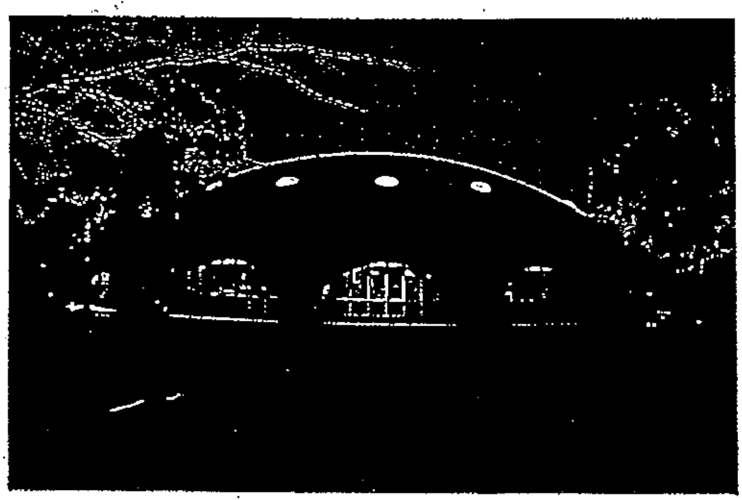
Public Lectures

1991-92



AUSTRALIAN ACADEMY OF SCIENCE

Control Engineering from the 17th to the 21st Century

by Professor Brian D O Anderson, FTS, FAA, FRS

Flinders Lecture

30 April 1992

Professor Brian Anderson, FTS, FAA, FRS, is Head of the Department of Systems Engineering, Research School of Physical Sciences and Engineering, at the Australian National University.

Professor Anderson's achievements lie in the area of control systems, signal processing, network theory and telecommunications. He is particularly recognised for his work in adaptive systems which includes the development and application of the idea of persistent excitation and the explanation of the associated adverse phenomenon of bursting (the apparent spontaneous commencement of oscillations and their later dying out in adaptive systems).

Professor Anderson's work in control theory also includes a major study of controller reduction which has been incorporated into commercial software. His applied work includes the development of towed array shape estimation which has undergone sea trials in the last two years. Professor Anderson is co-author of a number of texts, several of which are recognised as classics.

The Matthew Flinders Lecture acknowledges the work of Australia's pioneer scientists and was established in 1956. The lecture is given every two years at an Annual General Meeting of the Academy, alternating with the Burnet Lecture.

Abstract

Control engineering is a discipline that has in part been driven by practice, in part by theory. The earliest drivers were applications problems in the field of time measurement, mills and steam engine speed control. Major control tasks included control with zero steady state error, and achieving fast response to a step change, without instability or excessive overshoot.

Work in the late 19th century provided the first formal solution to the stability problem and an understanding of the value of integral control. A seventh order water turbine system had been successfully and scientifically controlled by 1900.

In the first half of the 20th century, electronic amplifier design and then the military demands of the Second World War gave much impetus to the development of control engineering. The methods developed for design were predominantly graphical and involved adjustment of only a few parameters. The role of high gain, proportional, integral and derivative control all became understood and control engineering ideas found applications throughout chemical and mineral industries.

Theoretical developments in the second half of this century have been substantial. Many took some years to be translated into practice, such as LQG design, adaptive control and sampled data control. Aerospace applications requirements drove some of these developments, many of which are now finding their place in materials processing and handling systems as diverse as sugar cane mills and chemical process control.

Future developments will arise from applications pressure and theoretical work. Applications pressure is strong in the areas of robotics, automobiles, discrete-event systems and environmental control. Replacement of existing nonadaptive systems by adaptive systems will be widespread. Theoretical developments will occur in many areas including nonlinear systems, robust control design and, perhaps, use of time-varying controllers for time-invariant plants.

Introduction

The earliest uses of control engineering probably go back thousands of years, but for the purposes of this paper, I have taken a starting point in the 17th century. As explained below, there occurred a significant application of control engineering at this time (although it was not regarded as such). At the other end of the scale, I have chosen to set a limit of the year 2000. Predictions of the future are notoriously unsafe, and the further one seeks to see, the less confident one can be of the outcome.

For the sake of presentation, but also to reflect several historical trends in control engineering, I have chosen to divide the period of the 17th to the 21st centuries into four epochs:

Pre-scientific control (17th-19th century)

The classical period (1900-1955)

The period of major developments (1955-1990)

The future (1991-2000)

Pre-scientific control (17th-19th century)

The period of pre-scientific control was primarily driven by applications which, of course, were associated with economic activities. Yet the resolution of the associated imperfectly understood control tasks brought with it a realisation of the existence of at least two generic control problems (control science problems). Towards the end of the 19th century, attempts were made to formally resolve these scientific problems.

I begin, however, with a discussion of applications and the driving effect they had on the development of control. I have chosen four examples:

- time-measurement,
- windmills,
- steam engines,
- telescopes.

The economic significance of the first three is unquestioned. The economic significance of the last is very slight indeed, but as described in more detail below, the work on telescopes gave rise to a major scientific advance.

The four applications areas, each of which is described further below, together resulted in the identification of at least two control problems:

- securing dynamic stability of a feedback system, and
- securing zero steady state error given constant disturbances.

To understand what these phrases mean, it is necessary to understand the prototypical control problem.

The prototypical control problem

Imagine a physical entity to which one can apply some form of excitation (or control) and which generates some sort of response.

For example:

Physical entity	Excitation	Response
Electric heater	Electric power	Room temperature
Car engine	Fuel flow	Engine speed
Aircraft	Hydraulic power to control surface	Attitude of aircraft

In each of these three examples, there is very frequently a desired response, which may or may not be constant. For example, one may want the aircraft to change its attitude from horizontal flight to a descending glide path. The achieving of a given response is made possible by the introduction of an appropriate excitation. Consider the following table, which highlights how the appropriate excitation is achieved.

Physical entity	Excitation	Excitation achieved by
Electric heater	Electric power	Turning switch on or off depending whether you are cold or hot
Car engine	Fuel flow	Linking throttle to a foot control (accelerator) and then operating accelerator
Aircraft	Hydraulic power to control surface	Linking control surfaces via actuators to control stick

These are all examples of open-loop control. That is, the system (as opposed to the human operating it) does not sense what the response is and therefore cannot take corrective action by way of adjusting the excitation. Closed-loop control is also possible:

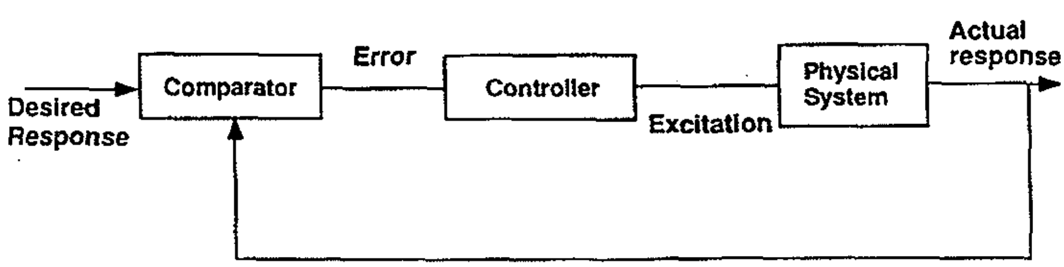
Physical entity	Excitation	Closed-loop control approach
Electric heater	Electric power	Thermostat compares desired and actual temperature and switches power accordingly
Car engine	Fuel flow	Cruise control compares desired and actual speed, and adjusts fuel flow and even braking
Aircraft	Hydraulic power to control surface	Automatic pilot compares desired flight trajectory (level, landing etc) with actual, and adjusts control surfaces.

Closed-loop control is summed up in Figure 1.

Now we can interpret the two key control problems. Securing dynamic stability means that the corrective action taken by the controller should not overcompensate and drive the system into oscillation or some catastrophe. Securing steady state error given constant disturbance means that if the desired response is constant, the actual response should (after some transient) exactly match the desired response even with constant disturbance such as constant heat loss through windows (room heating problem) or constant head-wind (cruise control or aircraft control).

It is very common to consider the response of a closed-loop system to a step change (thermostat dial is adjusted, for example). This is the so-called step response. It can be sluggish or fast, exhibit great overshoot before becoming correct or almost no overshoot, exhibit oscillatory behaviour before settling down, or just a fluctuation or two. It can achieve zero steady state error, or non-zero steady state error.

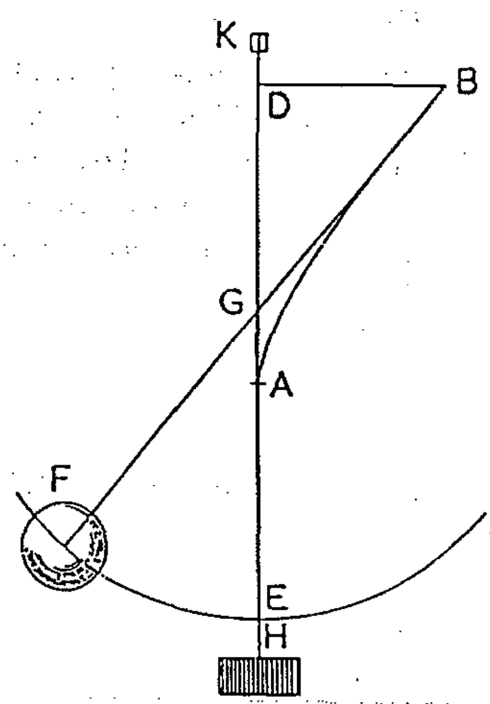
Figure 1. Closed-loop control



Time measurement

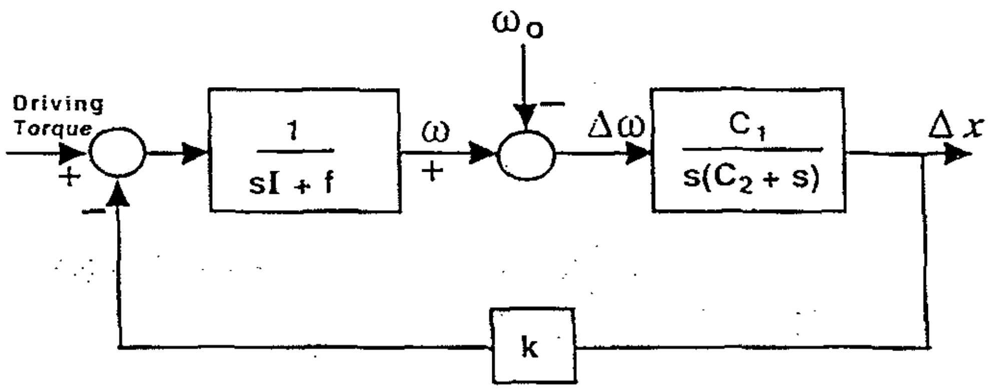
Both for its intrinsic worth, and for its help in maintaining accurate navigation for ships at sea, the accurate measurement of time was highly valued. Most clocks of the period relied on the motion of a pendulum. The great 17th century physicist Huygens, perhaps better known for his contributions to the theory of light, turned his mind to the question of improving the accuracy of time measurement by the devices of his day. He conceived the idea of arranging for the pendulum, whose period governed the basic advance of the clock hands, to become longer when it speeded up. The details of how this is done are rather intricate, and therefore will not be described fully here. (For further details, see Fuller, 1976a.) The crucial concept he employed was to use a conical pendulum (Figure 2). *Kh* is a vertical axis, *DBGF* forms a plane, and AB is a carefully designed curved surface. When the pendulum moves, *GF* traces out a conical surface. The effective length of the pendulum is governed by the point where *BF* is tangential to the surface *AB*. The faster the pendulum swings, and thus the higher point *F* is above point *E*, the longer the effective length of the pendulum, and thus the lower the natural frequency. [Digression:

Figure 2. Part of the conical pendulum geometry



For those who have some control systems knowledge, Figure 3 can be regarded as a representation of what is happening. The frequency ω_0 is a consequence of the mechanical design, and particularly the design of the curved surface *AB*, and the frequency ω is the instantaneous angular frequency of the pendulum. The presence of the integrator in the loop ensures that in steady state, the error between ω and ω_0 must go to zero. Evidently, Huygens was addressing the problem of securing steady state error in the presence of disturbances.

Figure 3. Modern viewpoint of Huygens conical pendulum



Windmills

A second major application was drawn from the area of windmills. It was necessary for windmills, their prime purpose being to crush grain, to be controllable in several respects. First, it was important to be able to turn the massive windmill structure to make its sails face the wind. The solution was to use auxiliary sails (a fantail) at right angles to the main sails, with the power developed in the fantail being used to drive the whole main structure around to its correct orientation (see Figures 4 and 5). In Figure 5, the block labelled *k/s* is a conceptual representation of determining the error between the wind angle and the fantail angle, viz θ and then integrating it, i.e. adjusting the fantail angle at a rate proportional to the error. The equation description is $\dot{\theta} + k\theta = 0$, and with $k > 0$, θ goes to zero). This is an extremely simple set-up in terms of the control science involved. Other uses of control in windmills included:

- using a centrifugal governor employing fly-balls to control the speed of the windmill sails (with speed-up of the governor, the fly-balls would rise, inducing partial furling of the sails and then a decrease in speed);

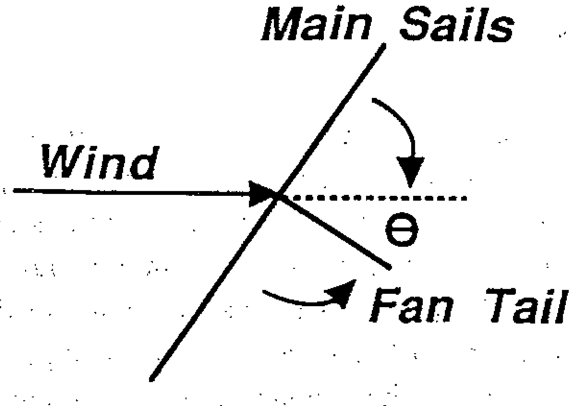
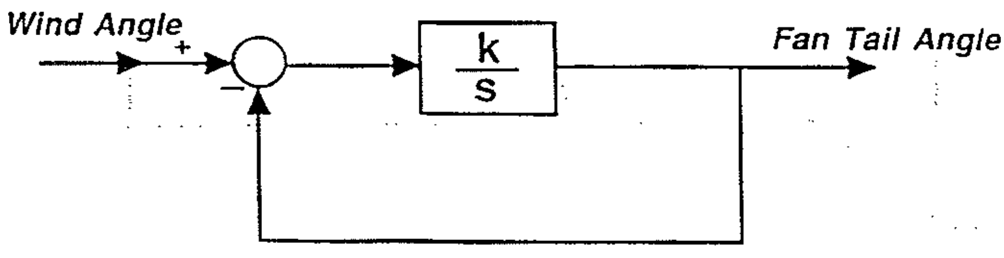


Figure 5. Modern viewpoint of fantail action



- using a governor to adjust a gap between the crushing stones; and
- using a governor to adjust the rate of grain supply to the stones.

Some elements of a multivariable design problem can even be discerned here.

Speed control in steam engines

The next major advance involved the adaptation of fly-ball governors which had been used on windmills to steam engines. This was a practical technology, full of art, which reached its zenith in the United Kingdom, the home of the steam engine. By the mid-19th century, there were between 50,000 and 100,000 Watt governors in use in Great Britain. All these governors had an adjustment capability, and the extensive experience of their operation soon highlighted a fundamental trade-off. Offset error (that is, error between desired speed and actual speed) could only be reduced at the expense of increased overshoot in responding to a step change in level. These days of course, we see much of control engineering, particularly classical control, as embodying the task of picking the right trade-off point between a number of conflicting phenomena. It was the use of Watt governors that first highlighted this trade-off aspect of control engineering.

[Digression: A rough explanation in control engineering terms is provided by noting that the centrifugal fly-ball governor has a transfer function $k/(c_0 + s^2)$. The design could be modified to secure integral action, thus offering a zero offset error in the steady state response to a step change, in which case the effective transfer function becomes k/s^2. But now the double, rather than the single, pole at the origin brings with it the likelihood of a stability problem.]

Telescopes

The fourth major applications area (but, as noted above, not one driven by economic considerations) was associated with the telescope. Use and development of telescopes in Britain in the 19th century were under the charge of the Astronomer Royal, at that time a man called G. B. Airy. (He was the father-in-law of Routh, famous for a later major contribution to control.) Airy was a major scientific figure of his age, with some 500 papers and 11 books to his credit. His contributions are still recognised today in areas such as mechanical engineering. His particular problem with telescopes was to find a way of rotating them at a uniform rate, so that once a telescope was aligned with a heavenly body, it would automatically track the apparent motion of that heavenly body across the sky. The technology he proposed to use was the fly-ball governor, and he quickly became aware of the trade-off which had to be faced between low offset error and the tendency to instability. He then set out to obtain a scientific understanding of the instability (Airy, 1840). To do this, he brought to bear his considerable knowledge of celestial mechanics to model mathematically the phenomenon he was observing, and this led him to the following equation:

$$[\frac{d\theta}{dt}]^2 + \frac{a}{\sin^2\theta} - \frac{2q}{b}\cos\theta = c \qquad (1)$$

Even today, with a sophisticated knowledge of control, we might find this equation somewhat overpowering, and certainly so in terms of its non-linearity. Be that as it may, Airy was able to describe the instability phenomenon with this equation and explain how the dynamics could systematically be adjusted (i.e. the knobs set) so as to ensure stability.

Not only was this a considerable tour de force in the scientific sense, but it was the first illustration that a control problem was susceptible to analysis via a differential equation. Indeed, more than this was true: control design could be regarded as adjusting coefficients in a differential equation to secure properties for its solutions. {Digression: Figure 6 shows various Nyquist loci for Airy's system, with various values of damping coefficient. Of course, Airy did not use Nyquist loci to analyse his system, and the figure is associated with a linearisation of the basic equation.}

Another great scientist/engineer of the day was J. C. Maxwell who attempted a systematic analysis of governor stability, having previously analysed the stability of the rings of the planet Saturn which is defined by a fourth order system (i.e. the underlying differential equation involves derivatives up to order 4). His analysis of governor stability led him to consider a number of third order equations (Maxwell, 1868). (A block diagram illustrating one of these systems is shown in Figure 7.) Maxwell also set himself the task of establishing criteria for stability of higher order systems, but the problem defeated him. Given his great powers, one must wonder whether or not he devoted his full energies to the problem.

Figure 6. Nyquist loci for Airy's telescope control with various damping coefficient values

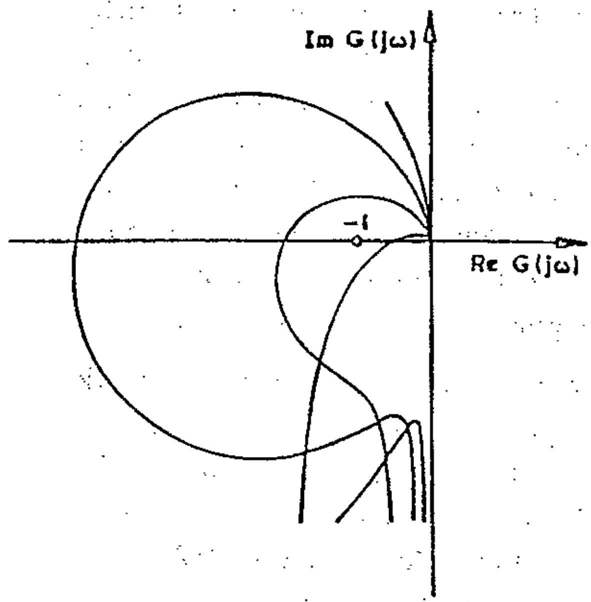
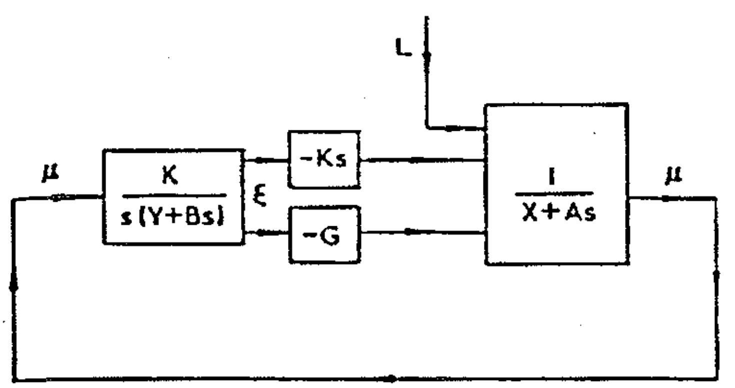


Figure 7. Block diagram representation of one governor system analysed by Maxwell



The stability problem in mathematical terms

The problem which defeated Maxwell can be posed in mathematical terms as follows. Given a polynomial

$$p(s) = s^n + a_1 s^{n-1} + \ldots + a_n \tag{2}$$

when does this polynomial have all its roots with negative real parts? This problem goes to the heart of one of the scientific problems of control. It is interesting to reflect on three different solutions which were presented to this problem. The solutions appeared as a result of three streams of work, all of which seem to have been independent of one another.

The first stream of work was that of French mathematicians, Cauchy (1831), Sturm (1836) and Hermite (1856). Hermite's paper actually gave a nice solution of the stability problem, nice in the sense that there was a closed form procedure for manipulating the coefficients a_i *of* $p(s)$ to give a yes/no answer to the question on the roots. Certainly, the roots did not have to be found, and of course for high order polynomials, the methods for finding the roots were at best primitive. Hermite's work was published in French, and was uninterpreted by engineers.

Maxwell, as part of the second stream, did not know of Hermite's work and conducted some of his work on stability after Hermite had published his work. It was not till 1877 that E. J. Routh in England, drawing on the work of Cauchy, Sturm, Maxwell and his father-in-law Airy, published a solution to the problem, embracing the Routh table (Routh, 1877). The Routh table requires one to manipulate the coefficients a_i in a systematic way, and to check for the positivity of the leading entries of the table. Positivity of the entries is equivalent to stability. Figure 8 illustrates the construction of the Routh table.

Figure 8. Routh Table

1	a_2	a_4	a_6	.	.
a_1	a_3	a_5	a_7	.	.
$b_1 = \frac{a_1 a_2 - a_3}{a_1}$	$b_2 = \frac{a_1 a_4 - a_5}{a_1}$	$b_3 = \frac{a_1 a_6 - a_7}{a_1}$	b_4	.	.
$c_1 = \frac{b_1 a_3 - a_1 b_2}{b_1}$	$c_2 = \frac{b_1 a_5 - a_1 b_3}{b_1}$	.	.		
.					
.					

Stream three in the stability problem was associated largely with Swiss scientists and engineers. The two most important names are Stodola and Hurwitz who worked at the very end of the 19th century. Stodola was arguably the first control engineering academic, although he drew on the work of a Russian, Vishnegradsky, who chose to present his results in typical engineering fashion using charts. Stodola had an interest in the control of water turbines, the equation descriptions of which ranged in order from three to seven. He recognised the nature of the stability problem and turned to his mathematician friend Hurwitz for advice, as a result of which Hurwitz developed the so-called Hurwitz criterion, which requires the checking for positivity of a number of determinants, easily constructed from the a_i (Hurwitz, 1895). In fact, the polynomial $p(s)$ has all its roots in the left half plane if, and only if, the following determinantal conditions are satisfied:

$$a_1 > 0; \quad \begin{vmatrix} a_1 & a_3 \\ 1 & a_2 \end{vmatrix} > 0; \quad \begin{vmatrix} a_1 & a_3 & a_5 \\ 1 & a_2 & a_4 \\ 0 & a_1 & a_3 \end{vmatrix} > 0;$$

$$\begin{vmatrix} a_1 & a_3 & a_5 & a_7 \\ 1 & a_2 & a_4 & a_6 \\ 0 & a_1 & a_3 & a_5 \\ 0 & 1 & a_2 & a_4 \end{vmatrix} > 0 \ldots \qquad (3)$$

In summary then, by the end of the 19th century, two control science problems had arisen out of the control applications problems; and, in the case of the stability problem when posed with differential equations, a solution had been identified.

The classical period (1900-1955)

In this section of the paper, I shall indicate very briefly a number of the advances which occurred in the classical period, and then focus on several. I shall then point out some of the shortcomings.

There were two major driving forces for control during this period. One driving force was indirect, namely electronic amplifier design. A great many control advances came about because people were trying to understand not how to design a control system, but how to design an electronic amplifier when the active element in the amplifier (at that time a vacuum tube) could have characteristics which varied very substantially over the useful life of the device. It was probably not until the Second World War that control applications needs became the real driving force for the development of control.

Over the period 1900-1955, there were a number of theoretical advances and a recognition of the applicability of analysis tools. These included:

- formal recognition of the feedback concept;
- system description via the mathematical tools of transfer functions and Fourier transforms;
- the Nyquist graphical criterion for stability;
- the use of graphical tools (Bode diagrams and Nichols charts) as a way of representing system behaviour;
- the Routh test (actually available since 1877);
- the use of another graphical tool, the root locus, for studying the effect of a design parameter variation;
- an understanding of the benefits and costs of high loop gain; and
- the general recognition that much design was a matter of trade-offs.

I shall say more about the Nyquist criterion and high loop gain later.

Particular design ideas evolving in the period included:

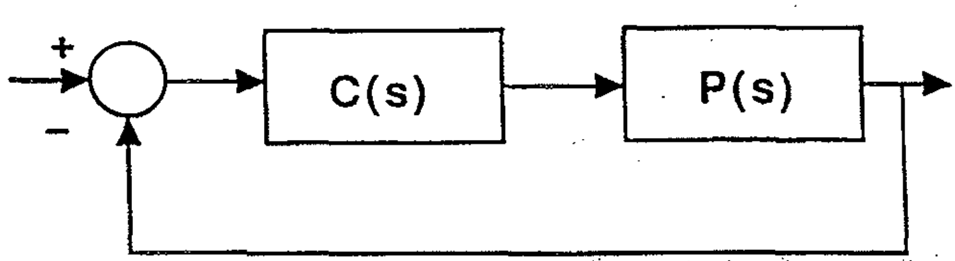
- position feedback;
- rate or velocity feedback;
- integral feedback, with its implications for zero offset error;
- PID (propositional, integral, derivative) controllers, in effect a
- combination of position, rate and integral feedback;
- lead and lag compensation (and varieties thereof); and
- general graphical procedures.

The thinking at the end of this period, including discussion of the Nyquist criterion and high loop gain, is well reflected in Thaler and Brown (1960).

Nyquist's contribution

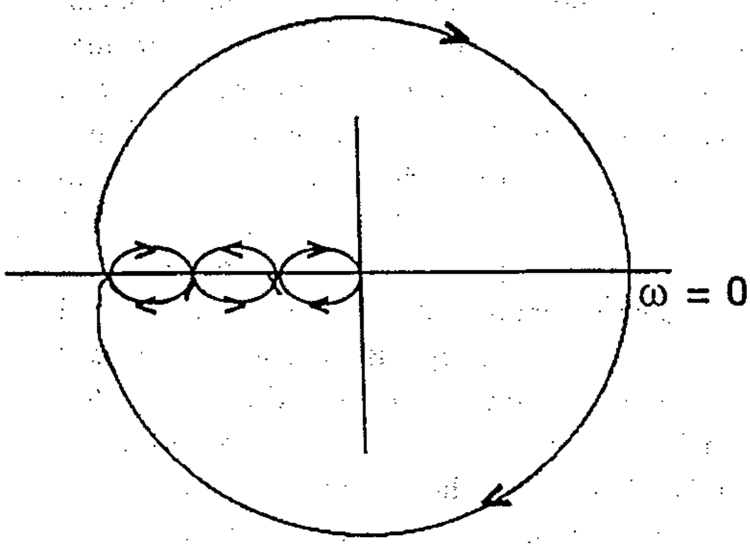
A truly outstanding development in the period was the Nyquist criterion. The Nyquist criterion represented a massive piece of lateral thinking, in that its starting point for the determination of stability was nothing like that used in the only other approach to stability available to that time, namely the approach based on differential equations. Instead, the Nyquist criterion took as its starting point the availability of a system description obtained using physical measurements; that is, a frequency domain description of the system, and in fact one represented in particular graphical form. No differential equation was needed, and there was in fact no restriction to systems which could be described by an ordinary differential equation. Not only was the system description totally different to that which had been used before, but the way of describing the Nyquist test (involving as it does a topological property of a graph, together with a way of proving it based on complex variable theory) represented a total departure from the past. To recall what the Nyquist criterion is, consider the set-up of Figure 9, in which P is notionally a plant and C a controller. Imagine that sinusoidal signals of frequency ω are applied to the input of C and the resulting response is measured at the output of P. The ratio of response to input, depending on ω, is called the value of the loop transfer function at ω, and is written as $P(\omega j)\ C(j\omega)$, where $j = \sqrt{-1}$. With an appropriate definition of 'ratio', it is a complex number and it can be plotted in the complex plane as a function of ω. The number of encirclements of the point -1 made by the graph in the counter-clockwise direction is counted,

Figure 9. Feedback System



with the arrows on the graph pointing in the direction of increasing ω (see Figure 10). If, and only if, the number of unstable poles (or independent instability modes existing prior to loop closure) of *PC* is equal to the number of encirclements of the -1 point, the closed loop is stable.

Figure 10. Nyquist plot



High loop gain

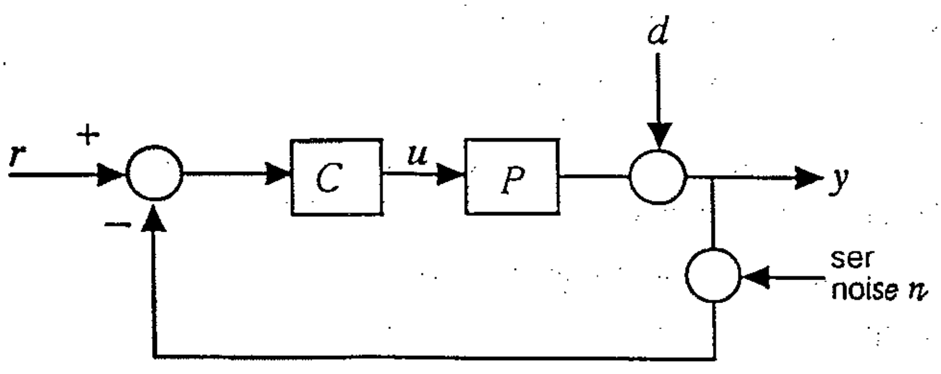
Another one of the major areas in which conceptual understanding was reached during the classical control era concerned the trade-offs available through high loop gain. Consider Figure 11. The loop gain is *PC* which is of course a frequency dependent gain. The output y is related to the input r, the disturbance d and the sensor noise n in accordance with

$$y = \frac{PC}{1+PC}r + \frac{1}{1+PC}d - \frac{PC}{1+PC}n \qquad (4)$$

and the plant input u is related to r, d and n by

$$u = \frac{1}{P}\,\frac{PC}{1+PC}\,(r-d-n) \qquad (5)$$

Figure 11. Feedback system including disturbance and noise



The positive effects of high loop gain were identified as

- the ability to suppress some effects of plant gain variation (in today's jargon, this would be termed securing robust control, but the original concept was to secure insensitivity of overall performance to the gain variations of a vacuum tube in an electronic amplifier);
- the ability to reduce the effect seen at the output of additive disturbances, *d*. (reference to equation (4) above shows that the bigger PC is the smaller will be the contribution of d to the output y with the point of the feedback being to measure the changes in the output introduced by d and feed them back so as to exert a countervailing effect); and
- the ability to promote better tracking by the output of a reference input. Again, reference to equation (4) shows that as *PC* becomes very large, *y* must become closer and closer to *r*, neglecting for the moment the effects of *d* and *n*.

High loop gain is not without its disadvantages. In particular, it can induce high gain instability or very oscillatory behaviour, worsen sensor noise problems, and cause the plant input to saturate (i.e. overdrive the plant).

In relation to this last point, consider the equation for *u* above, and suppose *PC* is made very large at a frequency at which *P* has become small. Then, neglecting the effect of *d* and *n*, *u* will be approximately P^{-1} r, and thus a very large quantity; of course, in the event of plant input saturation, the linear analysis giving rise to the above equations is no longer valid.

The overall design flavour of the classical period can be summed up as being graphical, having only very limited numbers of design parameters, and being based on rule of thumb (e.g. such and such a gain margin is desirable, such and such a phase margin is desirable, etc.).

The shortcomings were reasonably clear. It was clearly a disadvantage that at any one time one could only study the variation of a limited number of design parameters, and the extensive use of graphs for representing systems carried an inherent limitation. More broad criticisms included:

- the impossibility of systematic multivariable design,
- the impossibility of design for time-varying systems;
- the normal inability to perform optimisation, and
- the inability to handle (other than on the most rudimentary basis) stochastic or noise problems.

There were some other developments in the classical period. Pre-figuring the computer age, a start had been made on the development of sampled data theory, but there were no text books by 1955. Attempts were made at handling non-linear systems by describing functions (a Procrustean approach and, as such, one for which it proved very hard ever to get adequate theoretical justification), and phase plane analysis (with its inherent limitation on the dimensionality of problems which could be considered). Relay control was also attempted, and actually used in the German V-weapons of the Second World War. Wiener filtering represented a major advance, the full exploitation of which had not occurred really by 1955.

The period of major developments (1955-1990)

During the period 1955 to 1990, a number of subfields of control were substantially developed and new applications found. New viewpoints for description, analysis and synthesis of control systems were found, and it became necessary for control engineers to use new background tools. What were the driving forces during this period? For most of the time, the strongest driving forces were probably those associated with defence and the Cold War. Vast amounts of research work were supported by military or quasi-military agencies, and academics themselves played a significant role in setting the research agenda. Applications of control in civilian industries may have occurred more as a result of fall-out from the defence-driven work rather than because those industries themselves drove forward the development of control engineering.

Four subfields stand out from among the many which achieved major development during the three and a half decades. These are:

- sampled-data control,
- (LQG) Linear-Quadratic-Gaussian design including multivariable system design,
- adaptive control (including identification), and
- non-linear and time-varying systems.

Also during this period, the use of state-variable descriptions came to play a prominent role. After early ideological discussions that sought to argue that it was better to describe a system in time-domain terms than frequency-domain terms or vice-versa, or better to describe it in state-variable terms than transfer function terms, or vice-versa, it was recognised that it was best to work with a multiplicity of descriptions. Viewpoints were also conditioned by the availability of computers and subsequently by the availability of very sophisticated design packages. These design packages today include extensive simulation capability, and capability for switching of system descriptions from one type to another.

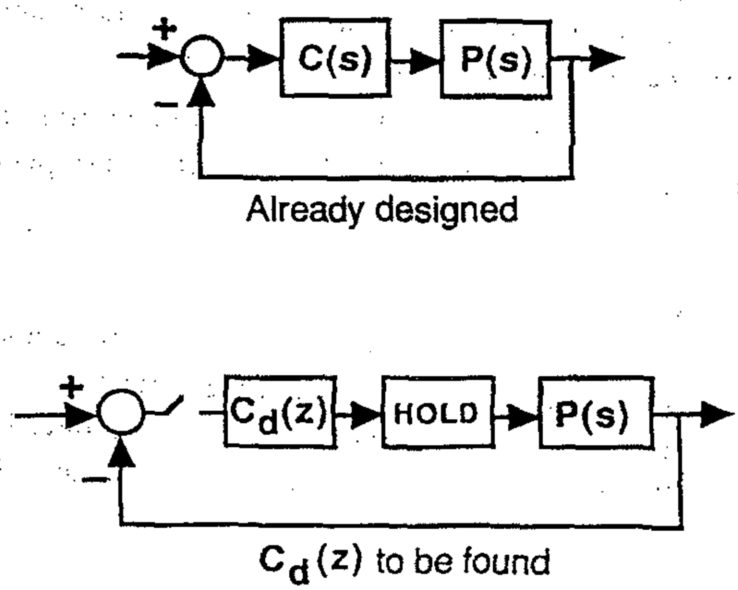
To understand the textbooks, and in particular the sophisticated design packages, control engineers needed to learn some matrix algebra, needed to understand some properties of differential equations, and preferably needed to acquire some understanding of random processes.

I will now highlight some particular problems which arose in three of the subfields mentioned above and with which I have had a close personal association.

Sampled-data control

In sampled-data control, one seeks to avoid using a continuous-time controller, and instead one plans to use a discrete-time controller (together with a sampling element and hold element; an anti-aliasing filter is also normally used, but this is inessential for the present discussion). Quite frequently, it can be the case that a continuous-time controller is designed and it is then to be replaced for the purposes of implementation by a discrete-time controller (see Figure 12). In a discrete-time controller, the input signal is sampled in time every T seconds; calculations are performed on it (using a computer), and a value generated. This value is used as the plant input over a T second duration. Thus the plant is driven by a piecewise constant signal, changing values every T seconds. Moreover, the plant input values are determined by sampling, rather than continuous monitoring, of the error between the derived response and actual response. The question then arises as to how the discrete-time controller $C_d(z)$ should be found from the continuous-time controller $C(s)$. Many methods can be found in textbooks for answering this question (see, for example, Aström and Wittenmark, 1990; Franklin, Powell and Workman, 1990). But it is the wrong question. What is the right question? The right question is: How should $C_d(z)$ be found from $C(s)$ <u>and</u> $P(s)$? Why is this the right question? The answer is that we are seeking, in replacing a continuous-time controller by a discrete-time controller, to preserve as far as possible the closed-loop properties. These closed-loop properties depend on the plant $P(s)$ as well as the controller. It follows that the plant has to affect the definition of what the best discrete-time controller is. This simple difference in viewpoint in a way goes to the heart of what distinguishes control from signal processing; all the time, it is closed-loop behaviour that is relevant, and not the behaviour of an entity by itself.

Figure 12. Controller discretisation



One of the first applications of 'the right question' was in the design of controllers for the Australia Telescope. Conventional (textbook) methods for the generation of a discrete-time controller from a continuous-time controller were found to fail with the Australia Telescope. The design engineers determined what the right question was, and then developed a way of solving it (Evans, Cooper and Kennedy, 1996; Kennedy and Evans, 1990). More recently, general theoretical tools have been developed for answering the right question (Keller and Anderson, 1990a, 1990b and forthcoming).

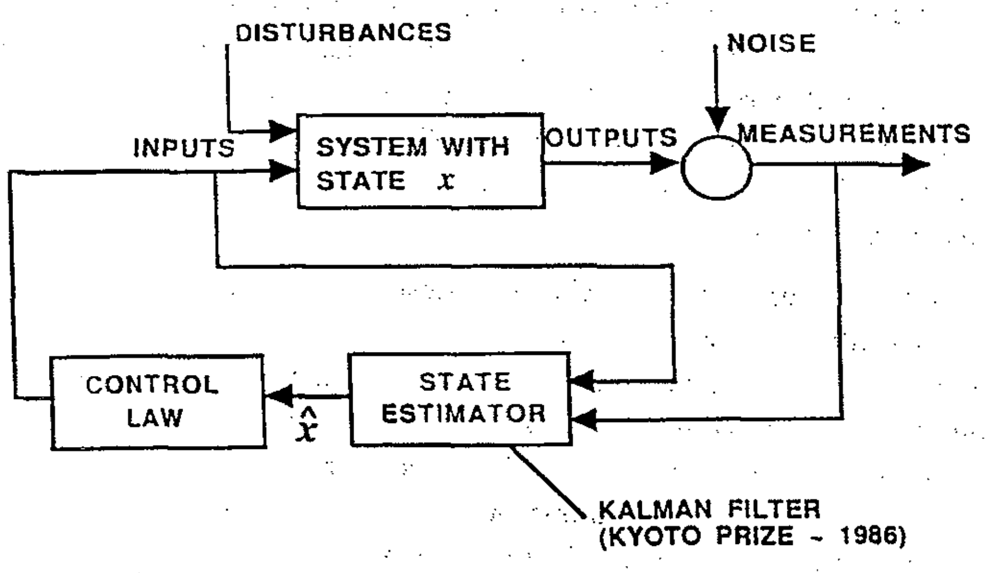
LQG design

Our second example is drawn from the field of Linear-Quadratic-Gaussian design (Anderson and Moore, 1989). This design procedure allows treatment of high dimension multivariable plants with noise. For example, a pitch control system for a commercial aeroplane has two inputs (the flaps and the aileron settings) and two outputs (the attitude and the angular velocity). The differential equations contain some 40 to 50 states, and there is a stochastic disturbance in the form of wind, for which a good model is available, as well as noise on the sensors

The key theoretical idea of Linear-Quadratic-Gaussian design is embodied in Figure 13. The controller consists of a state estimator (a device for estimating the internal state of the system, in this case the aeroplane) together with a control law, which constructs values for the input based on the state estimate. The state estimator, or Kalman Filter, won for its originator the Kyoto Prize in 1986. Linear-Quadratic-Gaussian design is a marvellous tool which has required some time for people to understand it. One of the difficult issues is how one should tune the software knobs, that is the design parameters. A second, and hitherto not fully resolved issue, is how one should design to obtain a controller which will cope with plant parameter variations. The third issue is that the design procedure in its raw form leads to a controller with the same complexity as the system. Thus, in the aeroplane example above, the controller would contain 40 to 50 states. In many situations, this is simply unacceptable, and the question arises as to how a simple controller could be obtained.

One aircraft company with which I have been associated, several years ago indicated that an LQG design could be obtained efficiently. It proved to be satisfactory in all respects except for the order of the controller. Because the order of the controller was too high, an alternative design method had to be obtained yielding a simple controller. This design took 200 work-years and was obtained largely by trial and error. The importance of obtaining an algorithm which would allow systematic simplification of a complicated controller is evident. A survey of work on this problem is to be found in Anderson and Liu, (1989), and within the last year, my colleagues and I at the Australian National University have written some commercial software which will be included in the premium CAD control systems package MATRIXx marketed by Integrated Systems Incorporated, to achieve model and controller reduction.

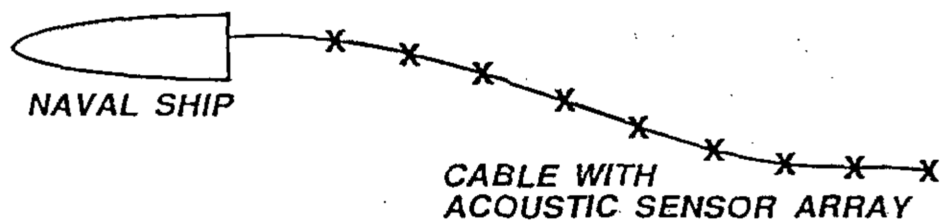
Figure 13. Key theoretical idea of Linear-Quadratic-Gaussian design



Kalman filter

The Kalman filter is not just a constituent of an LQG design, but an important and versatile tool in its own right. Other applications of its use are found in Gray, Anderson and Bitmead (1988) and in Riley, Gray and Hollingsworth (1990). Submarines can be searched for using a towed array of acoustic sensors (see Figure 14). The array is mounted on a cable, the motion of which is described by a fourth order nonlinear partial differential equation, with some random excitation due to currents. Accurate knowledge of array shape is necessary to obtain the advantage of having an array of acoustic sensors. By mounting compasses on the cable and using Kalman filter theory, an array shape estimator can be obtained.

Figure 14. Towed array



Adaptive control

Another example is drawn from the field of adaptive control, and reminds us of the old maxim that there is nothing so practical as a good theory. One of the original questions of adaptive control, now some 30 years old, is depicted in Figure 15. The plant *P(s)* is known, but the gain k_p is not. We are faced with the question of designing a controller that learns k_p, either explicitly, or implicitly. An early approach to this problem was provided by the MIT rule (defined below). In Figure 16, k_m is a known gain, and it is clear that the error will be zero for all inputs *r* if and only if $k_c k_p = k_m$. The gain k_c is adjustable and known, so that if an adjustment process can be found which results in *e* being identically equal to zero for all inputs, k_p will have been effectively identified. The MIT rule is a suggestion for a procedure for adjusting k_c, and is:

$$\dot{k}_c = -g[y_p - y_m]y_m \tag{6}$$

Figure 15. Plant with unknown gain

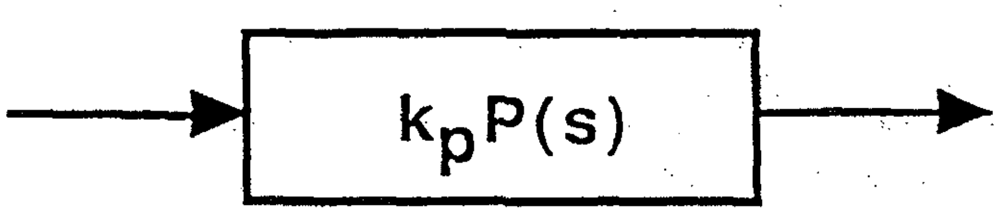
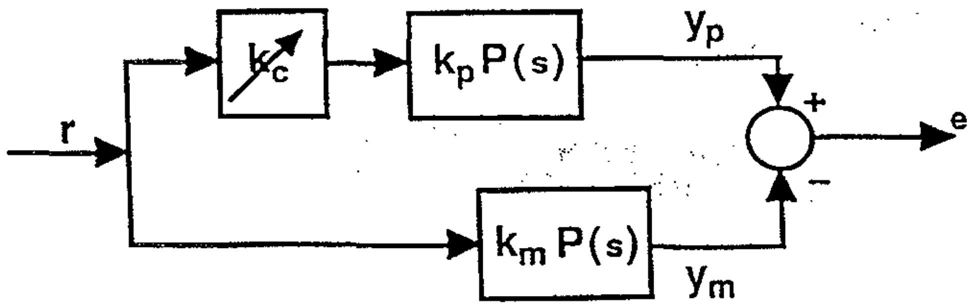


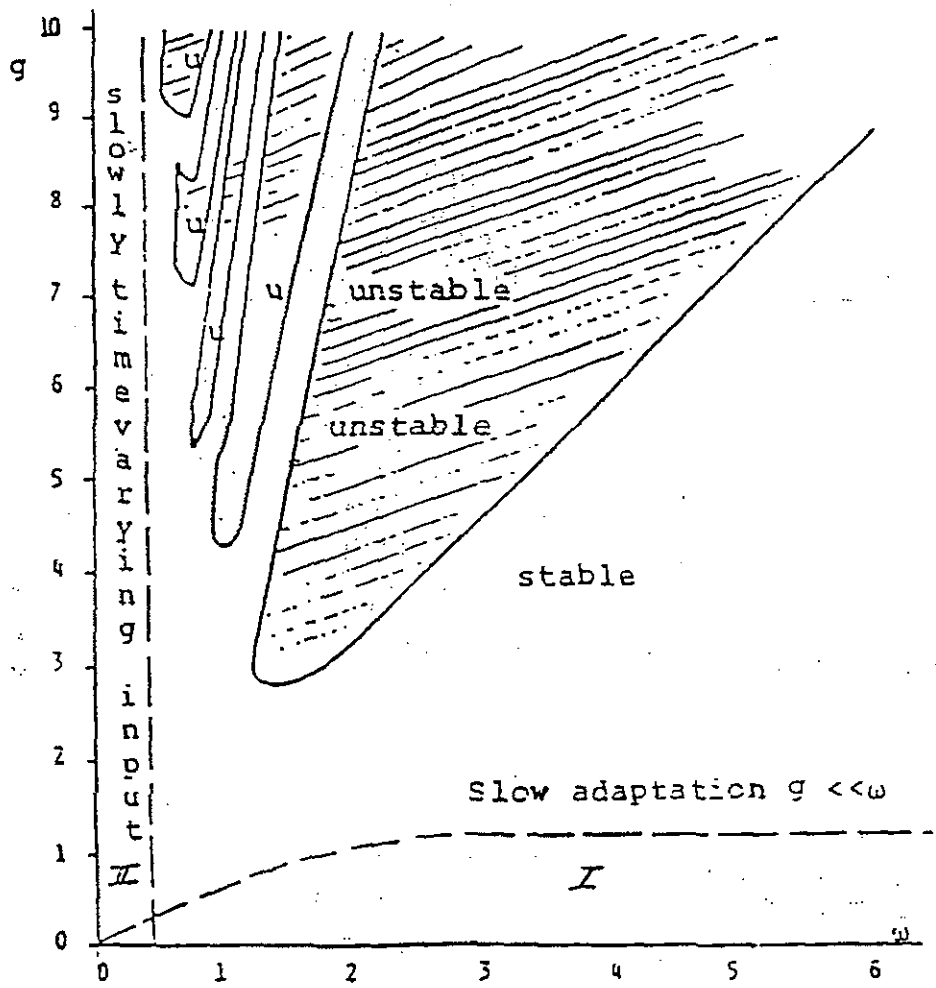
Figure 16. System for learning k_p



In this equation, g is a positive gain (termed the adaptive gain), and $y_p - y_m$ will be recognised as the error e. Thus, if the error is identically zero, the gain k_c will remain constant. There is a heuristic justification of the MIT rule, but more important is the question of how it performs. Its performance can be reviewed for sinusoidal inputs r and for different levels of adaptive gain g and frequency (Figure 17). The result is very surprising in that there are combinations of gain and input frequency for which one has stability and other combinations for which one has instability, without any clear pattern or apparent logic as to whether a given gain-frequency pair will be stable or unstable. If time delay is introduced into the plant, the situation is different again, with major changes to the regions. Workers were unable to explain why this happened, and because they were unable to explain why this happened, they were unable to predict the performance of the MIT rule and its derivatives in similar but different and sometimes more sophisticated situations. Because then there was no theory, there was effectively no use of adaptive control for some 15 to 20 years. The subject lay dormant for many years until new approaches to adaptive control were found; the theory for the MIT rule first became available in 1986 which thereby enabled consideration of many other adaptive schemes (Mareels *et al.*, 1986; Anderson *et al.*, 1986).

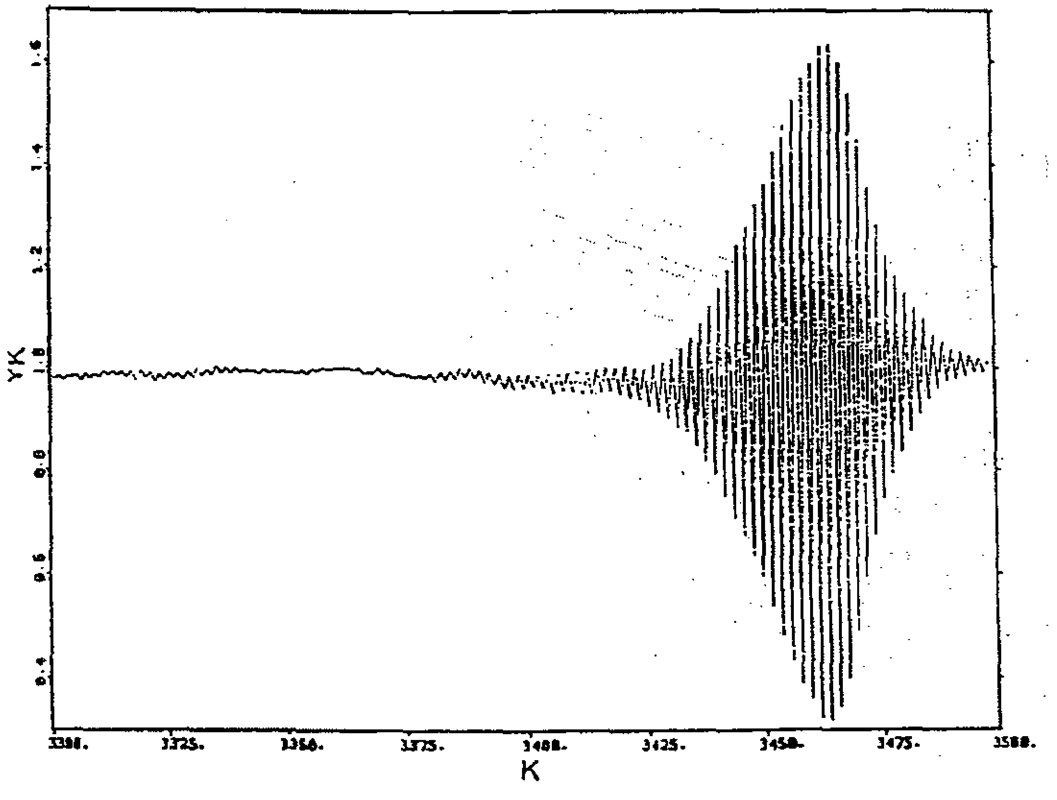
The new approaches first referred to were not without their surprises. In 1983, there were several reports of adaptive control implementations in which, after a very long period of satisfactory behaviour (a week, for

Figure 17. Behaviour of MIT rule for differing gain and frequency



example) oscillatory behaviour apparently spontaneously occurred, but then died down again (see Figure 18). This phenomenon became known as 'bursting'. Fortunately, a theoretical or scientific explanation for bursting was found without great difficulty (Anderson, 1985). The underlying cause is as follows. An adaptive controller usually attempts, implicitly or explicitly, to identify the plant to which it is connected. When a plant is subjected to a constant input (as was typical when bursting was encountered), it is impossible to identify more than one piece of information about the plant (its DC gain, or the amplification factor applying to constant signals). Despite this, the identifying part of the adaptive controller

Figure 18. Illustration of bursting in adaptive system with insufficient excitation



attempts to identify the whole plant, and that part of it identifying other than the DC gain is driven in effect just by noise. Accordingly, the identifier is likely to be wrong about everything except the DC gain; errors in estimating the plant then lead to an inappropriate controller and eventually to instability. With instability, the signals entering the plant suddenly become richer, its accurate identification becomes possible, the controller becomes correct, and stability is reencountered.

The need to ensure proper excitation ('persistent excitation') when identifying a plant is now accepted as one of the standard requirements in any adaptive system (Anderson *et al.*, 1986).

Adaptive control has now reached a certain stage of maturity, which means that applications are now becoming widespread. Figure 19 illustrates an application to a sugar mill crushing system, with which my colleagues have been involved. The system has effectively two inputs and two outputs, the inputs being the turbine governor setting and torque, with the turbine driving the crushing mill. The controlled signals are the feed chute height and the chute aperture which governs the feed rate to the crushers. Better extraction comes from better height control, and very sharp variations in the physical parameters of the feedstock of sugar cane occur. Adaptive generalised predictive control is possible for one loop, with fixed control for another loop.

An even more sophisticated application (Mills, Lee and McIntosh, forthcoming) is provided by an aluminium calciner (Figure 20). The control variables are the discharge alumina temperature, which governs product quality; the temperature fluctuation in the kiln, which governs the maintenance cost; and the energy consumption. The controlling variables are

Figure 19. Sugar cane crushing mill

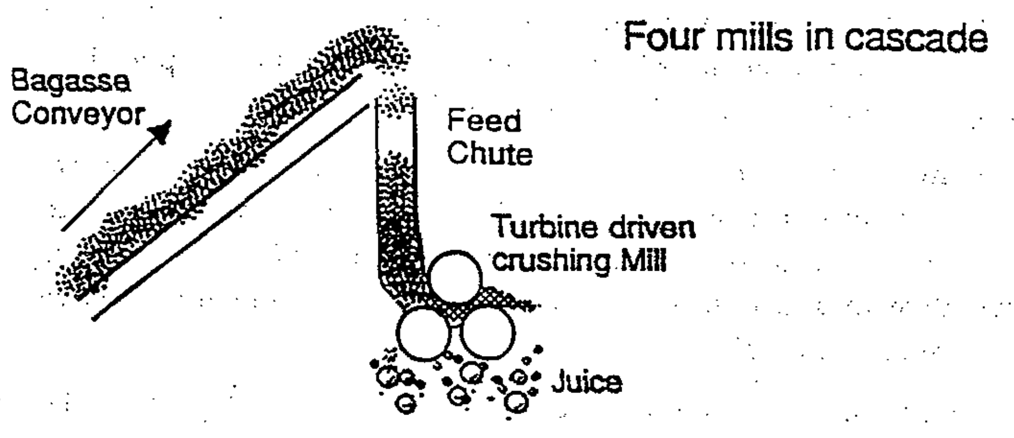
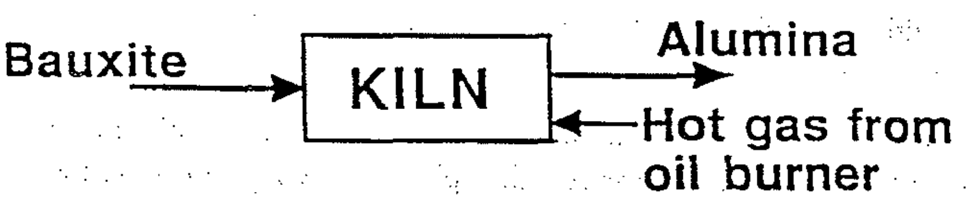


Figure 20. Alumina calciner



the bauxite feed rate, the oil mass feed rate for the oil burner and the air mass feed rate for the burner. Besides the obvious measurements, the temperature at the bauxite end of the kiln can be obtained, as can the carbon monoxide and oxygen content of the exhaust gas. The system contains time lag, is multivariable, and only the crudest of physical models is available. Nevertheless, adaptive control using a scheme called a Smith predictor achieves very effective results.

There is almost always a time lag between the generation of theory and its use and practice. Adaptive control has proved no exception, but the number of successful practical applications of the theory leave no doubt now as to its great usefulness as a particular control technology.

The future (1991-2000)

The future will be driven by applications challenges and the carrying forward of the current directions of theoretical development. There is enormous argument about which are the most important applications challenges, and I list some of these only tentatively:

- The environment. Legislation the world over is requiring industrial units to control their waste and legislation the world over is likely to increase demands for the more efficient use of energy. Both these legislative thrusts translate themselves into a demand for effective application of control.
- Automobiles. Automobiles represent a mature technology, but the application of control in automobiles has been comparatively primitive Engine control, braking control and suspension control represent three possibilities.
- Robots. Robots have already captured the attention of control engineers. They will continue to do so as control engineers attempt to cope with flexibility in the robots, associated adaptive control problems, maximization of speed, and so on.
- Discrete-event systems. For example, what are the control problems for an airport with freight, passengers and planes arriving and departing in a stochastic fashion, with all sorts of costs applying to different stages of their activity?

- Adaptive control. There is enormous scope for the application of adaptive control systems in situations where at present non-adaptive control is used. To squeeze an improvement in productivity of several per cent in a plant can translate to millions of dollars of savings in a year.

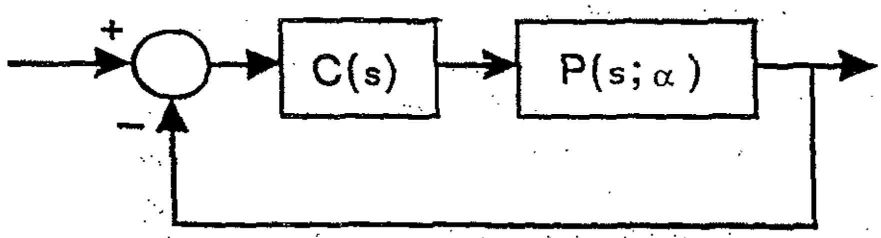
Theoretical developments are also moving in a number of exciting directions. These include:

- The use of so-called H^{∞} control as a design tool. This shows as much promise as Linear-Quadratic-Gaussian design, appears suited for similar problems, but at the same time, it does seem more closely tied to classical control ideas and the sorts of constraints that come up in classical control than does LQG design.
- Progress on time-delay systems, especially adaptive time-delay systems.
- Non-linear control. A very interesting survey of the applications of major new theoretical developments in non-linear control to the process control industry can be found in Henson and Seborg (1990).
- Robust control (see below).
- Time-varying control for time-invariant systems (see below).

Robust control

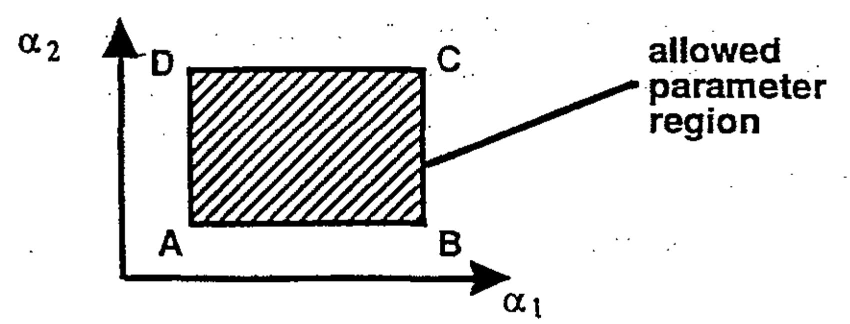
Consider Figure 21. The plant is designated by $P(s; \alpha)$, because the transfer function of the plant depends on some parameter, typically a physical parameter α, which can vary during the course of the operation of plant. Thus α could be an air pressure, a temperature, a dryness, or a friction

Figure 21. Robust control



coefficient, and may indeed be a vector. In other words, the plant depends on several scalar physical parameters. Analysis is the first question that may be faced. Consider Figure 22. One could ask: Is it enough to check for stability at the parameter settings corresponding to points *A*, *B*, *C* and *D* in order to conclude stability for the whole of the allowed parameter region? More generally, one could ask: If a controller *C(s)* gives adequate performance at *A*, *B*, *C* and *D*, will it give adequate performance for all allowed values of the parameters? And once the analysis problem is solved, the design problem comes up. How can one design a controller that will work satisfactorily for all allowed parameter settings? (Before this though, there comes the question: Can such a controller exist, or must one necessarily turn to an adaptive control approach to adequately control the plant?)

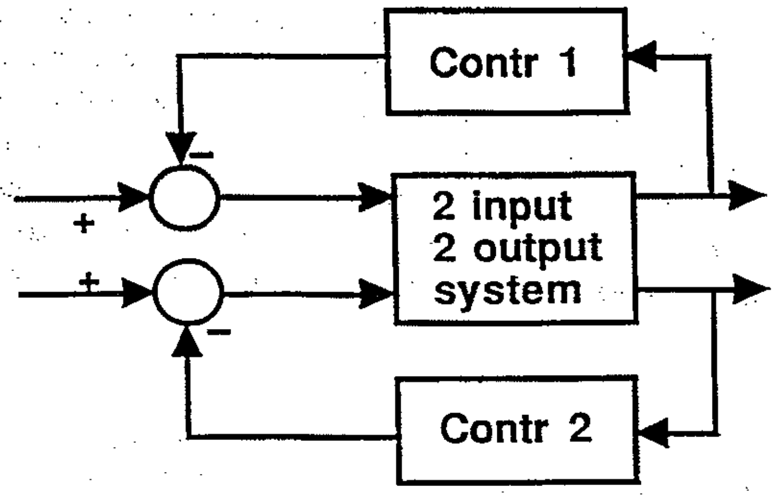
Figure 22. Parameter variations



Time-varying control

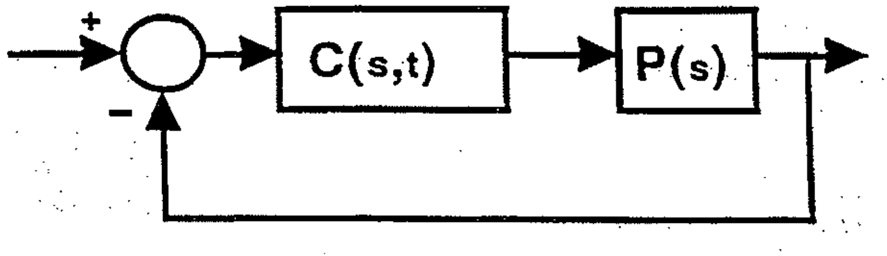
The second example is tied to time-varying control. Consider the arrangement in Figure 23. The example is artificial, but serves to make a particular point. The control structure in the figure is termed decentralised, because input 1 can only be affected through feedback from output 1 and input 2 can only be affected through feedback from output 2. Now it is a fact that there exist some 2 input, 2 output, linear time-invariant systems which cannot be stabilized by any choice of decentralised linear time-invariant controllers. Nevertheless, such systems can be stabilized by decentralised linear controllers which are periodically time-varying (Anderson and Moore, 1991). The feedback controllers switch at periodic intervals between one transfer function and another.

Figure 23. Decentralised control



This is a remarkable fact, because it shows that one can do strictly more with time-varying controllers than one can with time-invariant controllers, even for a time-invariant plant. It naturally then raises the question of what can usefully be done in practice with time-varying controllers that cannot be done with time-invariant controllers, all for a time-invariant plant (see Figure 24). Almost no answers are available to this question at the moment. One fact which can be stated is that if one compares time-invariant controller and time-variant controller designs which achieve the same level of disturbance suppression, it is normally always possible to get a better gain margin with the time-variant controller than the time-invariant controller (Yan, Anderson and Bitmead, in preparation). But such an isolated statement is an enormous distance from a full understanding of the possibilities, and there is at this stage no cohesive design theory.

Figure 24. Use of time-varying controller for a time-invariant plant



Conclusions

Control engineering has come a long way in four centuries. For most of that journey it has been applications driven, and this will continue in the future. Nevertheless, it is quite clear from the significant theoretical work going on at the moment that the scientific content of control will also develop substantially in the future. It will interact with the applications demands to solve problems more effectively than we could ever have dreamt of, and to solve problems that up to now we have assumed to be insoluble.

References

Airy, G.B., 1840 'On the regulator of the clock-work for effecting uniform movement of equatoreals', *Memoirs of The Royal Astronomical Society*, vol. 11, pp. 249-267.

Anderson, B.D.O., 1985. 'Adaptive systems, lack of persistency of excitation and bursting phenomena', *Automatica*, vol. 21, no. 21, pp. 247-258.

Anderson, B.D.O. and Liu, Y., 1989. 'Controller reduction: concepts and approaches', *IEEE Trans Auto Control*, vol. 34, no. 8, pp. 802-812.

Anderson, B.D.O. and Moore, J.B., 1981. 'Time-varying feedback laws for decentralised control', *IEEE Trans Auto Control*, vol. AC-26, no. 5, pp. 1133-1139.

Anderson, B.D.O. and Moore, J.B., 1989. *Optimal Control: Linear Quadratic Methods*, Prentice Hall Inc, Englewood Cliffs, New Jersey.

Anderson, B.D.O., Bitmead, R.R., Johnson, C.R. Jr., Kokotovic, P.J. *et al.*, 1986. *Stability of Adaptive Systems: Passivity and Averaging Analysis*, MIT Press.

Aström, K.J. and Wittenmark, B., 1990. *Computer-controlled Systems, Theory and Design*, Prentice Hall Inc, Englewood Cliffs, New Jersey.

Bennett, S., 1978. *A History of Control Engineering, 1800-1930*, Peter Peregrinus Ltd, London.

Evans, R.J., Cooper, D.N. and Kennedy, R.A., 1986. 'Servo system design for the Australia telescope', *Proc Third Conf on Control Eng*, Sydney, Australia, pp. 122-126, May.

Franklin, G.F., Powell, J.D. and Workman, M.L., 1990. *Digital Control of Dynamic Systems*, 2nd edn, Addison Wesley, Reading, Massachusetts.

Fuller, A.T., 1976a. 'The early development of control theory', *J Dynamic Systems, Measurement and Control*, vol. 98, pp. 109-117.

Fuller, A.T., 1976b. 'The early development of control theory II', *J Dynamic Systems, Measurement and Control*, vol. 98, pp. 224-235.

Gray, D.A., Anderson, B.D.O. and Bitmead, R.R., 1988. 'Models for the application of Kalman filtering to the estimation of the shape of towed arrays', *Proc NATO Adv Study Institute on Underwater Acoustic Data Processing*, Kingston, Ontario, Canada.

Henson, M.A. and Seborg, D.E., 1990. 'A critique of differential geometry control strategies for process control', (preprint) *11th IFAC World Congress*, Tallinn, Estonia, vol. 8, pp. 1-8.

Hermite, C., 1856. 'Extrait d'une lettre sur le nombre des racines d'une équation algébrique compris de limites données', *J Reine Angew Math*, vol. 52, pp. 39-51.

Hurwitz, A. 1895. 'Uber die Bedingungen, unter welchen eine Gleichung nur Wurzeln mit negativen reelen Teilen besitzt', *Mathematische Annalen*, vol. 46, pp. 273-280.

Keller, J. and Anderson, B.D.O., 1990a. 'Hα Optimal controller discretisation', *Proc 29th IEEE Conf Decision and Control*, Hawaii, pp. 1781-1785.

Keller, J. and Anderson, B.D.O., 1990b. 'A new approach to the discretisation of continuous-time controllers', *Proc ACC 1990*, San Diego, pp. 1127-1132.

Keller, J. and Anderson, B.D.O., forthcoming. 'Optimierung Abgetasteter Regelsysteme', *Automatisierungs-technik*.

Kennedy, R.A. and Evans, R.J., 1990. 'Digital redesign of a continuous controller based on closed loop performance', *Proc 29th IEEE Conf on Decision and Control*, Hawaii, pp. 1898-1901.

Mareels, I.M.Y., Anderson, B.D.O., Bitmead, R.R., Bodson, M. and Sastry, S., 1986. 'Revisiting the MIT rule for adaptive control', *Proc 2nd IFAC Workshop on Adaptive Control and Signal Processing*, Lund, pp. 67-72.

Maxwell, J.C., 1859. *On the Stability of Motion of Saturn's Rings*, MacMillan, Cambridge.

Maxwell, J.C., 1868. 'On governors', *Proc Royal Society London*, vol. 16, pp. 270-283.

Mills, P.M., Lee, P.L. and McIntosh, P., forthcoming. 'A practical study of adaptive control of an alumina calciner', *Automatica*.

Osburn, P.V., Whitaker, H.P. and Kezer, A., 1961. 'New developments in the design of model reference adaptive control systems', *Inst Aero Sciences*, Paper 61-39.

Riley, J.L., Gray, D.A. and Holdsworth, D.A., 1990. 'Estimating the position of an array of receivers using Kalman filtering techniques', *Proc 2nd IASTED Symp on Signal Processing and its Applications*, Gold Coast, Australia, pp. 364-367.

Routh, E.J., 1877. *Treatise on the Stability of a Given State or Motion*, MacMillan, London.

Sturm, C., 1836. 'Autres demonstrations de même théorême', *Journal de Mathematique Pures et Appliquees*, vol. 1, pp. 290-308.

Thaler, G.J. and Brown, R.G., 1960. *Analysis and Design of Feedback Control Systems*, McGraw Hill, New York.

Whitaker, H.P., 1959. 'An adaptive system for control of the dynamic performance of aircraft and spacecraft', *Inst Aero Sciences*, Paper 59-100.

Yan, Y-W, Anderson, B.D.O. and Bitmead, R.R., in preparation. 'On the gain margin improvement using dynamic compensation based on generalised sampled-data hold functions', *IEEE Trans Auto Control*.