tsypkin showed that a self-tuning approach was pursued, resulting in the self-tuning approach.

Interest in adaptive control resurfaced in the 1970s. There was significant research on model reference adaptive control (MRAC) [679]. MRAC automatically adjusts the parameters of a controller so that the response to command signals is close to that given by a reference model. The original MRACC which was based on a gradient scheme called the MIT Rule, was improved by applying Lyapunov theory to derive adaptation laws with guaranteed stability [127,521,411]. Variations of the algorithm were introduced using the augmented error [479,482]. The MRAC was extended to nonlinear systems using back-stepping [394]; Lyapunov stability and passivity were essential ingredients in developing the control laws.

One motivation for using adaptation for process control is that system identification experiments on real plants are tedious and time consuming, besides also requiring skilled personnel. It was therefore attractive to explore if an adaptive controller could be used instead. The self-tuning regulator (STR) estimates the process parameters and finds controller parameters that minimize a criterion, for example the variance of the process output. Steady state regulation is a typical problem which can be modeled by an ARMAX process. Estimation of parameters in such a model is a complex nonlinear problem. A surprising result in [36] showed that a controller based on least squares estimation and minimum variance control could converge to the desired controller. Industrial use was demonstrated [36,356,411,64] and a number of applications ensued, autopilots for ship steering, rolling mills, continuous casting, distillation columns, chemical reactors, distillation columns and ore crushers [64,259,37,25,206]. Many variations and generalizations evolved to consider different control objectives for noisy systems.

The self-tuning regulator stimulated a great deal of theoretical work. The problem was complicated by both the nonlinearity and the stochastic nature of the overall system. Similar issues had arisen in analysis of recursive algorithms such as stochastic approximation and recursive identification of ARMAX systems; the prior work paved the way for the analysis of the stochastic adaptive control systems [436,613,407,409,140,406]. Proofs of stability, convergence, self-optimality and self-tuning took several years to come [258,260,54,273]. The similarities between MRA and STR were also studied [187].

Early on, Egardt [187] had shown that even small bounded disturbances can cause adaptive controllers to lose stability. Ioannou and Kokotovic analyzed the effects of unmodeled high frequency dynamics and bounded disturbances on adaptive control schemes [335]. An investigation by Rohrs of robustness to unmodeled dynamics [566] stimulated much research that provided insight into modified algorithms. Stability proofs required bounded estimates. Normalization of signals [542,541] was proved to guarantee stability. Stability could also be achieved by projection alone [693,491].

Adaptive control was extended to feedback linearizable nonlinear systems [369]. It was also extended to include nonlinearities of the type commonly encountered in applications, such as dead-zone, backlash and hysteresis [635]. Adaptive control design methodologies such as backstepping became an integral part of the design of nonlinear control systems [395]. The increased knowledge in adaptive control that came from all this work is well documented in books [187,259,11,400,37,494,580,334].

Variations of adaptive algorithms are still appearing. The \( L_1 \) adaptive controller is one example; it inherits features of both the STR and the MRAC. The model-free controller by Fliess [210] is another example which is related to the self-tuning regulator.

Products use MRAC and STR to tune controllers on demand, to generate gain schedules and for continuous adaptation. There are systems that have been in operation for more than 30 years, for example for ship steering and rolling mills [206,270]. Automatic tuning of PID controllers is widely used; virtually all new single loop controllers have some form of automatic tuning. Automatic tuning is also used to build gain schedules semi-automatically [35].

There are strong similarities between adaptive filtering and adaptive control. Gabor worked on adaptive filtering [236] and Widrow developed an analog neural network (Adaline) for adaptive control [680,681], The adaptation mechanisms were inspired by Hebbian learning in
biological systems [291]. Today noise cancellation and adaptive equalization are widespread implementations of adaptation in consumer electronics products.

There is a renewed interest in adaptive control in the aerospace industry, both for aircraft and missiles. Good results in flight tests have been reported both using MRAC [416] and the $L_1$ adaptive controller [313]. In the future, adaptive control may be an important component of emerging autonomous systems.

4.13 Robust Control

Horowitz followed Bode to explore feedback in the context of linear systems in the book [308]. He investigated the cost of feedback and introduced a controller architecture having two degrees of freedom.

Bode had designed feedback systems that were robust to variations in the amplifier gain. He had shown that the open loop gain had to be much larger than its closed loop gain in order to obtain a robust amplifier. Robustness is thus obtained at the cost of a gain reduction. Horowitz, who was Bode’s intellectual grandson via Guillemin, extended this observation and introduced the notion of cost of feedback in general feedback systems [308, p 280–284]. Horowitz also generalized Bode’s robust design technique to more general process variations. The method is called QFT (Quantitative Feedback Theory) [311,309]. It is based on graphical constructs using Nyquist or Nichols plots.

There was a significant development of robust control in the state space framework, which had the advantage of leading to techniques that are well suited to numerical computations. The LQ controller, with state feedback, has amazing robustness properties, as noted in Section 4.8. In the 1970s much research was devoted to explore if the robustness could be extended to the LQG controller, which employs output feedback. The only condition required for solvability is that the system is reachable and observable. Researchers schooled in servomechanism theory did not understand why the classical limitations imposed by non-minimum phase dynamics did not show up [568,310]. Much work was done at the MIT Electronic Systems Laboratory and at Honeywell. An insightful summary is given by Safonov [571,572]. A key observation was that robustness measures should be based on the singular values of the loop transfer function and not on the eigenvalues. The main result is that the LQG controller is not robust. A simple counter example is given in the paper by Doyle entitled “Guaranteed Margins for LQG Regulators” [170] with the somewhat provocative abstract “There are none.” Several attempts were made to impose constraints on LQG control design but the real solution would come later from a different direction.

In 1981 George Zames published a paper [700] which laid the foundation for $H_\infty$ control. Following Bode’s ideas he considered input-output descriptions and designed controllers that minimized the $H_\infty$-norm of the sensitivity function for systems with right half plane zeros. Zames used functional analysis and interpolation theory to solve the problem. Zames’s work has a strong following, with many extensions and generalizations. The so-called four-block problem, consisting of addressing all four sensitivity functions became a standard formulation. The paper [172] by Doyle, Glover, Khargonekar and Francis was a major advance because it showed that the $H_\infty$ problem could be solved by state space methods, and that feedback and observer gains were given by Riccati equations. The controller obtained has the same architecture as the LQG controller but with different filter and feedback gains gains. McFarlane and Glover generalized classic loop shaping to multivariable systems and showed the relations to $H_\infty$ control [462], $H_\infty$ control developed into a standard design method with books [171,266,604,701,383] and Matlab toolboxes.

A side effect of $H_\infty$ control was a renewed interest in fundamental limitations [604,591]. It was shown that a system with right half plane zeros and time delays could not be controlled robustly if the bandwidth is too high, that robust control of a system with right half plane poles requires high bandwidth, and that systems with right half plane poles and zeros could not be controlled robustly if the poles and zeros were too close. A striking example of the difficulties is given in [373]; a practical illustration is a bicycle with rear-wheel steering. The paper [373] illustrates the need and importance of carefully investigating to what extent the end result of any design is fragile.

Zames also investigated the problem of finding norms that are suitable for comparing systems. The problem is straightforward for stable systems; simply compare the outputs for a given input. For unstable systems he introduced the gap metric [189] which only admits inputs that do not result in unbounded outputs. Vidyasagar provided an alternative graph metric [664]. Georgiou and Smith showed that robustness optimization in the gap metric is equivalent to robustness optimization for normalized coprime factor perturbations [248]; they also obtained results for nonlinear systems [249]. Vinnicombe introduced the $\nu$-gap metric that was adapted to robust stabilization [666].

Doyle and co-workers introduced the structured singular value (nu-analysis) to demonstrate that conservatism of gain arguments can be drastically reduced by optimization of frequency weights [173]. They used this effectively for analysis of systems with both parametric uncertainty and uncertain linear dynamics. The work was a pioneering application of convex optimization in control. It was extended to nonlinear components in the work on Integral Quadratic Constraints by Megretski and Rantzer [467]. This generalized the methods of Zames,
Yakubovich and Willems from the 1960s and 70s and integrated them with mu-analysis and semi-definite programming.

Linear matrix inequalities became a useful design tool when efficient computational procedures based on interior point methods were developed [498]. Many design problems can be captured by convex optimization and LMI’s, as was shown by Boyd and others [105,237,519,468,370,129].

Basar and Bernhard [48] formulated the problem of robust control as a game problem. The task of the controller is to deliver good performance even against an opponent who tries to perturb the system in the worst possible way. They showed that in the case of linear systems the optimal controller is the $H_\infty$ controller.

4.14 Control in Operations Research: Inventory, Manufacturing and Queuing Systems

Control is widely used in dynamic system problems that arise in operations research. Many applications can be modeled as problems involving the control of Markov chains over an infinite horizon with a discounted cost or long term average cost criterion, called Markov Decision Processes. One way to solve them is by the “value iteration method” that consists of determining the infinite horizon optimal cost as the limit of finite horizon costs [57].

In the late 1950s, when confronted with the problem of optimizing which customers should be mailed Sears catalogs based on profits from previous purchase history, Howard [314] developed the policy iteration method that converges in finite time for finite state and control sets [315]:

This all took place in the days when computers still had vacuum tubes. And so the runs were fairly time-consuming ... The optimum policy balanced ... return with the effect on future state transitions. The net result was a predicted few percent increase in the profitability of the catalog operation, which, however, amounted to several million dollars per year.

Dynamic programming has been very useful in inventory problems. A celebrated result of Scarf [581], generalized the work of Arrow, Harris and Marschak [18]. It analyzed a general model where the cost of an order is affine (i.e., constant plus linear) in the number of units ordered, and when there are costs both for holding inventory as well as shortages. They showed that if the demand is random, and there is a lag in fulfilling orders, then the optimal policy is of the $(S, s)$–type: if the level of inventory is less than $s$ then order up to inventory level $S$. Extension of this type of result is still an ongoing area of operations research [691].

Of great recent research interest is supply chain management of material flow over a network, coupling several agents who order from upstream suppliers and deliver to downstream customers, possibly also involving assembly, with the goal of minimizing cost of holding inventory or cost of shortages; see [672] for a recent review. Interestingly, an early investigator in this area was Forrester (see Section 3.3), who moved to the MIT Sloan School of Management and started a research program in System Dynamics in 1956. His book Industrial Dynamics [220] explored the dynamics of storage of goods in the chain from manufacturer to consumer via wholesalers. He developed the simulator Stella [220,565], which is still available [565,220]. Motivated by “what if” questions, Ho and coworkers developed the perturbation analysis method to obtain sensitivities to parameters of queuing, inventory, and other discrete-event systems, from simulations or traces of evolution [303,302].

It is interesting to note that Forrester continued to explore dynamics in broader contexts; in 1969 he published Urban Dynamics [221] that modeled population housing and industry in an urban area, and in 1971 he published the book World Dynamics [222] that modeled population, energy and pollution in the whole world. The book caught the attention of the newly founded Club of Rome [525] which funded a more detailed study “Limits to Growth” [466]. Forrester’s original model consisting of four differential equations was expanded to about 1000. The book predicted that growth was limited by natural resources. It was controversial because of many unvalidated assumptions; however, more than 12 million copies were sold, boosted by the 1973 oil crisis. Its central contention though is currently of great topical importance with respect to global warming as well as other environmental and ecological matters.

In an influential paper, Kimemia and Gershwin [382] formulated the problem of short-term scheduling of flexible manufacturing systems where machines are subject to random failures and repairs as a stochastic control problem, and exhibited interesting switching structure of the solution. In some cases the resulting stochastic optimal control problems have been explicitly solved to determine the optimal hedging point policies [6,75]. Kimemia and Gershwin also articulated a dynamic system approach to manufacturing systems, and proposed a hierarchical time-scale decomposition of the overall manufacturing problem ranging from long term capacity planning at the higher end to very short term part loading issues at lower end.

The dynamic systems viewpoint was further developed in [526], emphasizing the importance of scheduling policies that maintain stability of buffer levels. Counterexamples showed that even simple networks could be destabilized by scheduling policies when there was effective two-way interaction between machines, i.e., “feedback” [399,441,588,106]. There was much effort to
understand the stability of manufacturing systems and queuing networks. A powerful approach to establishing the stability of queuing networks, the fluid limit approach, was developed [570,161] as a complement to the direct Lyapunov-type analysis of the original stochastic system via Foster’s criterion for positive recurrence of Markov chains [224]. Another powerful approach to study performance, Brownian network models, was developed based on Brownian motion models of queuing networks [284]. They can be used to approximate heavy traffic behavior [325,326] of queuing networks. Fluid limits are analogous to the law of large numbers that provides information on the mean, while Brownian limits are analogous to the central limit theorem that provides information on the variance. A particular motivating system for this work was semiconductor manufacturing plants that feature re-entrant material flow [675,398], i.e., loops that create feedback effects. Policies based on the approach of viewing manufacturing systems as dynamic stochastic systems [440] were implemented on IBM’s 200mm wafer fab [481]. There is much current interest in stochastic processing networks [285]. They allow modeling of more general systems than queuing networks, allowing complex interactions between buffers, resources and activities. They encompass models not only of manufacturing systems but also of packet switches, call centers, etc.

The cumulative impact of all these control related developments was transformative in terms of emphasizing the dynamic stochastic nature of manufacturing and other such systems in contrast to static deterministic models. With respect to queuing systems, the first wave of work in the early 1900s due to Erlang [197,119] was motivated by problems of telephony, the second wave in the 1950s due to Jackson [341] was motivated by problems of job shops, and the third wave was motivated by problems of computer systems [51]. The fourth wave, motivated by problems of semiconductor manufacturing, and the most recent wave aiming to integrate very general problems of semiconductor manufacturing, has been heavily influenced by control theory.

There are also significant advantages in integrating the business systems for supply chain management and enterprise resource planning (ERP) with the process control systems at the job floor. This makes it possible to match process control with business objectives. Typical objectives are increased throughput, reduced energy consumption, improved capacity utilization, and reduced quality variability. The process control systems DCS and PLC systems are used for process control, and business systems like ERP (Enterprise Resource Planning) MRP (Material Resource planning) and master planning systems, delivered by companies like SAP and IBM, are used for plant management and business operations. To support interoperability between business systems and the process control system, an intermediate layer referred to as MES (Manufacturing Execution System) is often used. The international standard IEC 62264 [323], also known as ISA95, is providing support for Enterprise-Control System integration [107,584].

4.15 Simulation, Computing and Modeling

Simulation is useful because it allows exploration of the behavior of complex systems in a safe setting. The mechanical differential analyzer was driven by the need to understand power systems, and the electronic analog computer was invented to simulate control systems, as noted in Section 2.7. By 1960 analog computing was available at a number of industries and at some universities. At the turn of the century simulation was readily available on the desks of all engineers and students. Simulators were also combined with hardware to test controllers before they were delivered, so called hardware-in-the-loop simulation.

When digital computers appeared it was natural to use them for simulation [560]. The development was triggered by a paper of Selfridge [589] that showed how a digital computer could emulate a differential analyzer. Intense activity [642,111] was stimulated by advances in numerical integration of ordinary differential equations [160,294,201]. By 1967 there were more than 20 different programs available, e.g., CSMP [112] from IBM. The Simulation Council Inc (SCI) created the CSSL standard [630], a major milestone because it unified concepts and notation. The program ACSL from Mitchell and Gauntier Associates [476], which was based on CSSL, became the defacto standard. Like its predecessors, ACSL was implemented as a preprocessor to Fortran; the code for integration was interleaved with the code representing the model. It was easy to include Fortran statements as part of the model but documentation and maintenance were difficult. Another limitation was that computations were represented using the low level elements of analog computing, ACSL was a batch program. Recomputation was required when initial conditions or parameters were changed. The human-machine interaction was significantly inferior to traditional analog computing. The system Simmon [192], from 1972, admitted changes of parameters and initial conditions interactively without recomputation. The model was described in a special language with a formal definition, a simple language was also used for the interaction. Many other simulators appeared with the personal computer.

The general availability of computers in the 1970s inspired the development of tools for analysis and design of control systems. Computer-Aided Control System Design became a subspecialty with symposia and conferences. Initially, industry and university developed in-house systems. The appearance of personal computers and graphics in the mid 1980s stimulated a new generation of software. The state of the art circa 1985 is well summarized in the book [344].
Since design calculations are based on numerical algorithms, collaboration with researchers in numerical mathematics emerged. Two areas of particular importance were numerical linear algebra and integration of differential and differential-algebraic equations. Numerical analysts developed reliable computer code for solving Lyapunov and Riccati equations [415], and for integrating differential and differential-algebraic equations [244,277,276,278,24,274].

The advent of Matlab, created by Cleve Moler in 1981, was a game changer. Moler participated in the development of LINPACK and EISPACK software libraries for numerical linear algebra, and he wanted to have a simple way to test the programs. He designed an interpretive programming language in which it was very easy to enter matrices and perform the calculations by typing simple commands. Moler also added functions and macros (scripts) which allowed the user to extend the language.

Matlab was picked up by the control community, and tools for control system design were developed. Pioneering work was done by two companies in Palo Alto. Systems Control developed CTRL-C [433] and Integrated Systems developed MatrixX and SystemBuild [595]; both systems were based on Moler’s Matlab. John Little, who worked for Systems Control, obtained the rights to develop a PC version and teamed up with Moler and Bangert to found the company MathWorks. MathWorks developed the simulator SIMULINK [263] (originally called SIMULAB) integrated with Matlab, and Stateflow, a simulator for finite state machines [281]. MATLAB and Simulink are the dominant products but there is other similar software. The program Sysquake [532] is highly interactive, and executable files can be distributed freely. There are two public domain products, Octave [183] and Scilab [330]. Tools for control system design are also being developed for the scripting language Python [612].

John Little encouraged control researchers to develop toolboxes for solving control problems, and much of the work on computer aided control system design migrated to MATLAB. The toolboxes provided a convenient way to package theory and make it widely available. Mathworks also developed software for generating code for embedded systems from SIMULINK.

National Instruments (NI) supplied computer interfaces for instrumentation. In 1986 Kodosky of NI developed the program LabVIEW which allowed flexible configuration of instruments with nice graphical panels [387,350]. The program was based on data flow programming. It was originally intended for emulation of electronic instruments but it also became popular for control applications. National Instruments acquired MatrixX and features from it were gradually migrated to LabVIEW.

Simulation requires models of processes and controllers. Because of the wide range of applications, control engineers need models in many different domains. Even if modeling softwares for specific domains are available it is difficult to combine them. It is therefore highly desirable to have a unified approach to modeling that cuts across different domains.

A simple and general approach to modeling is to split a system into subsystems, define interfaces, write the balance equations for the subsystems, and add constitutive equations. This approach yields a description that is general, close to physics, and convenient for building libraries. A drawback is that much manual work is required to assemble the subsystems into a model which is suitable for simulation or optimization. Much of the work can be automated using computer algebra and object oriented programming. The procedure results in models that are differential algebraic equations. In the 1980s there had been significant advances in numerical solution of such equations [244,110,24,278]. The modeling method had been used for electronic circuits [490]. The language Dymola, developed by Elmqvist [193] in 1978, extended the method to general physical domains. Dymola had a formally defined syntax and it was implemented in Simula [76], the only object oriented environment available at the time. Many other object-oriented modeling languages were developed later when more memory and computing power became available, for example [191,455,454,504,509,574,234,665,345,113,346,530]. In 1992 Elmqvist started the company Dynasim to market a modern implementation of Dymola. The program quickly gained industrial acceptance, it was, for example, used to develop the Toyota Prius. Dynasim was later acquired by Dassault Systèmes.

A collaborative effort to develop a language for physical modeling was started in Europe in 1996. It was carried out by a diverse group with a broad range of experiences; modelers from many domains, software engineers, computer scientists and numerical analysts. Practically all European modeling groups participated. The effort resulted in the formation of the Modelica Association [28]. The first task was a formal definition of a modeling language; the first version was available in 1978 [194]. The Modelica language has many useful features such as units of variables, matrices and matrix equations, functions, hybrid modeling features and class parameters. A significant effort has been devoted to developing model libraries. There are libraries for many different fields, e.g., control systems, multi-body systems, electrical circuits, hydraulic systems, and thermal systems. The open source Modelica Standard Library contains about 1000 model components and more than 500 functions from many domains. The Modelica activity expanded, there are groups for advanced development, language specification, and libraries. Textbooks have appeared [643,233]. The 80th design meeting was held in 2013 and the 10th Modelica conference was held in 2014.
Several Modelica simulation environments are available commercially and there are also open source versions [28]. Models developed in Modelica can be exported to SIMULINK.

4.16 The Organizations Promoting Control

The International Federation of Automatic Control (IFAC), see Section 3.7, provided a global arena for control. Since IFAC operated through national member organizations it strongly contributed to the global spread of control. The national member organizations also organized conferences locally [81,537,186]. IFAC maneuvered very skillfully to maintain a world-wide control community in spite of political tensions during the cold war. The triennial IFAC World Congress has been operating since 1960 [354]. IFAC also arranges workshops and symposia. Participation in IFAC activities and committees was a good training experience, particularly for control engineers from small countries. Automatica became an IFAC journal in 1969 [149] with George Axelby [42] as the editor. IFAC’s activities have expanded substantially and today there are IFAC meetings almost every week. Later IFAC started several journals: Annual Reviews of Control (1977), Engineering Applications of Artificial Intelligence (1988), Journal of Process Control (1991), Mechatronics (1991) and Control Engineering Practice (1993).

There are also significant activities organized by other engineering organizations. The Instrument Society of America (ISA) formed in 1946, was renamed International Society of Automation in 2000. They organize a yearly Automation Week as well as Conferences and Symposia. ISA also publishes books and the Journals InTech and ISA Transactions. The American Society of Mechanical Engineers (ASME) created a division for control engineers from small countries. Automatica became an IFAC journal in 1969 [149] with George Axelby [42] as the editor. IFAC’s activities have expanded substantially and today there are IFAC meetings almost every week. Later IFAC started several journals: Annual Reviews of Control (1977), Engineering Applications of Artificial Intelligence (1988), Journal of Process Control (1991), Mechatronics (1991) and Control Engineering Practice (1993).

The IEEE Control Systems Society was formed in 1971, see Section 3.7. The long running Symposium on Adaptive Processes (1963–1970) became the IEEE Conference on Decision and Control (CDC). Interestingly it did so just as research in adaptive control began to take off. The CDC had generous acceptance practices for conference papers that encouraged researchers to submit their latest research and attend the annual conference. It became a fertile meeting ground with a large umbrella. The IEEE Transactions on Automatic Control, with a dynamic editorial board organized along very topical areas and regularly rotated with fresh talent, became a major publisher of theoretical research papers.

The American Automatic Control Council, which is the national member organization of IFAC in USA, organizes the yearly American Control Conference in collaboration with many engineering societies: AIAA, AIChE, ASCE, ASME, IEEE, ISA, and SCS. The European Control Conference, which now meets every year, started with a meeting in Grenoble in 1991. The Asian Control Conference, launched in 1994, now meets regularly every other year. The organization MTNS focuses on theoretical issues in system theory and organizes biannual conferences.

There are also strong organizations in China, England, France, Germany, Japan and many other countries which organize symposia and published journals.

Some organizations created during the war like the Radiation Laboratory at MIT were dismantled, others like the LIDS at MIT [477], the Coordinated Science Laboratory at the University of Illinois, the Institute of Control Sciences in Moscow, and the institutes run by the academies of sciences in Hungary, China, Czechoslovakia and Poland flourished after 1960. New institutions were also created. In Japan there were large national programs for Fourth Generation Computers and Fuzzy Control.

The Institute for Research in Computer Science and Control (IRIA) was started by the French Ministries of Research and Industry in 1967 as part of General de Gaulle’s Plan Calcul. It was one of the first research institutes that combined control and computer science. The institute was originally in Roquencourt outside Paris. It became a national institute and was renamed INRIA in 1979 and has since expanded with 8 regional research centers. The institute employs close to 4000 people, among them about 1000 PhDs and 500 postdocs. It became a powerhouse for research under superb leaders, among them the mathematicians Jacques-Louis Lions and Alain Bensoussan. INRIA has strong interactions with industry and has spun off about 100 companies. It pioneered work on control of systems governed by partial differential equations and created software like Scilab and Esterel was carried out at INRIA.

After the Second World War, there was a major expansion of research world wide, and a great growth of major research universities. Research funding increased significantly. The National Science Foundation was created in the US, and its mode of peer review of proposals leveled the playing field for researchers irrespective of location. After the experience with the fire control efforts and the Manhattan Project during the Second World War, there was a great infusion of funding to universities by the Department of Defense in the USA, often operating in a peer review mode. The European Union started major research programs, as did the Japanese government. Control research was a major beneficiary of all these developments in the period after 1960. Research from
universities in the area of control grew tremendously. There was a great expansion in hiring of control faculty. There was also a strong internationalization; students and teachers moved between different countries. The US benefited strongly from immigration of students and scientific talent from other countries. EU established the Erasmus Programme in 1987 followed by the Socrates, the Lifelong Learning Program and the Marie Curie program for experienced researchers.

5 Widening the Horizon

Around 2000 there were indications that control was entering a new era. Traditional applications were exploding because of the shrinking cost of computing, while new applications were emerging. The applications ranged from micro- and nano-scale devices to large-scale systems such as smart national power-grids and global communication systems. The expansion of the Internet and the cellular networks were strong technology drivers, as was the desire for systems with increased autonomy. A sign of the importance is that the inaugural Queen Elizabeth Prize for Engineering was awarded to Louis Pouzin, Robert Cerf, Tim Berners Lee and Marc Andreessen in 2013 for “the ground-breaking work, starting in 1970, which led to the internet and worldwide web. The internet and worldwide web initiated a communications revolution which has changed the world” [225]. It is an educated guess that it will also have a very strong impact on automatic control.

There was also a pressing need to develop methodologies for mass producing complex control systems efficiently. In the Golden Age control had benefitted strongly from interactions with mathematics. In this next phase, stronger interaction with communication engineers and computer scientists started to develop. Interactions with physics, biology and economics are also increasing. In this section we provide an overview of some of the trends. Our treatment of what lies ahead is necessarily speculative.

5.1 Advances in Computing and Networks

Computer hardware, following Moore’s law, is incomparably more powerful now than it was in 1960. Cray 1, delivered to Los Alamos National Laboratory in 1976, weighed over five tons, but could only deliver 250 megaflops, while the current Mac Pro desktop is a thousand times faster, delivering 90 gigaflops. In the past fifty years, embedded computers have also proliferated. Indeed, already by 1998, only 2% of all processors were workstations while 98% were for embedded systems [620].

Software engineering has made great advances. Experience based on large and complex projects has been codified and made reusable into design patterns, software frameworks and development processes [238,543].

One of the most noticeable changes is the birth and growth of communication networking. Telephony, which had originated around 1877, was based on a circuit-switched network. In 1969, the US Advanced Research Projects Agency (ARPA) developed a packet switched network that connected four university computers [72]. A flexible architecture was developed in 1974 [135] that allowed different previously incompatible networks to be interconnected. It featured hierarchical addressing, gateway routers between networks, and TCP, a protocol to ensure reliable delivery of packets in an ordered fashion across networks. Later this was split into two protocols, together designated TCP/IP [134], with the TCP part running only on end hosts, while the IP part took care of packet passing between networks or within a network. This made it feasible to scale up the network.

At around the same time, packet radio networks were also emerging. In fact one of the goals of TCP was to interconnect packet radio networks such as PRNET and SATNET with ARPANET. In 1971 the ALOHAnet packet radio network was developed at the University of Hawaii. It was used to connect users across the Hawaiian islands with a computer in Oahu [2]. The key innovation was the random access protocol to resolve contention between several users for the shared wireless medium. This was later the central feature of Ethernet [470], which was developed around 1973. Much later, the random access protocol was also adopted for use in wireless local area networks (WLANs) in the IEEE 802.11 standard which has proliferated across offices and homes worldwide [154].

In a parallel development, cellular telephone systems have also proliferated. The first design for a US cellular telephony system, the Advanced Mobile Phone System (AMPS), was developed in 1971. The first mobile portable handset was developed in 1973. In Japan, the Nippon Telegraph and Telephone (NTT) company developed the integrated commercial cell phone system in 1979. By 1960 several Nordic countries had their own mobile systems. In 1981 the Nordic Mobile Telephone Network (NMT) made it possible to use mobile phones across the countries. NMT later in 1992 created the Global System for Mobile Communications (GSM) which permitted users to place and receive calls globally. Researchers from ATT, NMT, NTT, and Motorola were awarded the 2013 Draper Prize [487] for these developments. Cellular technology and the World Wide Web of interlinked hypertext documents developed in 1990 have created a revolution in terms of connectivity and information access across the globe.

Today there are more than 5 billion wirelessly connected mobile devices and the number is expected to increase by one or two orders of magnitude by 2020 [196,145]. The
end-to-end latency is of particular interest for control. In the current LTE/4G system it is around 100 milliseconds, but is expected to be down to a few milliseconds in the 5G system planned for 2020. Such small latencies will significantly expand opportunities for control applications over the network.

In 1998, the Smart Dust project at the University of California at Berkeley [353] developed tiny devices called “Motes.” that could compute, communicate wirelessly, and to which sensors could be connected. The Rene Mote developed by CrossBow Technologies in 1999 had an AT-MEL CPU, 512 Bytes of RAM, 8K of Flash memory and a packet radio that could communicate at about 10Kbps [300]. A key development was the TinyOS open source operating system [423,155] which has facilitated much experimentation by academic researchers. Since their original development there have been several generations of Motes. They can be used to form relatively large “sensor networks” with widely distributed nodes that communicate wirelessly and perform computation on the data they receive.

If one attaches actuators to sensor networks, then one obtains what in some computer science communities are called “sensor-actuator” networks, a notion familiar to control engineers. Of particular interest for control is WirelessHART, which is designed as a communication standard for process control [151,617]. Also of interest for control is ISA100.11a developed by the International Society of Automation (ISA) [336].

### 5.2 Control of and over Networks

Information between multiple sensors and actuators can be transported over a packet-based network. Thus, the advances in networking make it possible to deploy control systems on a large scale, giving rise to “control over networks,” or what is dubbed “networked control” [45]. Since loops are closed over the communication network, it plays an important role in overall stability and performance. Networks also require control to provide good performance, an area called “control of networks.”

Congestion control is an early example of use of feedback in a network. The rate of injection of packets into a network is regulated to avoid congestion while maintaining a high throughput [660]. In fact, TCP, later the TCP/IP protocol, which does this is at the heart of the Internet, and is one of the reasons why the Internet has proliferated so rapidly. More generally, control principles and loops are needed at several levels for the operation of networks. The early ALOHA protocol [2], a component of WiFi, is a feedback control scheme which attempts to throttle nodes when they are causing too many packet “collisions,” in a manner similar to James Watts’ Governor.

It is important for control systems to design communication networks that provide the “quality of service” that control loops need. The networks will have to not only deliver packets from sensors to actuators at a specified throughput, but will also have to deliver them within a specified delay. The current Internet is what is called “Best Effort”; it does not provide such guarantees, but they are important if one is to close loops over networks. The CANBus [144] and Field Bus system [138] have been designed for control applications. A major challenge is to design wireless networks that provide such quality of service. For example, it is of interest to replace current wired intra-vehicular networks connecting about 75 sensors and 100 switches with a wireless access point serving them, since that can save weight, reduce complexity of manufacture, permit easier upgrades, etc.. The problem of characterizing what types of quality of service access points can support, and how to do so, is of great interest [312].

Concerning control over networks, issues such as the design of the system, proofs of stability, or establishment of performance, need to take into account the characteristics of the imperfect network over which information from sensors to actuators or actuators to sensors may be transported. This problem can be addressed at different granularities to take into different aspects of the constraints posed by the network or communication channel involved [439].

Probably one of the earliest issues to confront with respect to the control system design is when to sample the system so as to reduce the data needing to be transported over the network. One could of course sample periodically or at given time points; this is reminiscent of the manner in which the Riemann integral is defined. However, it may result in the system being sampled unnecessarily even if nothing has changed. An alternative is to sample it on an event-driven basis; which is reminiscent of the manner in which the Lebesgue integral is defined [33].

From the viewpoint of transporting packets over the network, three important characteristics are the rate at which the network can handle incoming packets, the delay that packets may experience before they are delivered at the intended destination due to the traffic load on the network, and the probability or likelihood with which the network may drop packets. All three aspects are of interest vis-a-vis their impact on the control system. For an LQG system the stability of the associated Kalman filter depends on the probability that packets containing observations are lost [610]. For the control problem, it is of interest to determine the data rate needed to be provided by the channel in order to stabilize a given linear system [492]; this is also related to the problem of how to quantize measurements for the purpose of control [689,118]. At a more fundamental level, when one wishes to stabilize unstable systems over control loops contain-
There has been increasing interest in control by the computer science community, and vice-versa. One reason is the increasing employment of feedback in computing systems. Another is the proliferation of embedded computing systems. The large number of computer control systems in cars has driven the need to automate design, production and testing of control systems. Also, with the increased complexity of systems featuring computers in the feedback loop, there is a need for methodologies and tools for reliable system design. In many complex applications, it is important to design systems with guaranteed safety.

Similar to “control of networks,” control is also increasingly being applied to computing systems [293]. The main drivers are the increased flexibility and better quality of service, and the desire to save energy and cost. Feedback-based techniques have been considered for dynamic management and scheduling of resources such as CPU time, memory, IO bandwidth, and power, in systems ranging from embedded computers used in, e.g., smartphones, to data centers hosting server-based cloud applications. Successful applications have been developed in web storage systems, high-performance server systems, real-time databases, software performance tuning, and multimedia streaming, to name a few.

In the reverse direction, the interaction of some sub-communities in computer science with control has been longstanding. A prime example is the real-time community. In 1961, the IBM 1720 Process Control Computer System was installed in three plants [287]. The theoretical foundation of real-time systems started with the work of Liu and Layland [434]. This pioneering work considered the problem of how to schedule a CPU to serve several tasks, where jobs in each task are periodic and require a certain execution time. The Rate Monotonic policy developed by them prioritizes jobs according to the frequency with which jobs of that task arrive. It is a particularly simple static priority policy that has seen widespread implementation. For a large number of tasks rate monotonic scheduling guarantees that all tasks will be executed properly provided that the CPU utilization is less than \( \log 2 = 0.68 \). This conservatism can be reduced by applying scheduling algorithms based on feedback [22,594]. A prominent example of the importance of real-time computing considerations in control systems is the priority inversion problem in the real-time computation system that occurred in 1997 on the Mars Rover [348,562]. Researchers in real-time systems have also begun addressing the problem of robustness of control loops, where the robustness includes errors in implementation. The so called “Simplex” architecture of [592,593] addresses the problem of robustness to software bugs in the implementation of new control algorithms.

An important aspect anticipated of future control systems is the interaction between the physical world often modeled by differential equations and the logical dynamics of the computational world. Typically, one would like to establish the properties of the composite systems comprising both. The emerging field of hybrid systems is one attempt to address these challenges [451,297,419,71]. It is an interesting meeting place of control and computer science [70]. Hybrid automata models have been used for this purpose [295]. It is of interest to determine if the system is “safe,” i.e., it never reaches an unsafe state. Software tools for computing such quantities are useful, e.g., [413]. However, determining that is undecidable for general models, and it is of interest to characterize what is decidable [9,296]. More generally one would like to establish properties such as safety and liveness of an entire system such as an automated distributed transportation system [265,380].

More broadly, “time” is an essential matter for control systems, in contrast to, say, general purpose computing [549]. That is, for safety critical systems, timeliness of interactions is important for maintaining safety, stability, etc. The time-triggered architecture is an approach to developing distributed embedded systems that seeks to attain reliability by temporal coordination [390]. When closing loops over a wireless network, there are certain limitations to synchronizing clocks; certain combinations of delays and clock offsets cannot be resolved [231].

Control systems are often safety critical. Their security is therefore a major concern. They may be amenable to attacks occurring over the network to which they are connected. The recent Stuxnet worm specifically attacked control systems [200,463,141]. There have been other less reported attacks of a natural gas pipeline system [582], a water system [198], a Supervisory Control and Data Acquisition system [605], trains [426] and power utilities [267]. Defense of ordinary data networks is already problematic, and defense of complex control systems is even more so since the attacks can take advantage of complex interactions between the networking, computational, control systems, and physical layers. Much needs to be done in the area of security [499]. There are some standardization efforts under way [497,627,333].

Automatic control has a strong base in mathematics. The interaction goes in both directions; a wide range of mathematics has found its use in control, and control has occasionally stimulated the development of mathematics. Some examples are system theory, optimal control, stochastic control, and nonlinear control [207,485]. With the convergence of communication, control and computing, newer theoretical areas such as hybrid systems and
real-time information theory have emerged. Theories of stability or performance or safety will also need to straddle different areas, since it is the overall system that is ultimately the determinant of performance. In some specific systems one can provide holistic handcrafted proofs of performance of the overall system that includes discrete event dynamics, real-time scheduling, kinematics, etc. [265]. However as we build more complex systems such as automated air transportation systems, it is useful to be able to automate the proofs of safety. Complexity however is a major challenge for computational procedures, so control researchers will need to develop new theories that permit tractable ways of modeling.

5.4 Autonomy

Research on systems with adaptation and learning has been well developed for a long time, as noted in Section 4.12. However, higher levels of autonomy which include cognition and reasoning will be required in the future. It is pointed out in an NAE study [488] that:

\[\text{Everything will, in some sense, be smart; that is, every product, every service, and every bit of infrastructure will be attuned to the needs of the humans it serves and will adapt its behavior to those needs.}\]

Interesting experiments with robot cars were performed by Ernst Dickmanns at the end of the last century. He equipped cars with cameras and other sensors [168]. In 1994 he demonstrated autonomous riding on a highway near Paris, and in 1995 one of his cars drove autonomously (though with human supervision) from Munich to Copenhagen at speeds up to 175 km/hour. The car was able to overtake and change lanes. More recently in 2007 several autonomous cars competed in a deserted city in the DARPA Grand Challenge. One of the rules required that the cars follow the traffic rules of California. More recently, Google has developed a driverless car [195]. Carnegie Mellon’s Boss is another autonomous vehicle [555]. Even if full autonomy is not introduced on a massive scale, elements of it, such as collision avoidance and parking assist, are now available in new cars. Autonomous air vehicles are in operation and an unmanned cargo mission has recently been performed with a Black Hawk helicopter [673].

Humanoid robots are other examples of systems with a high degree of autonomy. Research in Japan has been particularly dominant in this area. Several generations of robots have been developed. Toyota recently announced a violin playing robot. Humanoid robots that act as patients have been developed for training dentists.

5.5 Model Based Design

The automotive industry started to have an impact on computer-aided design of control systems when computer control was introduced in cars in the 1970s, a development that accelerated when the systems became more complex. More than 80 million cars were produced in 2012; ordinary cars may have ten or more electronic control units (ECU) while advanced cars may have over 100 ECUs. With this scale it is important to have efficient engineering procedures for design and manufacturing of the systems. Often the controller is co-designed with the plant. There are similar needs in many other industries even if the numbers are smaller.

Development typically includes the following tasks: requirements, modeling, control design, code generation, implementation, hardware-in-the-loop simulation, commissioning, operation and reconfiguration. Validation, verification and testing are inserted between the different tasks since it is expensive to find errors late in the design process. Since models are key elements of the procedure it has been known as model based design (MBD). The advantage of using models is that fewer prototypes have to be built; particularly important when building new systems like hybrid cars. The aerospace and automotive industries have been early adopters of MBD which is currently developing rapidly [379,275]. Use of MBD is endorsed by the following quote from an NAE study [488]:

\[\text{There will be growth in areas of simulation and modeling around the creation of new engineering structures. Computer-based design-build engineering ... will become the norm for most product designs, accelerating the creation of complex structures for which multiple subsystems combine to form a final product.}\]

An example of the use of MBD is that suppliers of components for climate control systems for cars in Germany are now providing not only hardware but also validated dynamic models of their equipment [428]. This makes it possible for car manufacturers to simulate the complete system and to explore the consequences of using components from different suppliers on fuel consumption and comfort.

In system design it is desirable to explore design choices and to investigate several process configurations. A cardinal sin of automatic control is to believe that the system to be controlled is given a priori. Control problems that are difficult can be alleviated by modification of the process or the system architecture. Integrated design of a process and its controller is highly desirable. As mentioned in Section 2.5, the Wright brothers succeeded where others failed because they deliberately designed an unstable airplane that was maneuverable, with a pilot used for stabilization. There are still substantial advantages in having an unstable aircraft that relies on a control system for stabilization. Modern fighters obtain their performance in this way.

There are tools for some phases of the design process:
DOORS for requirements [321], e.g., CAD programs for equipment design, Modelica for modeling, MATLAB for control design, and SIMULINK for simulation and code generation. Systems for documentation and version control are also available. Even if it is desirable to have a complete design suite it is unlikely that a single software package can serve all the needs. Software tools therefore have to be designed so that they can be combined. To give one example, Dassault Systèmes are combining Dy-mola/Modelica with their CAD program CATIA. This means that 3D geometry data, masses and inertias are available directly from the CAD system. High quality 3D rendering is also available to animate the simulation results.

A recent development in the industrial simulation community is the introduction of the Functional Mock-up Interface (FMI) designed to facilitate tool interoperability at the level of compiled dynamic models. FMI specifies an XML schema for model meta data, such as names and units, and a C API for evaluation of the model equations. The first version of FMI was introduced in 2010 and since then a large number of tools have adopted the standard. Future versions of FMI will support communication of complete models in XML format, which is suitable for use with integration in symbolic tools that can explore the structure of models beyond evaluating model equations [522].

Car manufacturers typically buy systems consisting of sensors, actuators, computers and software as packages from suppliers. This approach works very well when there were only a few electronic systems with small interaction. The situation became complicated when more control functions were added because a sensor from one subsystem could be used in another system. The automotive industry, including their suppliers and tool developers, therefore created AUTomotive Open System ARchitecture (AUTOSAR), an open and standardized automotive software architecture, for automotive electrical and electronic systems [40].

There are a wide range of interesting problems that appear after a control system is designed and implemented. First, the system has to be commissioned and all control loops have to be brought into operation. Then it is necessary to continuously supervise and assess that the system is running properly. Many of the problems occurring during this phase have only recently begun to be addressed in a systematic fashion. Typical issues are fault detection and diagnosis, but there are also many other interesting problems, such as loop assessment and performance assessment. Developments in this area are strongly motivated by the drive for safety and higher quality. Commissioning can be influenced substantially by a proper control design. Automatic tuners, for example, can drastically simplify the commissioning procedure.

When the automatic control system becomes a critical part of the process it may also become mission critical, which means that the system will fail if the control system fails. This induces strong demands on the reliability of the control system. An interesting discussion of the consequences of this are found in the inaugural IEEE Bode lecture by Gunter Stein [623].

5.6 Cyber-Physical Systems

The increased use of communication, and the increased sophistication and complexity of the software both in control systems design as well as operation, has led to closer interaction between control, computing and communication. It is also fostering the development of control systems of large scale. As was the case for the previous eras, this third potential platform revolution, after analog and digital control, is also creating a need for a framework for rigorous design.

This interaction has also been stimulated by several research funding agencies. DARPA launched a research program called Software Embedded Control in 1999 [575], which was followed by an NSF project on Embedded and Hybrid Systems. The European Union launched in 2001 the ARTIST Accompanying Measure on Advanced Real-Time Systems [19], and in 2004 the Advanced Research and Technology for Embedded Intelligence Systems (ARTEMIS) program [496]. In 2006, a group of researchers and program managers in the US coined a name, “Cyber-Physical Systems (CPS),” to describe the increasingly tight coupling of control, computing, communication and networking. In the US, the National Science Foundation established a major research funding program in Cyber-Physical Systems [44]. In its strategic plan towards 2020 [331], INRIA emphasizes “the challenge of very large digital, embedded and buried systems, and of systems of systems.” The Robert Bosch Center for Cyber-Physical Systems was established at the Indian Institute of Science in 2011. In Sweden, the Strategic Research Foundation supported 10 year Linnaeus Grants for three centers that support control research [3,484,417].

In 2008, a week long annual event called “CPS Week” was launched, which has grown to include five collocated conferences, the ACM/IEEE International Conference on Cyber-Physical Systems (ICCP), the Conference on High Confidence Networked Systems (HiCoNS), Hybrid Systems: Computation and Control (HSCC), the ACM International Conference on Information Processing in Sensor Networks (IPSN), and IEEE Real-Time and Embedded Technology and Applications Symposium (RTAS). In 2013, the Proceedings of the IEEE celebrated the hundredth anniversary of IRE (the IEEE was formed from the union of IRE and AIEE in 1963), and published a 100th Anniversary Issue in which Cyber-Physical Systems was one of nineteen topics chosen for
inclusion [381]. There is also great interest in “The Internet of Things,” focusing on connecting large numbers of general physical objects, from cattle to cars to coffee makers, expected to number 50 billion by 2020. IEEE is starting two new journals, IEEE Transactions on Control of Network Systems to commence in 2013, and IEEE Internet of Things Journal to commence in 2014.

5.7 Complexity of Systems

There is a general tendency that engineering systems are becoming more complex. Complexity is created by many mechanisms: size, interaction and complexity of the subsystems are three factors that contribute.

Chemical process control systems can have many thousands of feedback loops. Recirculation schemes save energy and raw material and reduce pollution, but they introduce coupling from the output streams to the input streams, which generates interactions and complexity. Efficient systems for distributing goods globally using computer assisted supply chains use complicated networks for transport of goods and information. Astronomical telescopes with adaptive optics may have a large number of reflecting surfaces whose orientations are controlled individually by feedback systems. Even in a small system, due to the “curse of dimensionality,” the resulting size can be extremely large after discretization.

One of the great achievements of the last five decades has been the development of a rigorous foundation for the study of complexity of computing as size increases [152,372], though some basic questions remain open. A fundamental challenge pervading many applied domains, including control system design and analysis, is to develop models that can result in tractable algorithms whose complexity scales polynomially, preferably of low degree, in the size of the system.

The Internet and the electricity grid are perhaps among the most complex systems that have been engineered. Both depend critically on feedback at several levels for their operation. As noted in Section 5.2, in the case of the former, this is what is meant by control of networks. Algorithms suitable for small systems may not be suitable for large systems. It is also of interest to determine scaling laws that provide insight into how system performance changes as size grows.

Another question attracting great interest is how to take advantage of the increasing availability of large amounts of data, called “big data.” This can be especially useful in understanding the behavior of large socio-economic-technological systems, whose “physics” is not well understood. An example is the control of demand response in the emerging smart grid supplied by renewable energy and controlled by price signals.

Another factor that introduces complexity is that the systems are hybrid: continuous systems are mixed with logic and sequencing. Cruise control in cars is a simple example. Other examples are found in process control, where many continuous controllers are combined with systems for logic and sequencing. Such systems are very difficult to analyze and design. The modern car is an example of a complex system; it has several networks and up to 100 electronic control units. Specification, design, manufacturing, operation and upgrading of such systems is an increasingly complex task.

5.8 Physics

Interactions between physicists and engineers are increasing [53]. Feedback control systems have played a critical role in instruments for physics, more so with the increasing complexity of the experiments. For example, governors were used to track the motion of planets in early telescopes [456], feedback was an essential element of early spectrometers [502], and the 1912 Nobel Prize in Physics was awarded to Gustaf Dahlen for “invention of automatic regulators for use in conjunction with gas accumulators for illuminating lighthouses and boys” [505]. The Dutch engineer van der Meer shared the 1984 Nobel Prize in Physics for a clever feedback system for generating a high density beam in a particle accelerator [505]. Feedback has also proven crucial for physics experiments. Large telescopes use adaptive optics to reduce the disturbances caused by the density variations in the atmosphere. Control systems are also widely used at the micro and nano-scales [262,190]. Binning and Rohrer shared the 1986 Nobel Prize in Physics for the invention of the scanning tunneling microscope. A variation, the atomic force microscope, is now a standard tool for biologists and material scientists, capable of providing images with sub nanoscale resolution. The control systems in the instruments are critical; improved control gives immediate benefits in terms of sharper and faster imaging. Great challenges faced by modern control engineering were overcome in making the Large Hadron Collider (LHC) operational. Many interacting system components function on time scales that differ by several orders of magnitude, from nanoseconds for particle beam steering to months for cooling large electromagnets.

Control has also had impact at a more fundamental level. A long time ago it was attempted to explain shear flow turbulence by linearizing the Navier Stokes equation. The hypothesis was that the linearized equations would become unstable when the Reynolds number increased, which somewhat surprisingly did not happen. Another attempt based on linear analysis was made by Dahleh and Bamieh [46]. They computed the gain of the operator mapping surface irregularities to velocity based on a linearized model, and found that the gain increased with
the Reynolds number in a way compatible with experimental data.

Control has also been applied to quantum systems [317,377,227], NMR imaging [378] being one application. Another spectacular application is given in [120] where it is proposed to break molecules into ions by applying very fast laser pulses. The problem can be formulated as an optimal control problem for the Schrödinger equation for the molecule, where the cost criterion is to break the molecule with minimal energy.

5.9 Biology and medicine

In 1932 the physiologist Walter Cannon wrote the book *The Wisdom of the Body* [132], in the introduction of which he says:

> Our bodies are made of extraordinarily unstable material. Pulses of energy, so minute that very delicate methods are required to measure them, course along our nerves. ... The instability of bodily structure is shown also by its quick change when conditions are altered. ... The ability of living mechanism to maintain their own constancy has long impressed biologists. ... Hence, then, is a striking phenomenon.

He then went on to say:

> Organisms composed of material which is characterized by the utmost inconstancy and unsteadiness, have somehow learned the methods of maintaining constancy and keeping steady in the presence of conditions which might reasonably be expected to prove profoundly disturbing.

Cannon’s book is based on insights obtained by careful observations and experiments. In our terminology we can summarize the above statements as: the human body has fantastic control systems. It is, however, a long way from this qualitative statement to quantitative results based on models and analysis, illustrated by the following quote from the book *The Way Life Works* [306]:

> Feedback is a central feature of life. All organisms have the ability to sense how they are doing and to make necessary modifications. 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.

Feedback has been used extensively when investigating biological systems. Hodgkin and Huxley received the 1963 Nobel Prize in Medicine for “their discoveries concerning the ionic mechanisms involved in excitation and inhibition in the peripheral and central portions of the nerve cell membrane.” They used a clever feedback system to investigate the propagation of action potentials in the axon. The measurement technique was further refined by Neher and Sakmann who received the 1991 Nobel Prize in Medicine “for their discoveries concerning the function of single ion channels in cells.”

Today, there are many efforts to develop efficient tools for patients and doctors to augment the body’s natural feedback systems when they fail [329]. Robotics surgery is now well established. Mechanical hearts are already in use. Experiments with on-line control of blood sugar are performed [520,332], as is automatic control of anesthesia, to mention a few examples.

Control subsystems in animals are being explored. Behaviors of insects and birds are being investigated by wind-tunnel and free flight experiments. It has also been attempted to make artificial devices that mimic animals. Some of the more basic functions such as standing, walking and running can now be performed by robots [678]. Some of these functions are simple control tasks involving stabilization and regulation, but there are many more complicated tasks that require cognition. Interesting efforts in this direction have been made by Professor Hiroshi Ishiguro at Osaka University who has designed several humanoid robots [23]. Experiments with synthetic biology are also performed at the molecular level [13].

There has been increasing interest in control and systems biology [376,327,157,240]. Rather than taking a reductionist approach consisting of studying an isolated entity, systems biology which originated around 1988 aims to understand how the components interact as dynamical systems, whether at the cell or organ levels. It thereby aims to unravel the complexity of biological and disease networks. A quote from the Institute of Systems Biology [218] summarizes it thus:

> Even the simplest living cell is an incredibly complex molecular machine. It contains long strands of DNA and RNA that encode the information essential to the cells functioning and reproduction. Large and intricately folded protein molecules catalyze the biochemical reactions of life, including cellular organization and physiology. Smaller molecules shuttle information, energy, and raw materials within and between cells, and are chemically transformed during metabolism. Viewed as a whole, a cell is like an immense city filled with people and objects and buzzing with activity.

The interaction between control engineers and biologists is increasing and new academic departments and educational programs are being established at major universities.
5.10 Economics

There are common interests between economists and control engineers in game theory, input-output models, stochastic control, optimization and system identification (econometrics). The economists Simon, Nash and Arrow have already been mentioned in this paper.

Early work in economics was done by Adam Smith and and Maynard Keynes. Keynes’s work was largely conceptual but he also introduced simple models that were important for emerging out of the Great Depression in the 1930s, such as the notion of multipliers which indicate the impact of government investment on GDP. One way to assess the research in economics is to look at the works that have been awarded the Economics Prize since the first one in 1969. Several have strong connections to control: the 1978 and 1986 Prizes on decision-making, the 1994 and 2005 Prizes for game-theory, the 1997 Prize on evaluation of derivatives, the 2004 Prize for driving forces behind the business cycles, and the 1969, 1980, 1989 and 2003 Prizes for modeling and time series analysis.

Economics influences us all and requires our attention, and economists have been well aware of the role of control for a long time [374]. However, the economic system is a large, complex, global, distributed, dynamic system with many actors, governments, banks, investment banks, companies and individuals. Governments control by laws and taxes, the central banks set interest rates and control money supply, companies and individuals buy, sell and invest. The different actors have widely different goals. Behavioral economics shows that individual decisions can be based on emotional and social aspects. The system oscillates in business cycles and there are occasional crises, clearly pointing to control problems.

Krugman who received the 2008 Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel (the Economics Prize) says the following in his book [396]:

*We have magnetotrouble, said John Maynard Keynes at the start of the Great Depression. Most of the economic engine was in good shape, but a crucial component, the financial system, wasn’t working. He also said this, “We have involved ourselves in a colossal muddle, having blundered in the control of a delicate machine, the workings of which we do not understand.” Both statements are as true now as they were then.*

Comparing this with the quote by Wilbur Wright on the difficulties of balancing and steering airplanes in Section 2.5 it is clear that leading economists realize that there are severe control problems in economics.

6 The Interplay of Theory and Applications

Feedback control is a key component of an amazingly broad range of applications, in fact touching upon almost everything in the modern world. The theory of control is a similarly deep field, drawing upon a broad range of mathematics, and sometimes even contributing to it. One can ask how such a broad range of applications and deep theory came to be. The answer is rich with texture, as expounded on in this paper.

Control systems need enabling technologies in order to be implementable. The availability of such technologies, and the advances in technology, sometimes revolutionary advances, have played a key role in shaping the evolution of control. An important role has also been sporadically played by motivating grand challenge applications that have led to great societal, i.e., government and industrial, investment in development of both control technology and theory. The development of the modern research university system with systematic research funding has also played a key role in fostering academic research. And, of course, there have been visionaries and deep researchers who have been at the front-line.

The evolution of the field has not been smooth. There were several fits and starts in the process that cumulatively resulted over the long haul in advances in both practice and theory. The “gap between theory and applications” has been a dynamic process which has frequently raised controversy in the community. In a nuanced editorial in 1964 [41], Axelby observes that:

*Certainly some gap between theory and application should be maintained, for without it there would be no progress. ... It appears that the problem of the gap is a control problem in itself; it must be properly identified and optimized through proper action.*

There were periods of time when applications were ahead of theory, with success not resulting until key theoretical breakthroughs had been made. At other times, the gap was in the reverse direction. Theory was developed in an open-loop fashion without feedback from real-world implementation and application, investigating what may possibly be feasible applications in the future and suggesting imaginative possibilities, even though the technology for implementation was not yet ripe enough or constraints fully understood. The field has seen both application pull of theory as well as theory push of applications. One could say that this gap, in whichever direction, has been a source of creative tension that ultimately led to advances to the benefit of both theory and applications.

Understanding the interplay of these diverse interactions provides some insight into how engineering research and development and knowledge have evolved vis-a-vis control, a key pillar of the modern technological era.
There have been several periods when applications have been developed without much of theory. In 1933, one of the leading actors in process control, Ivannoff, said [65, p 49]:

*The science of the automatic regulation of temperature is at present in the anomalous position of having erected a vast practical edifice on negligible theoretical foundations.*

Even today, PID control is enormously successful; it is one of the simplest ways to benefit from the power of feedback. In ship steering, the tinkerer Sperry outdid the theoretician Minorsky, as recounted in [65]. A similar situation prevailed in early flight control [465, p 5]:

... they seemed to have made progress with a minimum amount of mathematics until after the end of the 1939-1945 war. ... During roughly the first 50 years of aviation’s history, the study of the dynamics of aircrafts and their control system was of negligible interest to designers, who learned to get by with rules of thumb ... This was in spite of the fact that a mathematical theory for the stability of the unattended motion and of the aircraft’s response to control was developed at an early date. Very fortunate, wartime pressures produced two developments that fundamentally altered techniques for the design of automatic flight control systems. The first of these was the theory of servomechanisms: the second was the electronic computer. Analysis and simulation are today the twin pillars on which the entablature of aircraft flight control system design stands.

The gradual evolution over several decades from an application of feedback to a broad and deep theoretical framework for its analysis and design can be clearly seen in the case of the centrifugal governor. Used early on in windmills in 1745 [460], and subsequently by Watt in steam engines in the 1780s, it was originally a proportional controller, with integral and derivative action subsequently added [65]. About a century later, Vyshnegradskii [670] and Maxwell [456] initiated a theoretical investigation. This led to the work of Hurwitz [320] and Routh [569] on stability analysis.

In a similar vein, the challenge of designing repeaters for long-distance telephony led to the invention of the feedback amplifier by Black [83] in 1927, though without a theory. In the absence of a fundamental theoretical understanding this was a very difficult technology to employ. Difficulties with instability encountered in the lab inspired Bode and Nyquist to develop the theory for the feedback amplifier [507,93]. There was one other extremely important factor in this case: the presence of a powerful corporation with a research and development lab: AT&T. All three, Black, Nyquist and Bode, were its employees. This is a supreme example of the success of a large concentrated effort, in this case by a monopoly, in bringing to bear sufficient resources over a long period to solve fundamental application challenges.

Another example in the same vein was the development of fire control, originally for naval warfare and subsequently for anti-aircraft fire. The latter received sustained support by the US Government in a project led by Vannevar Bush, and eventually led to the development of servomechanism theory, as described in Section 3. In his earlier work on power system networks at MIT, Bush [684] had observed that:

*Engineering can proceed no faster than the mathematical analysis on which it is based. Formal mathematics is frequently inadequate for numerous problems pressing for solution, and in the absence of radically new mathematics, a mechanical solution offers the most promising and powerful attack.*

Similarly, the cold war air defense network Semi-Automatic Ground Environment (SAGE) led eventually to work on sampled data systems of Ragazzini, Jury, Franklin [553,352] (Section 3.1). Later, flight control with more stringent requirements flowing from pushing the flight envelope led to work on multi-variable stability margins, i.e., robustness [571] (Section 4.13). More recently, the need to design safe and reliable embedded systems, e.g., for automobiles, is driving work on model based development, validation and verification.

The development of model predictive control is another interesting instance of interaction, in this case serendipitous, between theory and practice. Richalet [564] solved a range of practical problems in a discrete time setting by calculating the optimal trajectory, but only using the initial portion of it, and then recomputing the trajectory after each sample. This procedure was called *receding horizon control*; it had also been investigated in [408]. Later, two chemical engineers, Charlie Cutler and Brian Ramaker, were running a refinery during a strike. A key problem was that several control variables had to be set during a grade change. Cutler and Ramaker solved the problem by first determining the steady state gain experimentally. Given the desired changes in the output, the appropriate changes in the controls were then obtained by matrix inversion. To make faster changes they measured the multi-variable pulse response and computed a set of future controls that would change the state according to a desired trajectory. They applied the first control signal and repeated the procedure. Not being versed in control theory, they called the impulse response the “dynamic matrix,” and the design procedure was called *dynamic matrix control (DMC)* [158,546]. When the strike was over, Cutler and Ramaker returned to research and development, and started refining the method. They sponsored research and arranged workshops at Shell to interact with academia [545,544], which in turn initiated research on stability and robustness [241,480,457,60].
On the other hand, there was a long period in the second half of the twentieth century when theoretical research in control was exuberant. When the digital computer came to be introduced into the feedback loop, as noted in Section 4, it caused a platform revolution. It obviously needed a different kind of theory, state-space theory, from the the frequency domain theory that had been so appropriate for harnessing the feedback amplifier. In fact, for his paper [360], Kalman claimed that “This paper initiates study of the pure theory of control.” The state-space theory found immediate application. Swerling applied his filter [634] to the estimation of satellite trajectories using ground based measurements [269], and S. F. Schmidt of NASA’s Ames Research Center applied Kalman’s filter [368] to the circumlunar navigation problem, and developed it for real-time on-board navigation in the Apollo mission [269]. The development of state space theory became a very active research topic in academia from about 1960 for almost four decades. The mathematics and control group at the Research Institute for Advanced Studies (RIAS) that Solomon Lefschetz began leading in 1957 played a major role, the center having been founded in 1955 to conduct work similar to what was being done in the Soviet Union [269]. Control researchers delved fully into studying linear differential/difference equations modeling linear systems, and beyond, into stochastic systems, nonlinear systems, decentralized systems, and distributed parameter systems. This was facilitated by well funded research programs of the U.S. Government for university researchers, a legacy of Vannevar Bush’s wartime efforts in defense research and development and subsequent advocacy for the creation of the National Science Foundation in the U.S. Many imaginative possibilities were investigated. A rich theory of model-based systems began to emerge. There were some important applications that emerged, such as system identification and adaptive control, as described in Section 4. This was particularly the case in process control where there was a rich tradition of experimentation.

The theory at this time was in some respects ahead of technology, since many ideas explored could not yet be implemented, and had to await the development of powerful computing, networking, etc. The theory explored the limits of the feasible, whether due to the infinite-dimensionality of the resulting solution, or the curse of dimensionality of dynamic programming, or more broadly complexity of either the solution or its implementation. For example, the class of nonlinear filtering problems for which the optimal solution was finite dimensional was carefully investigated. Efforts such as this served a valuable purpose in calibrating what was feasible and were important in themselves, a la information theory, even though they did not result in major applications. However, theory that earlier showed the limits to explicit solution was revisited in subsequent decades after computational power available had greatly increased. An example is the control of partially observed systems, which is now one of the mainstays of machine learning and artificial intelligence [596], as noted in Section 4.10.

For the reasons noted above, this extremely fertile period for theoretical research also led to control theory developing in an “open-loop” fashion without constant feedback from real-world applications against which it could be tested. There was a time gap between theoretical ideas and their testing against reality, if they could be tested at all, for, in some cases, the implementation technology was not yet ripe. Important shortcomings were only discovered after the theory was tried in an application. An example is the lack of multivariable stability margins for linear quadratic Gaussian control [170] alluded to in Section 4.13, discovered in simulation testing of submarines [571]. In fact, robustness to model uncertainty was broadly one of the major shortcomings of early model-based state-space theories. That this was true even in adaptive control when the model class within which parameters are fitted does not contain the true system became a cause celebre [567] that subsequently resulted in frenetic activity in “robustifying” adaptive control. Eventually there developed a theory that encompassed model uncertainty, culminating in a paper that won the IEEE W.R.G. Baker Award [172].

In the golden age for control theory research, control became well established in academia. There was a critical mass of theoretical researchers to dig deeply into many areas, facilitating the formation of a strong theory. A paper from Berkeley [73], entitled “Moving from Practice to Theory: Automatic Control after World War II,” describes this phenomenon in detail. Inevitably, there was an attitude of l’art pour l’art. Sometimes a lot of effort was also devoted to less important problems forgetting or oblivious of the following words from John von Neumann [347]:

*I think that it is a relatively good approximation to truth - which is much too complicated to allow anything but approximations - that mathematical ideas originate in empirics. But, once they are conceived, the subject begins to live a peculiar life of its own and is... governed by almost entirely aesthetic motivations. In other words, at a great distance from its empirical source, or after much “abstract” inbreeding, a mathematical subject is in danger of degradation. Whenever this stage is reached the only remedy seems to me to be the rejuvenating return to its source: the reinjection of more or less directly empirical ideas...*

An important factor influencing the evolution of control has of course been the availability of technology for implementation. While the digital computer did spawn a revolution, computational power was initially limited, and there was no data networking to any appreciable extent. Thus, in many respects in this era, the theoretical research led technology, as noted above, and naturally
needed course corrections as technological possibilities and limitations became clearer. Nevertheless the imaginative research left the field in a good creative state to pursue the opportunities that have since opened up after revolutions in computational power, software, data networking, sensors and actuators, following the microelectronics revolution realized by four incessant decades of Moore’s Law.

It is impossible to list the applications of control in their entirety; a crude random sampling yields the following: process control, telephony, cellular phones, power systems, aircraft and spacecraft, the Internet, computer control of fuel injection, emission control, cruise control, braking and cabin comfort in automobiles, production and inventory control, missile guidance, robotics, appliances, semiconductor wafer fabs, active noise canceling, automated highways, atomic force microscopes, quantum control, mass spectroscopy, large space structures. At present, almost every technology has feedback control at its core. As an exemplar, a recent article [527] describing the efforts of the most recent awardee of the IEEE Medal of Honor, Irwin Jacobs, a co-founder of Qualcomm, has this to say:

... he envisioned a rapid-response system: CDMA phones would monitor the power of the signal coming in from the tower; if the signals suddenly dropped, say, when a user walked into a building, the phone would crank up its transmitting signal, figuring that if it was having trouble hearing the tower, then the tower would have trouble hearing the phone. Next, equipment at CDMA towers would take a handful of received bits and calculate an average signal strength; if that signal fell above or below a preset threshold, then the tower would prompt the phone to lower or raise its power. ...

“Someone else might have looked at all the complexities and the concerns and concluded that it just wasn’t possible.”

It is not for nothing that control, omnipresent everywhere, is called a hidden technology.

Besides engineering and technology, there are many uses of feedback and feedforward in other areas too. In economics, central planning could perhaps be regarded as an example of feedforward, while a market economy could be regarded as an example of feedback. Tustin wrote a book [663] on applications of control to the economy as early as 1953. At least in technical systems it is known that the best results are obtained by combining feedback and feedforward. Control is also entering unusual fields like internet advertising [371] and art [159].

And then there is biology, perhaps on the edge of a revolution, where the unraveling and harnessing of omnipresent feedback processes is the dream of mankind.

With all the aforementioned advances, the stage is set for large scale system building. The twenty-first century could well be such an age of large scale system building. Not only are we running into resource and environmental limitations, whether in energy or water, but at the same time we are also facing great demands for modern transportation, energy, water, health care services, etc., from large segments of the globe that did not previously have access to such services. Major efforts across the globe are targeted at grand challenges vis-a-vis the smart electricity grid, automated transportation, health care, etc., for all of which sensing and actuation, viz., control is key.

7 Concluding Remarks

Control is a field with several unique characteristics. Its evolution is a veritable microcosm of the history of the modern technological world. It provides a fascinating interplay of people, projects, technology, and research.

Control transcends the boundaries of traditional engineering fields such as aeronautical, chemical, civil, electrical, industrial, mechanical and nuclear engineering. Its development was triggered not by a sole technological area but by several technological projects, such as fire control, telephony, power systems, flight control, space exploration, and robotics, at different times in its evolution. Control has also had impact on several non-engineering fields such as biology, economics, medicine and physics. Concepts and ideas have migrated between the fields.

The development of control has benefited greatly from several grand challenges, e.g., transcontinental telephony, fire control, and landing a man on the moon. Its development also benefited from the concentrated power of monopolistic industries and the government. It further benefited from the great post-second world war boom of academic research.

Different application areas have emphasized different aspects of control, leading to the development of a rich framework for control. In turn, closing the loop, the evolution of control has influenced, radically in many cases, the development of each of these areas.

Control is the first systems discipline. It recognized the commonality of issues at the heart of many engineering problems. The systems viewpoint – make the output of “plant” or entity behave in a desirable manner by manipulating its input – is a unifying viewpoint that provides great clarity to the design process irrespective of the field of application.

The enabling technology for implementation had a major impact on the evolution of the techniques for control design and the underlying theory, as witnessed by the development of frequency domain theory in the age
of analog computation, and later by the development of the state-space approach and multi-stage decision making in the era of digital computing.

Control is a field whose progress has been punctuated by several key theoretical contributions. These have involved a variety of mathematical sub-disciplines, such as complex analysis, differential equations, probability theory, differential geometry, optimization and graph theory. As such, control is currently one of the most mathematized fields of engineering.

The research in the field has resulted in an exceedingly rich collection of advanced and specialized books covering several subfields. The range includes adaptive control, classical control, discrete-event systems, differential games, digital control, distributed parameter systems, dynamic programming, estimation, identification, linear systems, multi-variable control, networked systems, nonlinear systems, optimal control, robust control, sliding mode control, stability, stochastic control. There are even encyclopedias of control.

Control has become a central component of many modern technologies, even though often hidden from view. In fact it is hard to conceive of any technology dealing with dynamic phenomena that does not involve control.

There has been a dynamic gap between theory and practice. At times, applications consisted mainly of tinkering. At times it was the severe difficulties encountered in practice that led to dramatic theoretical breakthroughs which were extremely relevant and important in practice; an example being the work of Bode and Nyquist. At other times, incipient technological possibilities opened up new fields of theoretical research. This resulted in a broad exploration of systems theoretic concepts such as stability, controllability, information structures, optimality, complexity and robustness. At times, the exploration developed in an open-loop way without feedback from practical applications, and sometimes as a mathematical endeavor. In some cases, technology was not yet ripe to implement and test out some the concepts being explored.

Where are we now, and what may we learn from history? How may the past provide some guidance and feedback for the future? We present our viewpoint.

On the technological side, with dramatic evolution of sensors, actuators, networks, computational hardware and software, it has become feasible to deploy and implement large and complex systems. With this considerable strengthening of the implementation capabilities, the theory-practice gap needs to be narrowed. This may need to happen on both fronts – more theory to solve difficulties encountered in applications, as well as more experimentation to determine what are the difficulties and thereby identify problems that need a solution.

On the one hand, where the problems are well understood, there need to be strong theoretical efforts to develop solutions. An example is the need for formal methods. A good theory can obviate the need for massive simulation based testing that is very expensive both in cost and time. Design, implementation, maintenance and upgrading of complex systems cannot be done safely without formal methods that go all the way from requirements to the final product.

On the other hand, there needs to be greater experimentation to understand what are the bottlenecks, calibrate purported solutions, and to understand what works or improves performance and what does not. An example is the goal of building autonomous systems, where prior distributions of uncertainties or model classes are not well understood. Experimentation in such situations is important for learning about the real world, and is intended to be revelatory. It can lead to a relevant theory.

Experimentation is different from demonstrations. Experimentation involves two way dynamic interaction between theories or models and practice; i.e., a feedback loop. Demonstrations are on the other hand just that – they demonstrate that a particular solution performs as claimed. They are not a substitute for experiments or genuine laboratories.

It is important for research to investigate applications, being guided by them, by what works and what does not. Awareness of what are the real bottlenecks for performance, robustness, reliability and how to shorten the cycle of design and deployment is important.

Control systems researchers should take full systems responsibility. In fact, as history has shown, for example in the case of the feedback amplifier, the recognition of what is really the problem is itself a major accomplishment in research. Such awareness can then lead to relevant advances that have deep impact on practice.

Pedagogy also needs to play an important role. The field should educate students who are capable of solving the whole problem from conceptual design to implementation and commissioning. Due to the convergence of control, communication and computing, students will also need to be knowledgeable across a broad front of all these fields, as well as mathematics. We must also leave space for students to acquire knowledge of fields such as biology, where advances have been extraordinarily rapid. How all this can be accomplished within the time limited confines of an undergraduate curriculum requires a thorough examination. At the graduate level, one also has the additional challenge of preserving depth, since that is critical for research, and in fact has been an important strength of the field.

Book writing has an important role to play. The tremendous research advances of the past seven decades must
be distilled with the benefit of hindsight into compact books. We need to compress current knowledge, emphasizing the fundamentals. This needs to be done at both the undergraduate and graduate levels.

There are also challenges with respect to how control is dispersed in several engineering departments. Since education and research in engineering grew out of specific technologies such as mining, building of roads and dams, construction of machines, generation and transmission of electricity, industrial use of chemistry, etc., it led to an organization of engineering schools based on departments of mining, civil engineering, mechanical engineering, electrical engineering, chemical engineering, etc. This served very well at the end of the 19th century and the beginning of the 20th century. But is this the optimal structure in the twenty-first century to teach an increasingly powerful systems discipline such as control that cuts across these areas?

The field of control has a bright future since there are many grand challenges. There is great planet wide demand for advanced systems for transportation, health care, energy, water, etc., which have to be engineered in a resource limited environment. Biology is another major frontier of research. The twenty-first century could well be the age of large system building.

Acknowledgements

The authors are grateful to Leif Andersson, Karl-Erik Årzn, Tamer Basar, John Ballieu, Bo Bernhardsson, John Cassidy, Helen Gill, Aniruddha Datta, Johan Eker, Y. C. Ho, Charlotte Johnsson, David Mayne, Petar Kokotovic, Sanjoy Mitter, Richard Murray, Anders Rantzer, Pravin Varaiya, Eva Westin and the several anonymous reviewers for input and feedback on this paper.

References

[17] Anon. The Instrumentation Laboratory (Draper Lab).
References


