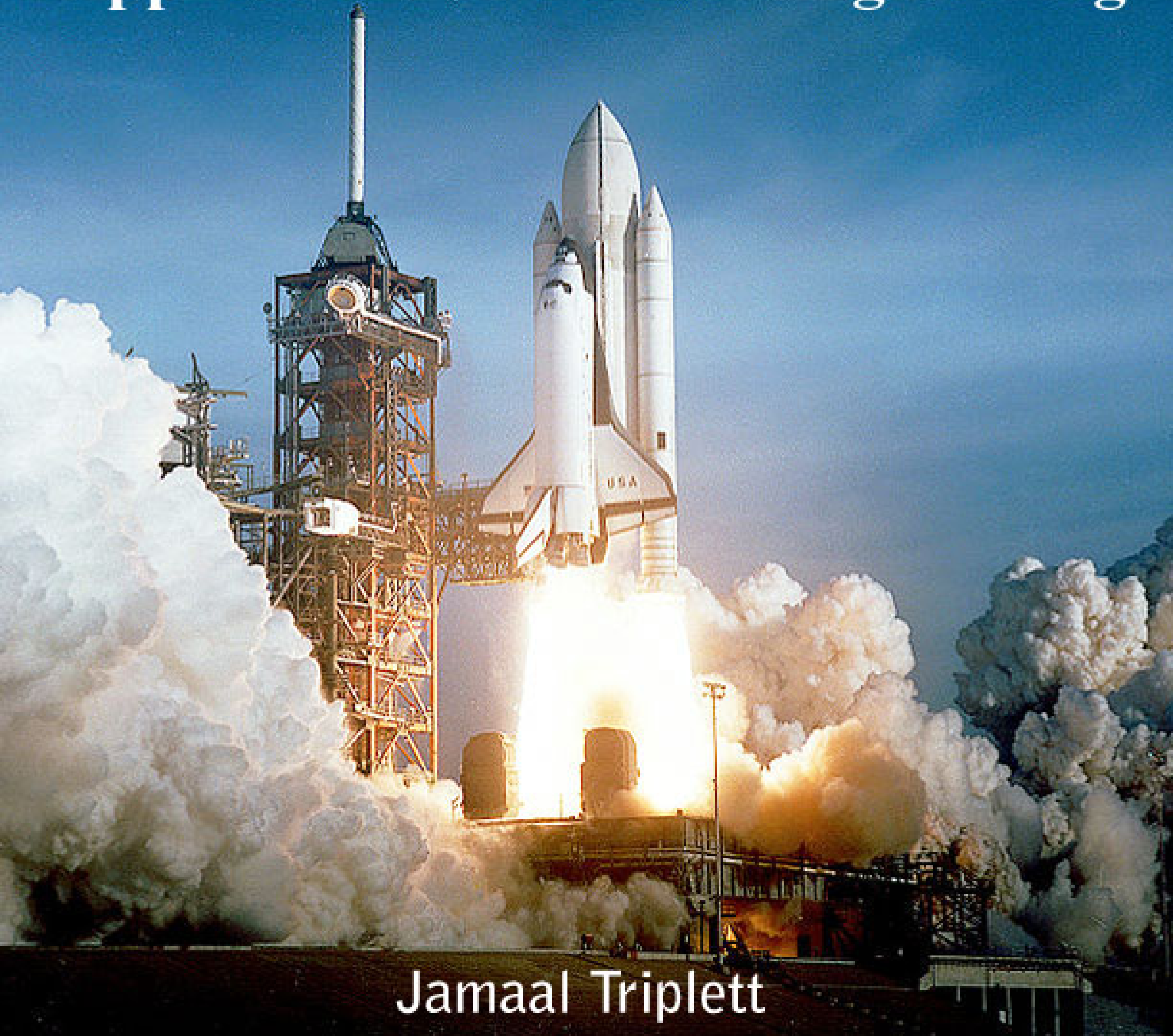
Applications of Control Engineering



Jamaal Triplett

First Edition, 2012

ISBN 978-81-323-4347-9

© All rights reserved.

Published by:
White Word Publications
4735/22 Prakashdeep Bldg,
Ansari Road, Darya Ganj,
Delhi - 110002
Email: info@wtbooks.com

Table of Contents

Chapter 1 - Aircraft Flight Control System

Chapter 2 - Electronic Flight Instrument System

Chapter 3 - Guidance System

Chapter 4 - Fire-Control System

Chapter 5 - Ignition System

Chapter 6 - Intelligent Flight Control System

Chapter 7 - Engine Control Unit

Chapter 8 - Embedded System

Chapter 9 - Electronic Stability Control

Chapter 10 - Control Reconfiguration

Chapter 11 - Networked Control System

Chapter 12 - Inverted Pendulum

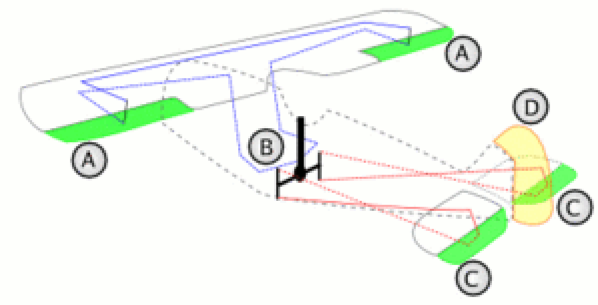
Chapter 13 - Distributed Control System

Chapter 14 - Control Center Solutions

Chapter 1

Aircraft Flight Control System

Aircraft flight control system



A typical aircraft's primary flight controls in motion

A conventional fixed-wing **aircraft flight control system** consists of flight control surfaces, the respective cockpit controls, connecting linkages, and the necessary operating mechanisms to control an aircraft's direction in flight. Aircraft engine controls are also considered as flight controls as they change speed.

The fundamentals of aircraft controls are explained in flight dynamics.

Cockpit controls

Primary controls

Generally the primary cockpit controls are arranged as follows:

- A control column or a control yoke attached to a column—for roll and pitch, which moves the ailerons when turned or deflected left and right, and moves the elevators when moved backwards or forwards
- Rudder pedals, or the earlier, pre-1919 "rudder bar", to control yaw, which move the rudder; left foot forward will move the rudder left for instance.
- Throttle controls to control engine speed or thrust for powered aircraft.

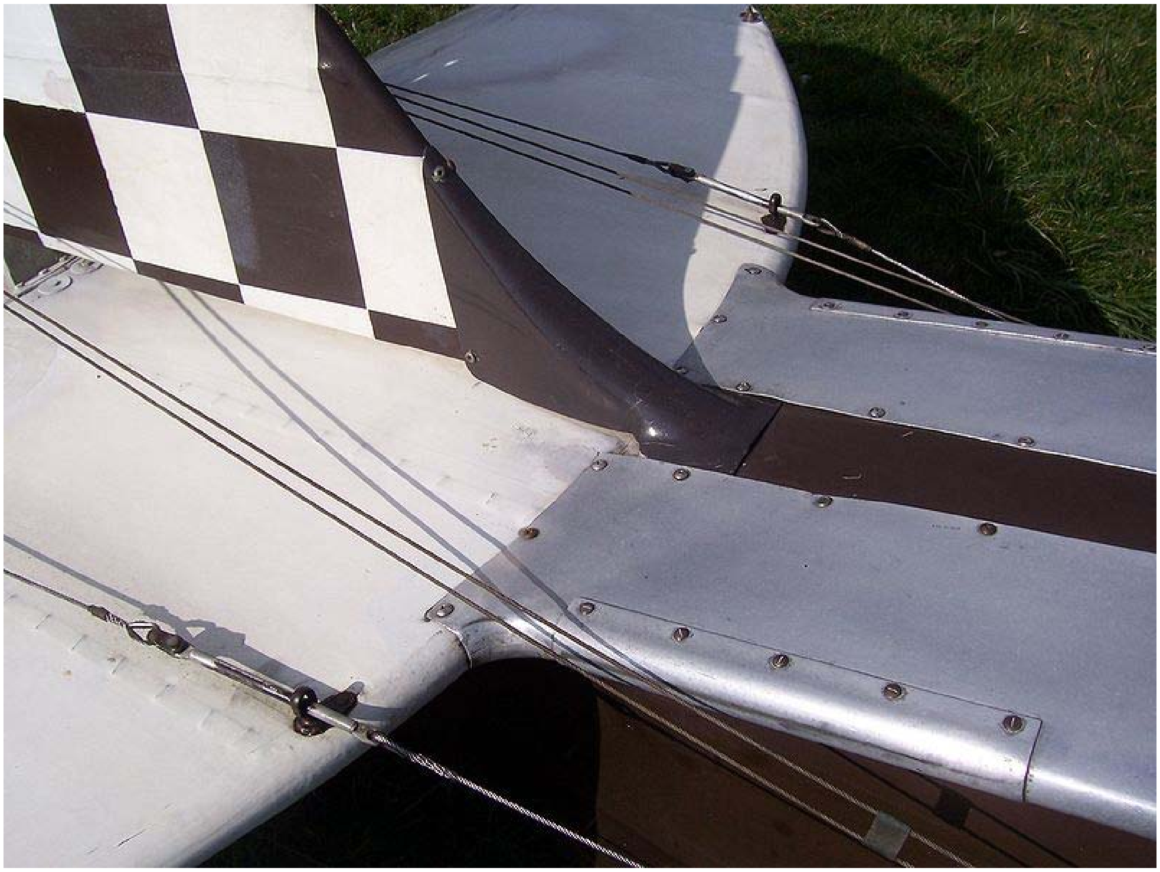
Even when an aircraft uses different kinds of surfaces, such as a V-tail/ruddervator, flaperons, or elevons, to avoid pilot confusion the aircraft will still normally be designed so that the yoke or stick controls pitch and roll in the conventional way, as will the rudder pedals for yaw. The basic pattern for modern flight controls was pioneered by French aviation figure Robert Esnault-Pelterie, with fellow French aviator Louis Blériot popularizing Esnault-Pelterie's control format initially on Louis' Blériot VIII monoplane, and standardizing the format on the July 1909 Channel-crossing Blériot XI.

Secondary controls

In addition to the primary flight controls for roll, pitch, and yaw, there are often secondary controls available to give the pilot finer control over flight or to ease the workload. The most commonly-available control is a wheel or other device to control elevator trim, so that the pilot does not have to maintain constant backward or forward pressure to hold a specific pitch attitude (other types of trim, for rudder and ailerons, are common on larger aircraft but may also appear on smaller ones). Many aircraft have wing flaps, controlled by a switch or a mechanical lever or in some cases are fully automatic by computer control, which alter the shape of the wing for improved control at the slower speeds used for takeoff and landing. Other secondary flight control systems may be available, including slats, spoilers, air brakes and variable-sweep wings.

Flight control systems

Mechanical



de Havilland Tiger Moth elevator and rudder cables

Mechanical or manually-operated flight control systems are the most basic method of controlling an aircraft. They were used in early aircraft and are currently used in small aircraft where the aerodynamic forces are not excessive. Very early aircraft, such as the Wright Flyer I, Blériot XI and Fokker Eindecker used a system of wing warping where no conventionally hinged control surfaces were used on the wing, and sometimes not even for pitch control as on the Wright Flyer I and original versions of the 1909 Etrich Taube, which only had a hinged/pivoting rudder in addition to the warping-operated pitch and roll controls. A manual flight control system uses a collection of mechanical parts such as pushrods, tension cables, pulleys, counterweights, and sometimes chains to transmit the forces applied to the cockpit controls directly to the control surfaces. Turnbuckles are often used to adjust control cable tension. The Cessna Skyhawk is a typical example of an aircraft that uses this type of system. Gust locks are often used on parked aircraft with mechanical systems to protect the control surfaces and linkages from damage from wind. Some aircraft have gust locks fitted as part of the control system.

Increases in the control surface area required by large aircraft or higher loads caused by high airspeeds in small aircraft lead to a large increase in the forces needed to move them, consequently complicated mechanical gearing arrangements were developed to extract maximum mechanical advantage in order to reduce the forces required from the pilots. This arrangement can be found on bigger or higher performance propeller aircraft such as the Fokker 50.

Some mechanical flight control systems use servo tabs that provide aerodynamic assistance. Servo tabs are small surfaces hinged to the control surfaces. The flight control mechanisms move these tabs, aerodynamic forces in turn move, or assist the movement of the control surfaces reducing the amount of mechanical forces needed. This arrangement was used in early piston-engined transport aircraft and in early jet transports. The Boeing 737 incorporates a system, whereby in the unlikely event of total hydraulic system failure, it automatically and seamlessly reverts to being controlled via servo-tab.

Hydro-mechanical

The complexity and weight of mechanical flight control systems increase considerably with the size and performance of the aircraft. Hydraulically powered control surfaces help to overcome these limitations. With hydraulic flight control systems, the aircraft's size and performance are limited by economics rather than a pilot's muscular strength. At first, only-partially boosted systems were used in which the pilot could still feel some of the aerodynamic loads on the control surfaces (feedback).

A hydro-mechanical flight control system has two parts:

- The *mechanical circuit*, which links the cockpit controls with the hydraulic circuits. Like the mechanical flight control system, it consists of rods, cables, pulleys, and sometimes chains.

- The *hydraulic circuit*, which has hydraulic pumps, reservoirs, filters, pipes, valves and actuators. The actuators are powered by the hydraulic pressure generated by the pumps in the hydraulic circuit. The actuators convert hydraulic pressure into control surface movements. The electro-hydraulic servo valves control the movement of the actuators.

The pilot's movement of a control causes the mechanical circuit to open the matching servo valve in the hydraulic circuit. The hydraulic circuit powers the actuators which then move the control surfaces. As the actuator moves, the servo valve is closed by a mechanical feedback linkage - one that stops movement of the control surface at the desired position.

This arrangement was found in the older-designed jet transports and in some high-performance aircraft. Examples include the Antonov An-225 and the Lockheed SR-71.

Artificial feel devices

With purely mechanical flight control systems, the aerodynamic forces on the control surfaces are transmitted through the mechanisms and are felt directly by the pilot. This gives tactile feedback of airspeed and aids flight safety.

With hydromechanical flight control systems however, the load on the surfaces cannot be felt and there is a risk of overstressing the aircraft through excessive control surface movement. To overcome this problem artificial feel systems are used. For example, for the controls of the RAF's Avro Vulcan jet bomber and the RCAF's Avro Canada CF-105 Arrow supersonic interceptor, both 1950's-era designs, the required force feedback was achieved by a spring device. The fulcrum of this device was moved in proportion to the square of the air speed (for the elevators) to give increased resistance at higher speeds. For the controls of the American Vought F-8 Crusader and the LTV A-7 Corsair II warplanes, a "bob-weight" was used in the pitch axis of the control stick, giving force feedback that was proportional to the airplane's normal acceleration.

Stick shaker

A stick shaker is a device (available in some hydraulic aircraft) which is fitted into the control column which shakes the control column when the aircraft is about to stall. Also in some aircraft like the McDonnell Douglas DC-10 there is/was a back-up electrical power supply which the pilot can turn on to re-activate the stick shaker in case the hydraulic connection to the stick shaker is lost.

Fly-by-wire control systems

A fly-by-wire (FBW) system replaces manual flight control of an aircraft with an electronic interface. The movements of flight controls are converted to electronic signals transmitted by wires (hence the fly-by-wire term), and flight control computers determine how to move the actuators at each control surface to provide the expected response.

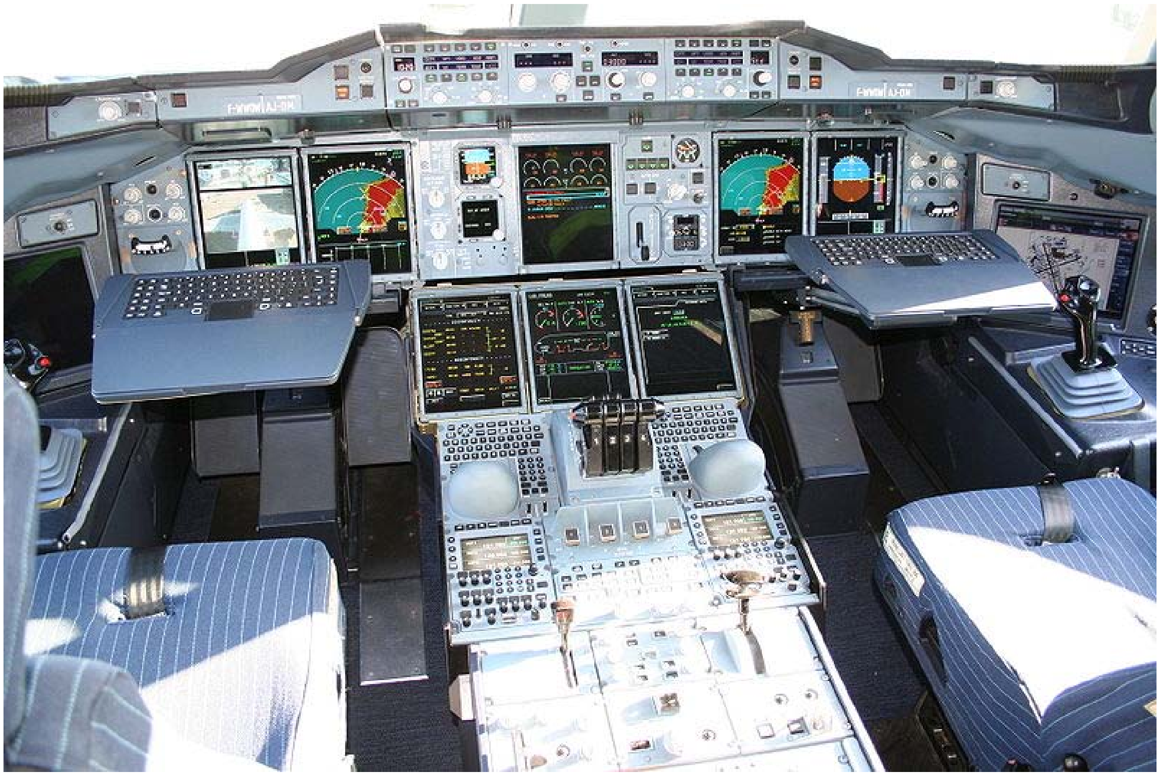
Commands from the computers are also input without the pilot's knowledge to stabilize the aircraft and perform other tasks.

Fluidic flight controls

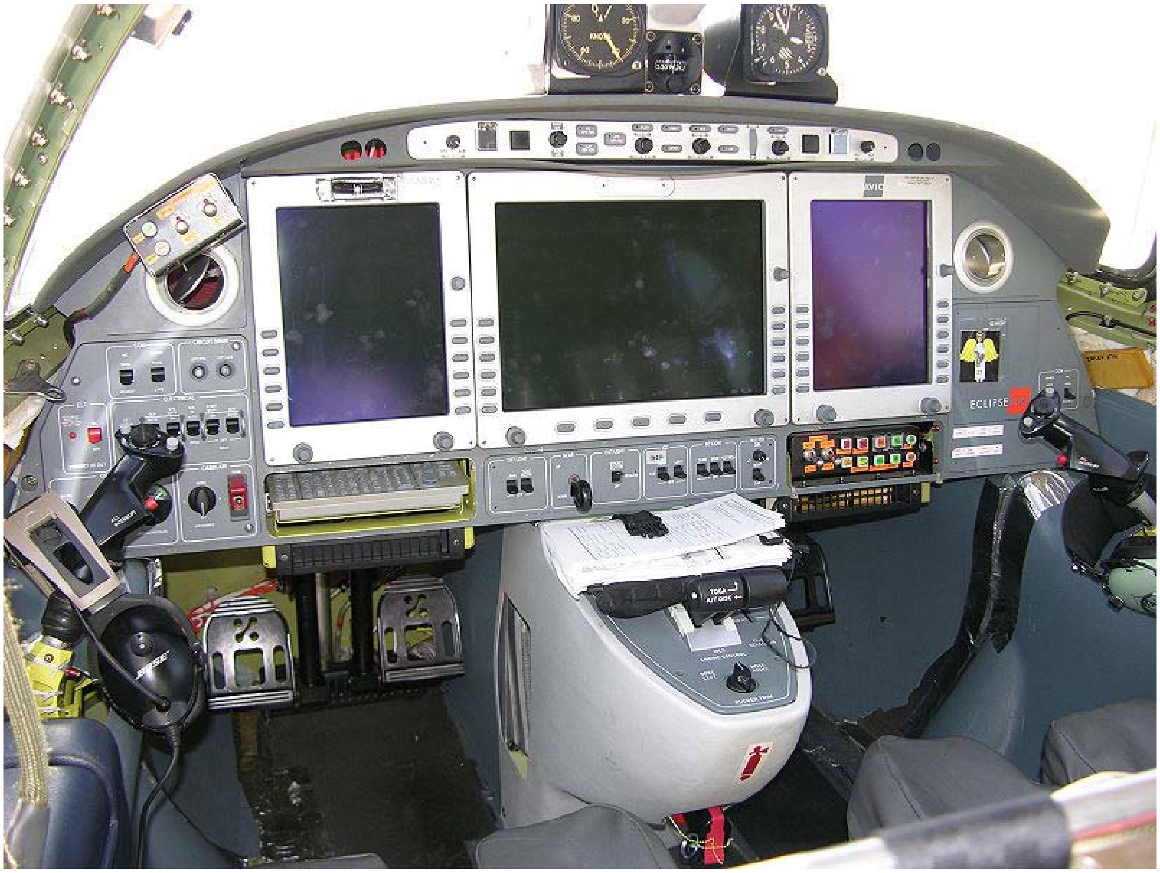
Conventional mechanical flight control surfaces may also be replaced completely, by a fluidic flight control system, provided by differential streams of blown air, such as with the Demon (UAV), aircraft which flew for the first time, in the UK, in September 2010.

Chapter 2

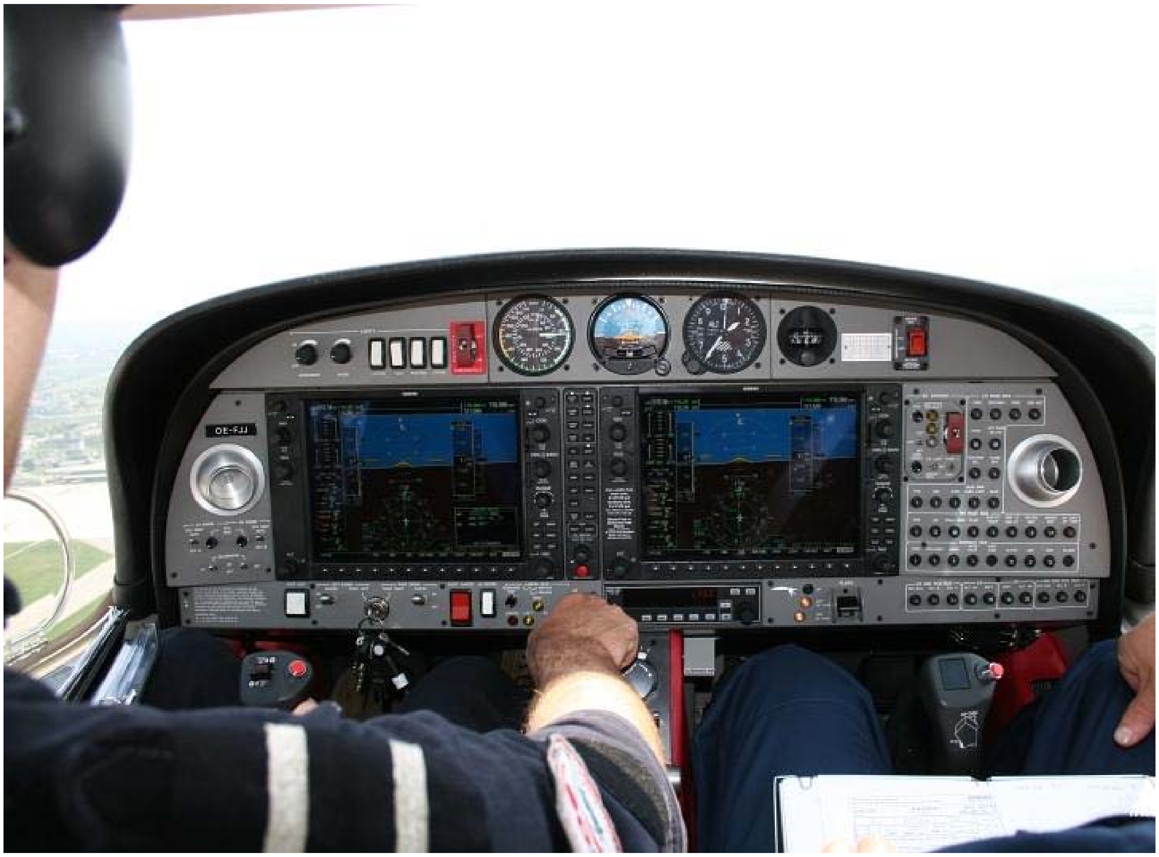
Electronic Flight Instrument System



EFIS on an Airbus A380



EFIS on an Eclipse 500



Garmin G1000 on a Diamond DA42

An **Electronic Flight Instrument System** (EFIS) is a flight deck instrument display system in which the display technology used is electronic rather than electromechanical. EFIS normally consists of a primary flight display (PFD), multi-function display (MFD) and Engine Indicating and Crew Alerting System (EICAS) display. Although cathode ray tube (CRT) displays were used at first, liquid crystal displays (LCD) are now more common.

The complex electromechanical attitude director indicator (ADI) and horizontal situation indicator (HSI) were the first candidates for replacement by EFIS. However, there are now few flight deck instruments for which no electronic display is available.

Overview

EFIS installations vary greatly. A light aircraft might be equipped with one display unit, on which are displayed flight and navigation data. A wide-body aircraft is likely to have six or more display units.

Typical EFIS displays and controls can be seen at this B737 technical information web site. The equivalent electromechanical instruments are also shown here.

An EFIS installation will have the following components:

- Displays
- Controls
- Data processors

A basic EFIS might have all these facilities in the one unit.

Display units

Primary Flight Display (PFD)

On the flight deck, the display units are the most obvious parts of an EFIS system, and are the features which give rise to the name "glass cockpit". The display unit taking the place of the ADI is called the primary flight display (PFD). If a separate display replaces the HSI, it is called the navigation display. The PFD displays all information critical to flight, including calibrated airspeed, altitude, heading, attitude, vertical speed and yaw. The PFD is designed to improve a pilot's situational awareness by integrating this information into a single display instead of six different analog instruments, reducing the amount of time necessary to monitor the instruments. PFDs also increase situational awareness by alerting the aircrew to unusual or potentially hazardous conditions — for example, low airspeed, high rate of descent — by changing the color or shape of the display or by providing audio alerts.

The names Electronic Attitude Director Indicator and Electronic Horizontal Situation Indicator are used by some manufacturers. However, a simulated ADI is only the centerpiece of the PFD. Additional information is both superimposed on and arranged around this graphic.

Multi-function displays can render a separate navigation display unnecessary. Another option is to use one large screen to show both the PFD and navigation display.

The PFD and navigation display (and multi-function display, where fitted) are often physically identical. The information displayed is determined by the system interfaces where the display units are fitted. Thus, spares holding is simplified: the one display unit can be fitted in any position.

LCD units generate less heat than CRTs; an advantage in a congested instrument panel. They are also lighter, and occupy a lower volume.

Multi-Function Display (MFD) / Navigation Display (ND)

The MFD (Multi-Function Display) displays navigational and weather information from multiple systems. MFDs are most frequently designed as "chart-centric", where the aircrew can overlay different information over a map or chart. Examples of MFD overlay information include the aircraft's current route plan, weather information from either on-board radar or lightning detection sensors or ground-based sensors, e.g., NEXRAD, restricted airspace and aircraft traffic. The MFD can also be used to view other non-overlay type of data (e.g., current route plan) and calculated overlay-type data, e.g., the glide radius of the aircraft, given current location over terrain, winds, and aircraft speed and altitude.

MFDs can also display information about aircraft systems, such as fuel and electrical systems. As with the PFD, the MFD can change the color or shape of the data to alert the aircrew to hazardous situations.

Engine Indications and Crew Alerting System (EICAS) / Electronic Centralized Aircraft Monitoring (ECAM)

EICAS (Engine Indications and Crew Alerting System) displays information about the aircraft's systems, including its fuel, electrical and propulsion systems (engines). EICAS displays are often designed to mimic traditional round gauges while also supplying digital readouts of the parameters.

EICAS improves situational awareness by allowing the aircrew to view complex information in a graphical format and also by alerting the crew to unusual or hazardous situations. For example, if an engine begins to lose oil pressure, the EICAS might sound an alert, switch the display to the page with the oil system information and outline the low oil pressure data with a red box. Unlike traditional round gauges, many levels of warnings and alarms can be set. Proper care must be taken when designing EICAS to ensure that the aircrew are always provided with the most important information and not overloaded with warnings or alarms.

ECAM is a similar system used by Airbus, which in addition to providing EICAS functions also recommend remedial action.

Control panels

The pilots are provided with controls, with which they select display range and mode (for example, map or compass rose) and enter data (such as selected heading).

Where inputs by the pilot are used by other equipment, data buses broadcast the pilot's selections so that the pilot only needs to enter the selection once. For example, the pilot selects the desired level-off altitude on a control unit. The EFIS repeats this selected altitude on the PFD and by comparing it with the actual altitude (from the air data computer) generates an altitude error display. This same altitude selection is used by the automatic flight control system to level off, and by the altitude alerting system to provide appropriate warnings.

Data processors

The EFIS visual display is produced by the symbol generator. This receives data inputs from the pilot, signals from sensors, and EFIS format selections made by the pilot. The symbol generator can go by other names, such as display processing computer, display electronics unit, etc.

The symbol generator does more than generate symbols. It has (at the least) monitoring facilities, a graphics generator and a display driver. Inputs from sensors and controls arrive via data buses, and are checked for validity. The required computations are performed, and the graphics generator and display driver produce the inputs to the display units.

Monitoring

Like personal computers, flight instrument systems need power-on-self-test facilities and continuous self-monitoring. Flight instrument systems, however, need additional monitoring capabilities:

- Input validation — verify that each sensor is providing valid data
- Data comparison — cross check inputs from duplicated sensors
- Display monitoring — detect failures within the instrument system

Former practice

Traditional (electromechanical) displays were equipped with synchro mechanisms which would transmit, to an instrument comparator, the pitch, roll and heading that were actually being shown on the Captain's and First Officer's instruments. The comparator warned of excessive differences between the Captain and First Officer displays. Even a fault as far *downstream* as a jam in, say, the roll mechanism of an ADI would trigger a comparator warning.

The instrument comparator thus provided both comparator monitoring and display monitoring.

Comparator monitoring

With EFIS, the comparator function is as simple as ever. Is the roll data (bank angle) from sensor 1 the same as the roll data from sensor 2? If not, put a warning caption (such as **CHECK ROLL**) on both PFDs. Comparison monitors will give warnings for airspeeds, pitch, roll and altitude indications. The more advanced EFIS systems, more comparator monitors will be enabled.

Display monitoring

An EFIS display allows no easy re-transmission of what is shown on the display. What is required is a new approach to display monitoring that provides safety equivalent to that of the traditional system. One solution is to keep the display unit as simple as possible, so that it is unable to introduce errors. The display unit either works or does not work. A failure is always obvious, never insidious. Now the monitoring function can be shifted *upstream* to the output of the symbol generator.

In this technique, each symbol generator contains two display monitoring channels. One channel, the internal, samples the output from its own symbol generator to the display unit and computes, for example, what roll attitude should produce that indication. This computed roll attitude is then compared with the roll attitude input to the symbol generator from the INS or AHRS. Any difference has probably been introduced by faulty processing, and triggers a warning on the relevant display.

The external monitoring channel carries out the same check on the symbol generator on the other side of the flight deck: the Captain's symbol generator checks the First Officer's, the First Officer's checks the Captain's. Whichever symbol generator detects a fault, puts up a warning on its own display.

The external monitoring channel also checks sensor inputs (to the symbol generator) for reasonableness. A spurious input, such as a radio height greater than the radio altimeter's maximum, results in a warning.

Human factors

Clutter

At various stages of a flight, a pilot uses different combinations of data. Ideally, only the data in use would be displayed, but an electromechanical instrument has to be in view all the time. To improve display clarity, intricate mechanisms are used on ADIs and HSIs to remove superfluous indications temporarily, e.g., removing the glide slope scale when it is not being used.

With EFIS, some indications, e.g., engine vibration, might not be displayed under normal conditions. If limits are exceeded, then the reading will be displayed. In similar fashion, EFIS is programmed to show the glideslope scale and pointer only during an ILS approach.

If a failure of input data is detected, electromechanical instruments add yet another indicator to the display. Typically, a bar is dropped across the erroneous data. EFIS, on the other hand, removes invalid data from the display and substitutes an appropriate warning.

A de-clutter mode is activated automatically when the pilot's attention is required to be focused on a specific item. For example, if the aircraft is pitched up or down above a specified pitch, usually 30 to 60 degrees, the attitude indicator will de-clutter items from sight until the pitch is brought to an acceptable level. This allows the pilot to focus on the most important matter of aircraft control.

Color

Although color has long been used in traditional instruments, it is restricted to aiding in identification of the data. There is no means of changing the color of any display component.

This restriction has been lifted with EFIS. For example, as an aircraft approaches the glideslope, a blue caption could indicate glide slope is armed; on capture the color might change to green.

On a typical EFIS system, the navigation needles are color coded to reflect the type of navigation being used. Green needles are used for ground based navigation such as VORs, Localizers and ILS systems. Magenta needles are used for GPS navigation.

Advantages

EFIS offers **versatility** by avoiding some of the physical limitations of traditional instruments. Thus, the same display which shows a course deviation indicator, can be switched to show the planned track provided by an area navigation or flight management system. If desired, the weather radar picture can then be superimposed on the displayed route.

The **flexibility** afforded by software modifications, minimises costs when new aircraft equipment and new regulations are introduced. The EFIS system can be updated with new software to extend its capabilities. Such updates introduced in the 1990s included enhanced GPWS, and TCAS.

A degree of **redundancy** is available even with the simple two-screen EFIS installation. Should the PFD fail, transfer switching repositions its vital information to the screen normally occupied by the navigation display.

Advances in EFIS

In the late 1980s, EFIS became standard equipment on most Boeing and Airbus airliners, and many business aircraft adopted EFIS in the 1990s.

Recent advances in computing power and reductions in the cost of liquid-crystal displays and navigational sensors (such as GPS and Attitude and Heading Reference Systems) have finally brought EFIS to general aviation aircraft. Notable examples are the Garmin G1000 and Chelton Flight Systems EFIS-SV.

Several EFIS manufacturers have focused on the experimental aircraft market, producing EFIS and EICAS systems for as little as US$1,000. The low cost is possible for several reasons, including steep drops in sensor prices and a lack of requirements to receive Federal Aviation Administration certification. This latter point restricts their use to experimental aircraft and certain other aircraft categories depending on local regulations. Uncertified EFIS systems are also found in Sport Pilot category aircraft, including factory built, microlight and ultralight aircraft. These systems can be fitted to certified aircraft in some cases as secondary or backup systems depending on local aviation authorities rules and regulations.

Chapter 3

Guidance System

A **guidance system** is a device or group of devices used to navigate a ship, aircraft, missile, rocket, satellite, or other craft. Typically, this refers to a system that navigates without direct or continuous human control. Systems that are intended to have a high degree of human interaction are usually referred to as a navigation system.

One of the earliest examples of a true guidance system is that used in the German V-1 during World War II. This system consisted of a simple gyroscope to maintain heading, an airspeed sensor to estimate flight time, an altimeter to maintain altitude, and other redundant systems.

A guidance system has three major sub-sections: Inputs, Processing, and Outputs. The input section includes sensors, course data, radio and satellite links, and other information sources. The processing section, composed of one or more CPUs, integrates this data and determines what actions, if any, are necessary to maintain or achieve a proper heading. This is then fed to the outputs which can directly affect the system's course. The outputs may control speed by interacting with devices such as turbines, and fuel pumps, or they may more directly alter course by actuating ailerons, rudders, or other devices.

History

Inertial navigation systems were originally developed for rockets. American rocket pioneer Robert Goddard experimented with rudimentary gyroscopic systems. Dr. Goddard's systems were of great interest to contemporary German pioneers including Wernher von Braun. The systems entered more widespread use with the advent of spacecraft, guided missiles, and commercial airliners.

US guidance history centers around 2 distinct communities. One driven out of Caltech and NASA JPL, the other from the German scientists that developed the early V2 rocket guidance and MIT. The GN&C system for V2 provided many innovations and was the most sophisticated military weapon in 1942 using self contained closed loop guidance. Early V2s leveraged 2 gyroscopes and lateral accelerometer with a simple analog computer to adjust the azimuth for the rocket in flight. Analog computer signals were used to drive 4 external rudders on the tail fins for flight control. Von Braun engineered the surrender of 500 of his top rocket scientists, along with plans and test vehicles, to the Americans. They arrived in Fort Bliss, Texas in 1945 and were subsequently moved to Huntsville, Al in 1950 (aka Redstone arsenal). Von Braun's passion was interplanetary space flight. However his tremendous leadership skills and experience with the V-2 program made him invaluable to the US military. In 1955 the Redstone team was selected to put America's first satellite into orbit putting this group at the center of both military and commercial space.

The Jet Propulsion Laboratory traces its history from the 1930s, when Caltech professor Theodore von Karman conducted pioneering work in rocket propulsion. Funded by Army Ordnance in 1942, JPL's early efforts would eventually involve technologies beyond those of aerodynamics and propellant chemistry. The result of the Army Ordnance effort was JPL's answer to the German V-2 missile, named Corporal, first launched in May 1947. On December 3, 1958, two months after the National Aeronautics and Space Administration (NASA) was created by Congress, JPL was transferred from Army jurisdiction to that of this new civilian space agency. This shift was due to the creation of a military focused group derived from the German V2 team. Hence, beginning in 1958, NASA JPL and the Caltech crew became focused primarily on unmanned flight and shifted away from military applications with a few exceptions. The community surrounding JPL drove tremendous innovation in telecommunication, interplanetary exploration and earth monitoring (among other areas).

In the early 1950s, the US government wanted to insulate itself against over dependency on the Germany team for military applications. Among the areas that were domestically "developed" was missile guidance. In the early 1950s the MIT Instrumentation Laboratory (later to become the Charles Stark Draper Laboratory, Inc.) was chosen by the Air Force Western Development Division to provide a self-contained guidance system backup to Convair in San Diego for the new Atlas intercontinental ballistic missile. The technical monitor for the MIT task was a young engineer named Jim Fletcher who later served as the NASA Administrator. The Atlas guidance system was to be a combination of an on-board autonomous system, and a ground-based tracking and command system. This was the beginning of a philosophic controversy, which, in some areas, remains unresolved. The self-contained system finally prevailed in ballistic missile applications for obvious reasons. In space exploration, a mixture of the two remains.

In the summer of 1952, Dr. Richard Battin and Dr. J. Halcombe ("Hal") Laning Jr., researched computational based solutions to guidance as computing began to step out of the analog approach. As computers of that time were very slow (and missiles very fast) it was extremely important to develop programs that were very efficient. Dr. J. Halcombe Laning, with the help of Phil Hankins and Charlie Werner, initiated work on MAC, an algebraic programming language for the IBM 650, which was completed by early spring of 1958. MAC became the work-horse of the MIT lab. MAC is an extremely readable language having a three-line format, vector-matrix notations and mnemonic and indexed subscripts. Today's Space Shuttle (STS) language called HAL, (developed by Intermetrics, Inc.) is a direct offshoot of MAC. Since the principal architect of HAL was

Jim Miller, who co-authored with Hal Laning a report on the MAC system, it is a reasonable speculation that the space shuttle language is named for Jim's old mentor, and not, as some have suggested, for the electronic superstar of the Arthur Clarke movie "2001-A Space Odyssey." (Richard Batin, AIAA 82-4075, April 1982)

Hal Laning and Richard Batin undertook the initial analytical work on the Atlas intertial guidance in 1954. Other key figures at Convair were Charlie Bossart, the Chief Engineer, and Walter Schweidetzky, head of the guidance group. Walter had worked with Wernher von Braun at Peenemuende during World War II.

The initial "Delta" guidance system assessed the difference in position from a reference trajectory. A velocity to be gained (VGO) calculation is made to correct the current trajectory with the objective of driving VGO to Zero. The mathematics of this approach were fundamentally valid, but dropped because of the challenges in accurate inertial navigation (e.g. IMU Accuracy) and analog computing power. The challenges faced by the "Delta" efforts were overcome by the "Q system" of guidance. The "Q" system's revolution was to bind the challenges of missile guidance (and associated equations of motion) in the matrix Q. The Q matrix represents the partial derivatives of the velocity with respect to the position vector. A key feature of this approach allowed for the components of the vector cross product (v, xdv,/dt) to be used as the basic autopilot rate signals-a technique that became known as "cross-product steering." The Q-system was presented at the first Technical Symposium on Ballistic Missiles held at the Ramo-Wooldridge Corporation in Los Angeles on June 21 and 22, 1956. The "Q System" was classified information through the 1960s. Derivations of this guidance are used for today's military missiles. The CSDL team remains a leader in the military guidance and is involved in projects for most divisions of the US military.

In Feb of 1961 NASA Awarded MIT a contract for preliminary design study of a guidance and navigation system for Apollo. Today's space shuttle guidance is named PEG4 (Powered Explicit Guidance). It takes into account both the Q system and the predictor-corrector attributes of the original "Delta" System (PEG Guidance). Although many updates to the shuttles navigation system have taken place over the last 30 years (ex. GPS in the OI-22 build), the guidance core of today's Shuttle GN&C system has evolved little. Within a manned system, there is a human interface needed for the guidance system. As Astronauts are the customer for the system, many new teams are formed that touch GN&C as it is a primary interface to "fly" the vehicle. For the Apollo and STS (Shuttle system) CSDL "designed" the guidance, McDonnell Douglas wrote the requirements and IBM programed the requirements.

Much system complexity within manned systems is driven by "redundancy management" and the support of multiple "abort" scenarios that provide for crew safety. Manned US Lunar and Interplanetary guidance systems leverage many of the same guidance innovations (described above) developed in the 1950s. So while the core mathematical construct of guidance has remained fairly constant, the facilities surrounding GN&C continue to evolve to support new vehicles, new missions and new hardware. The center of excellence for the manned guidance remains at MIT (CSDL) as well as the former McDonnell Douglas Space Systems (in Houston).

Guidance systems

Guidance systems consist of 3 esstential parts: **navigation** which tracks current location, **guidance** which leverages navigation data and target information to direct flight control "where to go", and **control** which accepts guidance commands to effect change in aerodynamic and/or engine controls.

Navigation is the art of determining where you are, a science that has seen tremendous focus in 1711 with the Longitude prize. Navigation aids either measure position from a *fixed* point of reference (ex. landmark, north star, LORAN Beacon), *relative* position to a target (ex. radar, infra-red, ...) or track *movement* from a known position/starting point (e.g. IMU). Today's complex systems use multiple approaches to determine current position. For example, today's most advanced navigation systems are embodied within the Anti-ballistic missile, the RIM-161 Standard Missile 3 leverages GPS, IMU and ground segment data in the boost phase and relative position data for intercept targeting. Complex systems typically have multiple redundancy to address drift, improve accuracy (ex. relative to a target) and address isolated system failure. Navigation systems therefore take multiple inputs from many different sensors, both internal to the system and/or external (ex. ground based update). Kalman filter provides the most common approach to combining navigation data (from multiple sensors) to resolve current position. Example navigation approaches:

- Celestial navigation is a position fixing technique that was devised to help sailors cross the featureless oceans without having to rely on dead reckoning to enable them to strike land. Celestial navigation uses angular measurements (sights) between the horizon and a common celestial object. The Sun is most often measured. Skilled navigators can use the Moon, planets or one of 57 navigational stars whose coordinates are tabulated in nautical almanacs. Historical tools include a sextant, watch and ephemeris data. Today's space shuttle, and most interplanetary spacecraft, use optical systems to calibrate inertial navigation systems: Crewman Optical Alignment Sight (COAS), Star Tracker.
- Long-range Navigation (LORAN) : This was the predecessor of GPS and was (and to an extent still is) used primarily in commercial sea transportation. The system works by triangulating the ship's position based on directional reference to known transmitters.
- Global Positioning System (GPS) : GPS was designed by the US military with the primary purpose of addressing "drift" within the inertial navigation of Submarine-launched ballistic missile(SLBMs) prior to launch. GPS transmits 2 signal types: military and a commercial. The accuracy of the military signal is classified but can be assumed to be well under 0.5 meters. GPS is a system of 24 satellites orbiting in unique planes 10.9-14.4 Nautical miles above the earth. The Satellites are in well defined orbits and transmit highly accurate time information which can be used to triangulate position.

- Inertial Measurement Units (IMUs) are the primary inertial system for maintaining current position (navigation) and orientation in missiles and aircraft. They are complex machines with one or more rotating Gyroscopes that can rotate freely in 3 degrees of motion within a complex gimbal system. IMUs are "spun up" and calibrated prior to launch. A minimum of 3 separate IMUs are in place within most complex systems. In addition to relative position, the IMUs contain accelerometers which can measure acceleration in all axis. The position data, combined with acceleration data provide the necessary inputs to "track" motion of a vehicle. IMUs have a tendency to "drift", due to friction and accuracy. Error correction to address this drift can be provided via ground link telemetry, GPS, radar, optical celestial navigation and other navigation aids. When targeting another (moving) vehicle, relative vectors become paramount. In this situation, navigation aids which provide updates of position *relative to the target* are more important. In addition to the current position, inertial navigation systems also typically estimate a predicted position for future computing cycles.
- Radar/Infrared/Laser : This form of navigation provides information to guidance *relative to a known target*, it has both civilian (ex rendezvous) and military applications.
    - active (employs own radar to illuminate the target),
    - passive (detects target’s radar emissions),
    - semiactive radar homing,
    - Infrared homing : This form of guidance is used exclusively for military munitions, specifically air-to-air and surface-to-air missiles. The missile’s seeker head homes in on the infrared (heat) signature from the target’s engines (hence the term “heat-seeking missile”),
    - Ultraviolet homing, used in FIM-92 Stinger - more resistive to countermeasures, than IR homing system
    - Laser designation : A laser designator device calculates relative position to a highlighted target. Most are familiar with the military uses of the technology on Laser-guided bomb. The space shuttle crew leverages a hand held device to feed information into rendezvous planning. The primary limitation on this device is that it requires a line of sight between the target and the designator.
    - Terrain contour matching (TERCOM). Uses a ground scanning radar to "match" topography against digital map data to fix current position. Used by cruise missiles such as the BGM-109 Tomahawk.

Guidance is the "driver" of a vehicle. It takes input from the navigation system (where am I) and uses targeting information (where do I want to go) to send signals to the flight control system that will allow the vehicle to reach its destination (within the operating constraints of the vehicle). The "targets" for guidance systems are one or more state vectors (position and velocity) and can be inertial or relative. During powered flight, guidance is continually calculating steering directions for flight control. For example the space shuttle targets an altitude, velocity vector, and gamma to drive main engine cut off. Similarly, an Intercontinental ballistic missile also targets a vector. The target vectors are developed to fulfill the mission and can be preplanned or dynamically created.

Mathematical foundations to today's guidance problems can be found in **An Introduction to the Mathematics and Methods of Astrodynamics**, Revised Edition (Aiaa Education Series) Richard Battin 1991.

Control. Flight control is accomplished either aerodynamically through powered controls such as engines. Guidance sends signals to flight control. A Digital Autopilot (DAP) is the common term used to describe the interface between guidance and control. Guidance and the DAP are responsible for calculating the precise instruction for each flight control. The DAP provides feedback to guidance on the state of flight controls.

Chapter 4

Fire-Control System

A German anti-aircraft 88 mm gun with its fire-control computer from World War II. Displayed in the Canadian War Museum.

A **fire-control system** is a number of components working together, usually a gun data computer, a director, and radar, which is designed to assist a weapon system in hitting its target. It performs the same task as a human gunner firing a weapon, but attempts to do so faster and more accurately.

History

The original fire-control systems were developed for ships. When gunnery ranges increased dramatically in the late 19th century it was no longer a simple matter of calculating the proper aim point given the flight times of the shells. Increasingly sophisticated mechanical calculators were employed for proper gunlaying, typically with various spotters and distance measures being sent to a central plotting station deep within the ship. There the fire direction teams fed in the location, speed and direction of the ship and its target, as well as various adjustments for Coriolis effect, weather effects on the air, and other adjustments. The resulting directions, known as a **firing solution**, would then be fed back out to the turrets for laying. If the rounds missed, an observer could work out how far they missed by and in which direction, and this information could be fed back into the computer along with any changes in the rest of the information and another shot attempted.

Submarines were also equipped with fire control computers for the same reasons, but their problem was even more pronounced; in a typical "shot", the torpedo would take one to two minutes to reach its target. Calculating the proper "lead" given the relative motion of the two vessels was very difficult, and torpedo data computers were added to dramatically improve the speed of these calculations.

The Dreyer Table fire control system was already fitted to British capital ships by mid 1916. It was claimed to been plagiarised from earlier work by Arthur Pollen, but given the differences between the two designs this is now disputed. Pollen's final "clock" (computer), the Argo Mark V was installed on ships of the Imperial Russian Navy. An improved development the "Admiralty Fire Control Table" was in use in 1927.

By the start of World War II, aircraft altitude performance had increased so much that anti-aircraft guns had similar predictive problems, and were increasingly equipped with fire-control computers. The main difference between these systems and the ones on ships was size and speed. The early versions of the High Angle Control System, or HACS, of Britain's Royal Navy were examples of a system that predicted based upon the assumption that target speed, direction, and altitude would remain constant during the prediction cycle, which consisted of the time to fuze the shell and the time of flight of the shell to the target. The USN Mk 37 system made similar assumptions except that it could predict based upon a constant rate of altitude change. The Kerrison Predictor is an example of a system that was built to solve laying in "real time", simply by pointing the director at the target and then aiming the gun at a pointer it directed. It was also deliberately designed to be small and light, in order to allow it to be easily moved along with the guns it served.

Simple systems, known as *lead computing sights* also made their appearance inside aircraft late in the war. These devices used a gyroscope to measure turn rates, and moved the gunsight's aim-point to take this into account. The only manual "input" to the sight was the target distance, which was typically handled by dialing in the size of the target's wing span at some known range. Small radar units were added in the post-war period to automate even this input, but it was some time before they were fast enough to make the pilots completely happy with them.

The United States Navy deployed the Mark I Fire Control Computer on many of its vessels constructed during World War II.

Modern fire-control computers, like all high-performance computers, are digital. The added performance allows basically any input to be added, from air density and wind, to wear on the barrels and distortion due to heating. These sorts of effects are noticeable for any sort of gun, and fire-control computers have started appearing on smaller and smaller platforms. Tanks were one early use, automating gun laying using a laser rangefinder and a barrel-distortion meter. Fire-control computers are not just useful for large cannons. They can be used to aim machine guns, small cannons, guided missiles, rifles, grenades, rockets—any kind of weapon which can have its launch or firing parameters varied. They are typically installed on ships, submarines, aircraft, tanks and even on some rifles, for example the Fabrique Nationale F2000. Fire-control computers have gone through all stages of technology that computers have, with some designs being based upon analogue technology and vacuum tubes which were later replaced with transistors.

Fire-control systems are often interfaced with sensors (such as sonar, radar, infra-red search and track, laser range-finders, anemometers, wind vanes, thermometers, etc.) in order to cut down or eliminate the amount of information which has to be manually inputted in order to calculate an effective solution. Sonar, radar, IRST and range-finders can give the system the direction to and/or distance of the target. Alternatively, an optical sight can be provided and an operator can point it at the target, which is easier than having someone input it using other methods and gives the target less warning that it is being tracked. Typically, weapons fired over long ranges need the environmental information—the longer a munition travels, the more the wind, temperature etc. will affect its trajectory, so the more important having accurate information becomes to getting a good solution. Sometimes, for very long-range rockets, environmental data has to be obtained at high altitudes or in between the launching point and the target. Often, satellites or balloons are used to gather this information.

Once the firing solution is calculated, many modern fire-control systems are also able to aim and fire the weapon(s). Once again, this is in the interest of speed and accuracy, and also in the case of a vehicle like an aircraft or tank, in order to allow the pilot/gunner/etc. to perform other actions simultaneously, such as tracking the target or flying the aircraft. Even if the system is unable to aim the weapon itself, for example the fixed cannon on an aircraft, it is able to give the operator cues on how to aim. Typically, the cannon points straight ahead and the pilot must maneuver the aircraft so that it points in the right direction before firing. In most aircraft the aiming cue takes the form of a "pipper" which is projected on the heads-up display (HUD). The pipper shows the pilot where the target must be relative to the aircraft in order to hit it. Once the pilot maneuvers the aircraft so that the target and pipper are superimposed, he or she fires the weapon, or on some aircraft the weapon will fire automatically at this point, in order to overcome the reaction delay of the pilot. In the case of a missile launch, the fire-control computer may give the pilot feedback about whether the target is in range of the missile and how likely the missile is to hit if launched at any particular moment. The pilot will then wait until the probability reading is satisfactorily high before launching the weapon.

Naval fire control

The situation for naval fire control was more complex because of the need to control the firing of several guns at once. In naval engagements both the firing guns and target are moving, and the variables are compounded by the greater distances and times involved. Naval gun fire control potentially involves three levels of complexity. Local control originated with primitive gun installations aimed by the individual gun crews. Director control aims all guns on the ship at a single target. Coordinated gunfire from a formation of ships at a single target was a focus of battleship fleet operations. Corrections are made for surface wind velocity, firing ship roll and pitch, powder magazine temperature, drift of rifled projectiles, individual gun bore diameter adjusted for shot-to-shot enlargement, and rate of change of range with additional modifications to the firing solution based upon the observation of preceding shots.

Rudimentary naval fire control systems were first developed around the time of World War I.

For the UK, their first system was built before the Great War. At the heart was an analogue computer designed by Commander (later Admiral Sir) Frederic Charles Dreyer that calculated rate of change of range. The Dreyer Table was to be improved and served into the interwar period at which point it was superseded in new and reconstructed ships by the Admiralty Fire Control Table.

The use of Director controlled firing together with the fire control computer moved the control of the gun laying from the individual turrets to a central position; although individual gun mounts and multi-gun turrets may retain a local control option for use when battle damage limits director information transfer (these would be simpler versions called "turret tables" in the RN). Guns could then be fired in planned salvos, with each gun giving a slightly different trajectory. Dispersion of shot caused by differences in individual guns, individual projectiles, powder ignition sequences, and transient distortion of ship structure was undesirably large at typical naval engagement ranges. Directors high on the superstructure had a better view of the enemy than a turret mounted sight, and the crew operating them were distant from the sound and shock of the guns. Gun directors were topmost, and the ends of their optical rangefinders protruded from their sides, giving them a distinctive appearance.

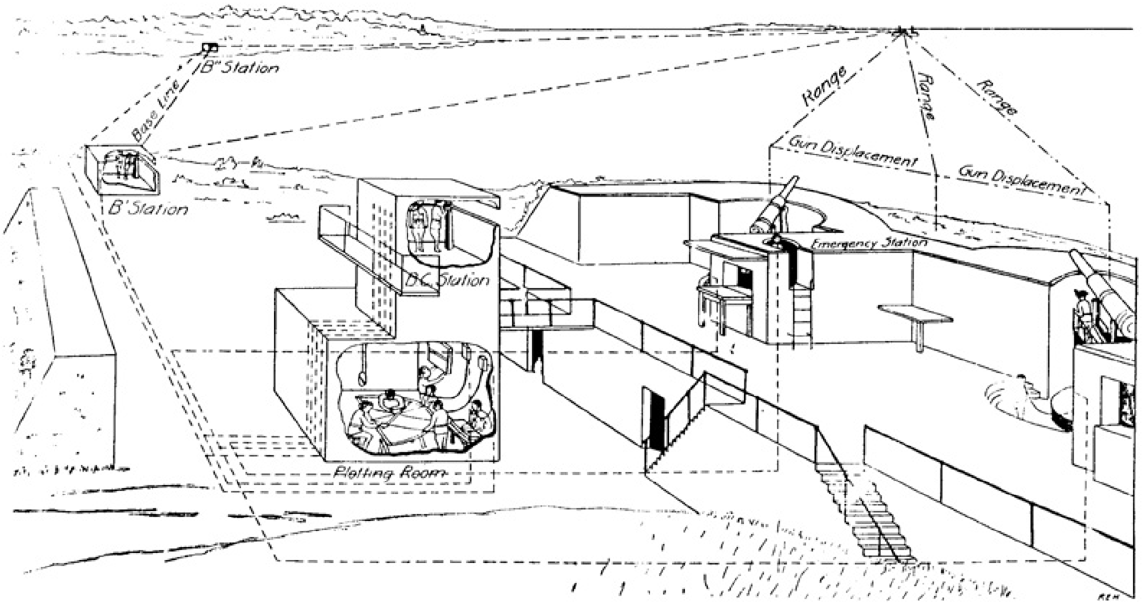
Unmeasured and uncontrollable ballistic factors like high altitude temperature, humidity, barometric pressure, wind direction and velocity required final adjustment through observation of fall of shot. Visual range measurement (of both target and shell splashes) was difficult prior to availability of Radar. The British favoured coincident rangefinders while the Germans favored the stereoscopic type. The former were less able to range on an indistinct target but easier on the operator over a long period of use, the latter the reverse.

In a typical WWII British ship the fire control system connected the individual gun turrets to the director tower (where the sighting instruments were) and the analogue computer in the heart of the ship. In the director tower, operators trained their telescopes on the target; one telescope measured elevation and the other bearing. Rangefinder telescopes on a separate mounting measured the distance to the target. These measurements were converted by the Fire Control Table into bearings and elevations for the guns to fire on. In the turrets, the gunlayers adjusted the elevation of their guns to match an indicator which was the elevation transmitted from the Fire Control table - a turret layer did the same for bearing. When the guns were on target they were centrally fired.

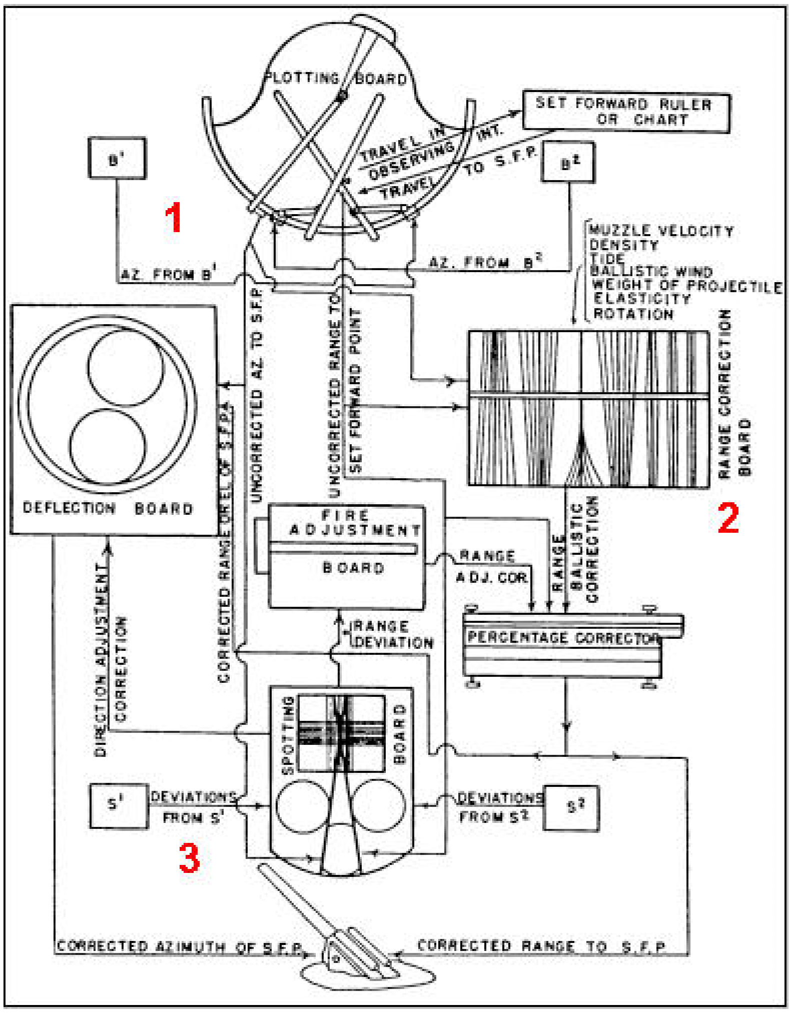
Even with as much mechanization of the process, it still required a large human element; the Transmitting Station (the room that housed the Dreyer table) for HMS *Hood*'s main guns housed 27 crew.

Directors were largely unprotected from enemy fire. It was difficult to put much weight into armour so high up on the ship and even if the armour did stop a shot the impact alone would likely knock the instruments out of alignment. Sufficient armour to protect from smaller shells and fragments from hits to other parts of the ship was the limit.

Coast Artillery Fire Control



A cut-away drawing of a Coast Artillery battery, showing its two base and stations (upper left), the plotting room (with plotting board), and the directing point (between the two guns), with their displacement from it.


PLOTTING BOARD
SET FORWARD RULER OR CHART
TRAVEL IN OBSERVING INT.
TRAVEL TO S.F.P.
B¹
B²
1
AZ. FROM B¹
AZ. FROM B²
MUZZLE VELOCITY
DENSITY
TIDE
BALLISTIC WIND
WEIGHT OF PROJECTILE
ELASTICITY
ROTATION
RANGE CORRECTION BOARD
DEFLECTION BOARD
UNCORRECTED AZ TO S.F.P.
UNCORRECTED RANGE TO SET FORWARD POINT
CORRECTED RANGE OR EL OF S.F.P.A
FIRE ADJUSTMENT BOARD
RANGE ADJ. COR.
RANGE
BALLISTIC CORRECTION
2
PERCENTAGE CORRECTOR
RANGE DEVIATION
DIRECTION ADJUSTMENT CORRECTION
SPOTTING BOARD
S¹
DEVIATIONS FROM S¹
DEVIATIONS FROM S²
S²
3
CORRECTED AZIMUTH OF S.F.P.
CORRECTED RANGE TO S.F.P.

A conceptual diagram of the flow of fire control data in the Coast Artillery (in 1940). The set forward point of the target was generated by using the plotting board (1). This position was then corrected for factors affecting range and azimuth (2). Finally, fire was adjusted for observations of the actual fall of the shells (3), and new firing data were sent to the guns.

In the United States, Coast Artillery fire control systems began to be developed at the end of the 19th Century and progressed on through WW2.

Early systems made use of multiple observation or base end stations to find and track targets attacking American harbors. Data from these stations were then passed to plotting rooms, where analog mechanical devices, such as the plotting board, were used to estimate targets' positions and derive firing data for batteries of coastal guns assigned to interdict them.

U.S. Coast Artillery forts bristled with a variety of armament, ranging from 12-inch coast defense mortars, through 3-inch and 6-inch mid-range artillery, to the larger guns, which included 10-inch and 12-inch barbette and disappearing carriage guns, 14-inch railroad artillery, and 16-inch cannon installed just prior to and up through WW2.

Fire control in the Coast Artillery became more and more sophisticated in terms of correcting firing data for such factors as weather conditions, the condition of powder used, or the Earth's rotation. Provisions were also made for adjusting firing data for the observed fall of shells. As shown in the figure at right, all of these data were fed back to the plotting rooms on a finely tuned schedule controlled by a system of time interval bells that rang throughout each harbor defense system.

It was only later in WW2 that electro-mechanical gun data computers, connected to coast defense radars, began to replace optical observation and manual plotting methods in controlling coast artillery. Even then, the manual methods were retained as a back-up through the end of the war.

Ground use

Once an engagement has begun, it is also possible for a fire control radar to track incoming fire, trace back the trajectories to their source, and produce the coordinates of an enemy unwise enough to fire ballistic rounds. This return-fire capability has been included in some systems since the 1970s. Returning fire to the location of the rounds' origin is known as counter-battery fire.

Aircraft use

World War II bomb sights

An early use of fire-control systems was in bomber aircraft, with the use of computing bombsights that accepted altitude and airspeed information to predict and display the impact point of a bomb released at that time. The best known US device was the Norden bombsight.

World War II aerial gunnery sights

Simple systems, known as *lead computing sights* also made their appearance inside aircraft late in the war. These devices used a gyroscope to measure turn rates, and moved the gunsight's aim-point to take this into account, with the aim point presented through a reflective sight. The only manual "input" to the sight was the target distance, which was typically handled by dialing in the size of the target's wing span at some known range. Small radar units were added in the post-war period to automate even this input, but it was some time before they were fast enough to make the pilots completely happy with them.

Post-World War II systems

By the start of the Vietnam War, a new computerized bombing predictor called the Low-Altitude Bombing System (LABS) began to be integrated into the systems of aircraft equipped to carry nuclear armaments. This new bomb computer was revolutionary in that the actual release command for the bomb was given by the computer, not the pilot; the pilot designated the target using the radar or other targeting system, then "consented" to release the weapon, and the computer then did so at a calculated "release point" some seconds later. This is very different from previous systems which, though they had also become computerized, still calculated an "impact point" showing where the bomb would fall if the bomb were released at that moment. The key advantage is that the weapon can be released accurately even when the plane is making a maneuver such as a climb or dive. Most bombsights until this time required that the plane maintain a constant attitude (usually level, though dive-bombing sights were also common).

The LABS system was originally designed to facilitate a tactic called toss bombing, to allow the aircraft to remain out of range of a weapon's blast radius. The principle of calculating the release point, however, was eventually integrated into the Fire Control Computers of later bombers and strike aircraft, allowing level, dive and toss bombing. In addition, as the fire control computer became integrated with ordinance systems, the computer can take the flight characteristics of the weapon to be launched into account.

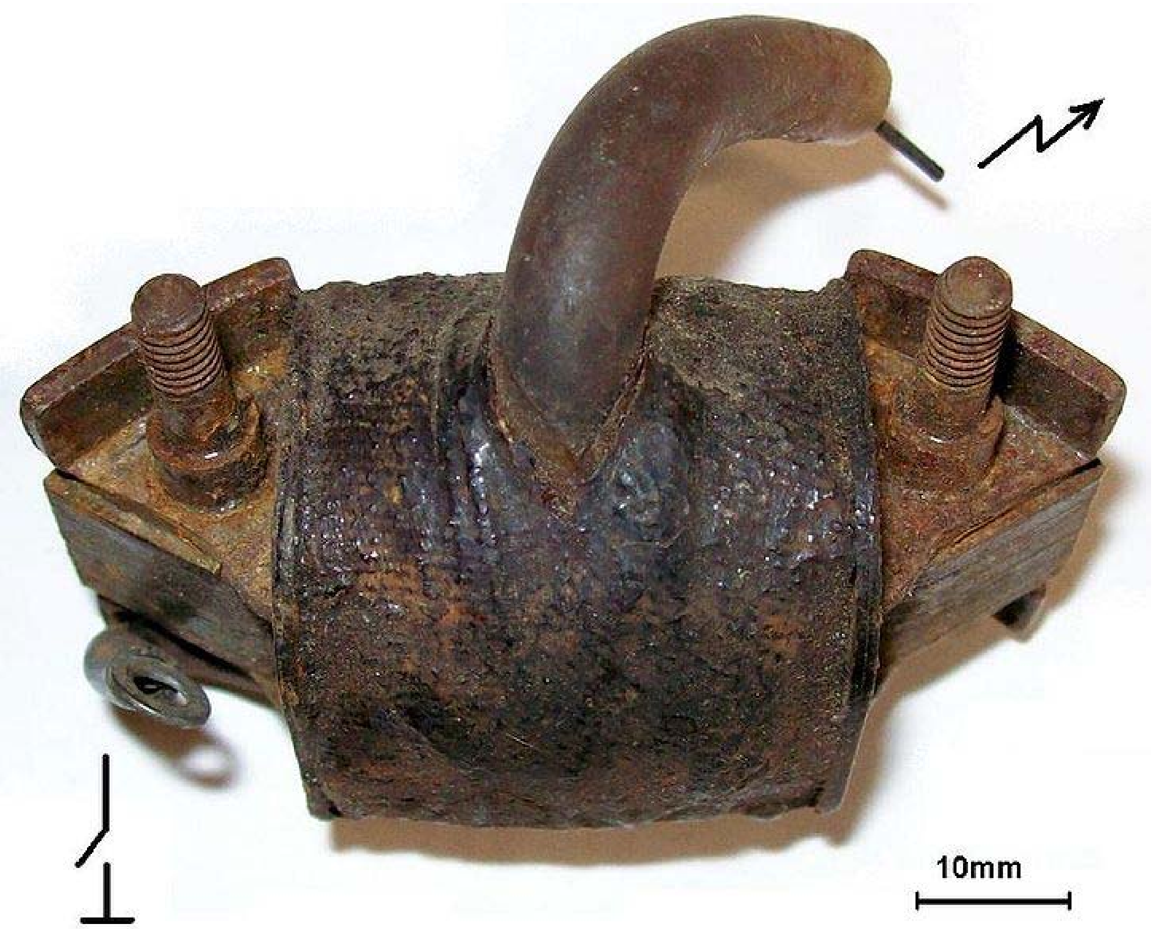
Chapter 5

Ignition System

An **ignition system** is a system for igniting a fuel-air mixture. It is best known in the field of internal combustion engines but also has other applications, e.g. in oil-fired and gas-fired boilers. The earliest internal combustion engines used a flame, or a heated tube, for ignition but these were quickly replaced by systems using an electric spark.

History

Magneto systems



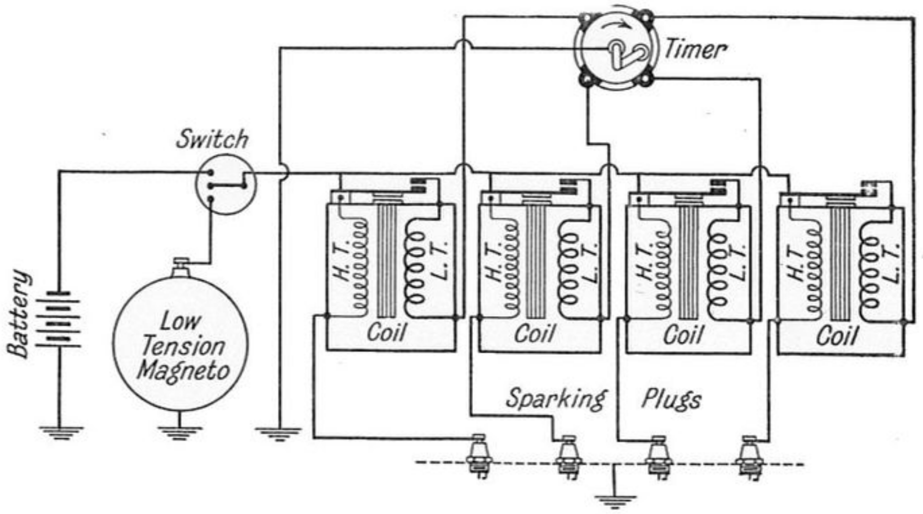
Magneto ignition coil

The simplest form of spark ignition is that using a magnet. The engine spins a magnet inside a coil, or, in the earlier designs, a coil inside a fixed magnet, and also operates a contact breaker, interrupting the current and causing the voltage to be increased sufficiently to jump a small gap. The spark plugs are connected directly from the magneto output. Early magnetos had one coil, with the contact breaker (sparking plug) inside the combustion chamber. In about 1902, Bosch introduced a double-coil magneto, with a fixed sparking plug, and the contact breaker outside the cylinder. Magnetos are not used in modern cars, but because they generate their own electricity they are often found on piston-engined aircraft engines and small engines such as those found in mopeds, lawnmowers, snowblowers, chainsaws, etc. where a battery-based electrical system is not present for any combination of necessity, weight, cost, and reliability reasons.

Magnetos were used on the small engine's ancestor, the stationary "hit or miss" engine which was used in the early twentieth century, on older gasoline or distillate farm tractors before battery starting and lighting became common, and on aircraft piston engines. Magnetos were used in these engines because their simplicity and self-contained operation was more reliable, and because magnetos weighed less than having a battery and dynamo or alternator.

Aircraft engines usually have multiple magnetos to provide redundancy in the event of a failure. Some older automobiles had both a magneto system and a battery actuated system running simultaneously to ensure proper ignition under all conditions with the limited performance each system provided at the time.This gave the benefits of easy starting (from the battery system) with reliable sparking at speed (from the magneto).

Switchable systems



Ford Model T ignition circuit

The output of a magneto depends on the speed of the engine, and therefore starting can be problematic. Some magnetos include an impulse system, which spins the magnet quickly at the proper moment, making easier starting at slow cranking speeds. Some engines, such as aircraft but also the Ford Model T, used a system which relied on non rechargeable dry cells, (similar to a large flashlight battery, and which was not maintained by a charging system as on modern automobiles) to start the engine or for starting and running at low speed. The operator would manually switch the ignition over to magneto operation for high speed operation.

In order to provide high voltage for the spark from the low voltage batteries, a 'tickler' was used, which was essentially a larger version of the once widespread electric buzzer. With this apparatus, the direct current passes through an electromagnetic coil which pulls open a pair of contact points, interrupting the current; the magnetic field collapses, the spring-loaded points close again, the circuit is reestablished, and the cycle repeats rapidly. The rapidly collapsing magnetic field, however, induces a high voltage across the coil which can only relieve itself by arcing across the contact points; while in the case of the buzzer this is a problem as it causes the points to oxidize and/or weld together, in the case of the ignition system this becomes the source of the high voltage to operate the spark plugs.

In this mode of operation, the coil would "buzz" continuously, producing a constant train of sparks. The entire apparatus was known as the 'Model T spark coil' (in contrast to the modern ignition coil which is *only* the actual coil component of the system). Long after the demise of the Model T as transportation they remained a popular self-contained source of high voltage for electrical home experimenters, appearing in articles in magazines such as *Popular Mechanics* and projects for school science fairs as late as the early 1960s. In the UK these devices were commonly known as trembler coils and were popular in cars pre-1910, and also in commercial vehicles with large engines until around 1925 to ease starting.

The Model T (built into the flywheel) differed from modern implementations by not providing high voltage directly at the output; the maximum voltage produced was about 30 volts, and therefore also had to be run through the spark coil to provide high enough voltage for ignition, as described above, although the coil would not "buzz" continuously in this case, only going through one cycle per spark. In either case, the low voltage was switched to the appropriate spark plug by the *'timer'* mounted on the front of the engine. This performed the equivalent function to the modern distributor, although by directing the low voltage, not the high voltage as for the distributor. The timing of the spark was adjustable by rotating this mechanism through a lever mounted on the steering column. As the precise timing of the spark depends on *both* the 'timer' and the trembler contacts within the coil, this is less consistent than the breaker points of the later distributor. However for the low speed and the low compression of such early engines, this imprecise timing was acceptable.

Battery-operated ignition

With the universal adaptation of electrical starting for automobiles, and the concomitant availability of a large battery to provide a constant source of electricity, magneto systems were abandoned for systems which interrupted current at battery voltage, used an ignition coil (a type of autotransformer) to step the voltage up to the needs of the ignition, and a distributor to route the ensuing pulse to the correct spark plug at the correct time.

The first reliable battery operated ignition was developed by the Dayton Engineering Laboratories Co. (Delco) and introduced in the 1910 Cadillac. This ignition was developed by Charles Kettering and was a wonder in its day. It consisted of a single coil, points (the switch), a capacitor and a distributor set up to allocate the spark from the ignition coil timed to the correct cylinder. The coil was basically an autotransformer set up to step up the low (6 or 12 V) voltage supply to the high ignition voltage required to jump a spark plug gap.

The points allow the coil to charge magnetically and then, when they are opened by a cam arrangement, the magnetic field collapses and a large (20 kV or greater) voltage is produced. The capacitor is used to absorb the back EMF from the magnetic field in the coil to minimize point contact burning and maximize point life. The Kettering system became the primary ignition system for many years in the automotive industry due to its lower cost, higher reliability and relative simplicity.

Modern ignition systems

The ignition system is typically controlled by a key operated Ignition switch.

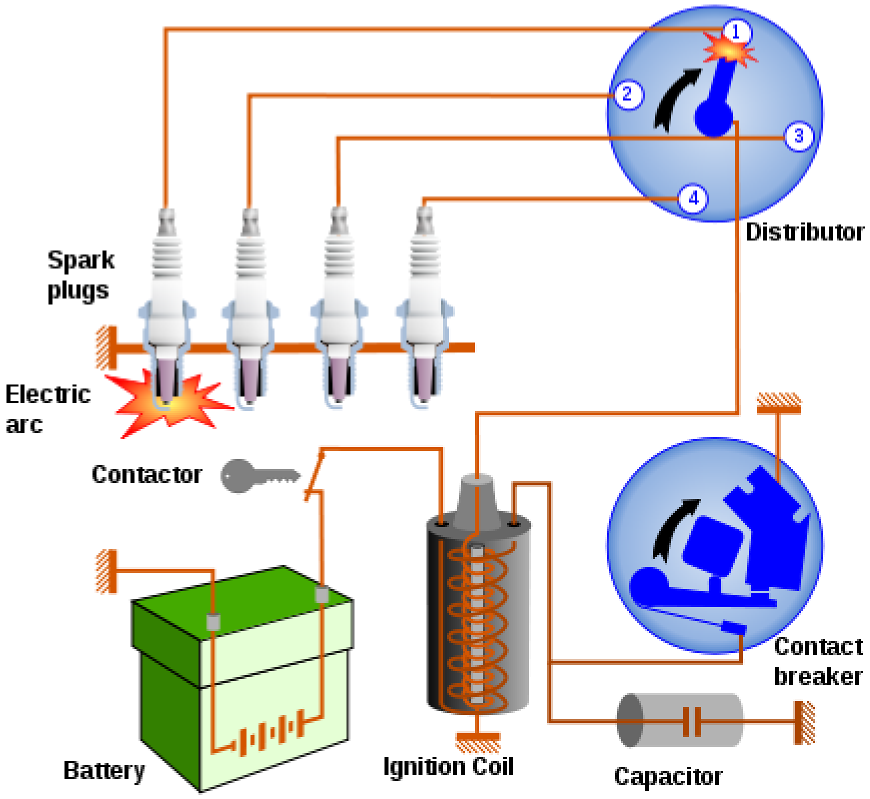
Mechanically timed ignition

Distributor cap

Most four-stroke engines have used a mechanically timed electrical ignition system. The heart of the system is the distributor. The distributor contains a rotating cam driven by the engine's drive, a set of breaker points, a condenser, a rotor and a distributor cap. External to the distributor is the ignition coil, the spark plugs and wires linking the distributor to the spark plugs and ignition coil.

The system is powered by a lead-acid battery, which is charged by the car's electrical system using a dynamo or alternator. The engine operates contact breaker points, which interrupt the current to an induction coil (known as the ignition coil).

The ignition coil consists of two transformer windings sharing a common magnetic core—the primary and secondary windings. An alternating current in the primary induces alternating magnetic field in the coil's core. Because the ignition coil's secondary has far more windings than the primary, the coil is a step-up transformer which induces a much higher voltage across the secondary windings. For an ignition coil, one end of windings of both the primary and secondary are connected together. This common point is connected to the battery (usually through a current-limiting ballast resistor). The other end of the primary is connected to the points within the distributor. The other end of the secondary is connected, via the distributor cap and rotor, to the spark plugs.



Ignition Circuit Diagram - Mechanically Timed Ignition

The ignition firing sequence begins with the points (or contact breaker) closed. A steady charge flows from the battery, through the current-limiting resistor, through the coil primary, across the closed breaker points and finally back to the battery. This steady current produces a magnetic field within the coil's core. This magnetic field forms the energy reservoir that will be used to drive the ignition spark.

As the engine turns, so does the cam inside the distributor. The points ride on the cam so that as the engine turns and reaches the top of the engine's compression cycle, a high point in the cam causes the breaker points to open. This breaks the primary winding's circuit and abruptly stops the current through the breaker points. Without the steady current through the points, the magnetic field generated in the coil immediately and rapidly collapses. This change in the magnetic field induces a high voltage in the coil's secondary windings.

At the same time, current exits the coil's primary winding and begins to charge up the capacitor ("condenser") that lies across the now-open breaker points. This capacitor and the coil's primary windings form an oscillating LC circuit. This LC circuit produces a damped, oscillating current which bounces energy between the capacitor's electric field and the ignition coil's magnetic field. The oscillating current in the coil's primary, which produces an oscillating magnetic field in the coil, extends the high voltage pulse at the output of the secondary windings. This high voltage thus continues beyond the time of the initial field collapse pulse. The oscillation continues until the circuit's energy is consumed.

The ignition coil's secondary windings are connected to the distributor cap. A turning rotor, located on top of the breaker cam within the distributor cap, sequentially connects the coil's secondary windings to one of the several wires leading to each cylinder's spark plug. The extremely high voltage from the coil's secondary -– often higher than 1000 volts—causes a spark to form across the gap of the spark plug. This, in turn, ignites the compressed air-fuel mixture within the engine. It is the creation of this spark which consumes the energy that was stored in the ignition coil's magnetic field.

High performance engines with eight or more cylinders that operate at high r.p.m. (such as those used in motor racing) demand both a higher rate of spark and a higher spark energy than the simple ignition circuit can provide. This problem is overcome by using either of these adaptations:

- **Two complete sets of coils, breakers and condensers** can be provided - one set for each half of the engine, which is typically arranged in V-8 or V-12 configuration. Although the two ignition system halves are electrically independent, they typically share a single distributor which in this case contains two breakers driven by the rotating cam, and a rotor with two isolated conducting planes for the two high voltage inputs.
- A single breaker driven by a cam and a return spring is limited in spark rate by the onset of contact bounce or float at high rpm. This limit can be overcome by substituting for the breaker a **pair of breakers** that are connected electrically in series but spaced on opposite sides of the cam so they are driven out of phase. Each breaker then switches at half the rate of a single breaker and the "dwell" time for current buildup in the coil is maximized since it is shared between the breakers. The Lamborghini V-12 engine has both these adaptations and therefore uses two ignition coils and a single distributor that contains 4 contact breakers.

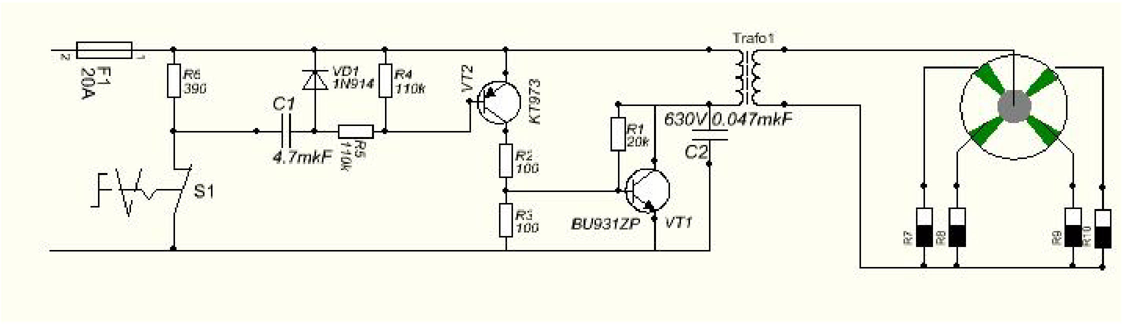
A distributor-based system is not greatly different from a magneto system except that more separate elements are involved. There are also advantages to this arrangement. For example, the position of the contact breaker points relative to the engine angle can be changed a small amount dynamically, allowing the ignition timing to be automatically advanced with increasing revolutions per minute (RPM) and/or increased manifold vacuum, giving better efficiency and performance.

However it is necessary to check periodically the maximum opening gap of the breaker(s), using a feeler gauge, since this mechanical adjustment affects the "dwell" time during which the coil charges, and breakers should be re-dressed or replaced when they have become pitted by electric arcing. This system was used almost universally until the late 1970s, when electronic ignition systems started to appear.

Electronic ignition

The disadvantage of the mechanical system is the use of breaker points to interrupt the low-voltage high-current through the primary winding of the coil; the points are subject to mechanical wear where they ride the cam to open and shut, as well as oxidation and burning at the contact surfaces from the constant sparking. They require regular adjustment to compensate for wear, and the opening of the contact breakers, which is responsible for spark timing, is subject to mechanical variations.

In addition, the spark voltage is also dependent on contact effectiveness, and poor sparking can lead to lower engine efficiency. A mechanical contact breaker system cannot control an average ignition current of more than about 3 A while still giving a reasonable service life, and this may limit the power of the spark and ultimate engine speed.



Example of a basic electronic ignition system

Electronic ignition (EI) solves these problems. In the initial systems, points were still used but they handled only a low current which was used to control the high primary current through a solid state switching system. Soon, however, even these contact breaker points were replaced by an angular sensor of some kind - either optical, where a vaned rotor breaks a light beam, or more commonly using a Hall effect sensor, which responds to a rotating magnet mounted on the distributor shaft. The sensor output is shaped and processed by suitable circuitry, then used to trigger a switching device such as a thyristor, which switches a large current through the coil.

The first electronic ignition (a cold cathode type) was tested in 1948 by Delco-Remy, while Lucas introduced a transistorized ignition in 1955, which was used on BRM and

Coventry Climax Formula One engines in 1962. The aftermarket began offering EI that year, with both the AutoLite Electric Transistor 201 and Tung-Sol EI-4 being available. Pontiac became the first automaker to offer an optional EI, the breakerless magnetic pulse-triggered Delcotronic, on some 1963 models; it was also available on some Corvettes. Ford fitted a Lucas system on the Lotus 25s entered at Indianapolis the next year, ran a fleet test in 1964, and began offering optional EI on some models in 1965. Beginning in 1958, Earl W. Meyer at Chrysler worked on EI, continuing until 1961 and resulting in use of EI on the company's NASCAR hemis in 1963 and 1964.

Prest-O-Lite's CD-65, which relied on capacitance discharge (CD), appeared in 1965, and had "an unprecedented 50,000 mile warranty." (This differs from the non-CD Prest-O-Lite system introduced on AMC products in 1972, and made standard equipment for the 1975 model year.) A similar CD unit was available from Delco in 1966, which was optional on Oldsmobile, Pontiac, and GMC vehicles in the 1967 model year. Also in 1967, Motorola debuted their breakerless CD system.

FIAT became the first company to offer standard EI, in 1968, followed by Chrysler (after a 1971 trial) in 1973 and by Ford and GM in 1975.

In 1967, Prest-O-Lite made a "Black Box" ignition amplifier, intended to take the load off of the distributor's breaker points during high r.p.m. runs, which was used by Dodge and Plymouth on their factory Super Stock Coronet and Belvedere and drag racers. This amp was installed on the interior-side of the cars' firewall, and had a duct which provided outside air to cool the amp. The rest of the system (distributor and spark plugs) remains as for the mechanical system. The lack of moving parts compared with the mechanical system leads to greater reliability and longer service intervals. Chrysler introduced breakerless ignition in mid-1971 as an option for its 340 V8 and the 426 Street Hemi. For the 1972 model year, the system became standard on its high-performance engines (the 340 cu in (5.6 l) and the four-barrel carburetor-equipped 400 hp (298 kW) 400 cu in (7 l)) and was an option on its 318 cu in (5.2 l), 360 cu in (5.9 l), two-barrel 400 cu in (6.6 l), and low-performance 440 cu in (7.2 l) . Breakerless Ignition was standardised across the model range for 1973. For older cars, it is usually possible to retrofit an EI system in place of the mechanical one. In some cases, a modern distributor will fit into the older engine with no other modifications, like the H.E.I. distributor made by General Motors, and the aforementioned Chrysler-built electronic ignition (with an "Orange Box" amplifier and a faster-advance curve distributor).

Other innovations are currently available on various cars. In some models, rather than one central coil, there are individual coils on each spark plug, sometimes known as direct ignition or coil on plug (COP). This allows the coil a longer time to accumulate a charge between sparks, and therefore a higher energy spark. A variation on this has each coil handle two plugs, on cylinders which are 360 degrees out of phase (and therefore reach TDC at the same time); in the four-cycle engine this means that one plug will be sparking during the end of the exhaust stroke while the other fires at the usual time, a so-called "wasted spark" arrangement which has no drawbacks apart from faster spark plug erosion; the paired cylinders are 1/4 and 2/3. Other systems do away with the distributor as a timing apparatus and use a magnetic crank angle sensor mounted on the crankshaft to trigger the ignition at the proper time.

During the 1980s, electronic ignition systems were developed alongside other improvements such as fuel injection systems. After a while it became logical to combine the functions of fuel control and ignition into one electronic system known as an engine control unit. However on older vehicles this was not possible and now a common electronic ignition system for classic cars is the Powerspark electronic ignition.

Digital electronic ignitions

At the turn of the 21st century digital electronic ignition modules became available for small engines on such applications as chainsaws, string trimmers, leaf blowers, and lawn mowers. This was made possible by low cost, high speed, and small footprint microcontrollers. Digital electronic ignition modules can be designed as either capacitor discharge ignition (CDI) or inductive discharge ignition (IDI) systems. Capacitive discharge digital ignitions store charged energy for the spark in a capacitor within the module that can be released to the spark plug at virtually any time throughout the engine cycle via a control signal from the microprocessor. This allows for greater timing flexibility, and engine performance; especially when designed hand-in-hand with the engine carburetor.

Engine management

In an Engine Management System (EMS), electronics control fuel delivery, ignition timing and firing order. Primary sensors on the system are engine angle (crank or Top Dead Center (TDC) position), airflow into the engine and throttle demand position. The circuitry determines which cylinder needs fuel and how much, opens the requisite injector to deliver it, then causes a spark at the right moment to burn it. Early EMS systems used analogue computer circuit designs to accomplish this, but as embedded systems became fast enough to keep up with the changing inputs at high revolutions, digital systems started to appear.

Some designs using EMS retain the original coil, distributor and spark plugs found on cars throughout history. Other systems dispense with the distributor and individual coils mounted directly atop each spark plug. This removes the need for both distributor and high-tension leads, both components with a poor record for long-term reliability.

Modern EMSs read in data from various sensors about the crank position, manifold temperature, manifold pressure (or air mass flow), throttle position, fuel mixture via the O2 sensor and sometimes the unit will read data from knock sensors and exhaust gas temperature sensors. The EMS then uses collected data to precisely determine how much fuel to deliver and when and thus how far to advance the ignition timing. With electronic ignition systems, individual cylinders can have their own individual ignition timing so that timing can be as aggressive as possible per cylinder without fuel detonation. As a result, sophisticated electronic ignition systems can be both more fuel efficient, and produce better performance, over their counterparts.

Turbine and jet engines

Turbine engines have a capacitor discharge ignition system using one or more ignitor plugs, which are only used at startup or in case the combustor(s) flame goes out. Rocket engines have particularly demanding ignitions systems- if prompt ignition does not occur the chamber can fill with excess fuel and oxidiser and significant overpressure can occur (a 'hard start'). Rockets often employ pyrotechnic devices that place flames across the face of the injector plate, or, alternatively, self-ignition chemicals.

Chapter 6

Intelligent Flight Control System

NASA's NF-15B was used for the project.

The **Intelligent Flight Control System** (IFCS) is a next-generation flight control system designed to provide increased safety for the crew and passengers of aircraft as well as to optimize the aircraft performance under normal conditions. The main benefit of this system is that it will allow a pilot to control an aircraft even under failure conditions that would normally cause it to crash. The IFCS is being developed under the direction of the NASA Dryden Flight Research Center with the collaboration of the NASA Ames Research Center, Boeing Phantom Works, the Institute for Scientific Research at West Virginia University, and the Georgia Institute of Technology.

Objectives of IFCS

The main purpose of the IFCS project is to create a system for use in civilian and military aircraft that is both adaptive and fault tolerant. This is accomplished through the use of upgrades to the flight control software that incorporate self-learning neural network technology. The goals of the IFCS neural network project are.

1. To develop a flight control system that can identify aircraft characteristics through the use of neural network technology in order to optimize aircraft performance.
2. To develop a neural network that can train itself to analyze the flight properties of the aircraft.
3. To be able to demonstrate the aforementioned properties on a modified F-15 ACTIVE aircraft during flight, which is the testbed for the IFCS project.

Theory of Operation

The neural network of the IFCS learns flight characteristics in real time through the aircraft's sensors and from error corrections from the primary flight computer, and then uses this information to create different flight characteristic models for the aircraft. The neural network only learns when the aircraft is in a stable flight condition, and will discard any characteristics that would cause the aircraft to go into a failure condition. If the aircraft's condition changes from stable to failure, for example, if one of the control surfaces becomes damaged and unresponsive, the IFCS can detect this fault and switch the flight characteristic model for the aircraft. The neural network then works to drive the error between the reference model and the actual aircraft state to zero.

Project History

Generation 1

Generation 1 IFCS flight tests were conducted in 2003 to test the outputs of the neural network. In this phase, the neural network was pre-trained using flight characteristics obtained for the F-15S/MTD in a wind tunnel test and did not actually provide any control adjustments in flight. The outputs of the neural network were run directly to instrumentation for data collection purposes only.

Generation 2

Generation 2 IFCS tests were conducted in 2005 and used a fully integrated neural network as described in the theory of operation. It is a direct adaptive system that continuously provides error corrections and then measures the effects of these corrections in order to learn new flight models or adjust existing ones. To measure the aircraft state, the neural network takes 31 inputs from the roll, pitch, and yaw axes and the control surfaces. If there is a difference between the aircraft state and model, the neural network adjusts the outputs of the primary flight computer through a dynamic inversion controller to bring the difference to zero before they are sent to the actuator control electronics which move the control surfaces.

Chapter 7

Engine Control Unit

An **engine control unit (ECU)**, also known as **power-train control module (PCM)**, or **engine control module (ECM)** is a type of electronic control unit that determines the amount of fuel, ignition timing and other parameters an internal combustion engine needs to keep running. It does this by reading values from multidimensional performance maps (so called LUTs), using input values (e.g. engine speed) calculated from signals coming from sensor devices monitoring the engine. Before ECU's, air/fuel mixture, ignition timing, and idle speed were directly controlled by mechanical and pneumatic sensors and actuators. One of the very first attempts to use such a unitized and automated "ECU" device to manage multiple engine control functions simultaneously was created by BMW in 1939, for their BMW 801 14-cylinder aviation engine, and known as the *Kommandogerät*, operated only by a single throttle lever.

Working of ECU

Control of fuel mixture

For an engine with fuel injection, an engine control unit (ECU) will determine the quantity of fuel to inject based on a number of parameters. If the throttle pedal is pressed further down, this will open the throttle body and allow more air to be pulled into the engine. The ECU will inject more fuel according to how much air is passing into the engine. If the engine has not warmed up yet, more fuel will be injected (causing the engine to run slightly 'rich' until the engine warms up). Mixture control on computer controlled carburetors works similarly but with a mixture control solenoid or stepper motor incorporated in the float bowl of the carburetor.

Control of ignition timing

A spark ignition engine requires a spark to initiate combustion in the combustion chamber. An ECU can adjust the exact timing of the spark (called ignition timing) to provide better power and economy. If the ECU detects knock, a condition which is potentially destructive to engines, and "judges" it to be the result of the ignition timing being too early in the compression stroke, it will delay (retard) the timing of the spark to prevent this. A second, more common source, cause, of knock/ping is operating the engine in too low of an RPM range for the "work" requirement of the moment. In this case the knock/ping results from the piston not being able to move downward as fast as the flame front is expanding, but this latter mostly applies only to manual transmission equipped vehicles. The ECU controlling an automatic transmission would simply downshift the transmission if this were the cause of knock/ping.

Control of idle speed

Most engine systems have idle speed control built into the ECU. The engine RPM is monitored by the crankshaft position sensor which plays a primary role in the engine timing functions for fuel injection, spark events, and valve timing. Idle speed is controlled by a programmable throttle stop or an idle air bypass control stepper motor. Early carburetor-based systems used a programmable throttle stop using a bidirectional DC motor. Early TBI systems used an idle air control stepper motor. Effective idle speed control must anticipate the engine load at idle. Changes in this idle load may come from HVAC systems, power steering systems, power brake systems, and electrical charging and supply systems. Engine temperature and transmission status, and lift and duration of camshaft also may change the engine load and/or the idle speed value desired.

A full authority throttle control system may be used to control idle speed, provide cruise control functions and top speed limitation.

Control of variable valve timing

Some engines have Variable Valve Timing. In such an engine, the ECU controls the time in the engine cycle at which the valves open. The valves are usually opened sooner at higher speed than at lower speed. This can optimize the flow of air into the cylinder, increasing power and economy.

Electronic valve control

Experimental engines have been made and tested that have no camshaft, but has full electronic control of the intake and exhaust valve opening, valve closing and area of the valve opening. Such engines can be started and run without a starter motor for certain multi-cylinder engines equipped with precision timed electronic ignition and fuel injection. Such a *static-start* engine would provide the efficiency and pollution-reduction improvements of a mild hybrid-electric drive, but without the expense and complexity of an oversized starter motor.

Programmable ECUs

A special category of ECUs are those which are programmable. These units do not have a fixed behavior, but can be reprogrammed by the user.

Programmable ECUs are required where significant aftermarket modifications have been made to a vehicle's engine. Examples include adding or changing of a turbocharger, adding or changing of an intercooler, changing of the exhaust system, and conversion to run on alternative fuel. As a consequence of these changes, the old ECU may not provide appropriate control for the new configuration. In these situations, a programmable ECU can be wired in. These can be programmed/mapped with a laptop connected using a serial or USB cable, while the engine is running.

The programmable ECU may control the amount of fuel to be injected into each cylinder. This varies depending on the engine's RPM and the position of the accelerator pedal (or the manifold air pressure). The engine tuner can adjust this by bringing up a spreadsheet-like page on the laptop where each cell represents an intersection between a specific RPM value and an accelerator pedal position (or the throttle position, as it is called). In this cell a number corresponding to the amount of fuel to be injected is entered. This spreadsheet is often referred to as a fuel table or fuel map.

By modifying these values while monitoring the exhausts using a wide band lambda probe to see if the engine runs rich or lean, the tuner can find the optimal amount of fuel to inject to the engine at every different combination of RPM and throttle position. This process is often carried out at a dynamometer, giving the tuner a controlled environment to work in. An engine dynamometer gives a more precise calibration for racing applications. Tuners often utilize a chassis dynamometer for street and other high performance applications.

Other parameters that are often mappable are:

- **Ignition:** Defines when the spark plug should fire for a cylinder.
- **Rev. limit:** Defines the maximum RPM that the engine is allowed to reach. After this fuel and/or ignition is cut. Some vehicles have a "soft" cut-off before the "hard" cut-off.
- **Water temperature correction:** Allows for additional fuel to be added when the engine is cold (choke) or dangerously hot.
- **Transient fueling:** Tells the ECU to add a specific amount of fuel when throttle is applied. The term is "acceleration enrichment"
- **Low fuel pressure modifier:** Tells the ECU to increase the injector fire time to compensate for a loss of fuel pressure.
- **Closed loop lambda:** Lets the ECU monitor a permanently installed lambda probe and modify the fueling to achieve stoichiometric (ideal) combustion. On traditional petrol powered vehicles this air:fuel ratio is 14.7:1.

Some of the more advanced race ECUs include functionality such as launch control, limiting the power of the engine in first gear to avoid burnouts. Other examples of advanced functions are:

- **Wastegate control:** Sets up the behavior of a turbocharger's wastegate, controlling boost.

- **Banked injection:** Sets up the behavior of double injectors per cylinder, used to get a finer fuel injection control and atomization over a wide RPM range.
- **Variable cam timing:** Tells the ECU how to control variable intake and exhaust cams.
- **Gear control:** Tells the ECU to cut ignition during (sequential gearbox) upshifts or blip the throttle during downshifts.

A race ECU is often equipped with a data logger recording all sensors for later analysis using special software in a PC. This can be useful to track down engine stalls, misfires or other undesired behaviors during a race by downloading the log data and looking for anomalies after the event. The data logger usually has a capacity between 0.5 and 16 megabytes.

In order to communicate with the driver, a race ECU can often be connected to a "data stack", which is a simple dash board presenting the driver with the current RPM, speed and other basic engine data. These race stacks, which are almost always digital, talk to the ECU using one of several proprietary protocols running over RS232 or CANbus, connecting to the DLC connector (Data Link Connector) usually located on the underside of the dash, inline with the steering wheel

History

Hybrid digital designs

Hybrid digital/analog designs were popular in the mid 1980s. This used analog techniques to measure and process input parameters from the engine, then used a look-up table stored in a digital ROM chip to yield precomputed output values. Later systems compute these outputs dynamically. The ROM type of system is amenable to tuning if one knows the system well. The disadvantage of such systems is that the precomputed values are only optimal for an idealised, new engine. As the engine wears, the system is less able to compensate than a CPU based system.

Modern ECUs

Modern ECUs use a microprocessor which can process the inputs from the engine sensors in real time. An electronic control unit contains the hardware and software (firmware). The hardware consists of electronic components on a printed circuit board (PCB), ceramic substrate or a thin laminate substrate. The main component on this circuit board is a microcontroller chip (CPU). The software is stored in the microcontroller or other chips on the PCB, typically in EPROMs or flash memory so the CPU can be re-programmed by uploading updated code or replacing chips. This is also referred to as an (electronic) Engine Management System (EMS).

Sophisticated engine management systems receive inputs from other sources, and control other parts of the engine; for instance, some variable valve timing systems are electronically controlled, and turbocharger wastegates can also be managed. They also may communicate with transmission control units or directly interface electronically-controlled automatic transmissions, traction control systems, and the like. The Controller Area Network or CAN bus automotive network is often used to achieve communication between these devices.

Modern ECUs sometimes include features such as cruise control, transmission control, anti-skid brake control, and anti-theft control, etc.

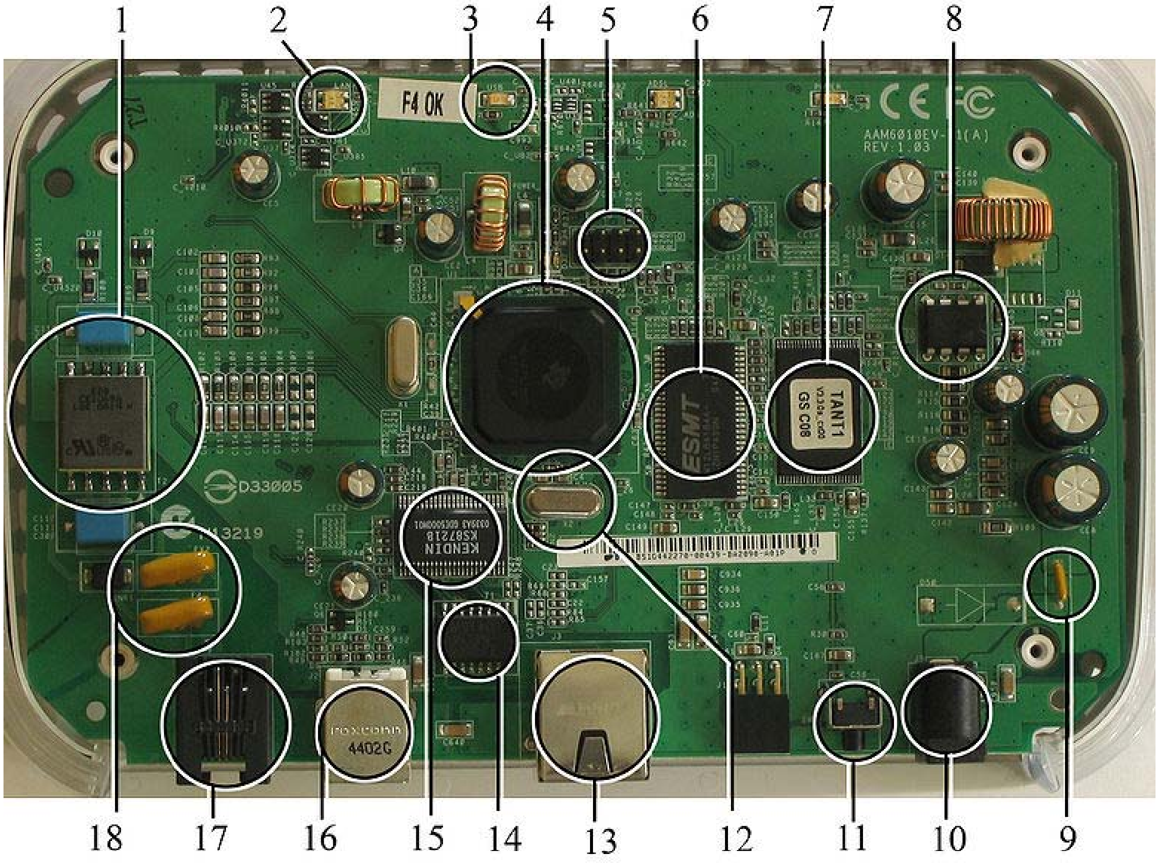
General Motors' first ECUs had a small application of hybrid digital ECUs as a pilot program in 1979, but by 1980, all active programs were using microprocessor based systems. Due to the large ramp up of volume of ECUs that were produced to meet the US Clean Air Act requirements for 1981, only one ECU model could be built for the 1981 model year. The high volume ECU that was installed in GM vehicles from the first high volume year, 1981, onward was a modern microprocessor based system. GM moved rapidly to replace carburetor based systems to fuel injection type systems starting in 1980/1981 Cadillac engines, following in 1982 with the Pontiac 2.5L "GM Iron Duke engine" and the Corvette Chevrolet L83 "Cross-Fire" engine. In just a few years all GM carburetor based engines had been replaced by throttle body injection (TBI) or intake manifold injection systems of various types. In 1988 Delco Electronics, Subsidiary of GM Hughes Electronics, produced more than 28,000 ECUs per day, the world's largest producer of on-board digital control computers at the time.

Other applications

Such systems are used for many internal combustion engines in other applications. In aeronautical applications, the systems are known as "FADECs" (Full Authority Digital Engine Controls). This kind of electronic control is less common in piston-engined aeroplanes than in automobiles, because of the large costs of certifying parts for aviation use, relatively small demand, and the consequent stagnation of technological innovation in this market. Also, a carbureted engine with magneto ignition and a gravity feed fuel system does not require electrical power generated by an alternator to run, which is considered a safety advantage.

Chapter 8

Embedded System



Picture of the internals of an ADSL modem/router. A modern example of an embedded system. Labelled parts include a microprocessor (4), RAM (6), and flash memory (7).

An **embedded system** is a computer system designed to perform one or a few dedicated functions often with real-time computing constraints. It is *embedded* as part of a complete device often including hardware and mechanical parts. By contrast, a general-purpose computer, such as a personal computer (PC), is designed to be flexible and to meet a wide range of end-user needs. Embedded systems control many devices in common use today.

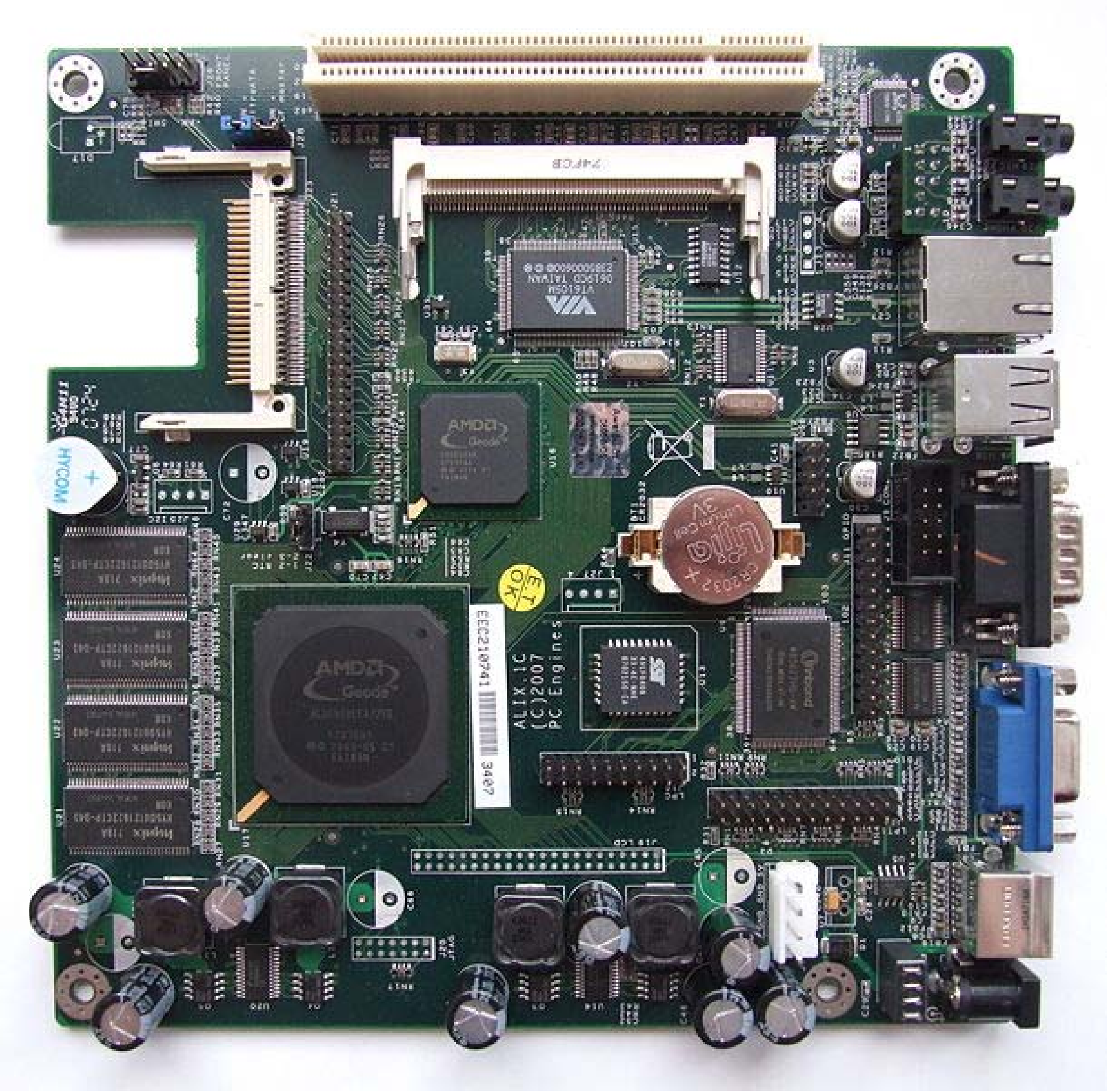
Embedded systems are controlled by one or more main processing cores that are typically either microcontrollers or digital signal processors (DSP). The key characteristic, however, is being dedicated to handle a particular task, which may require very powerful processors. For example, air traffic control systems may usefully be viewed as embedded, even though they involve mainframe computers and dedicated regional and national networks between airports and radar sites (each radar probably includes one or more embedded systems of its own).

Since the embedded system is dedicated to specific tasks, design engineers can optimize it to reduce the size and cost of the product and increase the reliability and performance. Some embedded systems are mass-produced, benefiting from economies of scale.

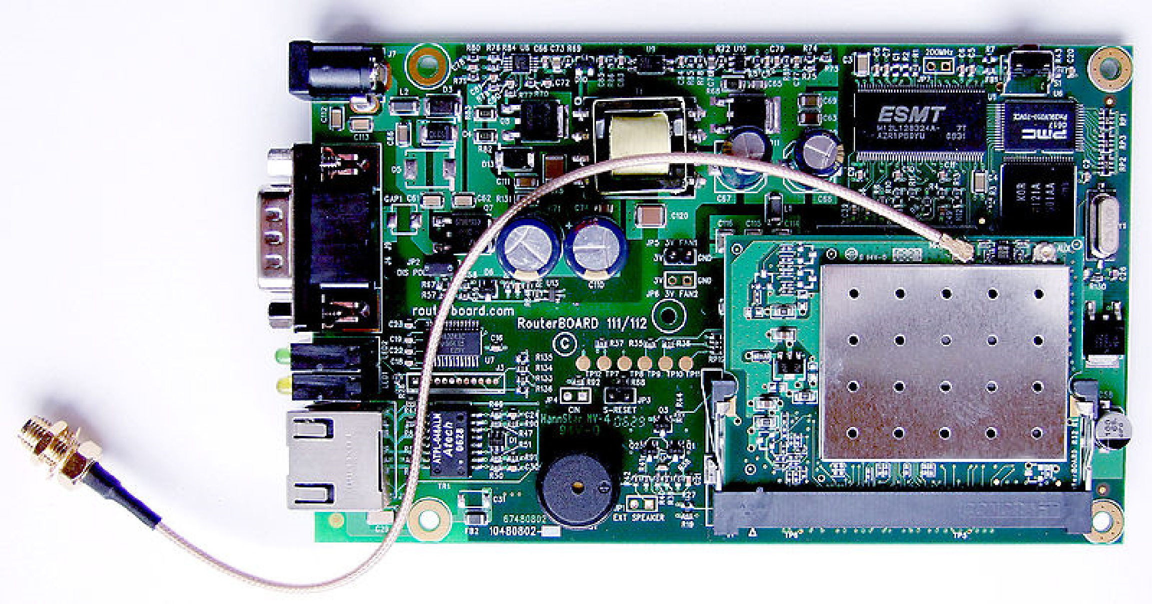
Physically, embedded systems range from portable devices such as digital watches and MP3 players, to large stationary installations like traffic lights, factory controllers, or the systems controlling nuclear power plants. Complexity varies from low, with a single microcontroller chip, to very high with multiple units, peripherals and networks mounted inside a large chassis or enclosure.

In general, "embedded system" is not a strictly definable term, as most systems have some element of extensibility or programmability. For example, handheld computers share some elements with embedded systems such as the operating systems and microprocessors which power them, but they allow different applications to be loaded and peripherals to be connected. Moreover, even systems which do not expose programmability as a primary feature generally need to support software updates. On a continuum from "general purpose" to "embedded", large application systems will have subcomponents at most points even if the system as a whole is "designed to perform one or a few dedicated functions", and is thus appropriate to call "embedded".

Variety of embedded systems



PC Engines' ALIX.1C Mini-ITX embedded board with an x86 AMD Geode LX 800 together with Compact Flash, miniPCI and PCI slots, 44-pin IDE interface, audio, USB and 256MB RAM



An embedded RouterBoard 112 with U.FL-RSMA pigtail and R52 miniPCI Wi-Fi card widely used by wireless Internet service providers (WISPs) in the Czech Republic.

Embedded systems span all aspects of modern life and there are many examples of their use.

Telecommunications systems employ numerous embedded systems from telephone switches for the network to mobile phones at the end-user. Computer networking uses dedicated routers and network bridges to route data.

Consumer electronics include personal digital assistants (PDAs), mp3 players, mobile phones, videogame consoles, digital cameras, DVD players, GPS receivers, and printers. Many household appliances, such as microwave ovens, washing machines and dishwashers, are including embedded systems to provide flexibility, efficiency and features. Advanced HVAC systems use networked thermostats to more accurately and efficiently control temperature that can change by time of day and season. Home automation uses wired- and wireless-networking that can be used to control lights, climate, security, audio/visual, surveillance, etc., all of which use embedded devices for sensing and controlling.

Transportation systems from flight to automobiles increasingly use embedded systems. New airplanes contain advanced avionics such as inertial guidance systems and GPS receivers that also have considerable safety requirements. Various electric motors — brushless DC motors, induction motors and DC motors — are using electric/electronic motor controllers. Automobiles, electric vehicles, and hybrid vehicles are increasingly using embedded systems to maximize efficiency and reduce pollution. Other automotive safety systems include anti-lock braking system (ABS), Electronic Stability Control (ESC/ESP), traction control (TCS) and automatic four-wheel drive.

Medical equipment is continuing to advance with more embedded systems for vital signs monitoring, electronic stethoscopes for amplifying sounds, and various medical imaging (PET, SPECT, CT, MRI) for non-invasive internal inspections.

Embedded systems are especially suited for use in transportation, fire safety, safety and security, medical applications and life critical systems as these systems can be isolated from hacking and thus be more reliable. For fire safety, the systems can be designed to be have greater ability to handle higher temperatures and continue to operate. In dealing with security, the embedded systems can be self sufficient and be able to deal with cut electrical and communication systems.

In addition to commonly described embedded systems based on small computers, a new class of miniature wireless devices called motes are quickly gaining popularity as the field of wireless sensor networking rises. Wireless sensor networking, WSN, makes use of miniaturization made possible by advanced IC design to couple full wireless subsystems to sophisticated sensors, enabling people and companies to measure a myriad of things in the physical world and act on this information through IT monitoring and control systems. These motes are completely self contained, and will typically run off a battery source for many years before the batteries need to be changed or charged.

History

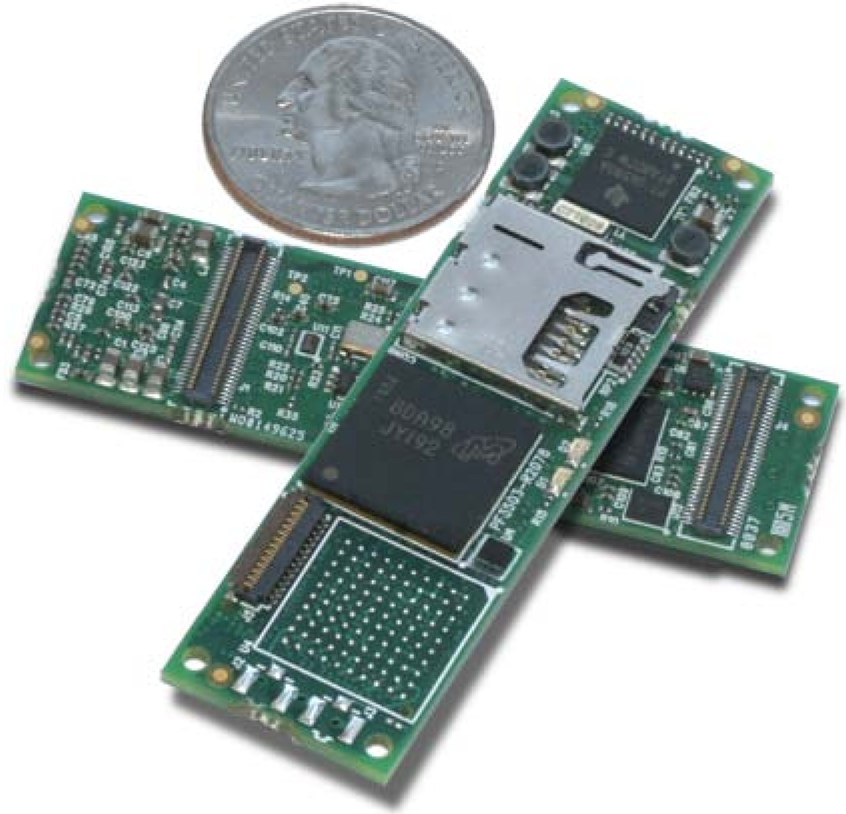
One of the first recognizably modern embedded systems was the Apollo Guidance Computer, developed by Charles Stark Draper at the MIT Instrumentation Laboratory. At the project's inception, the Apollo guidance computer was considered the riskiest item in the Apollo project as it employed the then newly developed monolithic integrated circuits to reduce the size and weight. An early mass-produced embedded system was the Autonetics D-17 guidance computer for the Minuteman missile, released in 1961. It was built from transistor logic and had a hard disk for main memory. When the Minuteman II went into production in 1966, the D-17 was replaced with a new computer that was the first high-volume use of integrated circuits. This program alone reduced prices on quad nand gate ICs from $1000/each to $3/each, permitting their use in commercial products.

Since these early applications in the 1960s, embedded systems have come down in price and there has been a dramatic rise in processing power and functionality. The first microprocessor for example, the Intel 4004, was designed for calculators and other small systems but still required many external memory and support chips. In 1978 National Engineering Manufacturers Association released a "standard" for programmable microcontrollers, including almost any computer-based controllers, such as single board computers, numerical, and event-based controllers.

As the cost of microprocessors and microcontrollers fell it became feasible to replace expensive knob-based analog components such as potentiometers and variable capacitors with up/down buttons or knobs read out by a microprocessor even in some consumer products. By the mid-1980s, most of the common previously external system components had been integrated into the same chip as the processor and this modern form of the microcontroller allowed an even more widespread use, which by the end of the decade were the norm rather than the exception for almost all electronics devices.

The integration of microcontrollers has further increased the applications for which embedded systems are used into areas where traditionally a computer would not have been considered. A general purpose and comparatively low-cost microcontroller may often be programmed to fulfill the same role as a large number of separate components. Although in this context an embedded system is usually more complex than a traditional solution, most of the complexity is contained within the microcontroller itself. Very few additional components may be needed and most of the design effort is in the software. The intangible nature of software makes it much easier to prototype and test new revisions compared with the design and construction of a new circuit not using an embedded processor.

Characteristics



Gumstix Overo COM, a tiny, OMAP-based embedded computer-on-module with Wifi and Bluetooth.

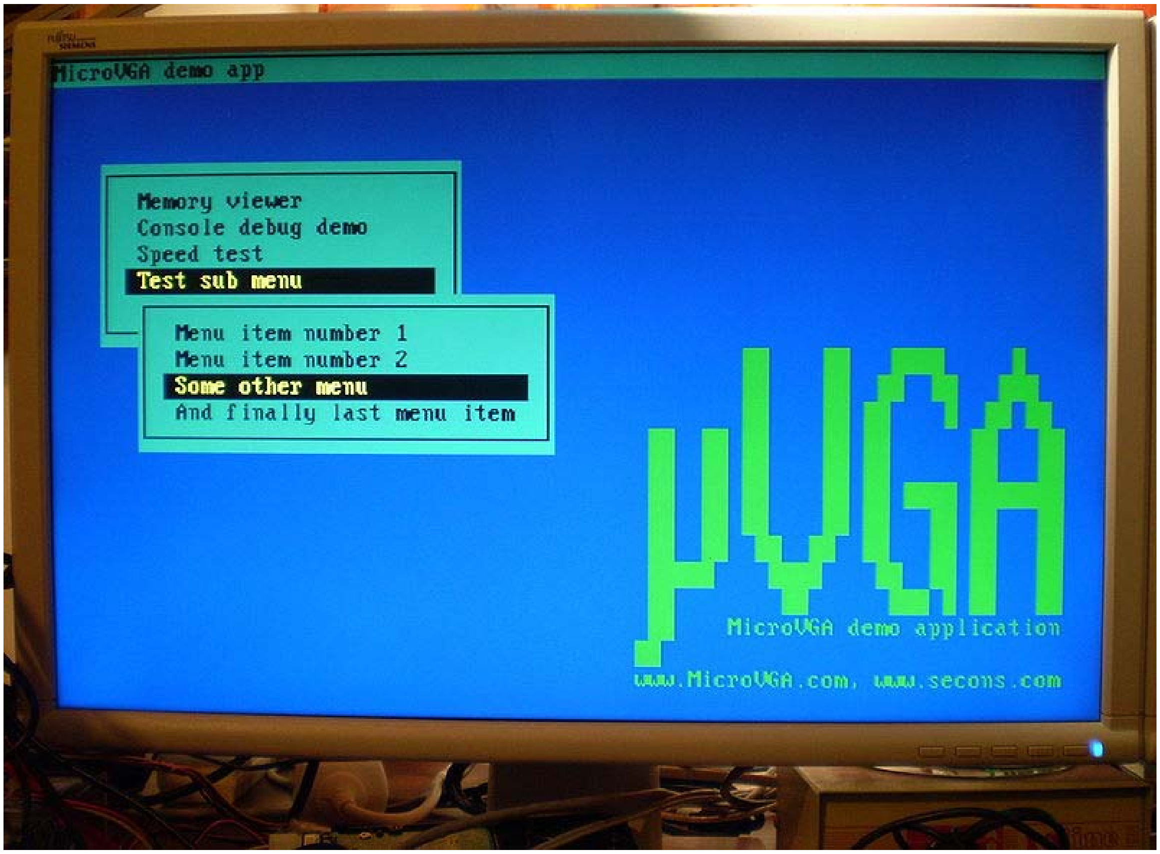
1. Embedded systems are designed to do some specific task, rather than be a general-purpose computer for multiple tasks. Some also have real-time performance constraints that must be met, for reasons such as safety and usability; others may have low or no performance requirements, allowing the system hardware to be simplified to reduce costs.

2. Embedded systems are not always standalone devices. Many embedded systems consist of small, computerized parts within a larger device that serves a more general purpose. For example, the Gibson Robot Guitar features an embedded system for tuning the strings, but the overall purpose of the Robot Guitar is, of course, to play music. Similarly, an embedded system in an automobile provides a specific function as a subsystem of the car itself.

e-con Systems eSOM270 & eSOM300 Computer on Modules

3. The program instructions written for embedded systems are referred to as firmware, and are stored in read-only memory or Flash memory chips. They run with limited computer hardware resources: little memory, small or non-existent keyboard and/or screen.

User interface


MicroVGA demo app
Memory viewer
Console debug demo
Speed test
Test sub menu
Menu item number 1
Menu item number 2
Some other menu
And finally last menu item
µVGA
MicroVGA demo application
www.MicroVGA.com, www.secons.com

Embedded system text user interface using MicroVGA

Embedded systems range from no user interface at all — dedicated only to one task — to complex graphical user interfaces that resemble modern computer desktop operating systems. Simple embedded devices use buttons, LEDs, graphic or character LCDs (for example popular HD44780 LCD) with a simple menu system.

More sophisticated devices use graphical screen with touch sensing or screen-edge buttons provide flexibility while minimizing space used: the meaning of the buttons can change with the screen, and selection involves the natural behavior of pointing at what's desired. Handheld systems often have a screen with a "joystick button" for a pointing device.

Some systems provide user interface remotely with the help of a serial (e.g. RS-232, USB, I²C, etc.) or network (e.g. Ethernet) connection. In spite of the potentially necessary proprietary client software and/or specialist cables that are needed, this approach usually gives a lot of advantages: extends the capabilities of embedded system, avoids the cost of a display, simplifies BSP, allows to build rich user interface on the PC. A good example of this is the combination of an embedded web server running on an embedded device (such as an IP camera) or a network routers. The user interface is displayed in a web browser on a PC connected to the device, therefore needing no bespoke software to be installed.

Processors in embedded systems

Secondly, Embedded processors can be broken into two broad categories: ordinary microprocessors (μP) and microcontrollers (μC), which have many more peripherals on chip, reducing cost and size. Contrasting to the personal computer and server markets, a fairly large number of basic CPU architectures are used; there are Von Neumann as well as various degrees of Harvard architectures, RISC as well as non-RISC and VLIW; word lengths vary from 4-bit to 64-bits and beyond (mainly in DSP processors) although the most typical remain 8/16-bit. Most architectures come in a large number of different variants and shapes, many of which are also manufactured by several different companies.

A long but still not exhaustive list of common architectures are: 65816, 65C02, 68HC08, 68HC11, 68k, 78K0R/78K0, 8051, ARM, AVR, AVR32, Blackfin, C167, Coldfire, COP8, Cortus APS3, eZ8, eZ80, FR-V, H8, HT48, M16C, M32C, MIPS, MSP430, PIC, PowerPC, R8C, RL78, SHARC, SPARC, ST6, SuperH, TLCS-47, TLCS-870, TLCS-900, Tricore, V850, x86, XE8000, Z80, AsAP etc.

Ready made computer boards

PC/104 and PC/104+ are examples of standards for *ready made* computer boards intended for small, low-volume embedded and ruggedized systems, mostly x86-based. These are often physically small compared to a standard PC, although still quite large compared to most simple (8/16-bit) embedded systems. They often use MSDOS, Linux, NetBSD, or an embedded real-time operating system such as MicroC/OS-II, QNX or VxWorks. Sometimes these boards use non-x86 processors.

In certain applications, where small size or power efficiency are not primary concerns, the components used may be compatible with those used in general purpose x86 personal computers. Boards such as the VIA EPIA range help to bridge the gap by being PC-compatible but highly integrated, physically smaller or have other attributes making them attractive to embedded engineers. The advantage of this approach is that low-cost commodity components may be used along with the same software development tools used for general software development. Systems built in this way are still regarded as embedded since they are integrated into larger devices and fulfill a single role. Examples of devices that may adopt this approach are ATMs and arcade machines, which contain code specific to the application.

However, most ready-made embedded systems boards are not PC-centered and do not use the ISA or PCI busses. When a System-on-a-chip processor is involved, there may be little benefit to having a standarized bus connecting discrete components, and the environment for both hardware and software tools may be very different.

One common design style uses a small system module, perhaps the size of a business card, holding high density BGA chips such as an ARM-based System-on-a-chip processor and peripherals, external flash memory for storage, and DRAM for runtime memory. The module vendor will usually provide boot software and make sure there is a selection of operating systems, usually including Linux and some real time choices. These modules can be manufactured in high volume, by organizations familiar with their specialized testing issues, and combined with much lower volume custom mainboards with application-specific external peripherals. Gumstix product lines are a Linux-centric example of this model.

ASIC and FPGA solutions

A common array of n configuration for very-high-volume embedded systems is the system on a chip (SoC) which contains a complete system consisting of multiple processors, multipliers, caches and interfaces on a single chip. SoCs can be implemented as an application-specific integrated circuit (ASIC) or using a field-programmable gate array (FPGA).

Peripherals

Embedded Systems talk with the outside world via peripherals, such as:

- Serial Communication Interfaces (SCI): RS-232, RS-422, RS-485 etc.
- Synchronous Serial Communication Interface: I2C, SPI, SSC and ESSI (Enhanced Synchronous Serial Interface)
- Universal Serial Bus (USB)
- Multi Media Cards (SD Cards, Compact Flash etc.)
- Networks: Ethernet, LonWorks, etc.
- Fieldbuses: CAN-Bus, LIN-Bus, PROFIBUS, etc.
- Timers: PLL(s), Capture/Compare and Time Processing Units
- Discrete IO: aka General Purpose Input/Output (GPIO)
- Analog to Digital/Digital to Analog (ADC/DAC)
- Debugging: JTAG, ISP, ICSP, BDM Port, BITP, and DP9 ports.

Tools

As with other software, embedded system designers use compilers, assemblers, and debuggers to develop embedded system software. However, they may also use some more specific tools:

- In circuit debuggers or emulators.
- Utilities to add a checksum or CRC to a program, so the embedded system can check if the program is valid.
- For systems using digital signal processing, developers may use a math workbench such as Scilab / Scicos, MATLAB / Simulink, EICASLAB, MathCad, Mathematica,or FlowStone DSP to simulate the mathematics. They might also use

libraries for both the host and target which eliminates developing DSP routines as done in DSPnano RTOS and Unison Operating System.
- Custom compilers and linkers may be used to improve optimisation for the particular hardware.
- An embedded system may have its own special language or design tool, or add enhancements to an existing language such as Forth or Basic.
- Another alternative is to add a real-time operating system or embedded operating system, which may have DSP capabilities like DSPnano RTOS.
- Modeling and code generating tools often based on state machines

Software tools can come from several sources:

- Software companies that specialize in the embedded market
- Ported from the GNU software development tools
- Sometimes, development tools for a personal computer can be used if the embedded processor is a close relative to a common PC processor

As the complexity of embedded systems grows, higher level tools and operating systems are migrating into machinery where it makes sense. For example, cellphones, personal digital assistants and other consumer computers often need significant software that is purchased or provided by a person other than the manufacturer of the electronics. In these systems, an open programming environment such as Linux, NetBSD, OSGi or Embedded Java is required so that the third-party software provider can sell to a large market.

Debugging

Embedded debugging may be performed at different levels, depending on the facilities available. From simplest to most sophisticated they can be roughly grouped into the following areas:

- Interactive resident debugging, using the simple shell provided by the embedded operating system (e.g. Forth and Basic)
- External debugging using logging or serial port output to trace operation using either a monitor in flash or using a debug server like the Remedy Debugger which even works for heterogeneous multicore systems.
- An in-circuit debugger (ICD), a hardware device that connects to the microprocessor via a JTAG or Nexus interface. This allows the operation of the microprocessor to be controlled externally, but is typically restricted to specific debugging capabilities in the processor.
- An in-circuit emulator (ICE) replaces the microprocessor with a simulated equivalent, providing full control over all aspects of the microprocessor.
- A complete emulator provides a simulation of all aspects of the hardware, allowing all of it to be controlled and modified, and allowing debugging on a normal PC.

Unless restricted to external debugging, the programmer can typically load and run software through the tools, view the code running in the processor, and start or stop its operation. The view of the code may be as HLL source-code, assembly code or mixture of both.

Because an embedded system is often composed of a wide variety of elements, the debugging strategy may vary. For instance, debugging a software- (and microprocessor-) centric embedded system is different from debugging an embedded system where most of the processing is performed by peripherals (DSP, FPGA, co-processor). An increasing number of embedded systems today use more than one single processor core. A common problem with multi-core development is the proper synchronization of software execution. In such a case, the embedded system design may wish to check the data traffic on the busses between the processor cores, which requires very low-level debugging, at signal/bus level, with a logic analyzer, for instance.

Reliability

Embedded systems often reside in machines that are expected to run continuously for years without errors, and in some cases recover by themselves if an error occurs. Therefore the software is usually developed and tested more carefully than that for personal computers, and unreliable mechanical moving parts such as disk drives, switches or buttons are avoided.

Specific reliability issues may include:

1. The system cannot safely be shut down for repair, or it is too inaccessible to repair. Examples include space systems, undersea cables, navigational beacons, bore-hole systems, and automobiles.
2. The system must be kept running for safety reasons. "Limp modes" are less tolerable. Often backups are selected by an operator. Examples include aircraft navigation, reactor control systems, safety-critical chemical factory controls, train signals, engines on single-engine aircraft.
3. The system will lose large amounts of money when shut down: Telephone switches, factory controls, bridge and elevator controls, funds transfer and market making, automated sales and service.

A variety of techniques are used, sometimes in combination, to recover from errors—both software bugs such as memory leaks, and also soft errors in the hardware:

- watchdog timer that resets the computer unless the software periodically notifies the watchdog
- subsystems with redundant spares that can be switched over to
- software "limp modes" that provide partial function
- Designing with a Trusted Computing Base (TCB) architecture ensures a highly secure & reliable system environment

- An Embedded Hypervisor is able to provide secure encapsulation for any subsystem component, so that a compromised software component cannot interfere with other subsystems, or privileged-level system software. This encapsulation keeps faults from propagating from one subsystem to another, improving reliability. This may also allow a subsystem to be automatically shut down and restarted on fault detection.
- Immunity Aware Programming

High vs low volume

For high volume systems such as portable music players or mobile phones, minimizing cost is usually the primary design consideration. Engineers typically select hardware that is just “good enough” to implement the necessary functions.

For low-volume or prototype embedded systems, general purpose computers may be adapted by limiting the programs or by replacing the operating system with a real-time operating system.

Embedded software architectures

There are several different types of software architecture in common use.

Simple control loop

In this design, the software simply has a loop. The loop calls subroutines, each of which manages a part of the hardware or software.

Interrupt controlled system

Some embedded systems are predominantly interrupt controlled. This means that tasks performed by the system are triggered by different kinds of events. An interrupt could be generated for example by a timer in a predefined frequency, or by a serial port controller receiving a byte.

These kinds of systems are used if event handlers need low latency and the event handlers are short and simple.

Usually these kinds of systems run a simple task in a main loop also, but this task is not very sensitive to unexpected delays.

Sometimes the interrupt handler will add longer tasks to a queue structure. Later, after the interrupt handler has finished, these tasks are executed by the main loop. This method brings the system close to a multitasking kernel with discrete processes.

Cooperative multitasking

A nonpreemptive multitasking system is very similar to the simple control loop scheme, except that the loop is hidden in an API. The programmer defines a series of tasks, and each task gets its own environment to “run” in. When a task is idle, it calls an idle routine, usually called “pause”, “wait”, “yield”, “nop” (stands for *no operation*), etc.

The advantages and disadvantages are very similar to the control loop, except that adding new software is easier, by simply writing a new task, or adding to the queue-interpreter.

Preemptive multitasking or multi-threading

In this type of system, a low-level piece of code switches between tasks or threads based on a timer (connected to an interrupt). This is the level at which the system is generally considered to have an "operating system" kernel. Depending on how much functionality is required, it introduces more or less of the complexities of managing multiple tasks running conceptually in parallel.

As any code can potentially damage the data of another task (except in larger systems using an MMU) programs must be carefully designed and tested, and access to shared data must be controlled by some synchronization strategy, such as message queues, semaphores or a non-blocking synchronization scheme.

Because of these complexities, it is common for organizations to use a real-time operating system (RTOS), allowing the application programmers to concentrate on device functionality rather than operating system services, at least for large systems; smaller systems often cannot afford the overhead associated with a *generic* real time system, due to limitations regarding memory size, performance, and/or battery life.The choice that a RTOS is required brings in its own issues however as the selection must be done prior to starting to the application development process. This timing forces developers to choose the embedded operating system for their device based upon current requirements and so restricts future options to a large extent. The restriction of future options becomes more of an issue as product life decreases. Additionally the level of complexity is continuously growing as devices are required to manage many variables such as serial, USB, TCP/IP, Bluetooth, Wireless LAN, trunk radio, multiple channels, data and voice, enhanced graphics, multiple states, multiple threads, numerous wait states and so on. These trends are leading to the uptake of embedded middleware in addition to a real time operating system.

Microkernels and exokernels

A microkernel is a logical step up from a real-time OS. The usual arrangement is that the operating system kernel allocates memory and switches the CPU to different threads of execution. User mode processes implement major functions such as file systems, network interfaces, etc.

In general, microkernels succeed when the task switching and intertask communication is fast, and fail when they are slow.

Exokernels communicate efficiently by normal subroutine calls. The hardware, and all the software in the system are available to, and extensible by application programmers.

Monolithic kernels

In this case, a relatively large kernel with sophisticated capabilities is adapted to suit an embedded environment. This gives programmers an environment similar to a desktop operating system like Linux or Microsoft Windows, and is therefore very productive for development; on the downside, it requires considerably more hardware resources, is often more expensive, and because of the complexity of these kernels can be less predictable and reliable.

Common examples of embedded monolithic kernels are Embedded Linux and Windows CE.

Despite the increased cost in hardware, this type of embedded system is increasing in popularity, especially on the more powerful embedded devices such as Wireless Routers and GPS Navigation Systems. Here are some of the reasons:

- Ports to common embedded chip sets are available.
- They permit re-use of publicly available code for Device Drivers, Web Servers, Firewalls, and other code.
- Development systems can start out with broad feature-sets, and then the distribution can be configured to exclude unneeded functionality, and save the expense of the memory that it would consume.
- Many engineers believe that running application code in user mode is more reliable, easier to debug and that therefore the development process is easier and the code more portable.
- Many embedded systems lack the tight real time requirements of a control system. Although a system such as Embedded Linux may be fast enough in order to respond to many other applications.
- Features requiring faster response than can be guaranteed can often be placed in hardware.
- Many RTOS systems have a per-unit cost. When used on a product that is or will become a commodity, that cost is significant.

Exotic custom operating systems

A small fraction of embedded systems require safe, timely, reliable or efficient behavior unobtainable with the one of the above architectures. In this case an organization builds a system to suit. In some cases, the system may be partitioned into a "mechanism controller" using special techniques, and a "display controller" with a conventional operating system. A communication system passes data between the two.

Additional software components

In addition to the core operating system, many embedded systems have additional upper-layer software components. These components consist of networking protocol stacks like CAN, TCP/IP, FTP, HTTP, and HTTPS, and also included storage capabilities like FAT and flash memory management systems. If the embedded devices has audio and video capabilities, then the appropriate drivers and codecs will be present in the system. In the case of the monolithic kernels, many of these software layers are included. In the RTOS category, the availability of the additional software components depends upon the commercial offering.

Chapter 9

Electronic Stability Control

Electronic stability control (**ESC**) is a computerized technology that improves safety of a vehicle's stability by detecting and minimizing skids. When ESC detects loss of steering control, it automatically applies the brakes to help "steer" the vehicle where the driver intends to go. Braking is automatically applied to wheels individually, such as the outer front wheel to counter oversteer or the inner rear wheel to counter understeer. Some ESC systems also reduce engine power until control is regained. ESC does not improve a vehicle's cornering performance; instead, it helps to minimize the loss of control. According to IIHS and NHTSA, one-third of fatal accidents could have been prevented by the technology.

History

In 1987, the earliest innovators of ESC, Mercedes-Benz and BMW, introduced their first traction control systems. Traction control works by applying individual wheel braking and throttle to keep traction while accelerating but, unlike the ESC, it is not designed to aid in steering.

Named simply TCL in 1990, the system has since evolved into Mitsubishi's modern *Active Skid and Traction Control* (ASTC) system. Developed to help the driver maintain the intended path through a corner, an onboard computer monitored several vehicle operating parameters through the use of various sensors. When too much throttle has been used, while taking a curve, engine output and braking are automatically regulated to ensure the proper path through a curve and to provide the proper amount of traction under various road surface conditions. While conventional traction control systems at the time featured only a slip control function, Mitsubishi developed a TCL system which had a preventive (active) safety feature. This improved the course tracing performance by automatically adjusting the traction force, thereby restraining the development of excessive lateral acceleration, while turning. Although not a ‘true’ modern stability control system, trace control monitors steering angle, throttle position and individual wheel speeds and there is no yaw rate input. The TCL system's standard wheel slip control function improves traction on slippery surfaces or during cornering. In addition to the TCL's traction control feature, it also works together with Diamante's electronic controlled suspension and four-wheel steering that Mitsubishi had equipped to improve total handling and performance.

BMW, working with Robert Bosch GmbH and Continental Automotive Systems, developed a system to reduce engine torque to prevent loss of control and applied it to the entire BMW model line for 1992. From 1987 to 1992, Mercedes-Benz and Robert Bosch GmbH co-developed a system called Elektronisches Stabilitätsprogramm (Ger. "Electronic Stability Programme" trademarked as ESP) a lateral slippage control system, the electronic stability control (ESC).

GM worked with Delphi Corporation and introduced its version of ESC called "StabiliTrak" in 1997 for select Cadillac models. StabiliTrak was made standard equipment on all GM SUVs and vans sold in the U.S. and Canada by 2007 except for certain commercial and fleet vehicles. While the "StabiliTrak" name is used on most General Motors vehicles for the U.S. market, the "Electronic Stability Control" identity is used for GM overseas brands, such as Opel, Holden and Saab, except in the case of Saab's 9-7X which also uses the "StabiliTrak" name. Ford's version of ESC, called AdvanceTrac, was launched in the year 2000. Ford later added Roll Stability Control to AdvanceTrac which was first introduced in Volvo XC90 in 2003 when Volvo Cars was fully owned by Ford and it is now being implemented in many Ford vehicles.

Introduction

In 1995, automobile manufacturers introduced ESC systems. Mercedes-Benz, supplied by Bosch, was the first to implement this with their W140 S-Class model. That same year BMW, supplied by Bosch and ITT Automotive (later acquired by Continental Automotive Systems). Volvo Cars began to offer their version of ESC called DSTC in 1998 on the new S80. Toyota's Vehicle Stability Control system (also in 2004, a preventive system called VDIM) appeared on the Crown Majesta in 1995. Meanwhile others investigated and developed their own systems.

During a moose test (swerving to avoid an obstacle) which became famous in Germany as "the Elk test" the Swedish journalist Robert Collin of Teknikens Värld (World of Technology) in October 1997 rolled a Mercedes A-Class (without ESC) at 78 km/h. Because Mercedes-Benz promotes a reputation for safety, they recalled and retrofitted 130,000 A-Class cars with ESC. This produced a significant reduction in crashes and the number of vehicles with ESC rose. Today virtually all premium brands have made ESC standard on all vehicles, and the number of models with ESC continues to increase. Ford and Toyota announced that all their North American vehicles would be equipped with ESC standard by the end of 2009 (it was standard on Toyota SUVs as of 2004, and after the 2011 model-year, All Lexus, Toyota, and Scion vehicles have ESC; the last one to get it was the 2011 model-year Scion tC). However, as recent as November 2010, Ford still sells models in North America without ESC. General Motors had made a similar announcement for the end of 2010. The NHTSA requires all passenger vehicles to be equipped with ESC by 2012 and estimates it will prevent 5,300-9,600 annual fatalities once all passenger vehicles are equipped with the system.

Operation

During normal driving, ESC works in the background and continuously monitors steering and vehicle direction. It compares the driver's intended direction (determined through the measured steering wheel angle) to the vehicle's actual direction (determined through measured lateral acceleration, vehicle rotation (yaw), and individual road wheel speeds).

ESC intervenes only when it detects loss of steering control, i.e. when the vehicle is not going where the driver is steering. This may happen, for example, when skidding during emergency evasive swerves, understeer or oversteer during poorly judged turns on slippery roads, or hydroplaning. ESC estimates the direction of the skid, and then applies the brakes to individual wheels asymmetrically in order to create torque about the vehicle's vertical axis, opposing the skid and bringing the vehicle back in line with the driver's commanded direction. Additionally, the system may reduce engine power or operate the transmission to slow the vehicle down.

ESC can work on any surface, from dry pavement to frozen lakes. It reacts to and corrects skidding much faster and more effectively than the typical human driver, often before the driver is even aware of any imminent loss of control. In fact, this led to some concern that ESC could allow drivers to become overconfident in their vehicle's handling and/or their own driving skills. For this reason, ESC systems typically inform the driver when they intervene, so that the driver knows that the vehicle's handling limits have been approached. Most activate a dashboard indicator light and/or alert tone; some intentionally allow the vehicle's corrected course to deviate very slightly from the driver-commanded direction, even if it is possible to more precisely match it.

Indeed, all ESC manufacturers emphasize that the system is not a performance enhancement nor a replacement for safe driving practices, but rather a safety technology to assist the driver in recovering from dangerous situations. ESC does not increase traction, so it does not enable faster cornering (although it can facilitate better-controlled cornering). More generally, ESC works within inherent limits of the vehicle's handling and available traction between the tires and road. A reckless maneuver can still exceed these limits, resulting in loss of control. For example, in a severe hydroplaning scenario, the wheels that ESC would use to correct a skid may not even initially be in contact with the road, reducing its effectiveness.

In July 2004, on the Crown Majesta, Toyota offered a Vehicle Dynamics Integrated Management (VDIM) system that incorporated formerly independent systems, including ESC. This worked not only after the skid was detected but also to prevent the skid from occurring in the first place. Using electric variable gear ratio steering power steering this more advanced system could also alter steering gear ratios and steering torque levels to assist the driver in evasive maneuvers.

Effectiveness

Numerous studies around the world confirm that ESC is highly effective in helping the driver maintain control of the car, thereby saving lives and reducing the severity of crashes. In the fall of 2004 in the U.S., the National Highway and Traffic Safety Administration confirmed the international studies, releasing results of a field study in the U.S. of ESC effectiveness. The NHTSA in United States concluded that ESC reduces crashes by 35%. Additionally, Sport utility vehicles (SUVs) with stability control are involved in 67% fewer accidents than SUVs without the system. The United States Insurance Institute for Highway Safety (IIHS) issued its own study in June 2006 showing that up to 10,000 fatal US crashes could be avoided annually if all vehicles were equipped with ESC The IIHS study concluded that ESC reduces the likelihood of all fatal crashes by 43%, fatal single-vehicle crashes by 56%, and fatal single-vehicle rollovers by 77-80%.

ESC is described as the most important advance in auto safety by many experts. including Nicole Nason, Administrator of the NHTSA, Jim Guest and David Champion of Consumers Union of the Fédération Internationale de l'Automobile (FIA), E-Safety Aware, Csaba Csere, editor of Car and Driver, and Jim Gill, long time ESC proponent of Continental Automotive Systems The European New Car Assessment Program (EuroNCAP) "strongly recommends" that people buy cars fitted with stability control.

The IIHS requires that a vehicle must have ESC as an available option in order for it to qualify for their *Top Safety Pick* award for occupant protection and accident avoidance.

Components and design

ESC incorporates yaw rate control into the anti-lock braking system (ABS). Yaw is a rotation around the vertical axis; i.e. spinning left or right. Anti-lock brakes enable ESC to brake individual wheels. Many ESC systems also incorporate a traction control system (TCS or ASR), which senses drive-wheel slip under acceleration and individually brakes the slipping wheel or wheels and/or reduces excess engine power until control is regained. However, ESC achieves a different purpose than ABS or Traction Control.

The ESC system uses several sensors to determine what the driver wants (input). Other sensors indicate the actual state of the vehicle (response). The control algorithm compares driver input to vehicle response and decides, when necessary, to apply brakes and/or reduce throttle by the amounts calculated through the state space (set of equations used to model the dynamics of the vehicle). The ESC controller can also receive data from and issue commands to other controllers on the vehicle such as an all wheel drive system or an active suspension system to improve vehicle stability and controllability.

The sensors used for ESC have to send data at all times in order to detect possible defects as soon as possible. They have to be resistant to possible forms of interference (rain, holes in the road, etc.). The most important sensors are:

- Steering wheel angle sensor: determines the driver's intended rotation; i.e. where the driver wants to steer. This kind of sensor is often based on AMR-elements.
- Yaw rate sensor : measures the rotation rate of the car; i.e. how much the car is actually turning. The data from the yaw sensor is compared with the data from the steering wheel angle sensor to determine regulating action.
- Lateral acceleration sensor: often based on the Hall effect. Measures the lateral acceleration of the vehicle.
- Wheel speed sensor : measures the wheel speed.

Other sensors can include:

- Longitudinal acceleration sensor: similar to the lateral acceleration sensor in design but can offer additional information about road pitch and also provide another source of vehicle acceleration and speed.
- Roll rate sensor: similar to the yaw rate sensor in design but improves the fidelity of the controller's vehicle model and correct for errors when estimating vehicle behavior from the other sensors alone.

ESC uses a hydraulic modulator to assure that each wheel receives the correct brake force. A similar modulator is used in ABS. ABS needs to reduce pressure during braking, only. ESC additionally needs to increase pressure in certain situations and an active vacuum brake booster unit may be utilized in addition to the hydraulic pump to meet these demanding pressure gradients.

The brain of the ESC system is the Electronic Control Unit (ECU). The various control techniques are embedded in it. Often, the same ECU is used for diverse systems at the same time (ABS, Traction control system, climate control, etc.). The input signals are sent through the input-circuit to the digital controller. The desired vehicle state is determined based upon the steering wheel angle, its gradient and the wheel speed. Simultaneously, the yaw sensor measures the actual state. The controller computes the needed brake or acceleration force for each wheel and directs via the driver circuits the valves of the hydraulic modulator. Via a CAN interface the ECU is connected with other systems (ABS, etc.) in order to avoid giving contradictory commands.

Many ESC systems have an "off" override switch so the driver can disable ESC, which may be desirable when badly stuck in mud or snow, or driving on a beach, or if using a smaller-sized spare tire which would interfere with the sensors. Some systems also offer an additional mode with raised thresholds so that a driver can utilize the limits of adhesion with less electronic intervention. However, ESC defaults to "On" when the ignition is re-started. Some ESC systems that lack an "off switch", such as on many recent Toyota and Lexus vehicles, can be temporarily disabled through an undocumented series of brake pedal and handbrake operations. Furthermore, unplugging a wheel speed sensor is another method of disabling most ESC systems. The ESC implementation on newer Ford vehicles cannot be completely disabled even through the use of the "off switch". The ESC will automatically reactivate at highway speeds, and below that if it detects a skid with the brake pedal depressed.

Availability and cost

ESC is built on top of an anti-lock brake (ABS) system, and all ESC-equipped vehicles are fitted with traction control. The ESC components include a yaw rate sensor, a lateral acceleration sensor, a steering wheel sensor, and an upgraded integrated control unit. According to National Highway Traffic Safety Administration research, ABS in 2005 cost an estimated US$368; ESC cost a further US$111. The retail price of ESC varies; as a stand-alone option it retails for as little as $250 USD. However, ESC is rarely offered as a sole option, and is generally not available for aftermarket installation. Instead, it is frequently bundled it with other features or more expensive trims, so the cost of a package that includes ESC could be several thousand dollars. Nonetheless, ESC is considered highly cost-effective and it might pay for itself in reduced insurance premiums. When new federal regulations requiring a safety tool called electronic stability control kick in during 2012, all cars will employ it.

Availability of ESC in passenger vehicles varies between manufacturers and countries. In 2007, ESC was available in roughly 50% of new North American models compared to about 75% in Sweden. However, consumer awareness affects buying patterns so that roughly 45% of vehicles sold in North America and the UK are purchased with ESC, contrasting with 78-96% in other European countries such as Germany, Denmark, and Sweden. While few vehicles had ESC prior to 2004, increased awareness will increase the number of vehicles with ESC on the used car market.

ESC is available on cars, SUVs and pickup trucks from all major auto makers. Luxury cars, sports cars, SUVs, and crossovers are usually equipped with ESC. Midsize cars are also gradually catching on, though the 2008 model years of the, Nissan Altima and Ford Fusion only offered ESC on their V6 engine-equipped cars. While ESC includes traction control, there are vehicles such as the 2008 Chevrolet Malibu LS and 2008 Mazda6 that have traction control but not ESC. ESC is rare among subcompact cars as of 2008. The 2009 Toyota Corolla in the United States (but not Canada) has stability control as a $250 option on all trims below that of the XRS which has it as standard. In Canada, for the 2010 Mazda3, ESC is as an option on the midrange GS trim as part of the moonroof package, and is standard on the top-of-the-line GT version. The 2009 Ford Focus has ESC as an option for the S and SE models, and standard on the SEL and SES models

ESC is also available on some motor homes. Elaborate ESC and ESP systems (including Roll Stability Control (RSC)) are available for many commercial vehicles, including transport trucks, trailers, and buses from manufacturers such as Bendix Corporation, WABCO Daimler, Scania AB, and Prevost, and light passenger vehicles.

The *ChooseESC!* campaign, run by the EU's *eSafetyAware!* project, provides a global perspective on ESC. One *ChooseESC!* publication shows the availability of ESC in EU member countries.

In the US, the Insurance Institute for Highway Safety (IIHS) website shows availability of ESC in individual US models and the National Highway Traffic Safety Administration (NHTSA website) lists US models with ESC.

In Australia, the *National Roads and Motorists' Association* NRMA shows the availability of ESC in Australian models.

Future

The market for ESC is growing quickly, especially in European countries such as Sweden, Denmark, and Germany. For example, in 2003 in Sweden the purchase rate on new cars with ESC was 15%. The Swedish road safety administration issued a strong ESC recommendation and in September 2004, 16 months later, the purchase rate was 58%. A stronger ESC recommendation was then given and in December 2004, the purchase rate on new cars had reached 69% and by 2008 it had grown to 96%. ESC advocates around the world are promoting increased ESC use through legislation and public awareness campaigns and by 2012, most new vehicles should be equipped with ESC.

Just as ESC is founded on the Anti-lock braking system (ABS), ESC is the foundation for new advances such as Roll Stability Control (RSC) that works in the vertical plane much like ESC works in the horizontal plane. When RSC detects impending rollover (usually on transport trucks or SUVs), RSC applies brakes, reduces throttle, induces understeer, and/or slows down the vehicle.

The computing power of ESC facilitates the networking of active and passive safety systems, addressing other causes of crashes. For example, sensors may detect when a vehicle is following too closely and slow down the vehicle, straighten up seat backs, and tighten seat belts, avoiding and/or preparing for a crash.

Regulation

While Sweden used public awareness campaigns to promote ESC use, others implemented or proposed legislation.

The Canadian province of Quebec was the first jurisdiction to implement an ESC law, making it compulsory for carriers of dangerous goods (without data recorders) in 2005.

The United States was next, requiring ESC for all passenger vehicles under 10,000 pounds (4536 kg), phasing in the regulation starting with 55% of 2009 models (effective 1 September 2008), 75% of 2010 models, 95% of 2011 models, and all 2012 models.

Canada will require all new passenger vehicles to have ESC from 1 September 2011.

The Australian Government announced on 23 June 2009 that ESC would be compulsory from 1 November 2011 for all new passenger vehicles sold in Australia, and for all new vehicles from November 2013.

The European Parliament has also called for the accelerated introduction of ESC. The European Commission has confirmed a proposal for the mandatory introduction of ESC on all new cars and commercial vehicle models sold in the EU from 2012, with all new cars being equipped by 2014.

The United Nations Economic Commission for Europe has passed a Global Technical Regulation to harmonize ESC standards.

Product names

Electronic stability control (ESC) is the generic term recognised by the European Automobile Manufacturers Association (ACEA), the North American Society of Automotive Engineers (SAE), the Japan Automobile Manufacturers Association, and other worldwide authorities. However, vehicle manufacturers may use a variety of different trade names for ESC:

- Acura: **Vehicle Stability Assist (VSA)**
- Alfa Romeo: **Vehicle Dynamic Control (VDC)**
- Audi: **Electronic Stability Program (ESP)**
- Bentley: **Electronic Stability Program (ESP)**
- BMW: **Dynamic Stability Control (DSC)** (including Dynamic Traction Control)
- Bugatti: **Electronic Stability Program (ESP)**
- Buick: **StabiliTrak**
- Cadillac: **StabiliTrak & Active Front Steering (AFS)**
- Chery Automobile: **Electronic Stability Program**
- Chevrolet: **StabiliTrak**; **Active Handling** (Corvette only)
- Chrysler: **Electronic Stability Program(ESP)**
- Citroën: **Electronic Stability Program (ESP)**
- Dodge: **Electronic Stability Program (ESP)**
- Daimler: **Electronic Stability Program (ESP)**
- Fiat: **Electronic Stability Program (ESP)** and **Vehicle Dynamic Control (VDC)**
- Ferrari: **Controllo Stabilità (CST)**
- Ford: **AdvanceTrac with Roll Stability Control (RSC)** and **Interactive Vehicle Dynamics (IVD)** and **Electronic Stability Program (ESP)**; **Dynamic Stability Control (DSC)** (Australia only)
- General Motors: **StabiliTrak**
- Honda: **Vehicle Stability Assist (VSA)**
- Holden: **Electronic Stability Program (ESP)**
- Hyundai: **Electronic Stability Program (ESP)**, **Electronic Stability Control (ESC)**, and **Vehicle Stability Assist (VSA)**
- Infiniti: **Vehicle Dynamic Control (VDC)**

- Jaguar: **Dynamic Stability Control (DSC)**
- Jeep: **Electronic Stability Program (ESP)**
- Kia: **Electronic Stability Control (ESC)'** and '**Electronic Stability Program (ESP)'**
- Lamborghini: **Electronic Stability Program (ESP)**
- Land Rover: **Dynamic Stability Control (DSC)**
- Lexus: **Vehicle Dynamics Integrated Management (VDIM)** with **Vehicle Stability Control (VSC)**
- Lincoln: **AdvanceTrac**
- Maserati: **Maserati Stability Program (MSP)**
- Mazda: **Dynamic Stability Control (DSC)** (including Dynamic Traction Control)
- Mercedes-Benz (co-inventor): **Electronic Stability Program (ESP)**
- Mercury: **AdvanceTrac**
- MINI: **Dynamic Stability Control**
- Mitsubishi: **Active Skid and Traction Control MULTIMODE** and **Active Stability Control (ASC)**
- Nissan: **Vehicle Dynamic Control (VDC)**
- Oldsmobile: **Precision Control System (PCS)**
- Opel: **Electronic Stability Program (ESP)**
- Peugeot: **Electronic Stability Program (ESP)**
- Pontiac: **StabiliTrak**
- Porsche: **Porsche Stability Management (PSM)**
- Proton: **Electronic Stability Program**
- Renault: **Electronic Stability Program (ESP)**
- Rover Group: **Dynamic Stability Control (DSC)**
- Saab: **Electronic Stability Program (ESP)**
- Saturn: **StabiliTrak**
- Scania: **Electronic Stability Program (ESP)**
- SEAT: **Electronic Stability Program (ESP)**
- Škoda: **Electronic Stability Program (ESP)**
- Smart: **Electronic Stability Program (ESP)**
- Subaru: **Vehicle Dynamics Control (VDC)**
- Suzuki: **Electronic Stability Program (ESP)**
- Toyota: **Either Vehicle Stability Control (VSC) -OR- Vehicle Dynamics Integrated Management (VDIM)**
- Vauxhall: **Electronic Stability Program (ESP)**
- Volvo: **Dynamic Stability and Traction Control (DSTC)**
- Volkswagen: **Electronic Stability Program (ESP)**

System manufacturers

ESC system manufacturers include:

- Robert Bosch GmbH
- Aisin Advics
- Bendix Corporation

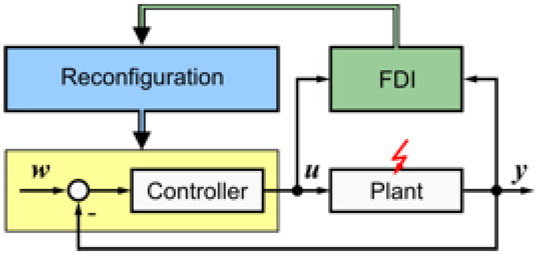
- Continental Automotive Systems
- Delphi
- Hitachi
- ITT Automotive, since 1998 part of Continental AG
- Johnson Electric
- Mando Corporation
- Nissin Kogyo
- Teves, now part of Continental AG
- TRW
- WABCO
- Hyundai Mobis
- Knorr-Bremse

Chapter 10

Control Reconfiguration

Control reconfiguration is an active approach in control theory to achieve fault-tolerant control for dynamic systems . It is used when severe faults, such as actuator or sensor outages, cause a break-up of the control loop, which must be restructured to prevent failure at the system level. In addition to loop restructuring, the controller parameters must be adjusted to accommodate changed plant dynamics. Control reconfiguration is a building block toward increasing the dependability of systems under feedback control .

Reconfiguration problem



Schematic diagram of a typical active fault-tolerant control system. In the nominal, i. e. fault-free situation, the lower control loop operates to meet the control goals. The fault detection (FDI) module monitors the closed-loop system to detect and isolate faults. The fault estimate is passed to the reconfiguration block, which modifies the control loop to reach the control goals in spite of the fault.

Fault modelling

The figure to the right shows a plant controlled by a controller in a standard control loop.

The nominal linear model of the plant is

$$\begin{cases} \dot{\mathbf{x}} & = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} \\ \mathbf{y} & = \mathbf{C}\mathbf{x} \end{cases}$$

The plant subject to a fault (indicated by a red arrow in the figure) is modelled in general by

$$\begin{cases} \dot{\mathbf{x}}_f &= \mathbf{A}_f \mathbf{x}_f + \mathbf{B}_f \mathbf{u} \\ \mathbf{y}_f &= \mathbf{C}_f \mathbf{x}_f \end{cases}$$

where the subscript *f* indicates that the system is faulty. This approach models multiplicative faults by modified system matrices. Specifically, actuator faults are represented by the new input matrix $\mathbf{B}_f$, sensor faults are represented by the output map $\mathbf{C}_f$, and internal plant faults are represented by the system matrix $\mathbf{A}_f$.

The upper part of the figure shows a supervisory loop consisting of *fault detection and isolation* (FDI) and *reconfiguration* which changes the loop by

1. choosing new input and output signals from {$\mathbf{u}, \mathbf{y}$} to reach the control goal,
2. changing the controller internals (including dynamic structure and parameters),
3. adjusting the reference input $\mathbf{w}$.

To this end, the vectors of inputs and outputs contain *all available signals*, not just those used by the controller in fault-free operation.

Alternative scenarios model faults as an additive external signal $\mathbf{f}$ influencing the state derivatives and outputs as follows:

$$\begin{cases} \dot{\mathbf{x}}_f &= \mathbf{A}\mathbf{x}_f + \mathbf{B}\mathbf{u} + \mathbf{E}\mathbf{f} \\ \mathbf{y}_f &= \mathbf{C}_f \mathbf{x}_f + \mathbf{F}\mathbf{f} \end{cases}$$

Reconfiguration goals

The goal of reconfiguration is to keep the reconfigured control loop performance sufficient for preventing plant shutdown. The following goals are distinguished:

1. Stabilisation
2. Equilibrium recovery
3. Output trajectory recovery
4. State trajectory recovery

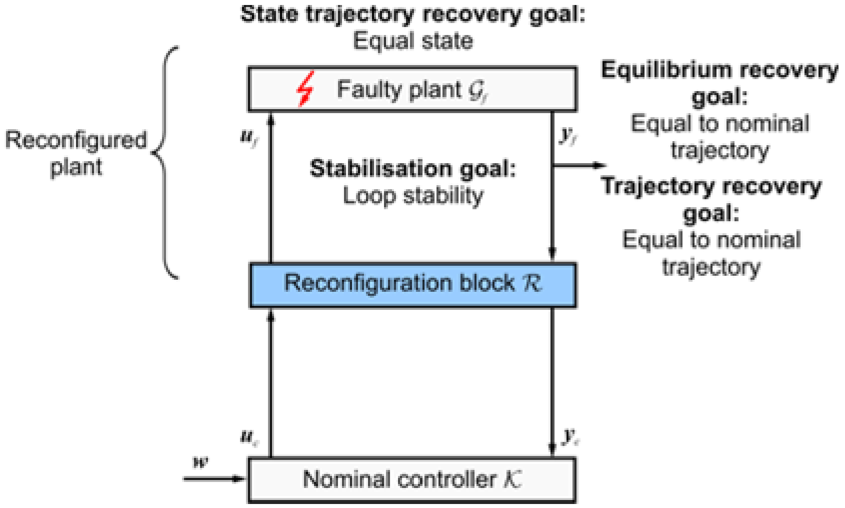
Internal stability of the reconfigured closed loop is usually the minimum requirement. The equilibrium recovery goal (also referred to as weak goal) refers to the steady-state output equilibrium which the reconfigured loop reaches after a given constant input. This equilibrium must equal the nominal equilibrium under the same input (as time tends to infinity). This goal ensures steady-state reference tracking after reconfiguration. The output trajectory recovery goal (also referred to as strong goal) is even stricter. It requires that the dynamic response to an input must equal the nominal response at all times. Further restrictions are imposed by the state trajectory recovery goal, which requires that the state trajectory be restored to the nominal case by the reconfiguration under any input.

Usually a combination of goals is pursued in practice, such as the equilibrium recovery goal with stability.

The question whether or not these or similar goals can be reached for specific faults is addressed by reconfigurability analysis.

Reconfiguration approaches

Fault hiding


State trajectory recovery goal:
Equal state
Faulty plant $\mathcal{G}_f$
Equilibrium recovery goal:
Equal to nominal trajectory
Reconfigured plant
u_f
y_f
Stabilisation goal:
Loop stability
Trajectory recovery goal:
Equal to nominal trajectory
Reconfiguration block $\mathcal{R}$
u_c
y_c
w
Nominal controller $\mathcal{K}$

Fault hiding principle. A reconfiguration block is placed between faulty plant and nominal controller. The reconfuigured plant behaviour must match the nominal behaviour. Furthermore, the reconfiguration goals are pointed out.

This paradigm aims at keeping the nominal controller in the loop. To this end, a reconfiguration block is placed between the faulty plant and the nominal controller. Together with the faulty plant, it forms the reconfigured plant. The reconfiguration block has to fulfill the requirement that the behaviour of the reconfigured plant matches the behaviour of the nominal, that is fault-free plant .

Linear model following

In linear model following, a formal feature of the nominal closed loop is attempted to be recovered. In the classical pseudo-inverse method, the closed loop system matrix $\bar{\mathbf{A}} = \mathbf{A} - \mathbf{B}\mathbf{K}$ of a state-feedback control structure is used. The new controller $\mathbf{K}_f$ is found to approximate $\bar{\mathbf{A}}$ in the sense of an induced matrix norm.

In perfect model following, a dynamic compensator is introduced to allow for the exact recovery of the complete loop behaviour under certain conditions.

In eigenstructure assignment, the nominal closed loop eigenvalues and eigenvectors (the eigenstructure) is recovered to the nominal case after a fault.

Optimisation-based control schemes

Linear-quadratic regulator design (LQR), model predictive control (MPC)

Probabilistic approaches

Learning control

Learning automata, neural networks etc.

Mathematical tools and frameworks

The methods by which reconfiguration is achieved differ considerably. The following list gives an overview of mathematical approaches that are commonly used .

- Adaptive control (AC)
- Disturbance decoupling (DD)
- Eigenstructure assignment (EA)
- Gain scheduling (GS)/linear parameter varying (LPV)
- Generalised internal model control (GIMC)
- Intelligent control (IC)
- Linear matrix inequality (LMI)
- Linear-quadratic regulator (LQR)
- Model following (MF)
- Model predictive control (MPC)
- Pseudo-inverse method (PIM)
- Robust control techniques

Chapter 11

Networked Control System

A **Networked Control System** (**NCS**) is a control system wherein the control loops are closed through a real-time network. The defining feature of an NCS is that control and feedback signals are exchanged among the system's components in the form of information packages through a network.

Overview

The functionality of a typical NCS is established by the use of four basic elements:

1. Sensors, to acquire information,
2. Controllers, to provide decision and commands,
3. Actuators, to perform the control commands and
4. Communication network, to enable exchange of information.

The most important feature of a NCS is that it connects cyberspace to physical space thus, enabling execution of several tasks from long distance. In addition, networked control systems eliminate unnecessary wiring thus, reducing the complexity and the overall cost in designing and implementing the control systems. They can also be easily modified or upgraded by adding sensors, actuators and controllers to them with relatively low cost and no major changes in their structure. Moreover, featuring efficient sharing of data between their controllers, NCS are able to easily fuse global information to make intelligent decisions over large physical spaces.

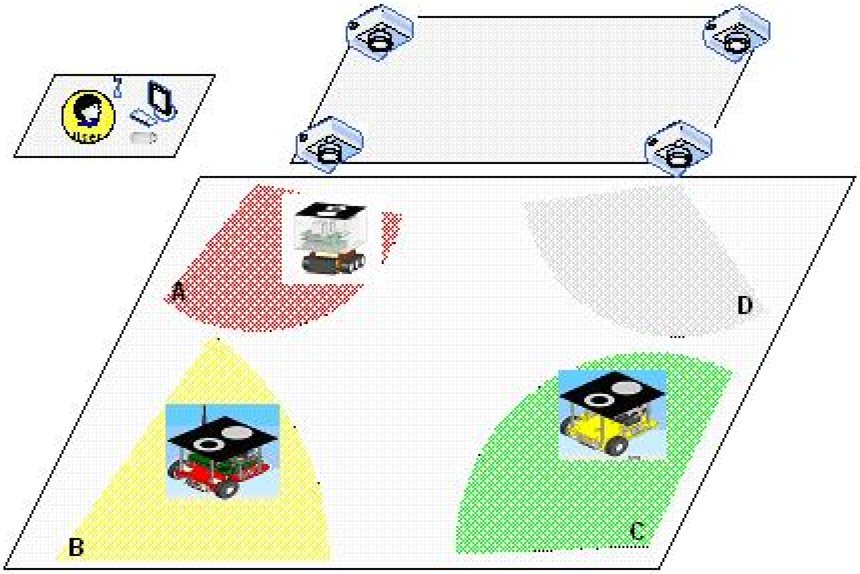
Their potential applications are numerous and cover a wide range of industries such as: space and terrestrial exploration, access in hazardous environments, factory automation, remote diagnostics and troubleshooting, experimental facilities, domestic robots, aircraft, automobiles, manufacturing plant monitoring, nursing homes and tele-operations.

Types of communication networks

- Fieldbuses, e.g. CAN, LON etc.
- Ethernet

- Wireless networks, e.g. Bluetooth or ZigBee. The term Wireless Networked Control System (WNCS) is often used in this connection.

Problems and solutions



iSpace concept

Advent and development of the Internet combined with the advantages provided by NCS attracted the interest of researchers around the globe. Along with the advantages, several challenges also emerged giving rise to many important research topics. New control strategies, kinematics of the actuators in the systems, reliability and security of communications, bandwidth allocation, development of data communication protocols, corresponding fault detection and fault tolerant control strategies, real-time information collection and efficient processing of sensors data are some of the relative topics studied in depth.

The insertion of the communication network in the feedback control loop makes the analysis and design of an NCS complex, since it imposes additional time delays in control loops or possibility of packages loss. Depending on the application, time-delays could impose severe degratation on the system performance.

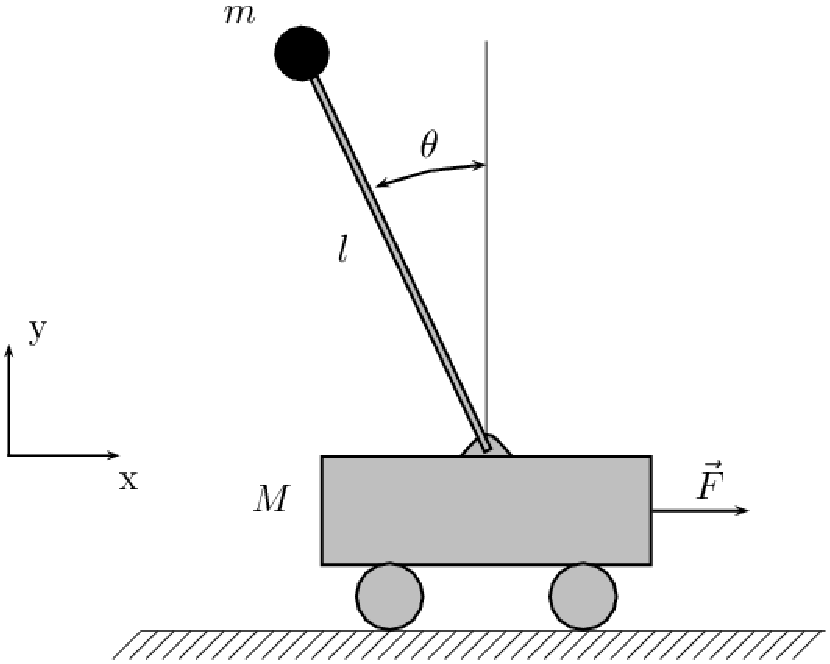
To alleviate the time-delay effect, Y. Tipsuwan and M-Y. Chow, in ADAC Lab at North Carolina State University, proposed the Gain Scheduler Middleware (GSM) methodology and applied it in iSpace. S. Munir and W.J. Book (Georgia Institute of Technology) used a Smith predictor, a Kalman filter and an energy regulator to perform teleoperation through the Internet.

K.C. Lee, S. Lee and H.H. Lee used a genetic algorithm to design a controller used in a NCS. Many other researchers provided solutions using concepts from several control areas such as robust control, optimal stochastic control, fuzzy logic etc.

Moreover, a most critical and important issue surrounding the design of distributed NCSs with the successively increasing complexity is to meet the requirements on system reliability and dependability, while guaranteeing a high system performance over a wide operating range. This makes network based fault detection and diagnosis techniques, which are essential to monitor the system performance, receive more and more attention.

Chapter 12

Inverted Pendulum



A schematic drawing of the inverted pendulum on a cart. The rod is considered massless. The mass of the cart and the pointmass at the end of the rod are denoted by M and m. The rod has a length l.

An **inverted pendulum** is a pendulum which has its mass above its pivot point. It is often implemented with the pivot point mounted on a cart that can move horizontally and may be called a **cart and pole**. Whereas a normal pendulum is stable when hanging downwards, an inverted pendulum is inherently unstable, and must be actively balanced in order to remain upright, either by applying a torque at the pivot point or by moving the pivot point horizontally as part of a feedback system.

Overview

The inverted pendulum is a classic problem in dynamics and control theory and is widely used as a benchmark for testing control algorithms (PID controllers, neural networks, fuzzy control, genetic algorithms, etc.). Variations on this problem include multiple links, allowing the motion of the cart to be commanded while maintaining the pendulum, and balancing the cart-pendulum system on a see-saw. The inverted pendulum is related to rocket or missile guidance, where the center of gravity is located behind the center of drag causing aerodynamic instability. The understanding of a similar problem is built in the technology of the Segway PT, a self-balancing transportation device. The largest implemented use are on huge lifting cranes on shipyards. When moving the shipping containers back and forth, the cranes move the box accordingly so that it never swings or sways. It always stays perfectly positioned under the operator even when moving or stopping quickly.

Another way that an inverted pendulum may be stabilized, without any feedback or control mechanism, is by oscillating the support rapidly up and down. If the oscillation is sufficiently strong (in terms of its acceleration and amplitude) then the inverted pendulum can recover from perturbations in a strikingly counterintuitive manner. If the driving point moves in simple harmonic motion, the pendulum's motion is described by the Mathieu equation.

Equations of motion

Stationary pivot point

The equation of motion is similar to that for a uninverted pendulum except that the sign of the angular position as measured from the vertical unstable equilibrium position:

$$\ddot{\theta} - \frac{g}{\ell} \sin\theta = 0$$

When added to both sides, it will have the same sign as the angular acceleration term:

$$\ddot{\theta} = \frac{g}{\ell} \sin\theta$$

Thus, the inverted pendulum will accelerate away from the vertical unstable equilibrium in the direction initially displaced, and the acceleration is inversely proportional to the length. Tall pendulums fall more slowly than short ones.

Pendulum on a cart

The equations of motion can be derived using Lagrange's equations. We refer to the drawing above where θ(*t*) is the angle of the pendulum of length *l* with respect to the vertical direction and the acting forces are gravity and an external force *F* in the x-direction. Define $x(t)$ to be the position of the cart. The Lagrangian $L = T - V$ of the system is:

$$L = \frac{1}{2} M v_1^2 + \frac{1}{2} m v_2^2 - m g \ell \cos\theta$$

where v_1 is the velocity of the cart and v_2 is the velocity of the point mass m. v_1 and v_2 can be expressed in terms of x and θ by writing the velocity as the first derivative of the position;

$$\begin{aligned} v_1^2 &= \dot{x}^2 \\ v_2^2 &= \left(\frac{d}{dt}(x - \ell\sin\theta)\right)^2 + \left(\frac{d}{dt}(\ell\cos\theta)\right)^2 \end{aligned}$$

Simplifying the expression for v_2 leads to:

$$v_2^2 = \dot{x}^2 - 2\ell\dot{x}\dot{\theta}\cos\theta + \ell^2\dot{\theta}^2$$

The Lagrangian is now given by:

$$L = \frac{1}{2}(M + m)\dot{x}^2 - m\ell\dot{x}\dot{\theta}\cos\theta + \frac{1}{2} m\ell^2\dot{\theta}^2 - m g \ell \cos\theta$$

and the equations of motion are:

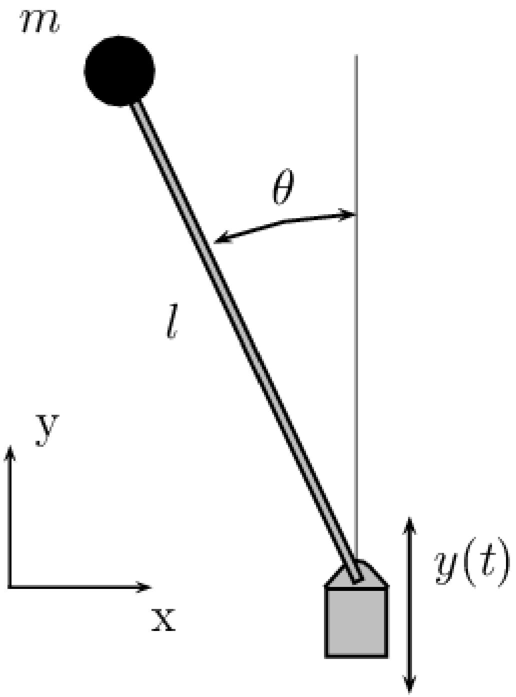
$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{x}} - \frac{\partial L}{\partial x} &= F \\ \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{\theta}} - \frac{\partial L}{\partial \theta} &= 0 \end{aligned}$$

substituting L in these equations and simplifying leads to the equations that describe the motion of the inverted pendulum:

$$\begin{aligned} (M + m)\ddot{x} - m\ell\ddot{\theta}\cos\theta + m\ell\dot{\theta}^2\sin\theta &= F \\ \ell\ddot{\theta} - g\sin\theta &= \ddot{x}\cos\theta \end{aligned}$$

These equations are nonlinear, but since the goal of a control system would be to keep the pendulum upright the equations can be linearized around $\theta \approx 0$.

Pendulum with oscillatory base



A schematic drawing of the inverted pendulum on an oscillatory base. The rod is considered massless. The pointmass at the end of the rod is denoted by *m*. The rod has a length *l*.

The equation of motion for a pendulum connected to a massless, oscillating base is derived the same way as with the pendulum on the cart. The position of the point mass is now given by:

$$(-\ell\sin\theta, y+\ell\cos\theta)$$

and the velocity is found by taking the first derivative of the position:

$$v^2=\dot y^2-2\ell\dot y\dot\theta\sin\theta+\ell^2\dot\theta^2.$$

The Lagrangian for this system can be written as:

$$L=\frac{1}{2}m\left(\dot y^2-2\ell\dot y\dot\theta\sin\theta+\ell^2\dot\theta^2\right)-mg\left(y+\ell\cos\theta\right)$$

and the equation of motion follows from:

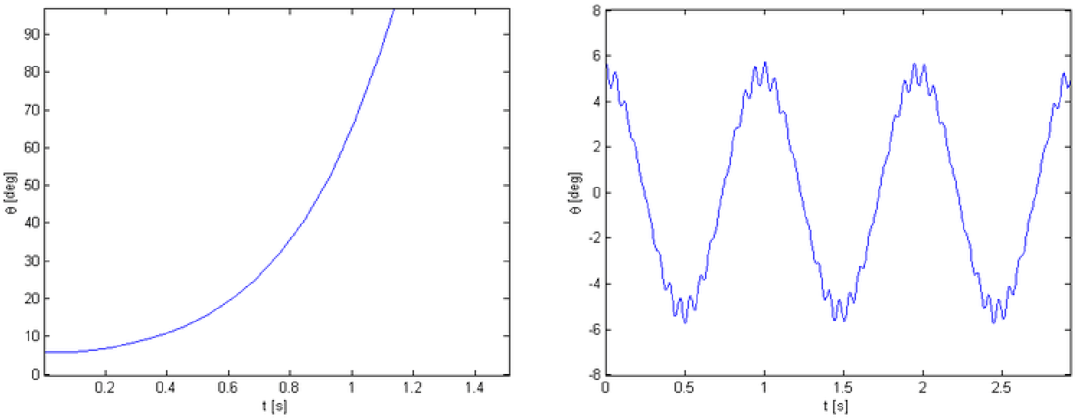
$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial\dot{\theta}}-\frac{\partial L}{\partial\theta}=0$$

resulting in:

$$\ell\ddot{\theta}-\ddot{y}\sin\theta=g\sin\theta.$$

If y represents a simple harmonic motion, $y = A\sin\omega t$, the following differential equation is:

$$\ddot{\theta}-\frac{g}{\ell}\sin\theta=-\frac{A}{\ell}\omega^2\sin\omega t\sin\theta.$$



Plots for the inverted pendulum on an oscillatory base. The first plot shows the response of the pendulum on a slow oscillation, the second the response on a fast oscillation

A solution for this equation will show that the pendulum stays upright for fast oscillations. The first plot shows that when y is a slow oscillation, the pendulum quickly falls over when disturbed from the upright position. The angle θ exceeds 90° after a short time, which means the pendulum has fallen on the ground.
If y is a fast oscillation the pendulum can be kept stable around the vertical position. The second plot shows that when disturbed from the vertical position, the pendulum now starts an oscillation around the vertical position ($\theta = 0$). The deviation from the vertical position stays small, and the pendulum doesn't fall over.

Applications

The inverted pendulum was a central component in the design of several early Seismometers.

Chapter 13

Distributed Control System

A **distributed control system** (DCS) refers to a control system usually of a manufacturing system, process or any kind of dynamic system, in which the controller elements are not central in location (like the brain) but are distributed throughout the system with each component sub-system controlled by one or more controllers. The entire system of controllers is connected by networks for communication and monitoring.

DCS is a very broad term used in a variety of industries, to monitor and control distributed equipment.

- Electrical power grids and electrical generation plants
- Environmental control systems
- Traffic signals
- radio signals
- Water management systems
- Oil refining plants
- Metallurgical Process Plants
- Chemical plants
- Pharmaceutical manufacturing
- Sensor networks
- Dry cargo and bulk oil carrier ships

Elements

A DCS typically uses custom designed processors as controllers and uses both proprietary interconnections and communications protocol for communication. Input and output modules form component parts of the DCS. The processor receives information from input modules and sends information to output modules. The input modules receive information from input instruments in the process (a.k.a. field) and transmit instructions to the output instruments in the field. Computer buses or electrical buses connect the processor and modules through multiplexer or demultiplexers. Buses also connect the distributed controllers with the central controller and finally to the Human-Machine Interface (HMI) or control consoles.

Elements of a distributed control system may directly connect to physical equipment such as switches, pumps and valves or may work through an intermediate system such as a SCADA system.

Applications

Distributed Control Systems (DCSs) are dedicated systems used to control manufacturing processes that are continuous or batch-oriented, such as oil refining, petrochemicals, central station power generation, pharmaceuticals, food & beverage manufacturing, cement production, steelmaking, and papermaking. DCSs are connected to sensors and actuators and use setpoint control to control the flow of material through the plant. The most common example is a setpoint control loop consisting of a pressure sensor, controller, and control valve. Pressure or flow measurements are transmitted to the controller, usually through the aid of a signal conditioning Input/Output (I/O) device. When the measured variable reaches a certain point, the controller instructs a valve or actuation device to open or close until the fluidic flow process reaches the desired setpoint. Large oil refineries have many thousands of I/O points and employ very large DCSs. Processes are not limited to fluidic flow through pipes, however, and can also include things like paper machines and their associated quality controls, variable speed drives and motor control centers, cement kilns, mining operations, ore processing facilities, and many others.

A typical DCS consists of functionally and/or geographically distributed digital controllers capable of executing from 1 to 256 or more regulatory control loops in one control box. The input/output devices (I/O) can be integral with the controller or located remotely via a field network. Today’s controllers have extensive computational capabilities and, in addition to proportional, integral, and derivative (PID) control, can generally perform logic and sequential control.

DCSs may employ one or several workstations and can be configured at the workstation or by an off-line personal computer. Local communication is handled by a control network with transmission over twisted pair, coaxial, or fiber optic cable. A server and/or applications processor may be included in the system for extra computational, data collection, and reporting capability.

History

Early minicomputers were used in the control of industrial processes since the beginning of the 1960s. The IBM 1800, for example, was an early computer that had input/output hardware to gather process signals in a plant for conversion from field contact levels (for digital points) and analog signals to the digital domain.

The first industrial control computer system was built 1959 at the Texaco Port Arthur, Texas, refinery with an RW-300 of the Ramo-Wooldridge Company

.

The DCS was introduced in 1975. Both Honeywell and Japanese electrical engineering firm Yokogawa introduced their own independently produced DCSs at roughly the same time, with the TDC 2000 and CENTUM systems, respectively. US-based Bristol also introduced their UCS 3000 universal controller in 1975. In 1980, Bailey (now part of ABB) introduced the NETWORK 90 system. Also in 1980, Fischer & Porter Company (now also part of ABB) introduced DCI-4000 (DCI stands for Distributed Control Instrumentation).

The DCS largely came about due to the increased availability of microcomputers and the proliferation of microprocessors in the world of process control. Computers had already been applied to process automation for some time in the form of both Direct Digital Control (DDC) and Set Point Control. In the early 1970s Taylor Instrument Company, (now part of ABB) developed the 1010 system, Foxboro the FOX1 system and Bailey Controls the 1055 systems. All of these were DDC applications implemented within minicomputers (DEC PDP-11, Varian Data Machines, MODCOMP etc.) and connected to proprietary Input/Output hardware. Sophisticated (for the time) continuous as well as batch control was implemented in this way. A more conservative approach was Set Point Control, where process computers supervised clusters of analog process controllers. A CRT-based workstation provided visibility into the process using text and crude character graphics. Availability of a fully functional graphical user interface was a way away.

Central to the DCS model was the inclusion of control function blocks. Function blocks evolved from early, more primitive DDC concepts of "Table Driven" software. One of the first embodiments of object-oriented software, function blocks were self contained "blocks" of code that emulated analog hardware control components and performed tasks that were essential to process control, such as execution of PID algorithms. Function blocks continue to endure as the predominant method of control for DCS suppliers, and are supported by key technologies such as **Foundation Fieldbus** today.

Midac Systems of Sydney Australia developed an objected-oriented distributed direct digital control system in 1982. The central system ran 11 microprocessors sharing tasks and common memory and connected to a serial communication network of distributed controllers each running two Z80s. The system was installed at the University of Melbourne.

Digital communication between distributed controllers, workstations and other computing elements (peer to peer access) was one of the primary advantages of the DCS. Attention was duly focused on the networks, which provided the all-important lines of communication that, for process applications, had to incorporate specific functions such as determinism and redundancy. As a result, many suppliers embraced the IEEE 802.4 networking standard. This decision set the stage for the wave of migrations necessary when information technology moved into process automation and IEEE 802.3 rather than IEEE 802.4 prevailed as the control LAN.

The Network Centric Era of the 1980s

The DCS brought distributed intelligence to the plant and established the presence of computers and microprocessors in process control, but it still did not provide the reach and openness necessary to unify plant resource requirements. In many cases, the DCS was merely a digital replacement of the same functionality provided by analog controllers and a panelboard display. This was embodied in The Purdue Reference Model (PRM) that was developed to define Manufacturing Operations Management relationships. PRM later formed the basis for ISA95 standards activities today.

In the 1980s, users began to look at DCSs as more than just basic process control. A very early example of a Direct Digital Control DCS was completed by the Australian business Midac in 1981-1982 using R-Tec Australian designed hardware. The system installed at the University of Melbourne used a serial communications network, connecting campus buildings back to a control room "front end". Each remote unit ran 2 Z80 microprocessors whilst the front end ran 11 in a Parallel Processing configuration with paged common memory to share tasks and could run up to 20,000 concurrent controls objects.

It was believed that if openness could be achieved and greater amounts of data could be shared throughout the enterprise that even greater things could be achieved. The first attempts to increase the openness of DCSs resulted in the adoption of the predominant operating system of the day: *UNIX*. UNIX and its companion networking technology TCP-IP were developed by the Department of Defense for openness, which was precisely the issue the process industries were looking to resolve.

As a result suppliers also began to adopt Ethernet-based networks with their own proprietary protocol layers. The full TCP/IP standard was not implemented, but the use of Ethernet made it possible to implement the first instances of object management and global data access technology. The 1980s also witnessed the first PLCs integrated into the DCS infrastructure. Plant-wide historians also emerged to capitalize on the extended reach of automation systems. The first DCS supplier to adopt UNIX and Ethernet networking technologies was Foxboro, who introduced the I/A Series system in 1987.

The Application Centric Era of the 1990s

The drive toward openness in the 1980s gained momentum through the 1990s with the increased adoption of Commercial off-the-shelf (COTS) components and IT standards. Probably the biggest transition undertaken during this time was the move from the UNIX operating system to the Windows environment. While the realm of the real time operating system (RTOS) for control applications remains dominated by real time commercial variants of UNIX or proprietary operating systems, everything above real-time control has made the transition to Windows.

The introduction of Microsoft at the desktop and server layers resulted in the development of technologies such as OLE for Process Control (OPC), which is now a de facto industry connectivity standard. Internet technology also began to make its mark in automation and the DCS world, with most DCS HMI supporting Internet connectivity. The '90s were also known for the "Fieldbus Wars", where rival organizations competed to define what would become the IEC fieldbus standard for digital communication with field instrumentation instead of 4-20 milliamp analog communications. The first fieldbus installations occurred in the 1990s. Towards the end of the decade, the technology began to develop significant momentum, with the market consolidated around Ethernet I/P, Foundation Fieldbus and Profibus PA for process automation applications. Some suppliers built new systems from the ground up to maximize functionality with fieldbus, such as Rockwell PAX System Honeywell with Experion & Plantscape SCADA systems, ABB with System 800xA, Emerson Process Management with the DeltaV control system, Siemens with the Simatic PCS7 and **azbil** from Yamatake with the Harmonas-DEO system.

The impact of COTS, however, was most pronounced at the hardware layer. For years, the primary business of DCS suppliers had been the supply of large amounts of hardware, particularly I/O and controllers. The initial proliferation of DCSs required the installation of prodigious amounts of this hardware, most of it manufactured from the bottom up by DCS suppliers. Standard computer components from manufacturers such as Intel and Motorola, however, made it cost prohibitive for DCS suppliers to continue making their own components, workstations, and networking hardware.

As the suppliers made the transition to COTS components, they also discovered that the hardware market was shrinking fast. COTS not only resulted in lower manufacturing costs for the supplier, but also steadily decreasing prices for the end users, who were also becoming increasingly vocal over what they perceived to be unduly high hardware costs. Some suppliers that were previously stronger in the PLC business, such as Rockwell Automation and Siemens, were able to leverage their expertise in manufacturing control hardware to enter the DCS marketplace with cost effective offerings, while the stability/scalability/reliability and functionality of these emerging systems are still improving. The traditional DCS suppliers introduced new generation DCS System based on the latest Communication and IEC Standards, which resulting in a trend of combining the traditional concepts/functionalities for PLC and DCS into a one for all solution—named "Process Automation System". The gaps among the various systems remain at the areas such as: the database integrity, pre-engineering functionality, system maturity, communication transparency and reliability. While it is expected the cost ratio is relatively the same (the more powerful the systems are, the more expensive they will be), the reality of the automation business is often operating strategically case by case. The current next evolution step is called Collaborative Process Automation Systems.

To compound the issue, suppliers were also realizing that the hardware market was becoming saturated. The lifecycle of hardware components such as I/O and wiring is also typically in the range of 15 to over 20 years, making for a challenging replacement market. Many of the older systems that were installed in the 1970s and 1980s are still in use today, and there is a considerable installed base of systems in the market that are approaching the end of their useful life. Developed industrial economies in North

America, Europe, and Japan already had many thousands of DCSs installed, and with few if any new plants being built, the market for new hardware was shifting rapidly to smaller, albeit faster growing regions such as China, Latin America, and Eastern Europe.

Because of the shrinking hardware business, suppliers began to make the challenging transition from a hardware-based business model to one based on software and value-added services. It is a transition that is still being made today. The applications portfolio offered by suppliers expanded considerably in the '90s to include areas such as production management, model-based control, real-time optimization, Plant Asset Management (PAM), Real Time Performance Management (RPM) tools, alarm management, and many others. To obtain the true value from these applications, however, often requires a considerable service content, which the suppliers also provide.

Chapter 14

Control Center Solutions

Control Center Solution is a generic term for different flavors of technical arrangement within command & control facilities. They represent more or less integrated installations that are used to manage resources in order to achieve results in complex environments. Common to all Control Center Solutions are underlying principles of control.

Control Center

Control Center follow a *Purpose* which - very generic - is to control and/or to apply control principles in order to affect a *Target* in a desired way. The purpose as such is outside the Control Center entity and defines its mission. Typically, the most advanced control centers operate in safety critical environments with complex missions:

- Air Traffic Control Centers (ATCC)
- Search and Rescue Centers (SAR)
- Space Mission Control Centers
- Harbor Control Centers,
- Police Dispatch and Control Centers,
- Fire fighter Dispatch and Control Centers.

Today Control Centers represent large facilities, impressive buildings with equally impressive infrastructure and potentially hundreds of controllers working.

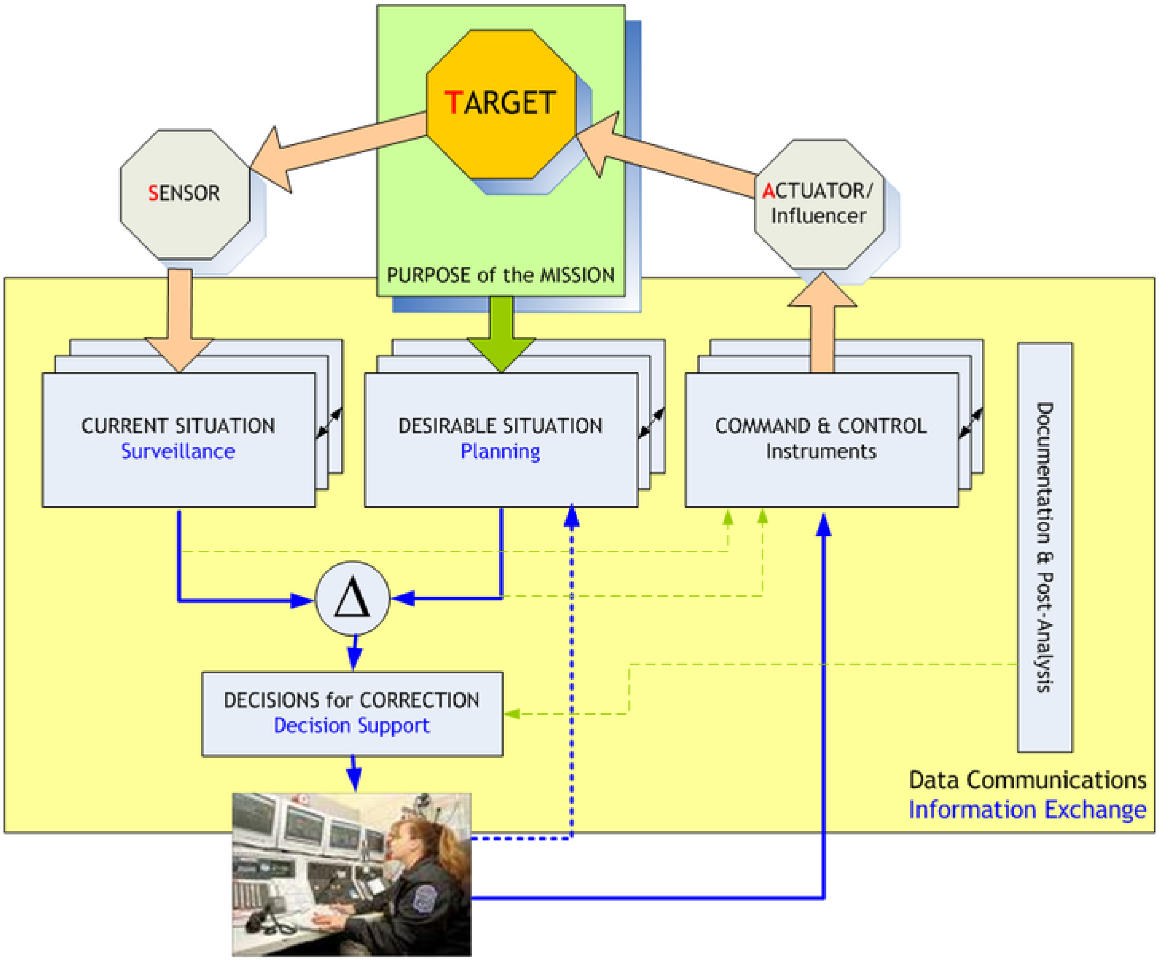
Dispatched Control Center

Control Centers are not necessarily fixed to a distinct location, although they typically are. In recent years remote missions with dispatched resources become more frequent in particular in Defense (Peacekeeping Missions) and Crisis Management. The "center" becomes a mobile unit.

Virtual Center

Even more progressively, Control Centers may not need a distinct space at all. Based on communications infrastructure with high availability figures controllers/commanders may operate on the spot with mobile devices retrieving all required information from the network (network centrics).

Control Principle


TARGET
SENSOR
ACTUATOR/ Influencer
PURPOSE of the MISSION
CURRENT SITUATION
Surveillance
DESIRABLE SITUATION
Planning
COMMAND & CONTROL
Instruments
Documentation & Post-Analysis
Δ
DECISIONS for CORRECTION
Decision Support
Data Communications
Information Exchange

Control Principle

Control is generally based on a **control loop** that consists of

- Sensors - generate data which represent (part of) the current environment (incl. human beings: e.g. observation reports)
- Actuators - affect the environment (incl. human beings: e.g. firefighter is an actuator)
- Controllers - forms a picture of the current situation, compares it with the intended situation and either decides on the adaptation of the intended situation (changing the plan) or formulates and decides on an action affecting the

environment. This decision making process may include human operators, but not necessarily.
- Communications - all functional elements of the Control Loop require adequate connectivity amongst each other

These entities together form a feedback loop which allows to detect changes on the affected target as a response to actions aimed for affect the target. The efficiency and effectiveness of this feedback loop depends on the quality of information generated and processed and the processing delays.

Decision Making

Control/Decision Making can be differentiated in

- Current Situation - representation of *Target* parameters relevant for fulfilling the purpose
- Desired/Intended Situation - representation of *Target* parameters relevant for fulfilling the purpose
- Decision Support (Deviation Analysis, Suggestions) and Decision Making
- Command & Control - elements that provide access to the Actuators (voice communication,..)
- Documentation & Analysis - generates database that feeds back into Decision Support (learning factor)

Sensors produce data which is transformed into information when presented as part of the surveillance picture. The closer the process gets to the decision making phase the more a transformation into knowledge can be observed. Completing the loop the characteristic of the content develops back into information and finally data which is operated by the actuator.

Control Center Solution

A **Control Center Solution** is a distinct compilation and integration of specialized services and components (equipment) which - each for itself or by mutual interaction - form part of a control loop required to follow the purpose.

Examples

Using a fire fighter mission as an example the whole control loop would operate as follows: The purpose is to remove the fire. Front line fire fighter observe the scene and act as sensors by reporting back to the mission command. With these reports and other sources of data a picture of the current situation is generated which deviates from the intended situation which may be "no fire". Considering the characteristics of the surveillance information and the quality of the deviation between *now* and *should be* mission command may decide to change the plan or order for an attack led by fire fighter. The target is the fire.