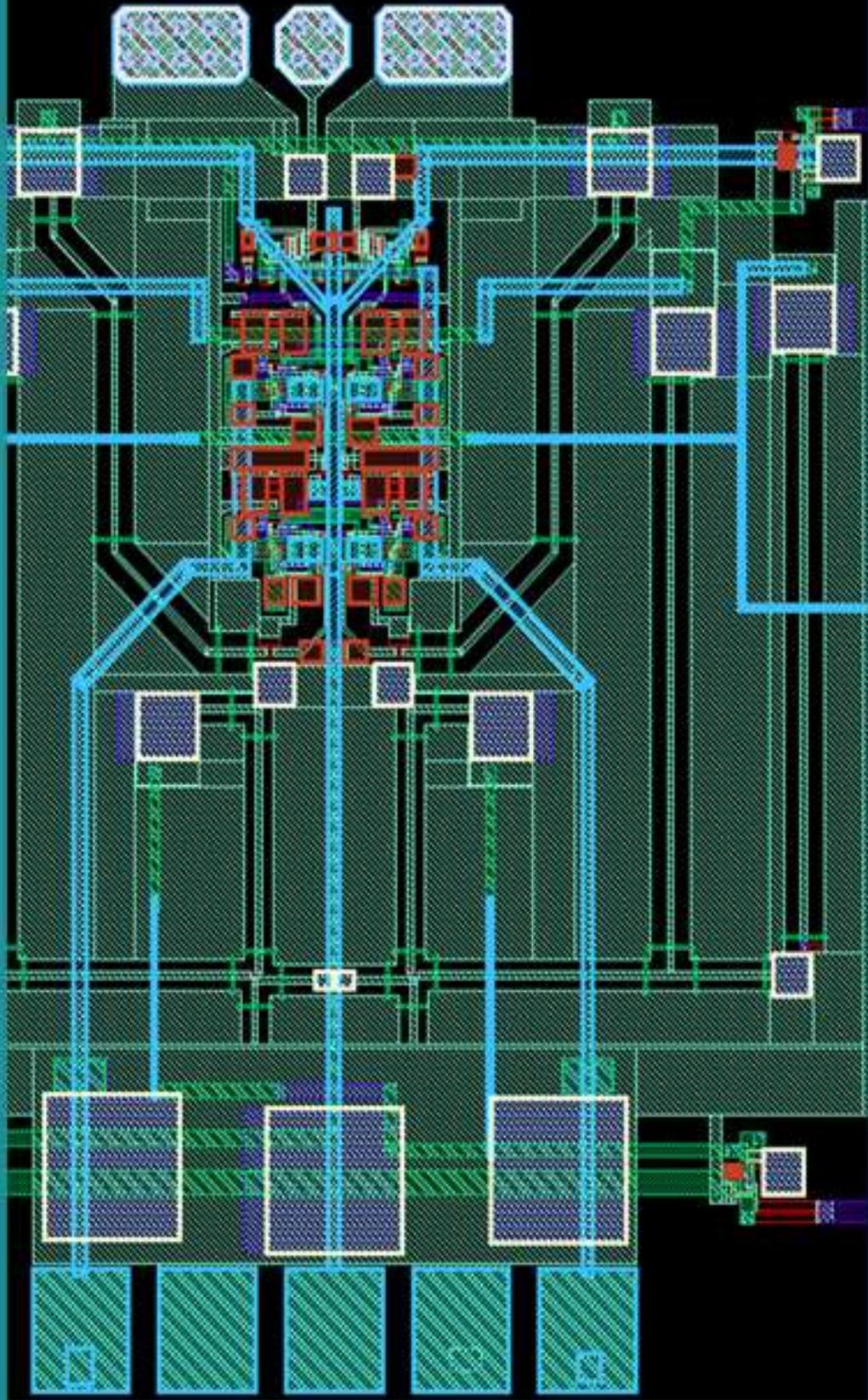


Handbook of Communication and Integrated Circuits



Rheba Shipley
Nichole Sell

First Edition, 2012

ISBN 978-81-323-1347-2



© All rights reserved.

Published by:

College Publishing House
4735/22 Prakashdeep Bldg,
Ansari Road, Darya Ganj,
Delhi - 110002
Email: info@wtbooks.com

WORLD TECHNOLOGIES

Table of Contents

Chapter 1 - Introduction to Integrated Circuits

Chapter 2 - 555 timer IC

Chapter 3 - 7400 series

Chapter 4 - 4000 series

Chapter 5 - Intel 8080

Chapter 6 - Intel 8088 & Intel 80286

Chapter 7 - Intel 80486

Chapter 8 - Frequency Multiplier

Chapter 9 - Frequency Synthesizer and Current Loop

Chapter 10 - Leased Line

Chapter 11 - Balanced Line and Balanced Circuit

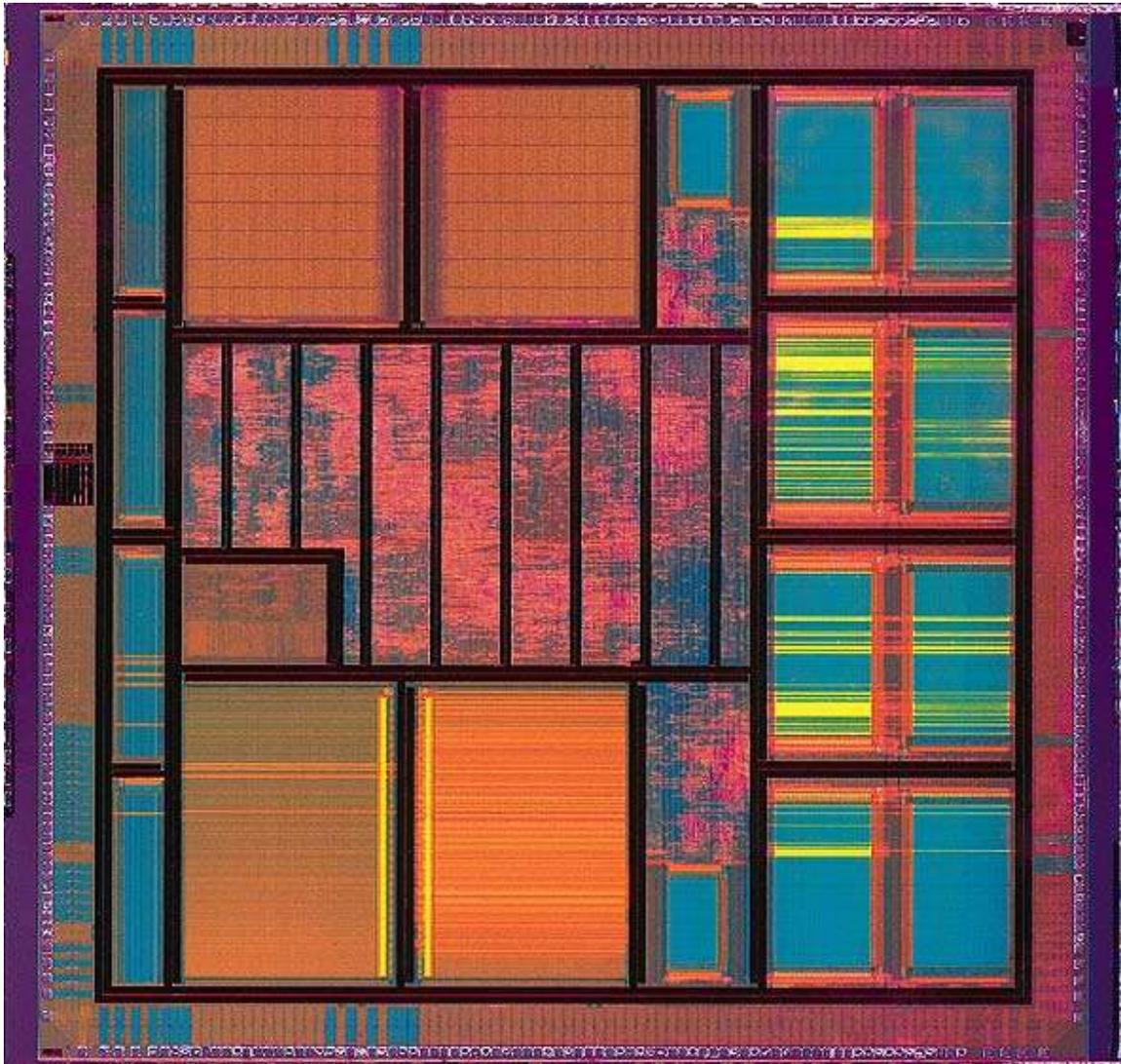
Chapter 12 - Loading Coil

Chapter 13 - Phantom Circuit and Frequency Mixer

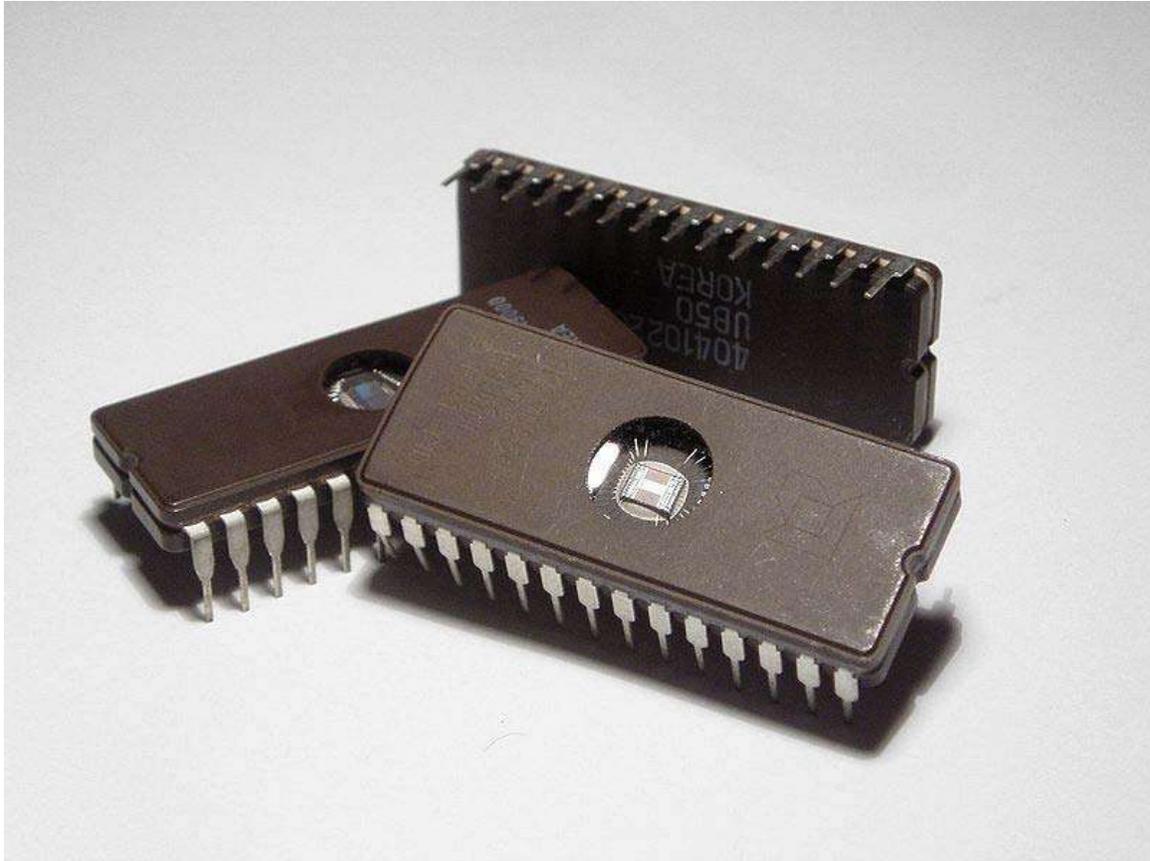
Chapter 14 - Phase-locked Loop

Chapter 1

Introduction to Integrated Circuits



Integrated circuit of Atmel Diopsis 740 System on Chip showing memory blocks, logic and input/output pads around the periphery

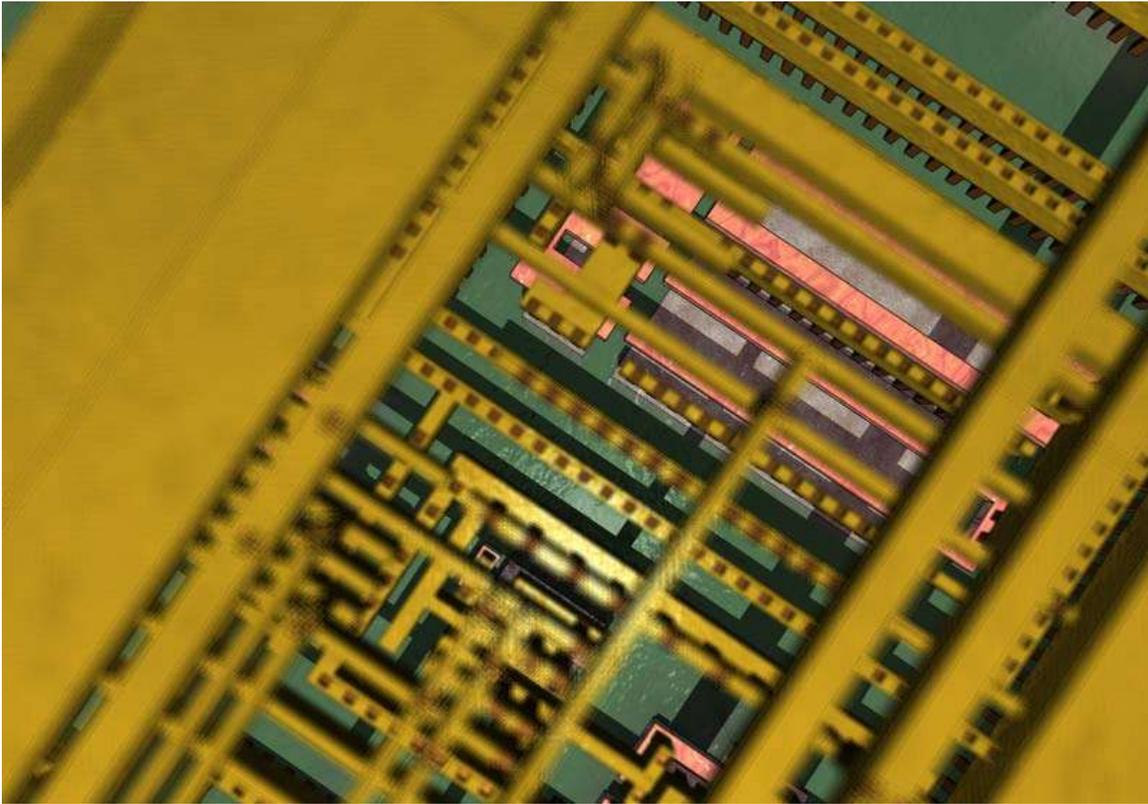


Microchips (EPROM memory) with a transparent window, showing the integrated circuit inside. Note the fine silver-colored wires that connect the integrated circuit to the pins of the package. The window allows the memory contents of the chip to be erased, by exposure to strong ultraviolet light in an eraser device.

In electronics, an **integrated circuit** (also known as **IC**, **chip**, or **microchip**) is a miniaturized electronic circuit (consisting mainly of semiconductor devices, as well as passive components) that has been manufactured in the surface of a thin substrate of semiconductor material. Integrated circuits are used in almost all electronic equipment in use today and have revolutionized the world of electronics. Computers, cellular phones, and other digital appliances are now inextricable parts of the structure of modern societies, made possible by the low cost of production of integrated circuits.

A hybrid integrated circuit is a miniaturized electronic circuit constructed of individual semiconductor devices, as well as passive components, bonded to a substrate or circuit board. A monolithic integrated circuit is made of devices manufactured by diffusion of trace elements into a single piece of semiconductor substrate, a chip.

Introduction



Synthetic detail of an integrated circuit through four layers of planarized copper interconnect, down to the polysilicon (pink), wells (greyish), and substrate (green).

Integrated circuits were made possible by experimental discoveries which showed that semiconductor devices could perform the functions of vacuum tubes and by mid-20th-century technology advancements in semiconductor device fabrication. The integration of large numbers of tiny transistors into a small chip was an enormous improvement over the manual assembly of circuits using electronic components. The integrated circuit's mass production capability, reliability, and building-block approach to circuit design ensured the rapid adoption of standardized ICs in place of designs using discrete transistors.

There are two main advantages of ICs over discrete circuits: cost and performance. Cost is low because the chips, with all their components, are printed as a unit by photolithography rather than being constructed one transistor at a time. Furthermore, much less material is used to construct a packaged IC die than a discrete circuit. Performance is high since the components switch quickly and consume little power (compared to their discrete counterparts) because the components are small and positioned close together. As of 2006, chip areas range from a few square millimeters to around 350 mm^2 , with up to 1 million transistors per mm^2 .

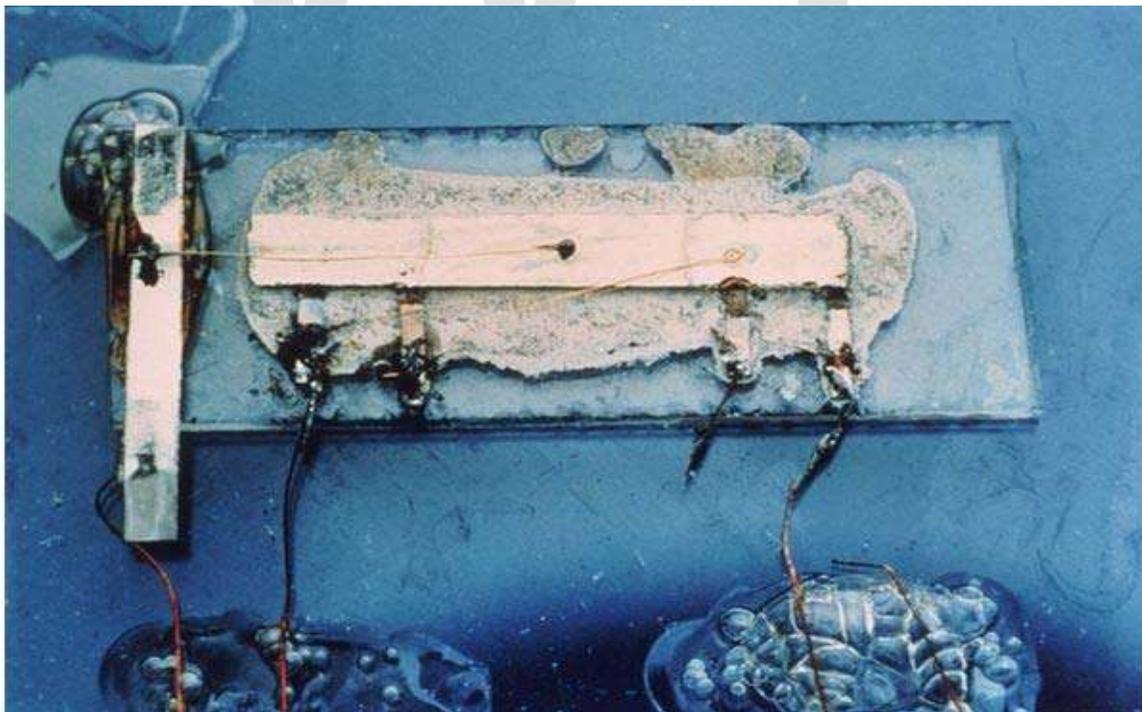
Invention

Early developments of the integrated circuit go back to 1949, when the German engineer Werner Jacobi (Siemens AG) filed a patent for an integrated-circuit-like semiconductor amplifying device showing five transistors on a common substrate arranged in a 2-stage amplifier arrangement. Jacobi discloses small and cheap hearing aids as typical industrial applications of his patent. A commercial use of his patent has not been reported.

The idea of the integrated circuit was conceived by a radar scientist working for the Royal Radar Establishment of the British Ministry of Defence, Geoffrey W.A. Dummer (1909–2002), who published it at the Symposium on Progress in Quality Electronic Components in Washington, D.C. on May 7, 1952. He gave many symposia publicly to propagate his ideas. Dummer unsuccessfully attempted to build such a circuit in 1956.

A precursor idea to the IC was to create small ceramic squares (wafers), each one containing a single miniaturized component. Components could then be integrated and wired into a bidimensional or tridimensional compact grid. This idea, which looked very promising in 1957, was proposed to the US Army by Jack Kilby, and led to the short-lived Micromodule Program (similar to 1951's Project Tinkertoy). However, as the project was gaining momentum, Kilby came up with a new, revolutionary design: the IC.

Robert Noyce credited Kurt Lehovec of Sprague Electric for the *principle of p-n junction isolation* caused by the action of a biased p-n junction (the diode) as a key concept behind the IC.



Jack Kilby's original integrated circuit

Jack Kilby recorded his initial ideas concerning the integrated circuit in July 1958 and successfully demonstrated the first working integrated circuit on September 12, 1958. In his patent application of February 6, 1959, Kilby described his new device as “a body of semiconductor material ... wherein all the components of the electronic circuit are completely integrated.” Kilby won the 2000 Nobel Prize in Physics for his part of the invention of the integrated circuit.

Robert Noyce also came up with his own idea of an integrated circuit half a year later than Kilby. Noyce's chip solved many practical problems that Kilby's had not. Noyce's chip, made at Fairchild Semiconductor, was made of silicon, whereas Kilby's chip was made of germanium.

Generations

In the early days of integrated circuits, only a few transistors could be placed on a chip, as the scale used was large because of the contemporary technology. As the degree of integration was small, the design was done easily. Later on, millions, and today billions, of transistors could be placed on one chip, and to make a good design became a task to be planned thoroughly. This gave rise to new design methods.

SSI, MSI and LSI

The first integrated circuits contained only a few transistors. Called "**Small-Scale Integration**" (SSI), digital circuits containing transistors numbering in the tens provided a few logic gates for example, while early linear ICs such as the Plessey SL201 or the Philips TAA320 had as few as two transistors. The term Large Scale Integration was first used by IBM scientist Rolf Landauer when describing the theoretical concept, from there came the terms for SSI, MSI, VLSI, and ULSI.

SSI circuits were crucial to early aerospace projects, and vice-versa. Both the Minuteman missile and Apollo program needed lightweight digital computers for their inertial guidance systems; the Apollo guidance computer led and motivated the integrated-circuit technology, while the Minuteman missile forced it into mass-production. The Minuteman missile program and various other Navy programs accounted for the total \$4 million integrated circuit market in 1962, and by 1968, U.S. Government space and defense spending still accounted for 37% of the \$312 million total production. The demand by the U.S. Government supported the nascent integrated circuit market until costs fell enough to allow firms to penetrate the industrial and eventually the consumer markets. The average price per integrated circuit dropped from \$50.00 in 1962 to \$2.33 in 1968. Integrated Circuits began to appear in consumer products by the turn of the decade, a typical application being FM inter-carrier sound processing in television receivers.

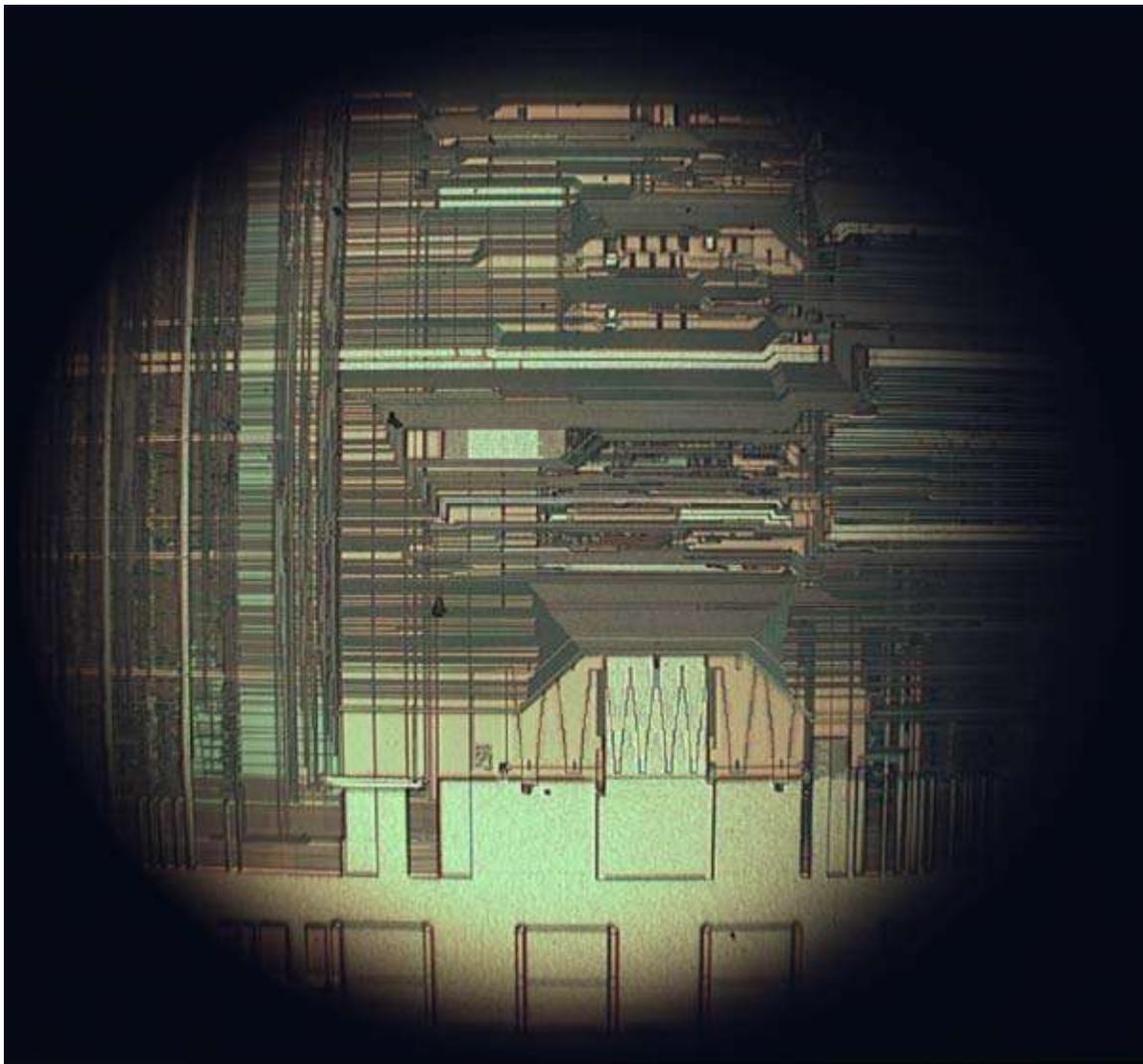
The next step in the development of integrated circuits, taken in the late 1960s, introduced devices which contained hundreds of transistors on each chip, called "**Medium-Scale Integration**" (MSI).

They were attractive economically because while they cost little more to produce than SSI devices, they allowed more complex systems to be produced using smaller circuit boards, less assembly work (because of fewer separate components), and a number of other advantages.

Further development, driven by the same economic factors, led to "**Large-Scale Integration**" (LSI) in the mid 1970s, with tens of thousands of transistors per chip.

Integrated circuits such as 1K-bit RAMs, calculator chips, and the first microprocessors, that began to be manufactured in moderate quantities in the early 1970s, had under 4000 transistors. True LSI circuits, approaching 10000 transistors, began to be produced around 1974, for computer main memories and second-generation microprocessors.

VLSI



Upper interconnect layers on an Intel 80486DX2 microprocessor die

The final step in the development process, starting in the 1980s and continuing through the present, was "very large-scale integration" (VLSI). The development started with hundreds of thousands of transistors in the early 1980s, and continues beyond several billion transistors as of 2009.

Multiple developments were required to achieve this increased density. Manufacturers moved to smaller rules and cleaner fabs, so that they could make chips with more transistors and maintain adequate yield. The path of process improvements was summarized by the International Technology Roadmap for Semiconductors (ITRS). Design tools improved enough to make it practical to finish these designs in a reasonable time. The more energy efficient CMOS replaced NMOS and PMOS, avoiding a prohibitive increase in power consumption. Better texts such as the landmark textbook by Mead and Conway helped schools educate more designers, among other factors.

In 1986 the first one megabit RAM chips were introduced, which contained more than one million transistors. Microprocessor chips passed the million transistor mark in 1989 and the billion transistor mark in 2005. The trend continues largely unabated, with chips introduced in 2007 containing tens of billions of memory transistors.

ULSI, WSI, SOC and 3D-IC

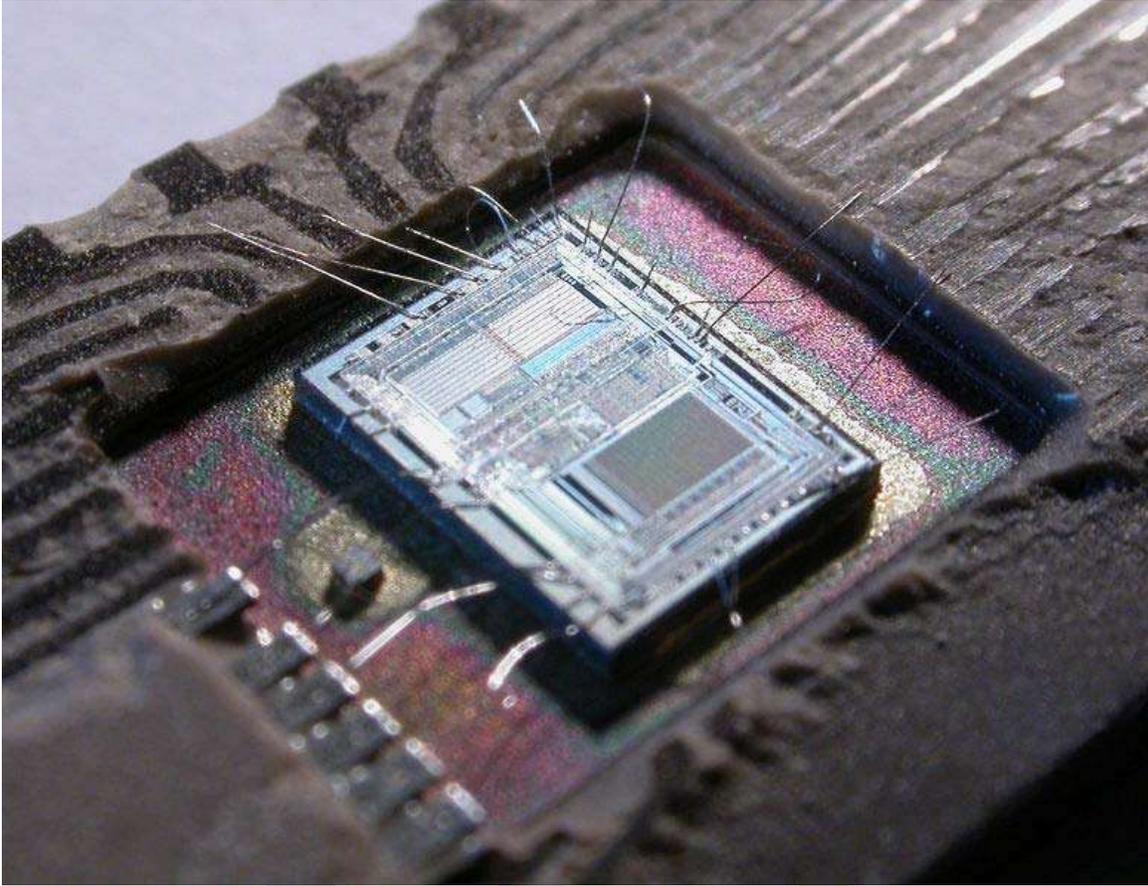
To reflect further growth of the complexity, the term *ULSI* that stands for "ultra-large-scale integration" was proposed for chips of complexity of more than 1 million transistors.

Wafer-scale integration (WSI) is a system of building very-large integrated circuits that uses an entire silicon wafer to produce a single "super-chip". Through a combination of large size and reduced packaging, WSI could lead to dramatically reduced costs for some systems, notably massively parallel supercomputers. The name is taken from the term Very-Large-Scale Integration, the current state of the art when WSI was being developed.

A system-on-a-chip (SoC or SOC) is an integrated circuit in which all the components needed for a computer or other system are included on a single chip. The design of such a device can be complex and costly, and building disparate components on a single piece of silicon may compromise the efficiency of some elements. However, these drawbacks are offset by lower manufacturing and assembly costs and by a greatly reduced power budget: because signals among the components are kept on-die, much less power is required.

A three-dimensional integrated circuit (3D-IC) has two or more layers of active electronic components that are integrated both vertically and horizontally into a single circuit. Communication between layers uses on-die signaling, so power consumption is much lower than in equivalent separate circuits. Judicious use of short vertical wires can substantially reduce overall wire length for faster operation.

Advances in integrated circuits



The die from an Intel 8742, an 8-bit microcontroller that includes a CPU running at 12 MHz, 128 bytes of RAM, 2048 bytes of EPROM, and I/O in the same chip.

Among the most advanced integrated circuits are the microprocessors or "**cores**", which control everything from computers and cellular phones to digital microwave ovens. Digital memory chips and ASICs are examples of other families of integrated circuits that are important to the modern information society. While the cost of designing and developing a complex integrated circuit is quite high, when spread across typically millions of production units the individual IC cost is minimized. The performance of ICs is high because the small size allows short traces which in turn allows low power logic (such as CMOS) to be used at fast switching speeds.

ICs have consistently migrated to smaller feature sizes over the years, allowing more circuitry to be packed on each chip. This increased capacity per unit area can be used to decrease cost and/or increase functionality—see Moore's law which, in its modern interpretation, states that the number of transistors in an integrated circuit doubles every two years. In general, as the feature size shrinks, almost everything improves—the cost per unit and the switching power consumption go down, and the speed goes up. However, ICs with nanometer-scale devices are not without their problems, principal among which

is leakage current, although these problems are not insurmountable and will likely be solved or at least ameliorated by the introduction of high-k dielectrics. Since these speed and power consumption gains are apparent to the end user, there is fierce competition among the manufacturers to use finer geometries. This process, and the expected progress over the next few years, is well described by the International Technology Roadmap for Semiconductors (ITRS).

In current research projects, integrated circuits are also developed for sensoric applications in medical implants or other bioelectronic devices. Particular sealing strategies have to be taken in such biogenic environments to avoid corrosion or biodegradation of the exposed semiconductor materials. As one of the few materials well established in CMOS technology, titanium nitride TiN turned out as exceptionally stable and well suited for electrode applications in medical implants.

Classification



A CMOS 4000 IC in a DIP

Integrated circuits can be classified into analog, digital and mixed signal (both analog and digital on the same chip).

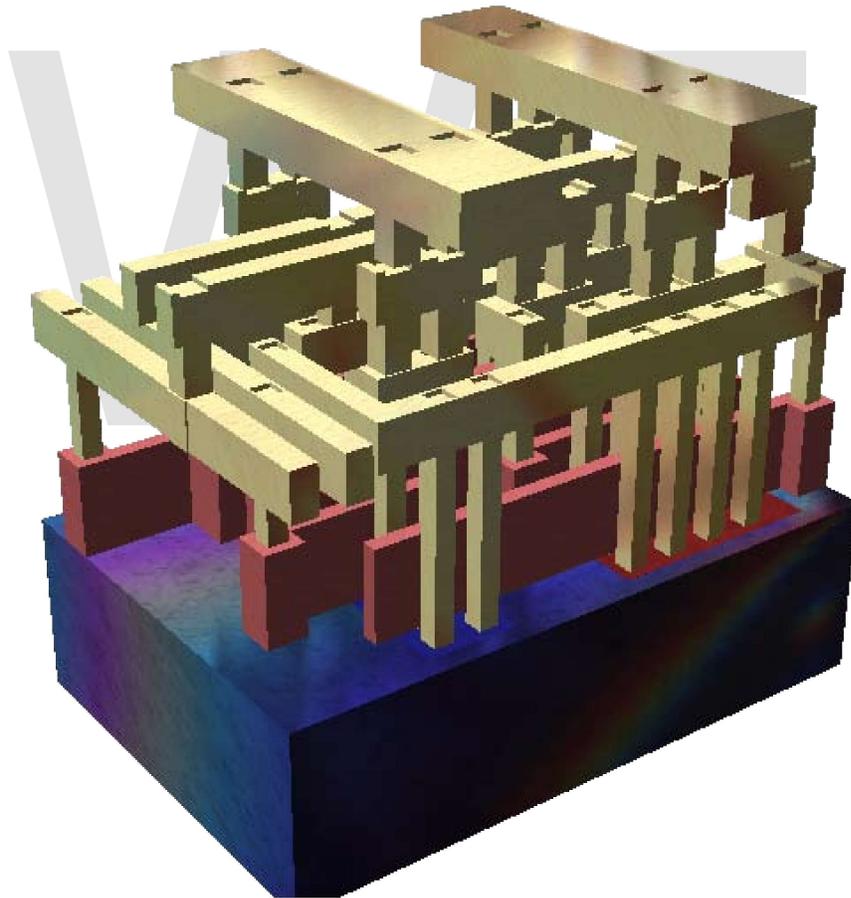
Digital integrated circuits can contain anything from one to millions of logic gates, flip-flops, multiplexers, and other circuits in a few square millimeters. The small size of these circuits allows high speed, low power dissipation, and reduced manufacturing cost compared with board-level integration. These digital ICs, typically microprocessors, DSPs, and micro controllers work using binary mathematics to process "one" and "zero" signals.

Analog ICs, such as sensors, power management circuits, and operational amplifiers, work by processing continuous signals. They perform functions like amplification, active filtering, demodulation, mixing, etc. Analog ICs ease the burden on circuit designers by having expertly designed analog circuits available instead of designing a difficult analog circuit from scratch.

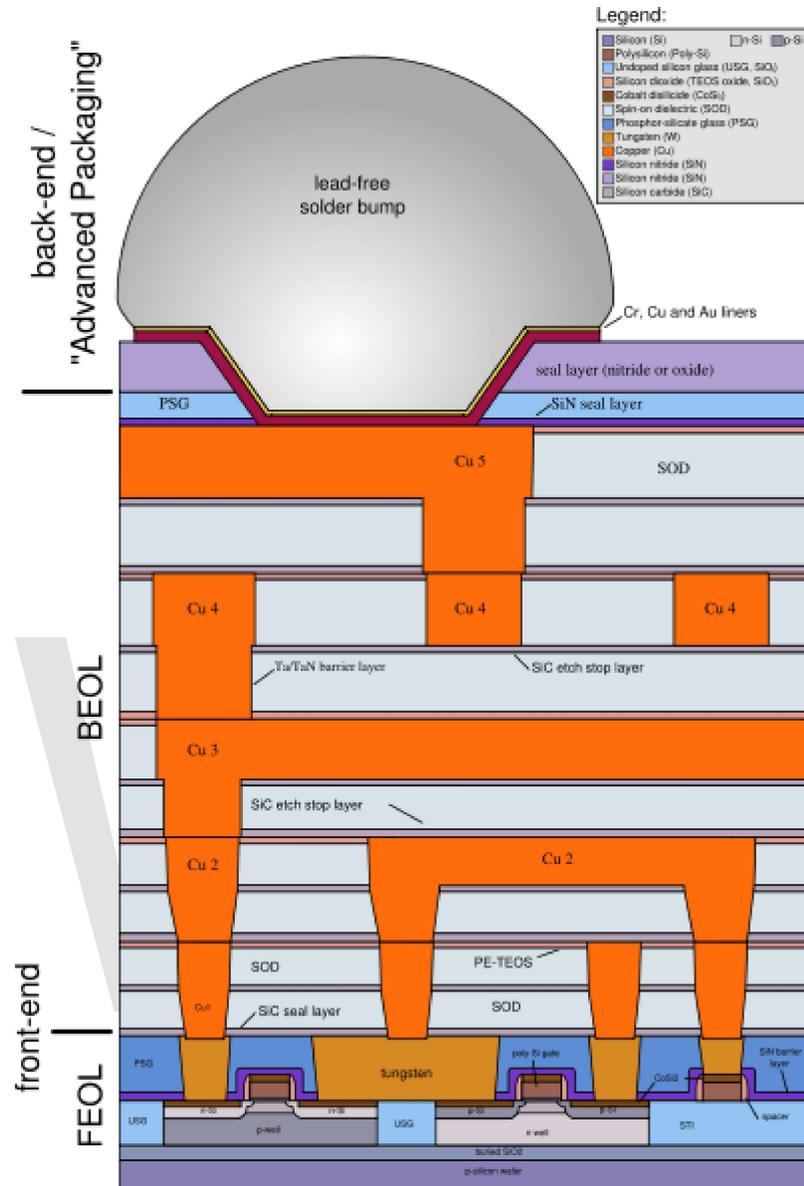
ICs can also combine analog and digital circuits on a single chip to create functions such as A/D converters and D/A converters. Such circuits offer smaller size and lower cost, but must carefully account for signal interference.

Manufacturing

Fabrication



Rendering of a small standard cell with three metal layers (dielectric has been removed). The sand-colored structures are metal interconnect, with the vertical pillars being contacts, typically plugs of tungsten. The reddish structures are polysilicon gates, and the solid at the bottom is the crystalline silicon bulk.



Schematic structure of a CMOS chip, as built in the early 2000s. The graphic shows LDD-MISFET's on an SOI substrate with five metallization layers and solder bump for flip-chip bonding. It also shows the section for FEOL (front-end of line), BEOL (back-end of line) and first parts of back-end process.

The semiconductors of the periodic table of the chemical elements were identified as the most likely materials for a *solid state vacuum tube*. Starting with copper oxide, proceeding to germanium, then silicon, the materials were systematically studied in the 1940s and 1950s. Today, silicon monocrystals are the main substrate used for *integrated circuits (ICs)* although some III-V compounds of the periodic table such as gallium arsenide are used for specialized applications like LEDs, lasers, solar cells and the highest-speed integrated circuits. It took decades to perfect methods of creating crystals without defects in the crystalline structure of the semiconducting material.

Semiconductor ICs are fabricated in a layer process which includes these key process steps:

- Imaging
- Deposition
- Etching

The main process steps are supplemented by doping and cleaning.

Mono-crystal silicon wafers (or for special applications, silicon on sapphire or gallium arsenide wafers) are used as the *substrate*. Photolithography is used to mark different areas of the substrate to be doped or to have polysilicon, insulators or metal (typically aluminium) tracks deposited on them.

- Integrated circuits are composed of many overlapping layers, each defined by photolithography, and normally shown in different colors. Some layers mark where various dopants are diffused into the substrate (called diffusion layers), some define where additional ions are implanted (implant layers), some define the conductors (polysilicon or metal layers), and some define the connections between the conducting layers (via or contact layers). All components are constructed from a specific combination of these layers.
- In a self-aligned CMOS process, a transistor is formed wherever the gate layer (polysilicon or metal) crosses a diffusion layer.
- Capacitive structures, in form very much like the parallel conducting plates of a traditional electrical capacitor, are formed according to the area of the "plates", with insulating material between the plates. Capacitors of a wide range of sizes are common on ICs.
- Meandering stripes of varying lengths are sometimes used to form on-chip resistors, though most logic circuits do not need any resistors. The ratio of the length of the resistive structure to its width, combined with its sheet resistivity, determines the resistance.
- More rarely, inductive structures can be built as tiny on-chip coils, or simulated by gyrators.

Since a CMOS device only draws current on the *transition* between logic states, CMOS devices consume much less current than bipolar devices.

A random access memory is the most regular type of integrated circuit; the highest density devices are thus memories; but even a microprocessor will have memory on the chip. Although the structures are intricate – with widths which have been shrinking for decades – the layers remain much thinner than the device widths. The layers of material are fabricated much like a photographic process, although light waves in the visible

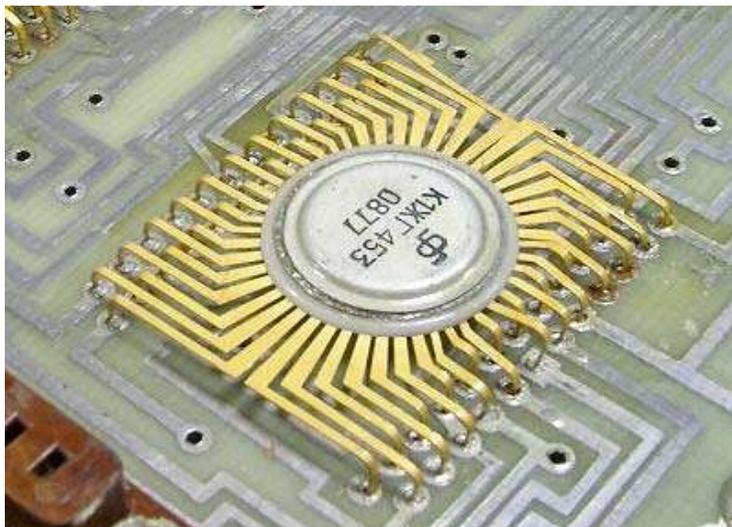
spectrum cannot be used to "expose" a layer of material, as they would be too large for the features. Thus photons of higher frequencies (typically ultraviolet) are used to create the patterns for each layer. Because each feature is so small, electron microscopes are essential tools for a process engineer who might be debugging a fabrication process.

Each device is tested before packaging using automated test equipment (ATE), in a process known as wafer testing, or wafer probing. The wafer is then cut into rectangular blocks, each of which is called a *die*. Each good die (plural *dice*, *dies*, or *die*) is then connected into a package using aluminium (or gold) bond wires which are welded and/or Thermosonic Bonded to *pads*, usually found around the edge of the die. After packaging, the devices go through final testing on the same or similar ATE used during wafer probing. Test cost can account for over 25% of the cost of fabrication on lower cost products, but can be negligible on low yielding, larger, and/or higher cost devices.

As of 2005, a fabrication facility (commonly known as a *semiconductor lab*) costs over \$1 billion to construct, because much of the operation is automated. The most advanced processes employ the following techniques:

- The wafers are up to 300 mm in diameter (wider than a common dinner plate).
- Use of 65 nanometer or smaller chip manufacturing process. Intel, IBM, NEC, and AMD are using 45 nanometers for their CPU chips. IBM and AMD are in development of a 45 nm process using immersion lithography.
- Copper interconnects where copper wiring replaces aluminium for interconnects.
- Low-K dielectric insulators.
- Silicon on insulator (SOI)
- Strained silicon in a process used by IBM known as strained silicon directly on insulator (SSDOI)

Packaging



Early USSR-made integrated circuit

The earliest integrated circuits were packaged in ceramic flat packs, which continued to be used by the military for their reliability and small size for many years. Commercial circuit packaging quickly moved to the dual in-line package (DIP), first in ceramic and later in plastic. In the 1980s pin counts of VLSI circuits exceeded the practical limit for DIP packaging, leading to pin grid array (PGA) and leadless chip carrier (LCC) packages. Surface mount packaging appeared in the early 1980s and became popular in the late 1980s, using finer lead pitch with leads formed as either gull-wing or J-lead, as exemplified by small-outline integrated circuit -- a carrier which occupies an area about 30 – 50% less than an equivalent DIP, with a typical thickness that is 70% less. This package has "gull wing" leads protruding from the two long sides and a lead spacing of 0.050 inches.

In the late 1990s, PQFP and TSOP packages became the most common for high pin count devices, though PGA packages are still often used for high-end microprocessors. Intel and AMD are currently transitioning from PGA packages on high-end microprocessors to land grid array (LGA) packages.

Ball grid array (BGA) packages have existed since the 1970s. Flip-chip Ball Grid Array packages, which allow for much higher pin count than other package types, were developed in the 1990s. In an FCBGA package the die is mounted upside-down (flipped) and connects to the package balls via a package substrate that is similar to a printed-circuit board rather than by wires. FCBGA packages allow an array of input-output signals (called Area-I/O) to be distributed over the entire die rather than being confined to the die periphery.

Traces out of the die, through the package, and into the printed circuit board have very different electrical properties, compared to on-chip signals. They require special design techniques and need much more electric power than signals confined to the chip itself.

When multiple dies are put in one package, it is called SiP, for *System In Package*. When multiple dies are combined on a small substrate, often ceramic, it's called an MCM, or Multi-Chip Module. The boundary between a big MCM and a small printed circuit board is sometimes fuzzy.

Chip labeling and manufacture date

Most integrated circuits large enough to include identifying information include four common sections: the manufacturer's name or logo, the part number, a part production batch number and/or serial number, and a four-digit code that identifies when the chip was manufactured. Extremely small surface mount technology parts often bear only a number used in a manufacturer's lookup table to find the chip characteristics.

The manufacturing date is commonly represented as a two-digit year followed by a two-digit week code, such that a part bearing the code 8341 was manufactured in week 41 of 1983, or approximately in October 1983.

Legal protection of semiconductor chip layouts

Like most of the other forms of intellectual property, IC layout designs are creations of the human mind. They are usually the result of an enormous investment, both in terms of the time of highly qualified experts, and financially. There is a continuing need for the creation of new layout-designs which reduce the dimensions of existing integrated circuits and simultaneously increase their functions. The smaller an integrated circuit, the less the material needed for its manufacture, and the smaller the space needed to accommodate it. Integrated circuits are utilized in a large range of products, including articles of everyday use, such as watches, television sets, washing machines, automobiles, etc., as well as sophisticated data processing equipment.

The possibility of copying by photographing each layer of an integrated circuit and preparing masks for its production on the basis of the photographs obtained is the main reason for the introduction of legislation for the protection of layout-designs.

A diplomatic conference was held at Washington, D.C., in 1989, which adopted a Treaty on Intellectual Property in Respect of Integrated Circuits (IPIC Treaty). The Treaty on Intellectual Property in respect of Integrated Circuits, also called Washington Treaty or IPIC Treaty (signed at Washington on May 26, 1989) is currently not in force, but was partially integrated into the TRIPs agreement.

National laws protecting IC layout designs have been adopted in a number of countries.

Other developments

In the 1980s, programmable integrated circuits were developed. These devices contain circuits whose logical function and connectivity can be programmed by the user, rather than being fixed by the integrated circuit manufacturer. This allows a single chip to be programmed to implement different LSI-type functions such as logic gates, adders and registers. Current devices named FPGAs (Field Programmable Gate Arrays) can now implement tens of thousands of LSI circuits in parallel and operate up to 1.5 GHz (Achronix holding the speed record).

The techniques perfected by the integrated circuits industry over the last three decades have been used to create microscopic machines, known as MEMS. These devices are used in a variety of commercial and military applications. Example commercial applications include DLP projectors, inkjet printers, and accelerometers used to deploy automobile airbags.

In the past, radios could not be fabricated in the same low-cost processes as microprocessors. But since 1998, a large number of radio chips have been developed using CMOS processes. Examples include Intel's DECT cordless phone, or Atheros's 802.11 card.

Future developments seem to follow the multi-core multi-microprocessor paradigm, already used by the Intel and AMD dual-core processors. Intel recently unveiled a prototype, "not for commercial sale" chip that bears 80 microprocessors. Each core is capable of handling its own task independently of the others. This is in response to the heat-versus-speed limit that is about to be reached using existing transistor technology. This design provides a new challenge to chip programming. Parallel programming languages such as the open-source X10 programming language are designed to assist with this task.

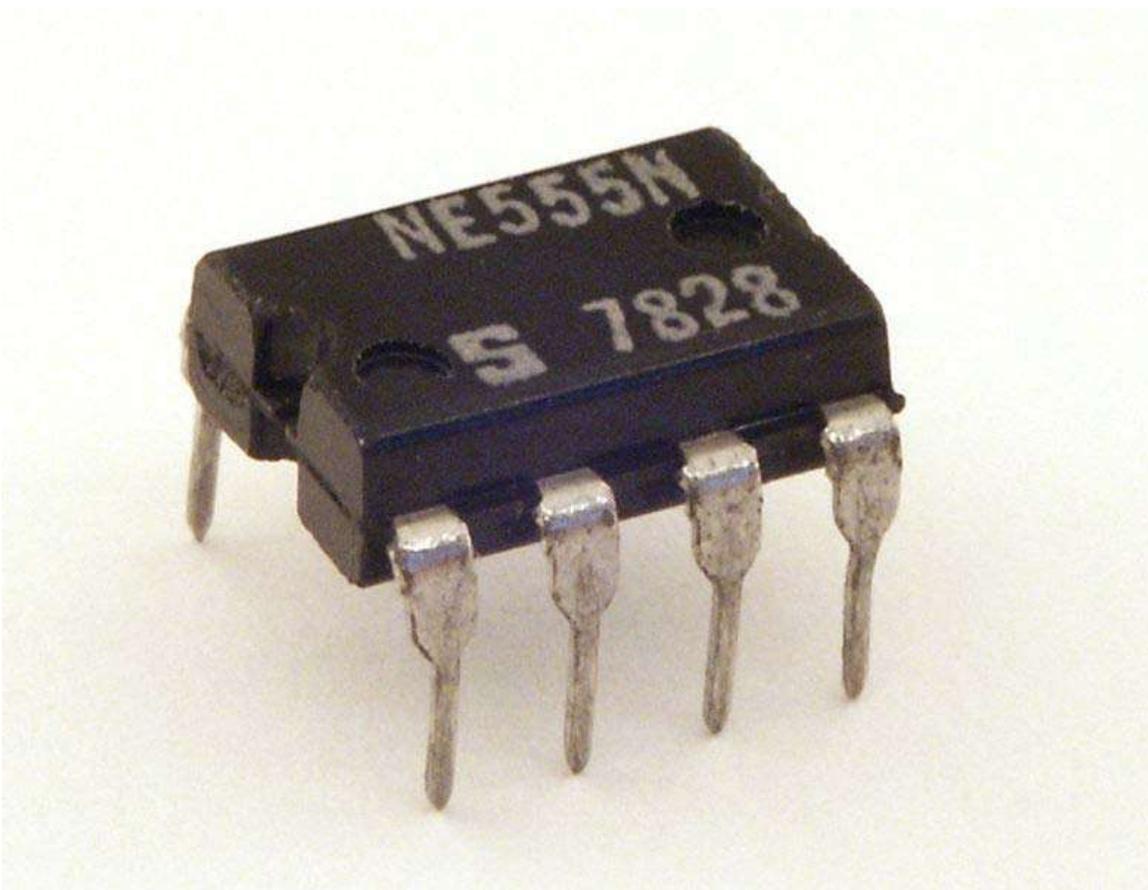
Silicon labelling and graffiti

To allow identification during production most silicon chips will have a serial number in one corner. It is also common to add the manufacturer's logo. Ever since ICs were created, some chip designers have used the silicon surface area for surreptitious, non-functional images or words. These are sometimes referred to as Chip Art, *Silicon Art*, *Silicon Graffiti* or *Silicon Doodling*.

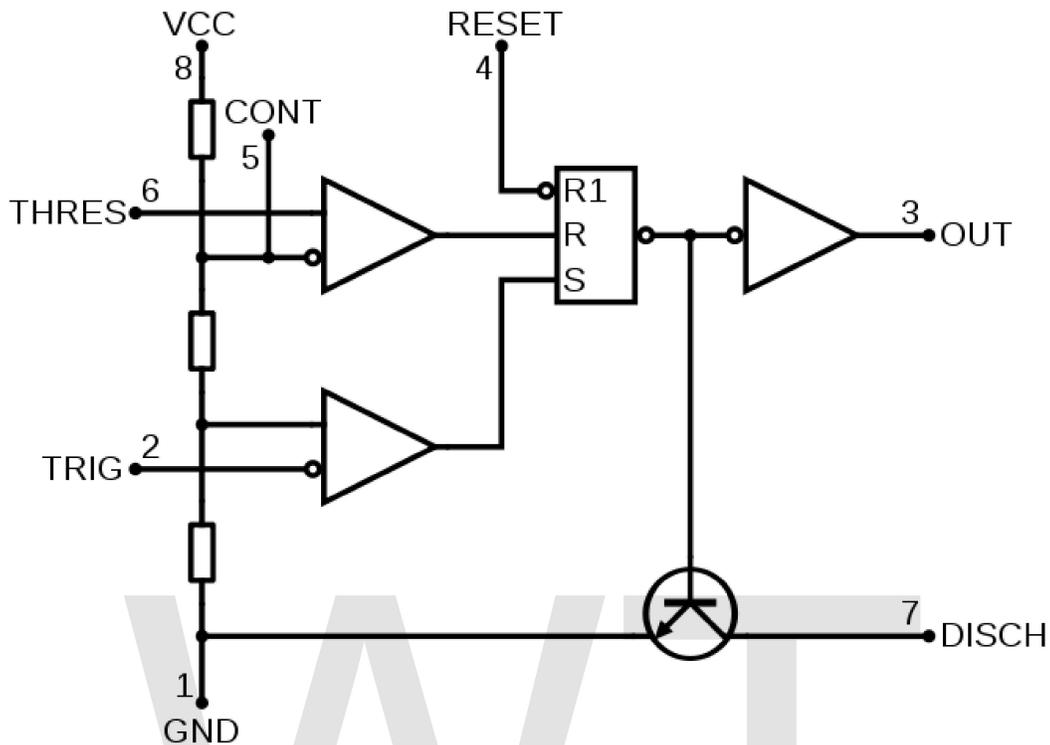


Chapter 2

555 timer IC



NE555 from Signetics in dual-in-line package



Internal block diagram

The **555 Timer IC** is an integrated circuit (chip) implementing a variety of timer and multivibrator applications. The IC was designed by Hans R. Camenzind in 1970 and brought to market in 1971 by Signetics (later acquired by Philips). The original name was the SE555 (metal can)/NE555 (plastic DIP) and the part was described as "The IC Time Machine". It has been claimed that the 555 gets its name from the three 5 kΩ resistors used in typical early implementations, but Hans Camenzind has stated that the number was arbitrary. The part is still in wide use, thanks to its ease of use, low price and good stability. As of 2003, it is estimated that 1 billion units are manufactured every year.

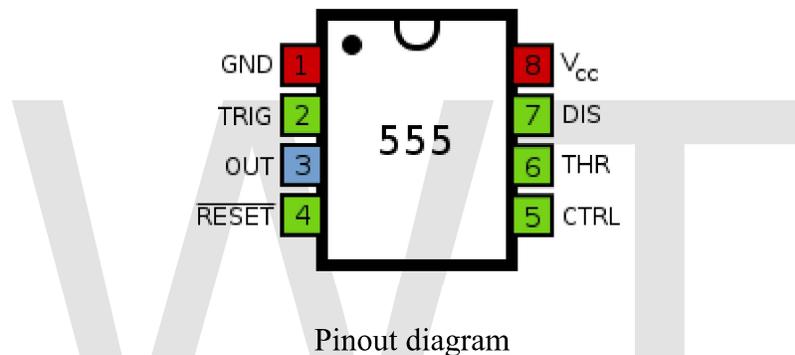
Depending on the manufacturer, the standard 555 package includes over 20 transistors, 2 diodes and 15 resistors on a silicon chip installed in an 8-pin mini dual-in-line package (DIP-8). Variants available include the 556 (a 14-pin DIP combining two 555s on one chip), and the 558 (a 16-pin DIP combining four slightly modified 555s with DIS & THR connected internally, and TR falling edge sensitive instead of level sensitive).

Ultra-low power versions of the 555 are also available, such as the 7555 and TLC555. The 7555 is designed to cause less supply glitching than the classic 555 and the manufacturer claims that it usually does not require a "control" capacitor and in many cases does not require a power supply bypass capacitor.

The 555 has three operating modes:

- Monostable mode: in this mode, the 555 functions as a "one-shot". Applications include timers, missing pulse detection, bouncefree switches, touch switches, frequency divider, capacitance measurement, pulse-width modulation (PWM) etc
- Astable - free running mode: the 555 can operate as an oscillator. Uses include LED and lamp flashers, pulse generation, logic clocks, tone generation, security alarms, pulse position modulation, etc.
- Bistable mode or Schmitt trigger: the 555 can operate as a flip-flop, if the DIS pin is not connected and no capacitor is used. Uses include bouncefree latched switches, etc.

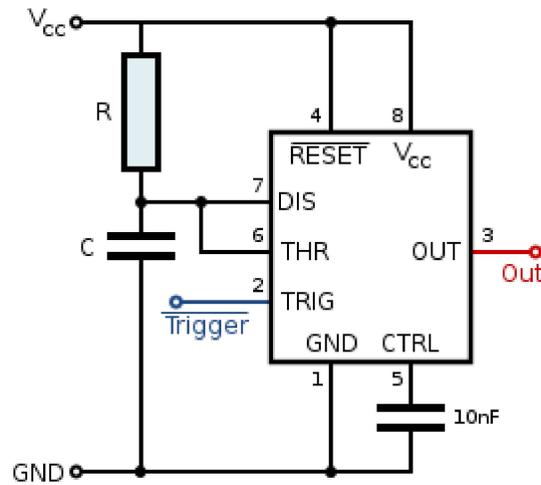
Usage



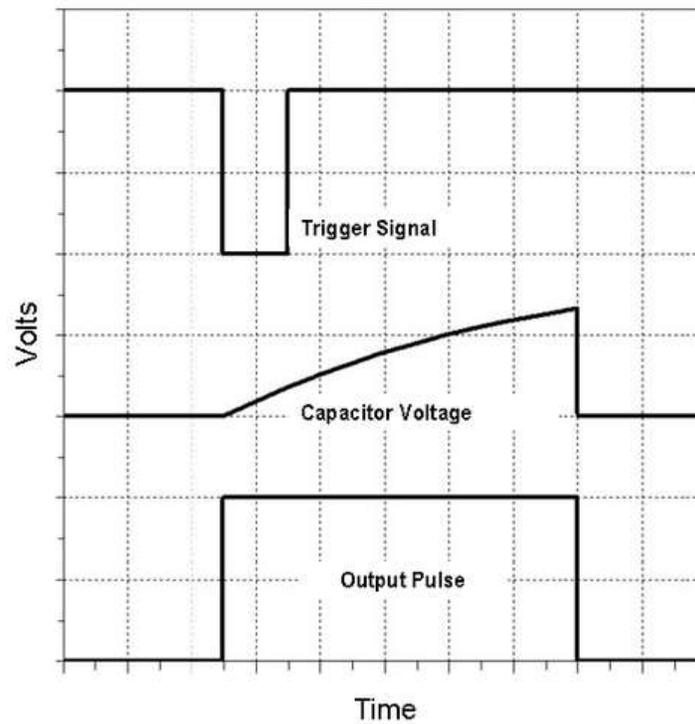
The connection of the pins is as follows:

| Pin | Name | Purpose |
|-----|-----------------|---|
| 1 | GND | Ground, low level (0 V) |
| 2 | TRIG | OUT rises, and interval starts, when this input falls below $1/3 V_{CC}$. |
| 3 | OUT | This output is driven to $+V_{CC}$ or GND. |
| 4 | RESET | A timing interval may be interrupted by driving this input to GND. |
| 5 | CTRL | "Control" access to the internal voltage divider (by default, $2/3 V_{CC}$). |
| 6 | THR | The interval ends when the voltage at THR is greater than at CTRL. |
| 7 | DIS | Open collector output; may discharge a capacitor between intervals. |
| 8 | $V+$, V_{CC} | Positive supply voltage is usually between 3 and 15 V. |

Monostable mode



Schematic of a 555 in monostable mode



The relationships of the trigger signal, the voltage on C and the pulse width in monostable mode

In the monostable mode, the 555 timer acts as a “one-shot” pulse generator. The pulse begins when the 555 timer receives a signal at the trigger input that falls below a third of the voltage supply. The width of the output pulse is determined by the time constant of an RC network, which consists of a capacitor (C) and a resistor (R). The output pulse ends when the charge on the C equals 2/3 of the supply voltage. The output pulse width can be lengthened or shortened to the need of the specific application by adjusting the values of R and C.

The output pulse width of time t , which is the time it takes to charge C to 2/3 of the supply voltage, is given by

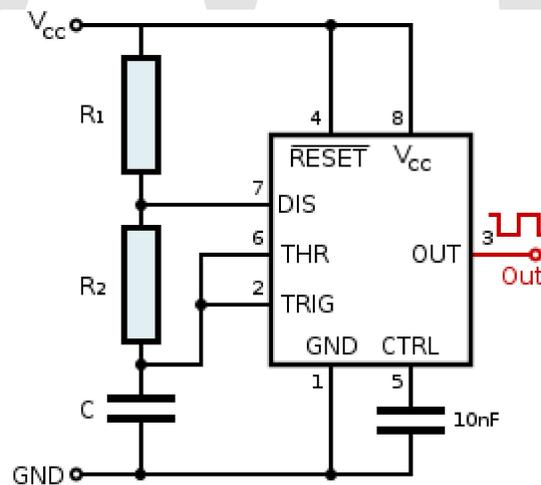
$$t = RC \ln(3) \approx 1.1RC$$

where t is in seconds, R is in ohms and C is in farads.

Bistable Mode

In bistable mode, the 555 timer acts as a basic flip-flop. The trigger and reset inputs (pins 2 and 4 respectively on a 555) are held high via pull-up resistors while the threshold input (pin 6) is simply grounded. Thus configured, pulling the trigger momentarily to ground acts as a 'set' and transitions the output pin (pin 3) to V_{CC} (high state). Pulling the reset input to ground acts as a 'reset' and transitions the output pin to ground (low state). No capacitors are required in a bistable configuration. Pins 5 and 7 (control and discharge) are left floating.

Astable mode



Standard 555 Astable Circuit

In astable mode, the 555 timer puts out a continuous stream of rectangular pulses having a specified frequency. Resistor R_1 is connected between V_{CC} and the discharge pin (pin 7)

and another resistor (R_2) is connected between the discharge pin (pin 7), and the trigger (pin 2) and threshold (pin 6) pins that share a common node. Hence the capacitor is charged through R_1 and R_2 , and discharged only through R_2 , since pin 7 has low impedance to ground during output low intervals of the cycle, therefore discharging the capacitor.

In the astable mode, the frequency of the pulse stream depends on the values of R_1 , R_2 and C :

$$f = \frac{1}{\ln(2) \cdot C \cdot (R_1 + 2R_2)}$$

The high time from each pulse is given by

$$\text{high} = \ln(2) \cdot (R_1 + R_2) \cdot C$$

and the low time from each pulse is given by

$$\text{low} = \ln(2) \cdot R_2 \cdot C$$

where R_1 and R_2 are the values of the resistors in ohms and C is the value of the capacitor in farads.

To achieve a duty cycle of less than 50% a diode can be added in parallel with R_2 towards the capacitor. This bypasses R_2 during the high part of the cycle so that the high interval depends only on R_1 and C_1 .

Specifications

These specifications apply to the NE555. Other 555 timers can have different specifications depending on the grade (military, medical, etc).

| | |
|---------------------------------------|----------------------|
| Supply voltage (V_{CC}) | 4.5 to 15 V |
| Supply current ($V_{CC} = +5$ V) | 3 to 6 mA |
| Supply current ($V_{CC} = +15$ V) | 10 to 15 mA |
| Output current (maximum) | 200 mA |
| Maximum Power dissipation | 600 mW |
| Power Consumption (minimum operating) | 30 mW@5V, 225 mW@15V |
| Operating temperature | 0 to 70 °C |

Derivatives

Many pin-compatible variants, including CMOS versions, have been built by various companies. Bigger packages also exist with two or four timers on the same chip. The 555 is also known under the following type numbers:

| Manufacturer | Model | Remark |
|--------------------------|---------------------|-------------------------------------|
| Custom Silicon Solutions | CSS555/CSS555C | CMOS from 1.2 V, $I_{DD} < 5 \mu A$ |
| Avago Technologies | Av-555M | |
| ECG Philips | ECG955M | |
| Exar | XR-555 | |
| Fairchild Semiconductor | NE555/KA555 | |
| Harris | HA555 | |
| IK Semicon | ILC555 | CMOS from 2 V |
| Intersil | SE555/NE555 | |
| Intersil | ICM7555 | CMOS |
| Lithic Systems | LC555 | |
| Maxim | ICM7555 | CMOS from 2 V |
| Motorola | MC1455/MC1555 | |
| National Semiconductor | LM1455/LM555/LM555C | |
| National Semiconductor | LMC555 | CMOS from 1.5 V |
| NTE Sylvania | NTE955M | |
| Raytheon | RM555/RC555 | |
| RCA | CA555/CA555C | |
| STMicroelectronics | NE555N/ K3T647 | |
| Texas Instruments | SN52555/SN72555 | |
| Texas Instruments | TLC555 | CMOS from 2 V |
| USSR | K1006ВИ1 | |
| Zetex | ZSCT1555 | down to 0.9 V |
| NXP Semiconductors | ICM7555 | CMOS |
| HFO / East Germany | B555 | |

Dual timer 556

The dual version is called 556. It features two complete 555s in a 14 pin DIL package.

Quad timer 558

The quad version is called 558 and has 16 pins. To fit four 555s into a 16 pin package the control voltage and reset lines are shared by all four modules. Also for each module the discharge and threshold are internally wired together and called *timing*.

Example applications

Joystick interface circuit using quad timer 558

The original IBM personal computer used a quad timer 558 in monostable (or "one-shot") mode to interface up to two joysticks to the host computer. In the joystick interface circuit of the IBM PC, the capacitor (C) of the RC network was generally a 10 nF capacitor. The resistor (R) of the RC network consisted of the potentiometer inside the joystick along with an external resistor of 2.2 kilohms. The joystick potentiometer acted as a variable resistor. By moving the joystick, the resistance of the joystick increased from a small value up to about 100 kilohms. The joystick operated at 5 V.

Software running in the host computer started the process of determining the joystick position by writing to a special address (ISA bus I/O address 201h). This would result in a trigger signal to the quad timer, which would cause the capacitor (C) of the RC network to begin charging and cause the quad timer to output a pulse. The width of the pulse was determined by how long it took the C to charge up to 2/3 of 5 V (or about 3.33 V), which was in turn determined by the joystick position.

Software running in the host computer measured the pulse width to determine the joystick position. A wide pulse represented the full-right joystick position, for example, while a narrow pulse represented the full-left joystick position.

Atari Punk Console

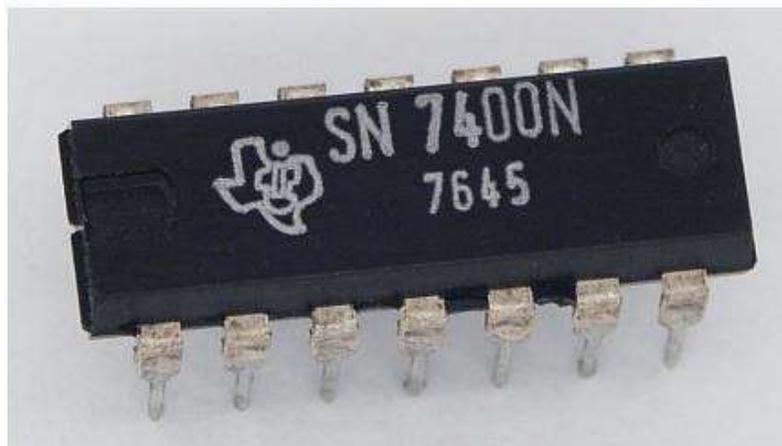
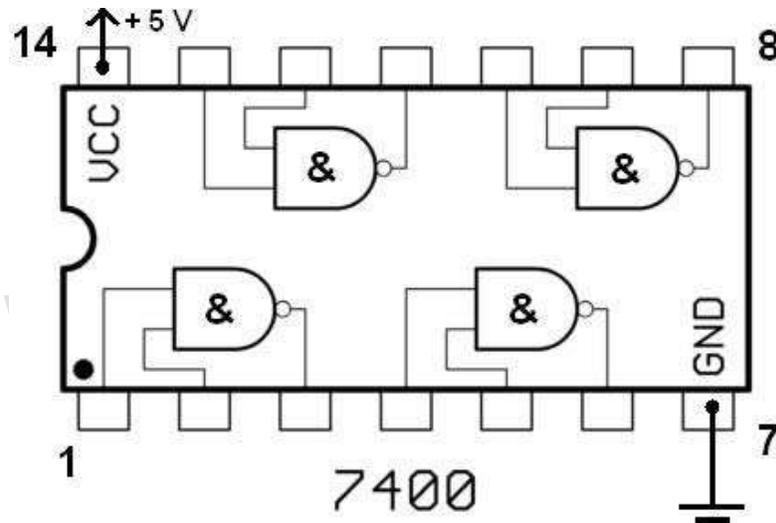
One of Forrest M. Mims III's many books was dedicated to the 555 timer. In it, he first published the "Stepped Tone Generator" circuit which has been adopted as a popular circuit, known as the Atari Punk Console, by circuit benders for its distinctive low-fi sound similar to classic Atari games.

Pulse Width Modulation

The 555 can be used to generate a variable PWM signal using a few external components. The chip alone can drive small external loads or an amplifying transistor for larger loads.

Chapter 3

7400 series



The 7400 chip, containing four NANDs. The second line of smaller numbers (7645) is a date code; this chip was manufactured in the 45th week of 1976. The N suffix on the part number is a vendor-specific code indicating PDIP packaging.

The **7400 series** of transistor-transistor logic (TTL) integrated circuits are historically important as the first widespread family of TTL integrated circuit logic. It was used to build the mini and mainframe computers of the 1960s and 1970s. Several generations of pin-compatible descendants of the original family have since been *de-facto* standard components.

Overview

The 7400 series contains hundreds of devices that provide everything from basic logic gates, flip-flops, and counters, to special purpose bus transceivers and Arithmetic Logic Units (ALU). Specific functions are described in a list of 7400 series integrated circuits.

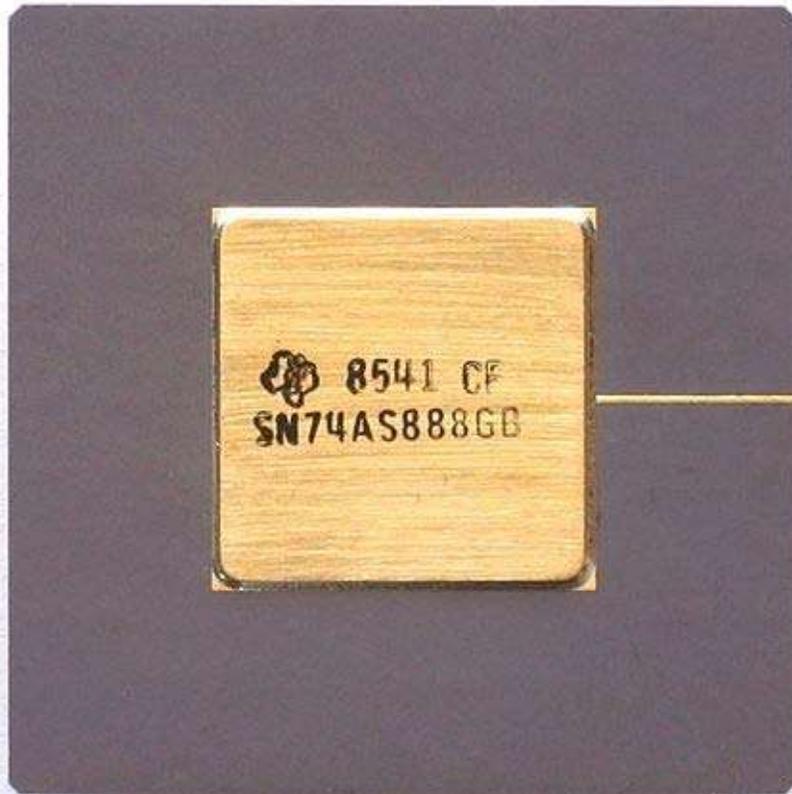
Today, surface-mounted CMOS versions of the 7400 series are used in various applications in electronics and for glue logic in computers and industrial electronics. The original through-hole devices in dual in-line packages (DIP/DIL), which were the mainstay of the industry for many decades, are very useful for rapid breadboard-prototyping and education and so remain available from most manufacturers. The fastest types and very low voltage versions are typically surface-mount only, however.

The 14-pin DIP shown to the right is an example of a 7400 part. The chip contains four two-input NAND gates. Each gate uses two pins for input and one pin for its output, and the remaining two contacts supply power (+5 V) and connect the ground.

While designed as a family of digital logic, it was not unusual to see TTL chips in analog circuits like Schmitt triggers. Like the 4000 series, the newer CMOS versions of the 7400 series are also usable as analog amplifiers using negative feedback (similar to operational amplifiers with only an inverting input).

The former Soviet Union manufactured the K155ЛA3 which was pin-compatible with the 7400 part available in the United States, except for using a metric spacing of 2.5mm between pins instead of the 1/10"-based (2.54mm) spacing used in the west.

7400 series derivative families



Part of the 7400 series:

Cascadable 8-bit ALU Texas Instruments SN74AS888

7400 series parts were constructed using bipolar transistors, forming what is referred to as transistor–transistor logic or **TTL**. Newer series, more or less compatible in function and logic level with the original parts, use CMOS technology or a combination of the two (BiCMOS). Originally the bipolar circuits provided higher speed but consumed more power than the competing 4000 series of CMOS devices. Bipolar devices are also limited to a fixed power supply voltage, typically 5 V, while CMOS parts often support a range of supply voltages.

Milspec-rated devices for use in extended temperature conditions are available as the 5400 series. Texas Instruments also manufactured radiation-hardened devices with the prefix *RSN*, and the company offered beam-lead bare dice for integration into hybrid circuits with a *BL* prefix designation.

Regular speed TTL parts were also available for a time in the 6400 series - these had an extended industrial temperature range of -40 C to $+85\text{ C}$. While companies such as Mullard listed 6400-series compatible parts in 1970 data sheets, by 1973 there was no mention of the 6400 family in the Texas Instruments *TTL Data Book*. Some companies have also offered industrial extended temperature range variants using the regular 7400 series part numbers with a prefix or suffix to indicate the temperature grade.

As integrated circuits in the 7400 series were made in different technologies, usually compatibility was retained with the original TTL logic levels and power supply voltages. Strictly, an integrated circuit made in CMOS is no longer a TTL chip since it uses field-effect transistors (FETs) and not bipolar junction transistors, but similar part numbers are retained to identify similar logic functions and electrical (power and I/O voltage) compatibility in the different subfamilies. Over 40 different logic subfamilies use this standardized part number scheme.

- Bipolar
 - 74 - the "standard TTL" logic family had no letters between the "74" and the specific part number.
 - 74L - Low power (compared to the original TTL logic family), very slow
 - H - High speed (still produced but generally superseded by the S-series, used in 1970s era computers)
 - S - Schottky (obsolete)
 - LS - Low Power Schottky
 - AS - Advanced Schottky
 - ALS - Advanced Low Power Schottky
 - F - Fast (faster than normal Schottky, similar to AS)
- CMOS
 - C - CMOS 4–15 V operation similar to buffered 4000 (4000B) series
 - HC - High speed CMOS, similar performance to LS, 12 nS
 - HCT - High speed, compatible logic levels to bipolar parts
 - AC - Advanced CMOS, performance generally between S and F
 - AHC - Advanced High-Speed CMOS, three times as fast as HC
 - ALVC - Low voltage - 1.65 to 3.3 V, Time Propagation Delay (TPD) 2 nS
 - AUC - Low voltage - 0.8 to 2.7 V, TPD < 1.9 nS@1.8 V
 - FC - Fast CMOS, performance similar to F
 - LCX - CMOS with 3 V supply and 5 V tolerant inputs
 - LVC - Low voltage – 1.65 to 3.3 V and 5 V tolerant inputs, tpd < 5.5 nS@3.3 V, tpd < 9 nS@2.5 V
 - LVQ - Low voltage - 3.3 V
 - LVX - Low voltage - 3.3 V with 5 V tolerant inputs

- VHC - Very High Speed CMOS - 'S' performance in CMOS technology and power
- BiCMOS
 - BCT - BiCMOS, TTL-compatible input thresholds, used for buffers
 - ABT - Advanced BiCMOS, TTL-compatible input thresholds, faster than ACT and BCT

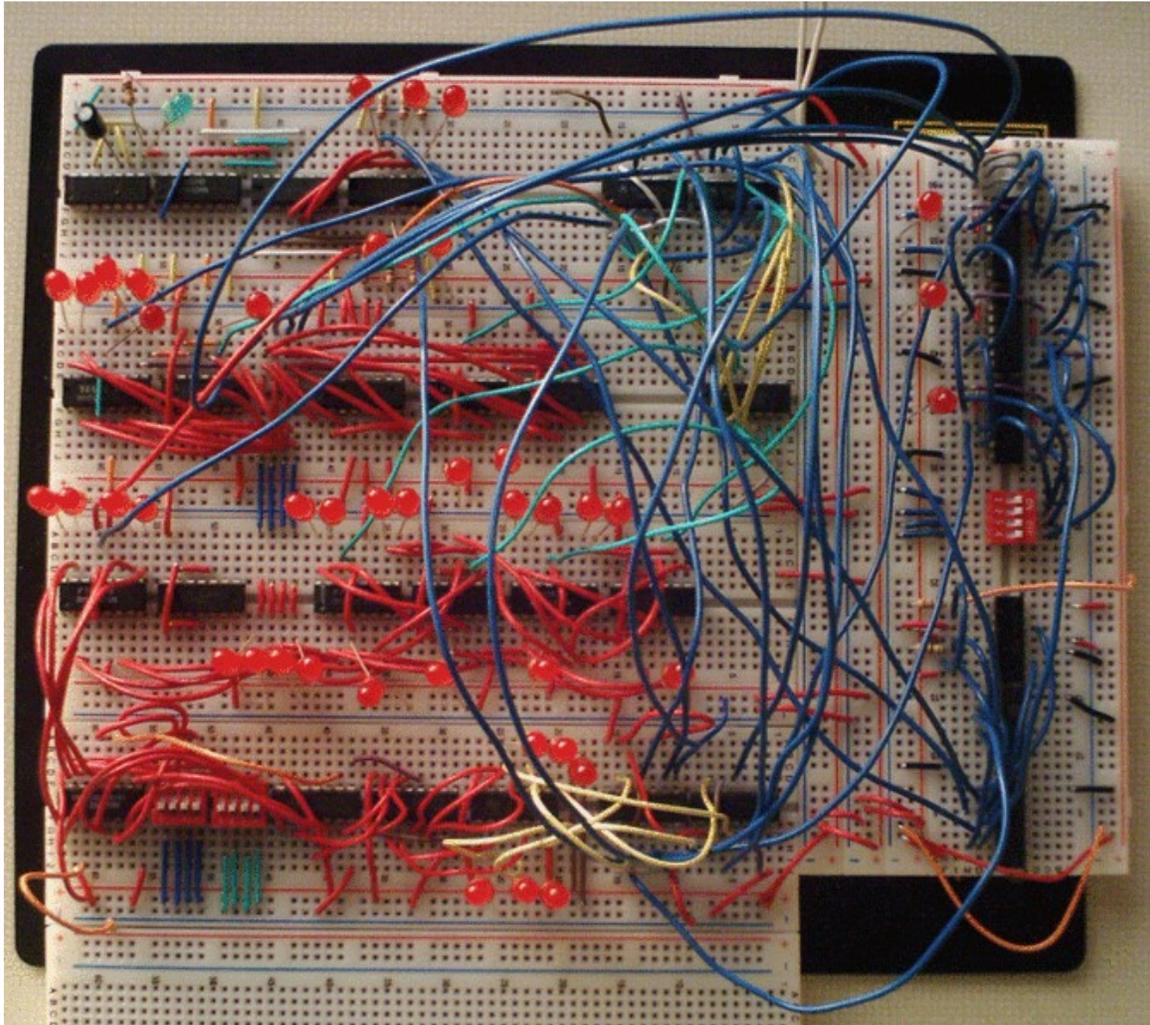
Many parts in the CMOS HC, AC, and FC families are also offered in "T" versions (HCT, ACT, and FCT) which have input thresholds that are compatible with both TTL and 3.3 V CMOS signals. The non-T parts have conventional CMOS input thresholds. The ACTQ family introduced by Fairchild utilizes Quiet Series technology to guarantee quiet output switching and improved dynamic threshold performance. FACT Quiet Series features GTO output control and undershoot corrector in addition to a split ground bus for superior performance. The 74H family is the same basic design as the 7400 family with resistor values reduced. This reduced the typical propagation delay from 9 ns to 6 ns but increased the power consumption. The 74H family provided a number of unique devices for CPU designs in the 1970s. Many designers of military and aerospace equipment used this family over a long period and as they need exact replacements, this family is still produced by Lansdale Semiconductor.

The 74S family, using Schottky circuitry, uses more power than the 74, but is faster. The 74LS family of ICs is a lower-power version of the 74S family, with slightly higher speed but lower power dissipation than the original 74 family; it became the most popular variant once it was widely available.

The 74F family was introduced by Fairchild Semiconductor and adopted by other manufacturers; it is faster than the 74, 74LS and 74S families.

Through the late 1980s and 1990s newer versions of this family were introduced to support the lower operating voltages used in newer CPU devices.

History



A 4-bit, 2 register, six assembly language instruction computer made entirely of 74-series chips.

Although the 7400 series was the first *de facto* industry standard TTL logic family, second-sourced by several semiconductor companies, there were earlier TTL logic families such as the Sylvania SUHL family, Motorola MC4000 MTTL family (not to be confused with RCA CD4000 CMOS), the National Semiconductor DM8000 family, Fairchild 9300 series, and the Signetics 8200 family.

The 7400N quad NAND gate was the first product in the series.

The 5400 and 7400 series were used in many popular minicomputers in the seventies and early eighties. The DEC PDP series 'minis' used the 74181 ALU as the main computing element in the CPU. Other examples were the Data General Nova series and Hewlett-Packard 21MX, 1000, and 3000 series.

Hobbyists and students equipped with wire wrap tools, a 'breadboard' and a 5-volt power supply could also experiment with digital logic referring to how-to articles in *Byte* magazine and *Popular Electronics* which featured circuit examples in nearly every issue. In the early days of large-scale IC development, a prototype of a new large-scale integrated circuit might have been developed using TTL chips on several circuit boards, before committing to manufacture of the target device in IC form. This allowed simulation of the finished product and testing of the logic before the availability of software simulations of integrated circuits.

In 1965, typical quantity-one pricing for the SN5400 (military grade, in ceramic welded flat-pack) was around 22 USD. As of 2007, individual commercial-grade chips in molded epoxy (plastic) packages can be purchased for approximately 0.25 USD each, depending on the particular chip. Purchased in bulk the price per unit falls even lower.

Part numbering scheme

The part numbers for 7400 series logic devices often use the following naming convention, though specifics vary between manufacturers.

- First, although sometimes omitted, a two or three letter prefix which indicates the manufacturer of the device (e.g. SN for Texas Instruments, DM for National Semiconductor) although these codes are no longer closely associated with a single manufacturer, for example Fairchild Semiconductor manufactures parts with MM and DM prefixes, and none.
- A two-figure secondary prefix, of which the two most common are "74", indicating a commercial temperature range device and "54", indicating an extended (military) temperature range
- Up to four letters describing the logic subfamily, as listed above (e.g. "LS" or "HCT").
- Two or more digits assigned for each device, e.g. 00 for a quad 2-input NAND gate. There are hundreds of different devices in each family. The allocation of device numbers (and, with a few exceptions, the pin-outs) of the original 7400 family was carried across to the later families, and new numbers allocated for new functions, plus some of the competing CD4000 numbers and pin-outs were included over time. There is no pattern to the allocation of these numbers. The function and pin-out of the chip is nearly always the same for the same device number regardless of subfamily manufacturer – exceptions are discussed below.
- Additional suffix letters and numbers may be attached to indicate the package type, quality grade, or other information, but this varies widely by manufacturer.

For example **SN74ALS245N** means this is a device probably made by Texas Instruments (SN), it is a commercial temperature range TTL device (74), it is a member of the "advanced low-power Schottky" family (ALS), and it is a *bi-directional eight-bit buffer* (245) in a plastic through-hole DIP package (N).

Many logic families maintain a consistent use of the device numbers, as an aid to designers. Often a part from a different 74x00 subfamily could be substituted ("drop-in replacement") in a circuit, with the same function and pin-out yet more appropriate characteristics for an application (perhaps speed or power consumption), which was a large part of the appeal of the 74C00 series over the competing CD4000B series, for example. But there are a few exceptions where incompatibilities (mainly in pin-out) across the subfamilies occurred, such as:

- some flat-pack devices (e.g. 7400W) and surface-mount devices,
- some of the faster CMOS series (for example 74AC)
- a few low-power TTL devices (e.g. 74L86, 74L9 and 74L95) have a different pin-out than the regular (or even 74LS) series part.
- five versions of the 74x54 (4-Wide AND-OR-INVERT gates IC), namely 7454(N), 7454W, 74H54, 74L54W and 74L54N/74LS54, are different from each other in pin-out and/or function!).

Some manufacturers such as Mullard and Siemens had pin-compatible TTL parts in the FJ series (e.g. an FJH131 was the equivalent of the 7400) but with a completely different numbering scheme, however, data sheets identified the *7400-compatible* number as an aid to recognition.

Second sources in Europe and the Eastern Block

At the time the 7400 series was being made, some European manufacturers (that traditionally followed the Pro Electron naming convention) such as Philips/Mullard produced a series of TTL integrated circuits with part names beginning FJ. Some examples of FJ series are:

- FJH101 (=7430) Single 8-input NAND gate,
- FJH131 (=7400) Quadruple 2-input NAND gate,
- FJH181 (=7454N or J) 2+2+2+2 input AND-OR-NOT gate.

The Soviet Union started manufacturing TTL ICs with 7400 series pin-out in late 1960s and early 1970s. Part numbering is somewhat different from the Western series:

- the technology modifications were considered different series, and were identified by different numbered prefixes - 155 series is equivalent to plain 74, 131 series is 74H, 158 series is 74L, 531 series is 74S, 555 series is 74LS, and 1530/1531/1533 are 74F/74AS/74ALS respectively. CMOS elements with TTL pin-out are also available, for example 1564 series is equivalent to 74HC.
- the function of the unit is described with a two-letter code followed by a number
 - the first letter represents the functional group - logical, triggers, counters, multiplexers, etc.
 - the second letter shows the functional subgroup making the distinction between logical NAND and NOR, D- and JK-triggers, decimal and binary counters, etc.

- the number distinguishes variants with different number of inputs, or different number of elements within a die - JIA1/JIA2/JIA3 (LA1/LA2/LA3) are 2 four-input / 1 eight-input / 4 two-input NAND elements respectively (equivalent to 7420/7430/7400)

Before July 1974 the two letters from the functional description were inserted after the first digit of the series. Examples: K1JIB551 and K155JIA1 (7420), K1TM552 and K155TM2 (7474) are the same ICs made at different times.

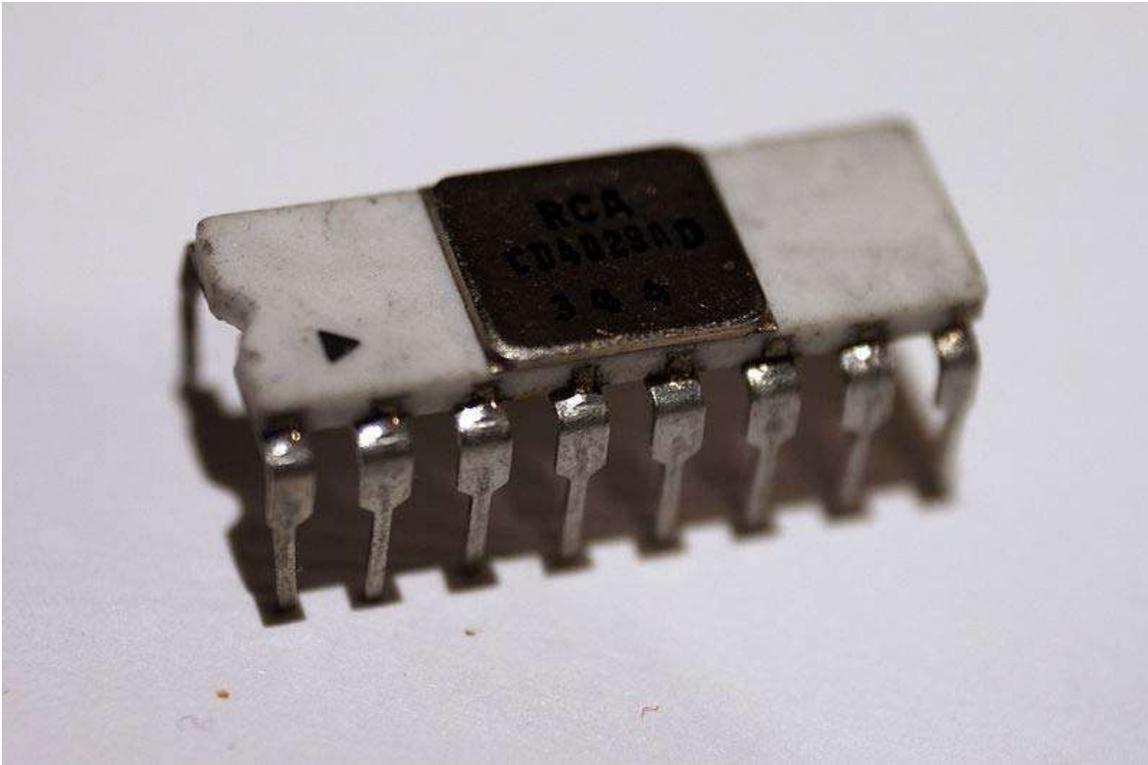
Clones of the plain 7400 series were also made in other Eastern Block countries

- Poland and Czechoslovakia used the 7400 numbering scheme with manufacturer prefixes UCY and MH respectively. Examples: UCY7400, and MH7400.
- Hungary also used the 7400 numbering scheme but with manufacturer suffix - 7400 is marked as 7400APC.
- Romania used trimmed 7400 numbering with own manufacturing prefix CDB. Example: CDB4123E is 74123.
- East Germany also used trimmed 7400 numbering without manufacturer prefix or suffix. The prefix D (or E) designates digital IC, and not the manufacturer. Example: D174 is 7474.

No information is available for technology modifications (74H, 74LS, etc.) manufactured outside USSR.

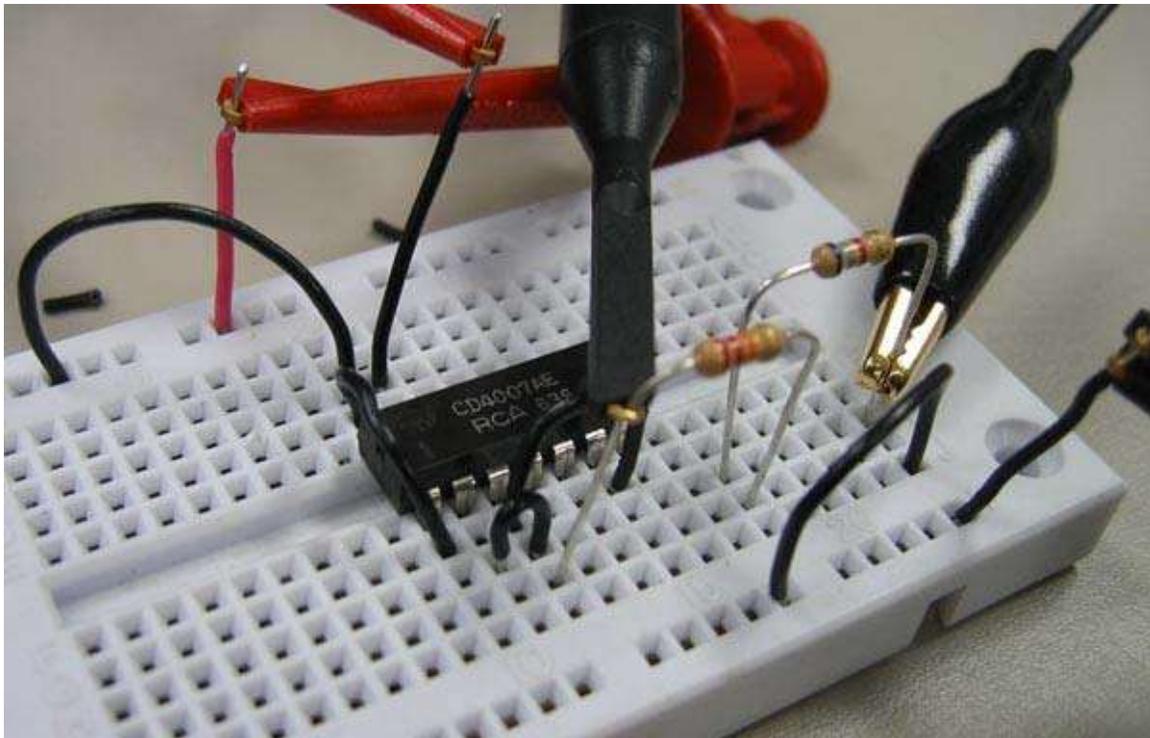
Chapter 4

4000 series



A very early CD4029 counter IC, manufactured by RCA.

The **4000 series** (originally: CD4000 series, but now includes HEF4000 series, etc.) is a family of industry standard integrated circuits which implement a variety of logic functions using Complementary Metal–Oxide–Semiconductor technology. They were introduced by RCA as *CD4000 COS/MOS* in 1968, as a lower power and more versatile alternative to the 7400 series of TTL logic chips. Almost all IC manufacturers active during the era fabricated chips from this series. RCA sometimes advertised the line as COSMOS, standing for Complementary Symmetry Metal-Oxide Semiconductor. The naming system followed the RCA convention of CA for analog, CD for digital, but did not relate to the Texas Instruments SN7400 series numbering scheme.



The CD4007 on a breadboard

For many years, the 4000 series devices could not operate at speeds as fast as the popular 7400 TTL chips, but had the advantage of much lower power consumption, the ability to operate over a much wider range of supply voltages (3V to 15V), and simpler circuit design due to the vastly increased fanout. However their slower speed (initially only capable of about 1 MHz operation, compared with TTL's 10 MHz) limited their applications to static or slow speed designs. Later, new fabrication technology largely overcame the speed problems, while retaining backward compatibility with most circuit designs. Although all semiconductors can be damaged by electrostatic discharge, the high impedance of CMOS inputs makes them more susceptible than bipolar transistor-based, TTL, devices. Eventually, the advantages of CMOS (especially the later series such as 74HC) edged out the older TTL chips, but at the same time ever increasing LSI techniques edged out the modular chip approach to design. The 4000 series is still widely available, but perhaps less important than it was two decades ago.

The series was extended in the late 1970s and 1980s to include new types which implemented new or more greatly integrated functions, or were better versions of existing chips in the 4000 series. Most of these newer chips were given 45xx and 45xxx designations, but are usually still regarded by engineers as part of the 4000 series.

In the 1990s, some manufacturers (e.g. Texas Instruments) ported the 4000 series to their newer HCMOS technology with devices such as the 74HCT4060 providing equivalent functionality to a 4060 IC but with greater speed.

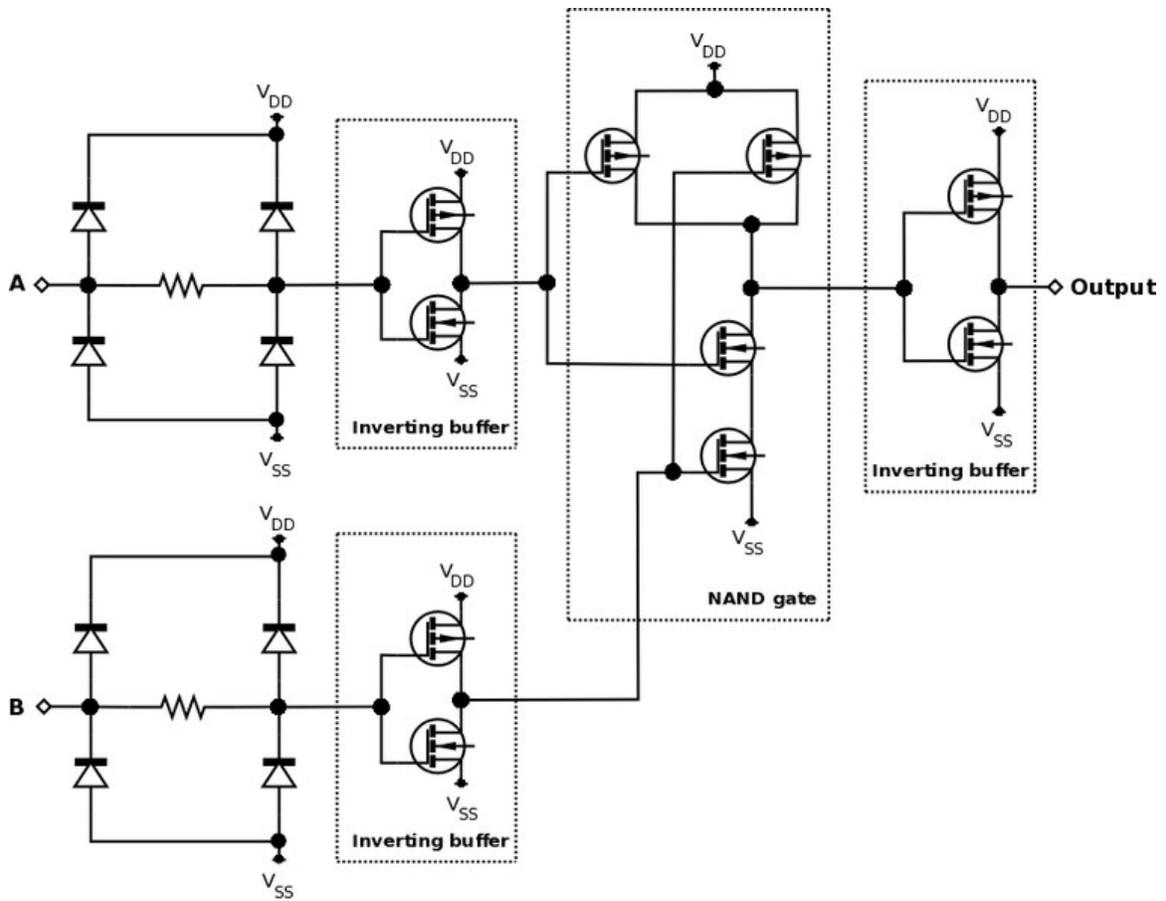
The 4000 series integrated circuits have been used in space satellites for many decades. As of 2000, people building satellites still use 4000 series integrated circuits.

Design considerations

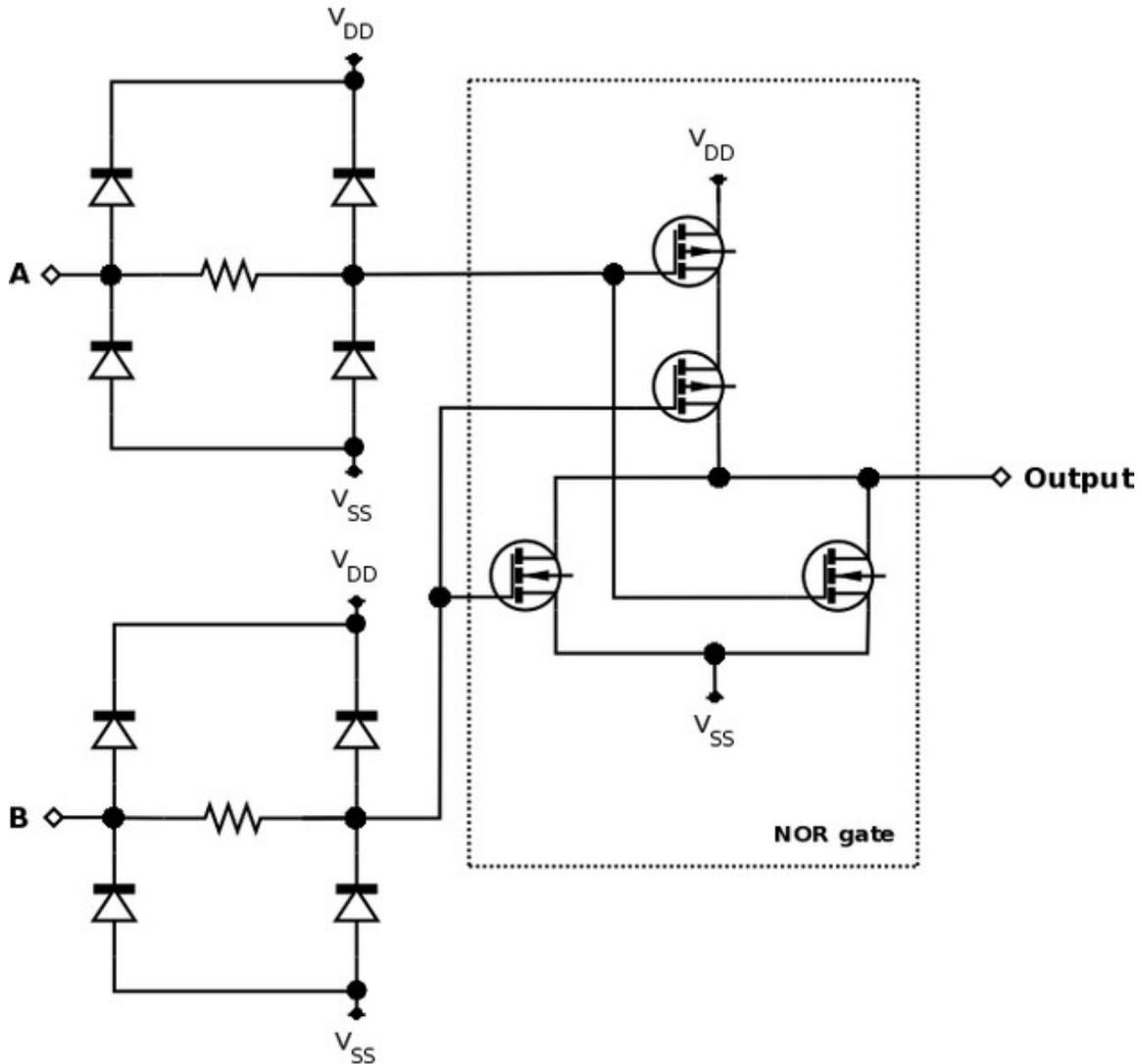
The original 4000 series was available in two versions: The A series was unbuffered, while the B series featured buffered inputs and outputs (in the form of additional, simple logic gates). The buffered outputs are able to source or sink more current than the unbuffered outputs, thus eliminating the need for discrete switching transistors in some designs. The buffered versions also have faster output switching times, as the signal rise time of the buffered output stage is faster than that of an unbuffered device, however the overall propagation delay through the buffered versions is higher due to the additional circuitry. The buffered devices are less susceptible to output oscillation with slow-changing inputs so designers have to weigh up the pros and cons of using buffered or unbuffered parts according to the nature of the circuit in which the devices are being used. The additional input and output gates on the buffered parts also make them marginally less susceptible to damage by electrostatic discharge (ESD).

Although the original designation for unbuffered and buffered parts was the addition of an 'A' or 'B' suffix to the part code (eg: 4000A = unbuffered, 4000B = buffered), some manufacturers (eg: Texas Instruments) later changed to using UB (unbuffered) and B (buffered) suffixes (eg: 4000UB and 4000B).

The diagrams below show the construction differences between a simple buffered and unbuffered CMOS NOR logic gate. Note that the logic gate at the core of the buffered part is actually a NAND gate, but the overall function of the complete circuit is a NOR gate due to the logic inversions performed by the buffers. (A negated NAND becomes a NOR as defined by De Morgan's laws in Boolean Algebra.) The clamping diodes on the inputs are to offer some protection against ESD.



Buffered CMOS two input NOR gate



Unbuffered CMOS two input NOR gate

The 4000 series permits the use of "cookbook design" at least for slow design, where standard circuit elements can be created, shared, and connected to other circuits with few, if any, connection difficulties. This greatly speeds the design of new hardware by reusing standard approaches to circuit design. In contrast, TTL circuits, while similarly modular, often require much more careful interfacing, since the limited fanout (and fan-in) require that the loading of each output be carefully considered. Some later TTL families, like 74LS reduce this problem with fanouts of 20. It is also much easier to prototype LSI designs using the 4000 series and get repeatable and transferable results when moving to the more integrated design.

Some care needs to be taken with the design of circuits using CMOS chips. Many parts offer multiple logic gates in a single package and it is common to not need or use all of them. An engineer who forgets to 'tie off' (connect the unused gate inputs to VSS or VDD

through pull-down or pull-up resistors) may find the chip draws excessive current. The problem is caused by biasing in each gate. With the inputs disconnected, the gate will bias itself into a linear mode where the outputs are partially switched; this leaves the output buffer drawing a great deal of current since it isn't fully on or off, creating a low resistance current path between the power supply rails.

Example common 4000 series chips

- 4000 - Dual 3-Input NOR Gate and Inverter
- 4001 - Quad 2-Input NOR Gate
- 4002 - Dual 4-Input NOR Gate OR Gate
- 4008 - 4-Bit Full Adder
- 4010 - hex non-inverting buffer
- 4011 - Quad 2-Input NAND Gate

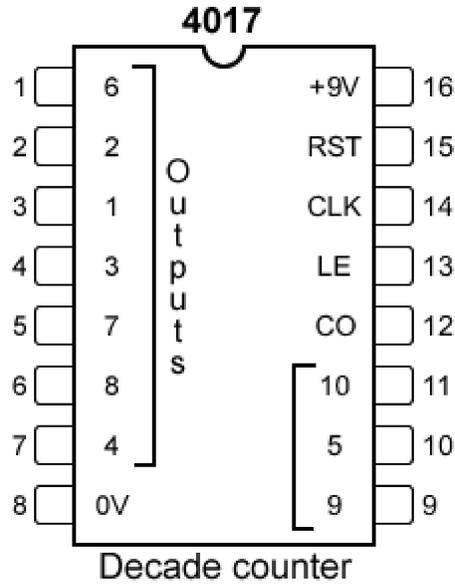
Notable parts

A few parts are notable in the 4000 series because of their level of integration compared to other chips. This list is intentionally incomplete and is meant to provide a sample of the more interesting parts in the series. Devices useful for switching analog signals (such as the 4066, and 4051 to 4053) have continued to enjoy popularity in some audio designs (although non-4000 series chips, often with less distortion, are now available).

4017 decade counter

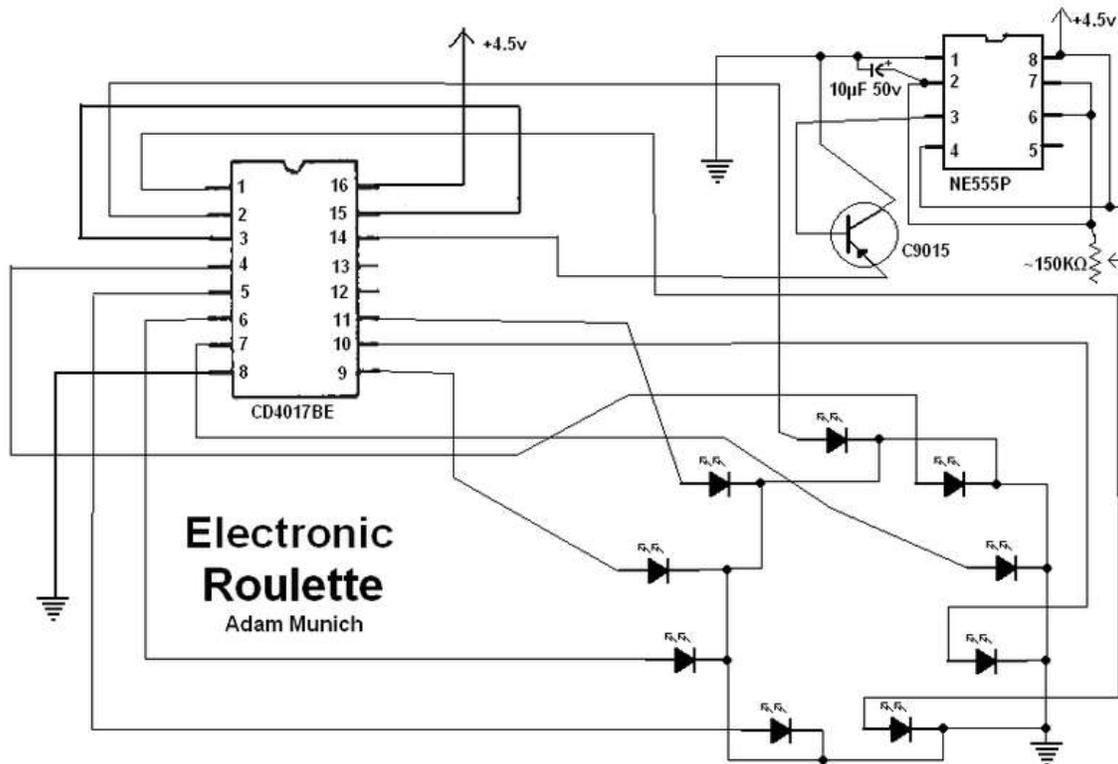
The **4017 IC** is a 16-pin CMOS decade counter from the 4000 series. It takes clock pulses from the clock input, and makes one of the ten outputs come on in sequence each time a clock pulse arrives.

Pinout



| Pin number | Name | Purpose |
|------------|-------------------------|--|
| 1 | 6 | The 6 th sequential output |
| 2 | 2 | The 2 nd sequential output |
| 3 | 1 | The 1 st sequential output |
| 4 | 3 | The 3 rd sequential output |
| 5 | 7 | The 7 th sequential output |
| 6 | 8 | The 8 th sequential output |
| 7 | 4 | The 4 th sequential output |
| 8 | 0V, V _{DD} | The connection to the 0V rail |
| 9 | 9 | The 9 th sequential output |
| 10 | 5 | The 5 th sequential output |
| 11 | 10 | The 10 th sequential output |
| 12 | CO | Carry out output - outputs high on counts 0 to 4, outputs low on counts 5 to 9 (thus a transition from low to high occurs when counting from 9 back to 0) |
| 13 | LE | Latch enable - latches on the current output when high (i.e. the chip counts when LE is low) |
| 14 | CLK | Clock in |
| 15 | RST | Reset - sets output 1 high and outputs 2 through 10 low, when taken high |
| 16 | +9V, V _{CC} | The connection to the +V _{CC} rail (voltage between +3V and +15V) |

Example: Electronic Roulette



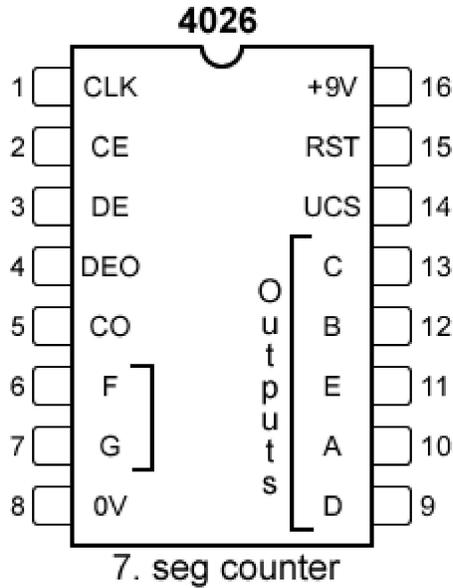
Electronic Roulette circuit diagram

The circuit diagram on the right shows how to create a game of roulette using the 4017 decade counter and various other electronic parts. The switch stops the roulette, and the resistor adjusts the spin speed.

4026 counter and display decoder

The **4026 IC** is a 16-pin CMOS seven-segment counter from the 4000 series. It counts clock pulses and returns the output in a form which can be displayed on a seven-segment display. This avoids using a binary-coded decimal to seven-segment decoder, but it can only be used to display the (decimal) digits 0-9.

Pinout



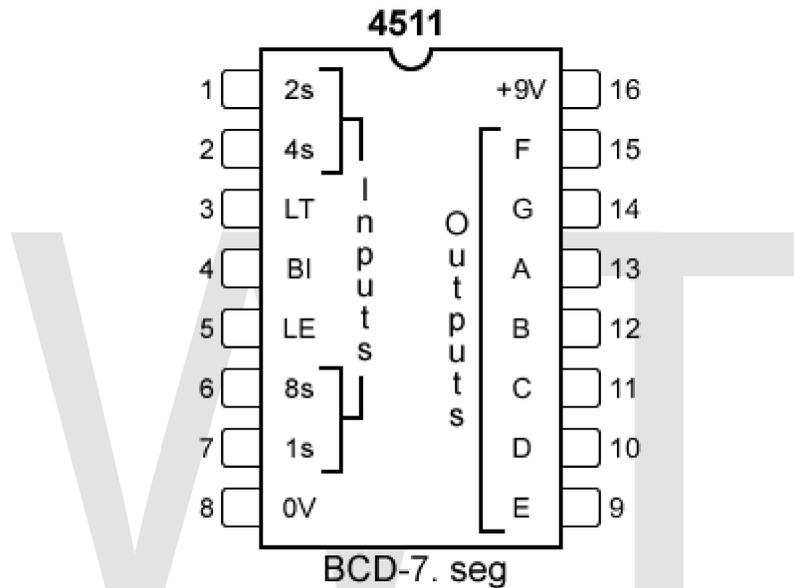
| Pin number | Name | Purpose |
|------------|-----------------|--|
| 1 | CLK | Clock in |
| 2 | CI | Clock inhibit - when low, clock pulses increment the seven-segment |
| 3 | DE | Display enable - the chip outputs to the seven-segment when this is high (i.e. when it's low, the seven-segment is off) - useful to conserve battery life, for instance |
| 4 | DEO | Display enable out - for chaining 4026s |
| 5 | CO | Carry out output - Is high when changing from 9 to 0. It provides an output at 1/10 of the clock frequency, to drive the clock input of another 4026 to provide multi-digit counting. |
| 6 | F | Output for the seven-segment's F input |
| 7 | G | Output for the seven-segment's G input |
| 8 | V _{DD} | The connection to the 0V rail |
| 9 | D | Output for the seven-segment's D input |
| 10 | A | Output for the seven-segment's A input |
| 11 | E | Output for the seven-segment's E input |
| 12 | B | Output for the seven-segment's B input |
| 13 | C | Output for the seven-segment's C input |
| 14 | UCS | Ungated C-segment - an output for the seven-segment's C input which isn't affected by the DE input. This output is high unless the count is 2, when it goes low. |

| | | |
|----|-----------------|--|
| 15 | RST | Reset - resets all outputs to low when taken high |
| 16 | V _{SS} | The connection to the +9V rail |

4511 BCD to seven-segment decoder

The **4511 IC** is a 16-pin CMOS BCD to seven-segment decoder from the 4000 series. It takes the binary-coded decimal from a binary counter and decodes it to drive a seven-segment display.

Pinout



| Pin number | Name | Purpose |
|------------|------------------------|---|
| 1 | 2s | Input for the 2s digit from the binary counter |
| 2 | 4s | Input for the 4s digit from the binary counter |
| 3 | LT | Lamp test - when low, the chip takes all the segments on the display high (to test connections, etc.) |
| 4 | BI | Blanking input - when low, the chip doesn't output to the display - to conserve battery life, for instance |
| 5 | LE | Latch enable - latches on the current output when high (i.e. the inputs change the output when LE is low) |
| 6 | 8s | Input for the 8s digit from the binary counter |
| 7 | 1s | Input for the 1s digit from the binary counter |
| 8 | 0V, V _{DD} | The connection to the 0V rail |
| 9 | E | Output for the seven-segment's E input |

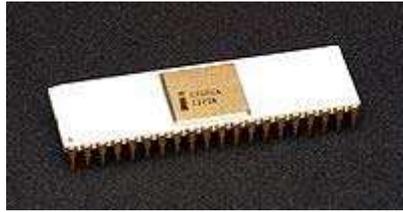
| | | |
|----|-------------------------|---|
| 10 | D | Output for the seven-segment's D input |
| 11 | C | Output for the seven-segment's C input |
| 12 | B | Output for the seven-segment's B input |
| 13 | A | Output for the seven-segment's A input |
| 14 | G | Output for the seven-segment's G input |
| 15 | F | Output for the seven-segment's F input |
| 16 | +9V, V _{CC} | The connection to the +9V rail |

WWT

Chapter 5

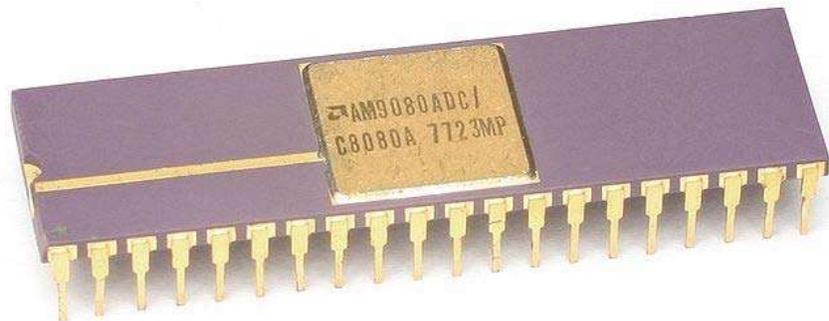
Intel 8080

Intel 8080



An Intel C8080A processor.

| | |
|-------------------------------|------------|
| Produced | mid 1974 |
| Common manufacturer(s) | Intel |
| Max. CPU clock rate | 2 MHz |
| Instruction set | pre x86 |
| Package(s) | 40-pin DIP |



AMD clone AM9080



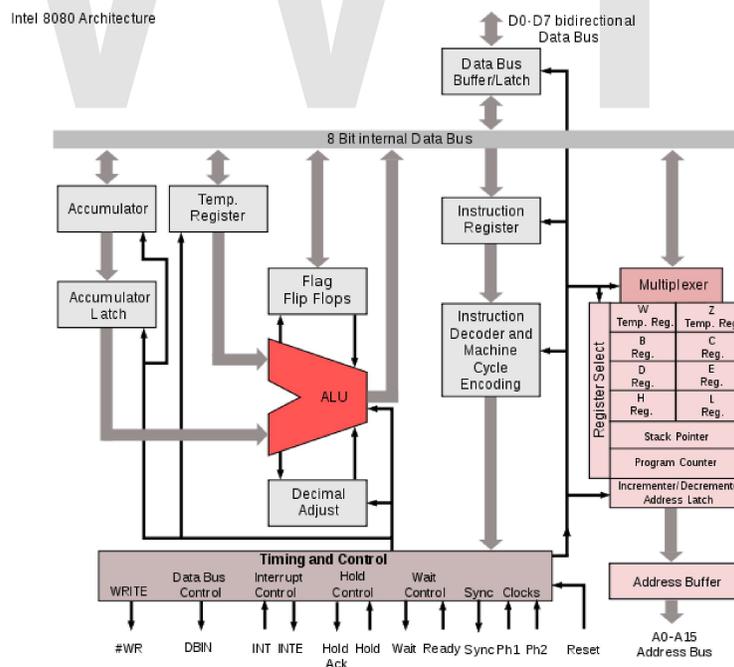
NEC 8080AF (second source)

The **Intel 8080** was an early microprocessor designed and manufactured by Intel. The 8-bit microprocessor was released in April 1974 running at 2 MHz (at up to 500,000 instructions per second), and is sometimes considered to be the first truly usable microprocessor.

The 8080 was implemented using non-saturated enhancement-load NMOS, demanding extra voltages.

Description

Programming model



i8080 microarchitecture

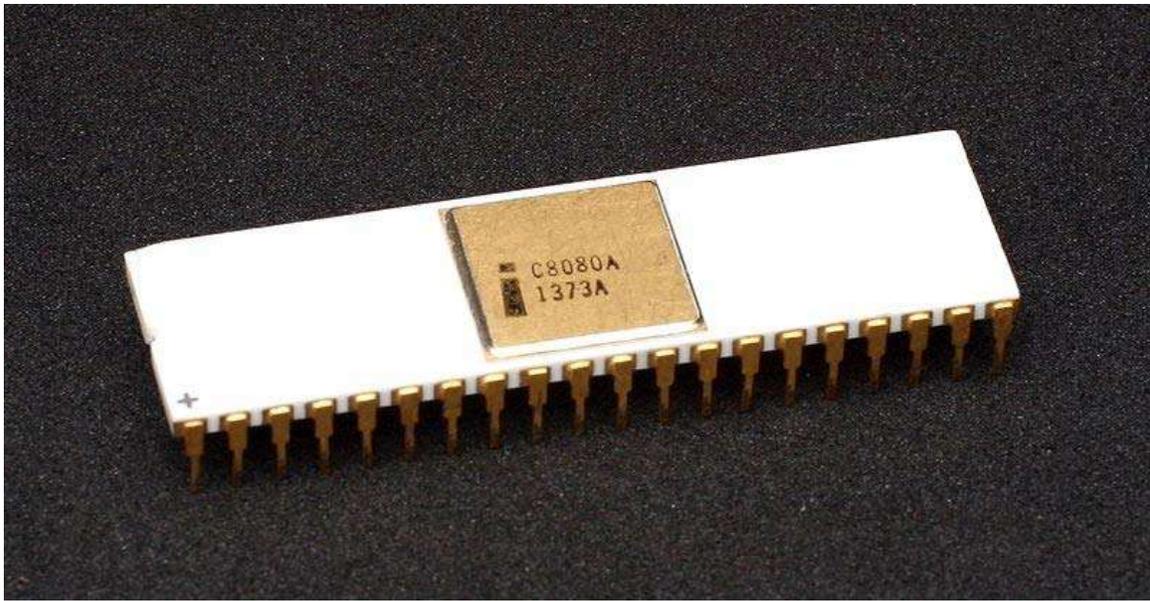
The Intel 8080 was the successor to the 8008. It used the same basic instruction set as the 8008 (developed by Computer Terminal Corporation) and was source code compatible with its predecessor, but added some handy 16-bit operations to the instruction set as well. The 8080's large 40-pin DIP packaging permitted it to provide a 16-bit address bus and an 8-bit data bus, allowing easy access to 64 kilobytes of memory.

Registers

The processor had seven 8-bit registers, (A, B, C, D, E, H, and L) where A was the 8-bit accumulator and the other six could be used as either byte-registers or as three 16-bit register pairs (BC, DE, HL) depending on the particular instruction. Some instructions also enabled HL to be used as (a limited) 16-bit accumulator, and a pseudoregister, M, could be used almost anywhere that any other register could be used and referred to the memory address pointed to by HL. It also had a 16-bit stack pointer to memory (replacing the 8008's internal stack), and a 16-bit program counter.

Commands/instructions

Like in many other 8-bit processors, all instructions were encoded in a single byte (including register-numbers, but excluding immediate data), for simplicity. Some of them were followed by one or two bytes of data, which could be an immediate operand, a memory address, or a port number. Like larger processors, it had automatic CALL and RET instructions for multi-level procedure calls and returns (which could even be conditionally executed, like jumps) and instructions to save and restore any 16-bit register-pair on the machine stack. There were also eight one-byte call instructions (RST) for subroutines located at the fixed addresses 00h, 08h, 10h,...,38h. These were intended to be supplied by external hardware in order to invoke a corresponding interrupt-service routine, but were also often employed as fast system calls. The most sophisticated command was XTHL, which was used for exchanging the register pair HL with the value stored at the address indicated by the stack pointer.



8-bit instructions

Most 8-bit operations could only be performed on the 8-bit accumulator (the A register). For dyadic 8-bit operations, the other operand could be either an immediate value, another 8-bit register, or a memory cell addressed by the 16-bit register pair HL. Direct copying was supported between any two 8-bit registers and between any 8-bit register and an HL-addressed memory cell. Due to the regular encoding of the MOV-instruction (using a quarter of available opcode space) there were redundant codes to copy a register into itself (MOV B,B, for instance), which was of little use, except for delays. However, what would have been a copy from the HL-addressed cell into itself (i.e., MOV M,M) was instead used to encode the HLT instruction (halting execution until an external reset or interrupt).

16-bit operations

Although the 8080 was generally an 8-bit processor, it also had limited abilities to perform 16-bit operations: Any of the three 16-bit register pairs (BC, DE, HL) or SP could be loaded with an immediate 16-bit value (using LXI), incremented or decremented (using INX and DCX), or added to HL (using DAD). The XCHG operation exchanged the values of HL and DE. By adding HL to itself, it was possible to achieve the same result as a 16-bit arithmetical left shift with one instruction. The only 16-bit instructions that affect any flag are DAD H/D/B, which sets the CY (carry) flag in order to allow for programmed 24-bit or 32-bit arithmetics (or larger), needed to implement floating point arithmetics, for instance.

Input/output scheme

Input output port space

The 8080 supported up to 256 input/output (I/O) ports, accessed via dedicated I/O instructions—taking port addresses as operands. This I/O mapping scheme was regarded as an advantage, as it freed up the processor's limited address space. Many CPU architectures instead use so-called memory mapped I/O, in which a common address space is used for both RAM and peripheral chips. This removes the need for dedicated I/O instructions, although a drawback in such designs may be that special hardware must be used to insert wait states as peripherals are often slower than memory. However, in some simple 8080 computers, I/O was indeed addressed as if they were memory cells, "memory mapped", leaving the I/O commands unused. I/O addressing could also sometimes employ the fact that the processor would output the same 8-bit port address to both the lower and the higher address byte (i.e. IN 05h would put the address 0505h on the 16-bit address bus). Similar I/O-port schemes were used in the backward compatible Zilog Z80 and Intel 8085 as well as the closely related x86 families of microprocessors.

Separate stack space

One of the bits in the processor state word (see below) indicates that the processor is accessing data from the stack. Using this signal, it is possible to implement a separate stack memory space. However, this feature was seldom used.

The internal state word

For more advanced systems, during one phase of its working loop the processor set its "internal state byte" on the data bus. This byte contains flags which indicate whether the memory or I/O port is accessed and whether it was necessary to handle an interrupt.

The interrupt system state (enabled or disabled) was also output on a separate pin. For simple systems, where the interrupts were not used, it is possible to find cases where this pin is used as an additional single-bit output port (the popular Radio86RK computer made in the former Soviet Union, for instance).

Applications and successors

The 8080 was used in many early microcomputers, such as the MITS Altair 8800 Computer, Processor Technology SOL-20 Terminal Computer and IMSAI 8080 Microcomputer, forming the basis for machines running the CP/M operating system (the later, almost fully compatible and more capable, Zilog Z80 processor would capitalize on this, with Z80 & CP/M becoming the dominant CPU & OS combination of the period circa 1976 to 1983 much like x86 & MS-DOS for the PC a decade later). Even in 1979 after introduction of the Z80 and 8085 processors, five manufacturers of the 8080 were selling an estimated 500,000 units per month at a price around \$3 to \$4 per unit. The first single-board microcomputer called the "dyna-micro" was based on the Intel

C8080A, and also used Intel's first EPROM, the C1702A. The dyna-micro was re-branded by E&L Instruments of Derby, CT in 1976 as the "MMD-1" (Mini-Micro Designer 1) and was made famous as the example microcomputer in the very popular 8080 "BugBook" series of the time. One of the early uses of the 8080 was made in the late 1970s by Cubic-Western Data of San Diego, CA in its Automated Fare Collection Systems custom designed for mass transit systems such as BART and others around the world. An early industrial use of the 8080 was as the "brain" of the DatagraphiX Auto-COM (Computer Output Microfiche) line of products which took large amounts of user data from reel-to-reel tape and imaged it onto microfiche. The Auto-COM instruments also included an entire automated film cutting, processing, washing, and drying subsystem — quite a feat, both then and in the 21st century, to all be accomplished successfully with only an 8-bit microprocessor running at a clock speed of less than 1 MHz with a 64 KB memory limit. In addition, several early arcade video games were built around the 8080 microprocessor. *Space Invaders* was perhaps the most popular such title.

Shortly after the launch of the 8080, the Motorola 6800 competing design was introduced, and after that, the MOS Technology 6502 variation of the 6800. Zilog introduced the Z80, which had a compatible machine-language instruction set and initially used the same assembly language as the 8080, but for legal reasons, Zilog developed a syntactically-different (but code compatible) alternative assembly language for the Z80. At Intel, the 8080 was followed by the compatible and electrically more elegant 8085, and later by the assembly language compatible 16-bit 8086 and then the 8/16-bit 8088, which was selected by IBM for its new PC to be launched in 1981. Later NEC made an NEC V20 processor (an 8088 clone) which supported 8080 emulation mode. Thus, the 8080, via its ISA, made a lasting impact on computer history.

In the Soviet Union, manufacturers cloned the 8080 microprocessor's layout geometry and started to produce them under the name КР580ИК80 (later marked as КР580ВМ80), where even the pins were placed identically. This processor was the base of the Radio86RK (Радио 86РК in Russian), probably the most popular amateur single-board computer in the Soviet Union. Radio86RK's predecessor was the Micro-80 (Микро-80 in Russian), and its successor the Orion-128 (Орион-128 in Russian) which had a graphical display. Both were built on the КР580 processor. According to some sources, the Soviet analog had two undocumented instructions, specific to itself; however, these were not widely known.

Another model compatible with Intel 8080A, named MMN8080, was produced at Microelectronica Bucharest in Romania. There was also a compatible Polish CPU named MCY7880 and the Czech-made Tesla MHB 8080A.

Chapter 6

Intel 8088 & Intel 80286

Intel 8088

Intel 8088



An Intel 8088 microprocessor

| | |
|-------------------------------|--|
| Produced | From 1979 to 1990s |
| Common manufacturer(s) | Intel, AMD, NEC, Fujitsu, Harris (Intersil), OKI, Siemens AG, Texas Instruments, Mitsubishi. |
| Max. CPU clock rate | 5 MHz to 10 MHz |
| Instruction set | x86-16 |
| Package(s) | 40 pin DIP, 44 pin PLCC |

The **Intel 8088** microprocessor was a variant of the Intel 8086 and was introduced on July 1, 1979. It had an 8-bit external data bus instead of the 16-bit bus of the 8086. The 16-bit registers and the one megabyte address range were unchanged, however. The original IBM PC was based on the 8088.

History and description

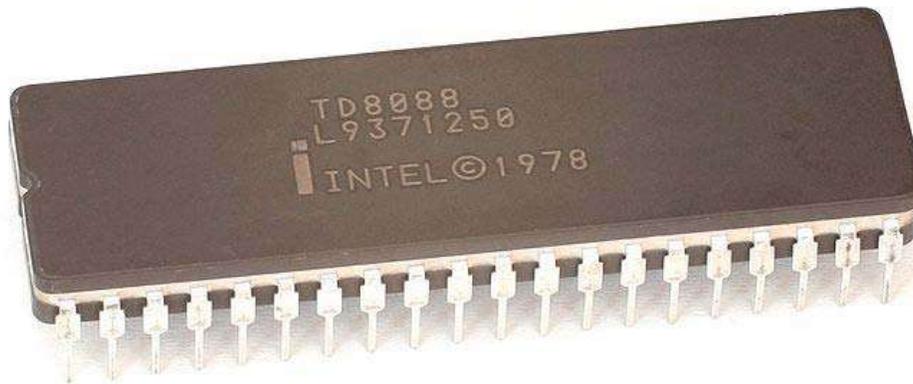
The 8088 was targeted at economical systems by allowing the use of an 8-bit data path and 8-bit support and peripheral chips; complex circuit boards were still fairly cumbersome and expensive when it was released. The prefetch queue of the 8088 was

shortened to four bytes, from the 8086's six bytes, and the prefetch algorithm was slightly modified to adapt to the narrower bus. These modifications of the basic 8086 design were one of the first jobs assigned to Intel's then new design office and laboratory in Haifa Israel.



The Intel 80C88

Variants of the 8088 with more than 5 MHz maximum clock frequency include the 8088-2, which was fabricated using Intel's new enhanced nMOS process called HMOS and specified for a maximum frequency of 8 MHz. Later followed the 80C88, a fully static CHMOS design, which could operate from DC to 8 MHz. There were also several other, more or less similar, variants from other manufacturers. For instance, the NEC V20 was a pin compatible and slightly faster (at the same clock frequency) variant of the 8088, designed and manufactured by NEC. Successive NEC 8088 compatible processors would run at up to 16 MHz.



Performance

Depending on the clock frequency, the number of memory wait states, as well as on the characteristics of the particular application program, the *average* performance for the Intel 8088 ranged from approximately 0.33 to 1 million instructions per second . Meanwhile, the **mov reg,reg** and **ALU reg,reg** instructions taking 2 and 3 cycles respectively yielded an *absolute peak* performance of between 1/3 and 1/2 MIPS per MHz, that is, somewhere in the range 3–5 MIPS at 10 MHz.

Selection for use in the IBM PC

The original IBM PC was the most influential microcomputer to use the 8088. It used a clock frequency of 4.77 MHz (4/3 the NTSC colorburst frequency). Some of IBM's engineers and other employees wanted to use the IBM 801 processor, some would prefer the new Motorola 68000, while others argued for a small and simple microprocessor similar to what had been used in earlier personal computers (such as TRS-80 or Apple II). However, IBM already had a history of using Intel chips in its products and had also acquired the rights to manufacture the 8086 family. Another factor was that the 8088 allowed the computer to be based on a modified 8085 design, as it could easily interface with most nMOS chips with 8-bit databuses, i.e. existing and mature, and therefore economical, components. This included ICs originally intended for support and peripheral functions around the 8085 and similar processors (not exclusively Intels) which were already well known by many engineers, further reducing cost.

The descendants of the 8088 include the 80188, 80186, 80286, 80386, and later software compatible processors, which are in use today.

Intel 80286

Intel 80286

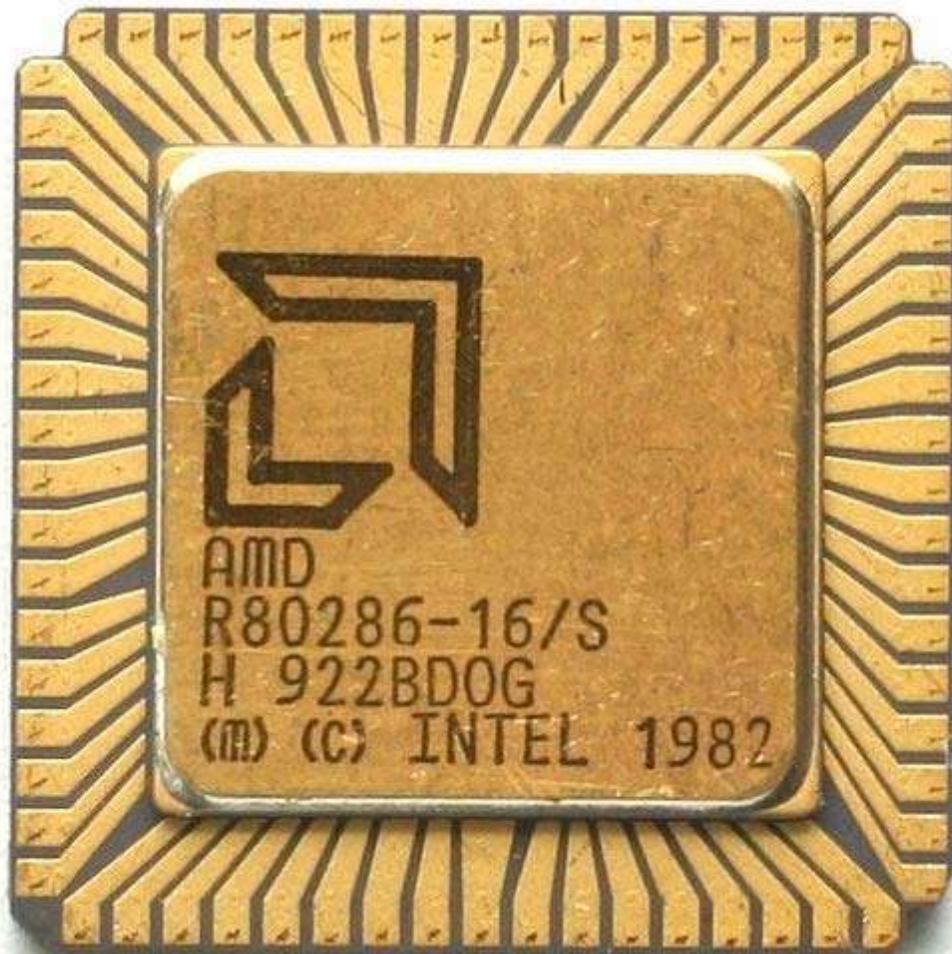


An 8MHz Intel 80286 Microprocessor

| | |
|-------------------------------|---|
| Produced | From 1982 to early 1990s |
| Common manufacturer(s) | Intel, IBM, AMD, Harris (Intersil), Siemens AG, Fujitsu |
| Max. CPU clock rate | 6 MHz (4 MHz for a short time) to 25 MHz |
| Min. feature size | 1.5 μ m |
| Instruction set | x86-16 (with MMU) |
| Package(s) | PGA, CLCC and PLCC 68-pin |

The Intel **80286** (also called **iAPX 286**), introduced on February 1, 1982, was a 16-bit x86 microprocessor with 134,000 transistors. Like its contemporary simpler cousin, the 80186, it could correctly execute most software written for the earlier Intel 8086 and Intel 8088. It was employed for the IBM PC/AT, introduced in 1984, and then widely used in most PC/AT compatible computers until the early 1990s.

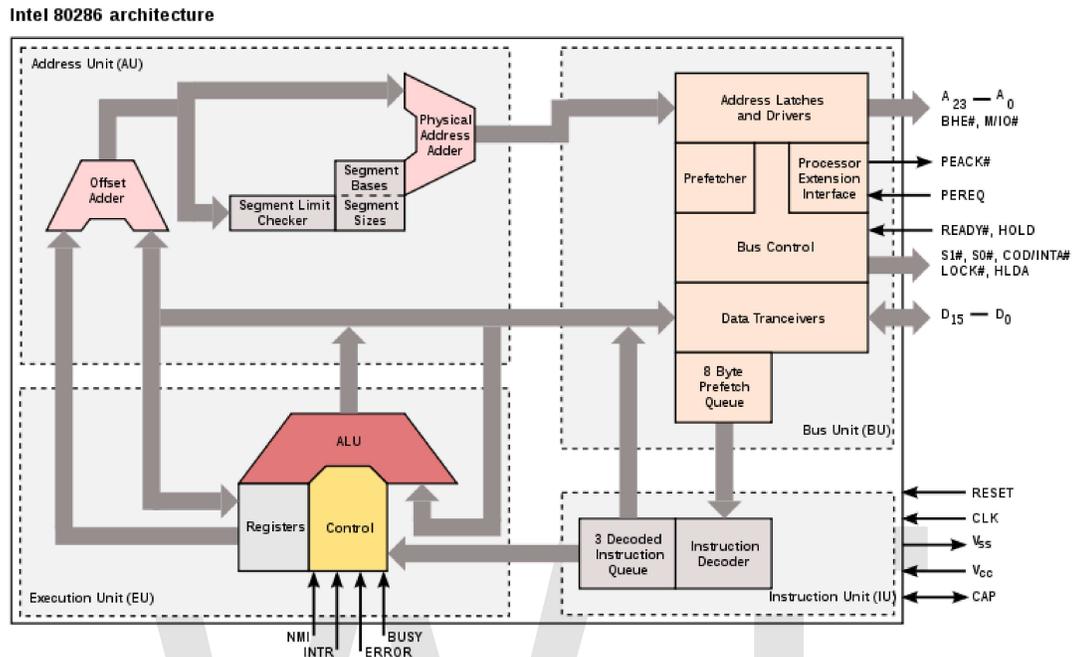
History



AMD 80286 (16 MHz version)

After the 6 and 8 MHz initial releases, it was subsequently scaled up to 12.5 MHz. (AMD and Harris later pushed the architecture to speeds as high as 20 MHz and 25 MHz, respectively.) On average, the 80286 had a speed of about 0.21 instructions per clock. The 6 MHz model operated at 0.9 MIPS, the 10 MHz model at 1.5 MIPS, and the 12 MHz model at 1.8 MIPS.

Design



80286 microarchitecture

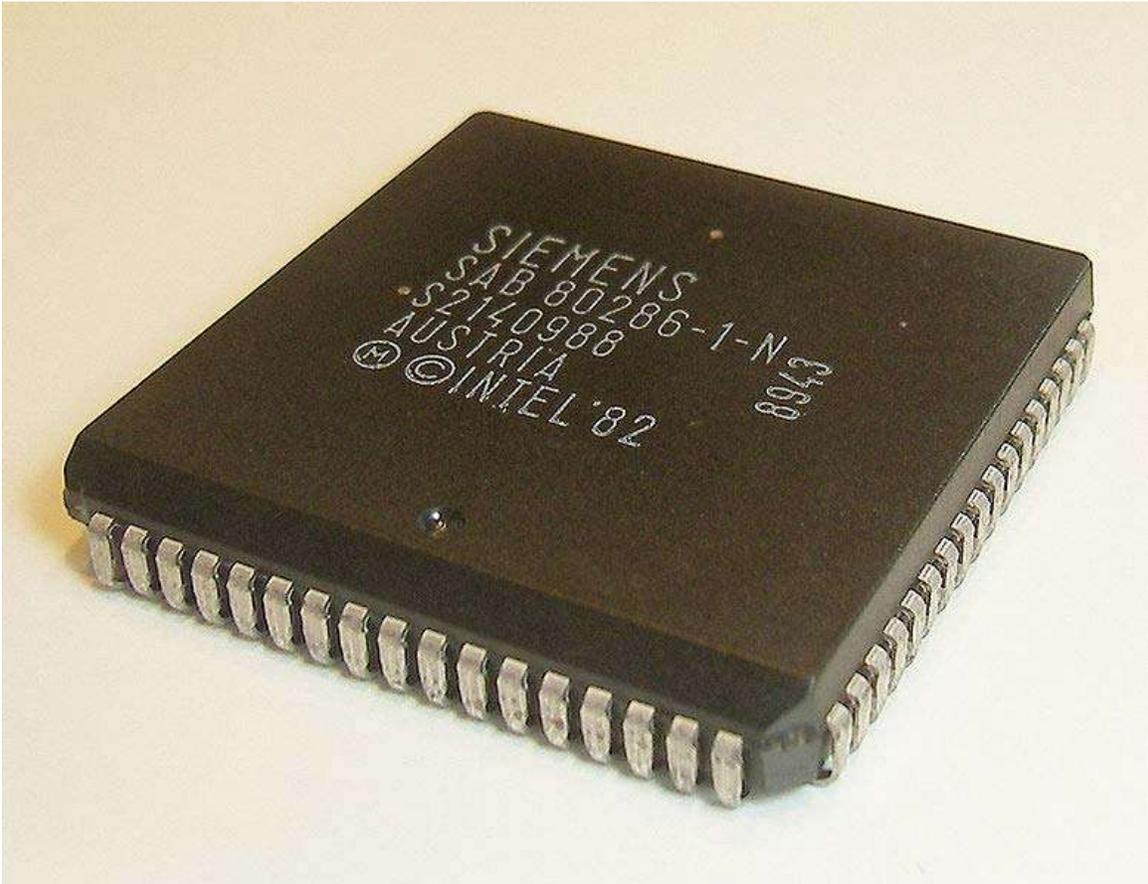
The 80286's performance per clock cycle was more than twice that of its predecessors, the Intel 8086 and Intel 8088. In fact, the performance increase per clock cycle of the 80286 over its immediate predecessor may be the largest among the generations of x86 processors. Calculation of the more complex addressing modes (such as base+index) had less clock penalty because it was performed by a special circuit in the 286; the 8086, its predecessor, had to perform effective address calculation in the general ALU, taking many cycles. Also, complex mathematical operations (such as MUL/DIV) took fewer clock cycles compared to the 8086.

Having a 24-bit address bus, the 286 was able to address up to 16 MB of RAM, in contrast to 1 MB that the 8086 could directly access. While DOS could utilize this additional RAM (extended memory) via BIOS call (INT 15h, AH=87h), or as RAM disk, or emulation of expanded memory, cost and initial rarity of software utilizing extended memory meant that 286 computers were rarely equipped with more than a megabyte of RAM. Additionally, there was a performance penalty involved in accessing extended memory from real mode, as noted below.

The 286 was designed to run multitasking applications, including communications (such as automated PBXs), real-time process control, and multi-user systems.

The later E-stepping level of the 80286 was a very clean CPU, free of the several significant errata that caused problems for programmers and operating system writers in the earlier B-step and C-step CPUs (common in the AT and AT clones).

Features



Siemens 80286 (10 MHz version)



IBM 80286 (8 MHz version)

An interesting feature of this processor is that it was the first x86 processor with protected mode. Protected mode enabled up to 16 MB of memory to be addressed by the on-chip linear memory management unit (MMU) with 1 GB logical address space. The MMU also provided protection from (crashed or ill-behaved) applications writing outside their allocated memory zones. By design, the 286 could not revert from protected mode to the basic 8086-compatible "real mode" without a hardware-initiated reset. In the PC-AT, IBM added external circuitry as well as specialized code in the ROM BIOS to enable special series of program instructions to cause the reset, allowing real-mode reentry (while retaining active memory and control). Though it worked correctly, the method imposed a huge performance penalty.

This limitation led to Bill Gates famously referring to the 80286 as a "brain dead chip", since it was clear that the new Microsoft Windows environment would not be able to run multiple MS-DOS applications with the 286. It was arguably responsible for the split between Microsoft and IBM, since IBM insisted that OS/2, originally a joint venture between IBM and Microsoft, would run on a 286 (and in text mode). To be fair, when Intel designed the 286, it was not designed to be able to multitask real-mode applications; real mode was intended to be a simple way for a bootstrap loader to prepare the system and then switch to protected mode.

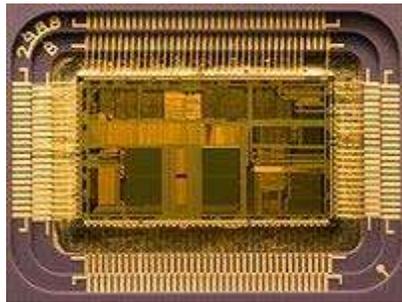
In theory, real mode applications could be directly executed in 16-bit protected mode if certain rules were followed; however, as many DOS programs broke those rules, protected mode was not widely used until the appearance of its successor, the 32-bit Intel 80386, which was designed to go back and forth between modes easily.



Chapter 7

Intel 80486

Intel 486



The exposed die of an Intel 80486DX2 microprocessor.

| | |
|-------------------------------|--|
| Produced | From 1989 to 2007 |
| Common manufacturer(s) | Intel, IBM, AMD, Texas Instruments, Harris Semiconductor, UMC, SGS Thomson |
| Max. CPU clock rate | 16 MHz to 100 MHz |
| FSB speeds | 16 MHz to 50 MHz |
| Min. feature size | 1 μ m to 0.6 μ m |
| Instruction set | x86 (including x87 floating point) |
| Package(s) | PGA (socket 1, 2, 3), 196-pin PQFP |

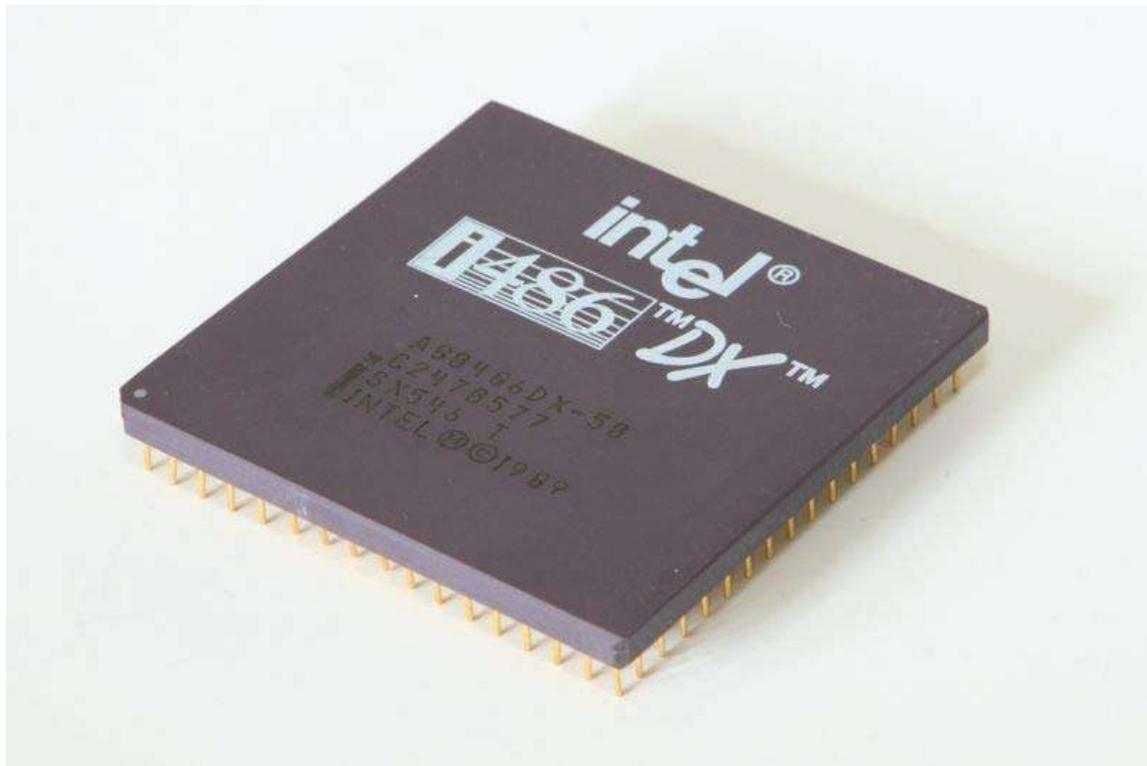
The Intel **80486** microprocessor (alias **i486** or **Intel486**) was a higher performance follow up on the Intel 80386. Introduced in 1989, it was the first tightly pipelined x86 design as well as the first x86 chip to use more than a million transistors, due to a large on-chip cache and an integrated floating point unit. It represents a fourth generation of binary compatible CPUs since the original 8086 of 1978.

A 50 MHz 80486 executed around 40 million instructions per second on average and was able to reach 50 MIPS peak.

The i486 was without the usual 80-prefix because of a court ruling that prohibited trademarking numbers (such as 80486). Later, with the introduction of the Pentium brand, Intel began branding its chips with words rather than numbers.

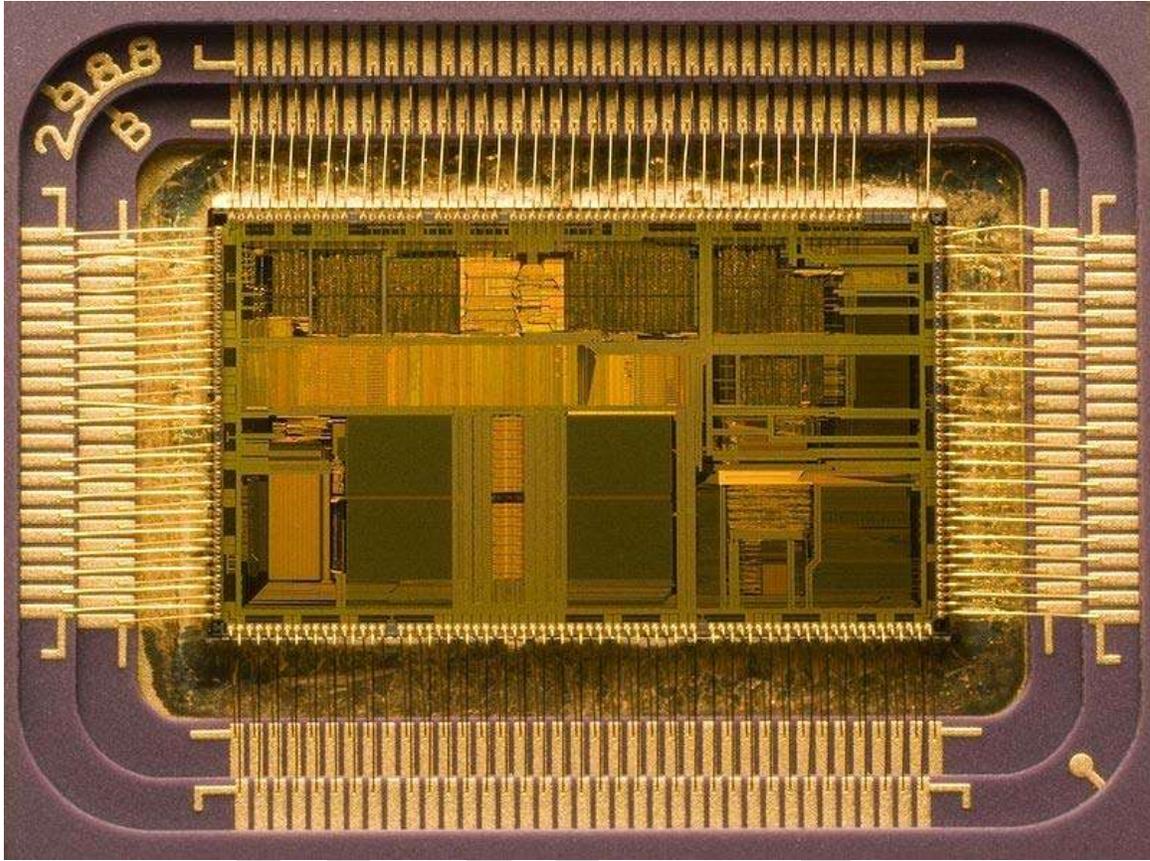
Background

The 80486 was announced at Spring Comdex in April 1989. At the announcement, Intel stated that samples would be available in the third quarter of 1989 and production quantities would ship in the fourth quarter of 1989. The first 80486-based PCs were announced in late 1989, but some advised that people wait until 1990 to purchase an 80486 PC because there were early reports of bugs and software incompatibilities.

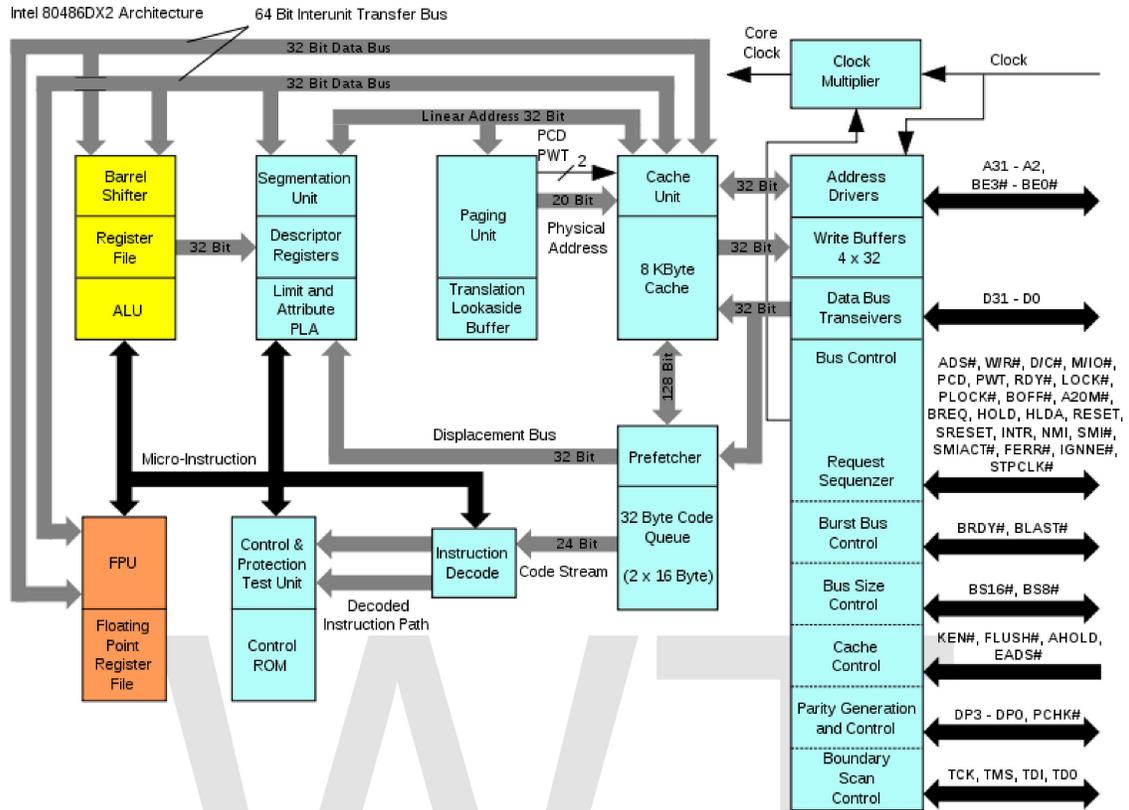


Improvements

The instruction set of the i486 is very similar to its predecessor, the Intel 80386, with the addition of only a few extra instructions, such as CMPXCHG which executes the compare-and-swap atomic operation and the XADD which executes the fetch-and-add atomic operation returning the original value, unlike the ADD instruction that only returned some flags.



From a performance point of view, the architecture of the i486 is a vast improvement over the 80386. It has an on-chip unified instruction and data cache, an on-chip floating-point unit (FPU), except in the SX and SL models, and an enhanced bus interface unit. Simple instructions (such as ALU reg, reg) execute in one clock cycle. These improvements yield a rough doubling in ALU performance over the 386 at the same clock rate. A 16-MHz 486 therefore has a performance similar to a 33-MHz 386 (or 286), and the older design has to reach 50 MHz to be comparable with a 25-MHz 486 part.



The 486DX2 architecture

Differences between i386 and i486

- An 8 KB on-chip SRAM cache stores the most recently used instructions and data (16 KB and/or write-back on some later models). The 386 had no such internal cache but supported a slower off-chip cache.
- Tightly coupled pipelining allows the 486 to complete a simple instruction like *ALU reg,reg* or *ALU reg,im* every clock cycle. The 386 needed two clock cycles for this.
- Integrated FPU (disabled or absent in SX models) with a dedicated local bus; together with faster algorithms on more extensive hardware than in the i387, this gives faster floating point calculations compared to the i386+i387 combination.
- Improved MMU performance.



The 486 has a 32-bit data bus and a 32-bit address bus. This required either four matched 30-pin (8-bit) SIMMs or one 72-pin (32-bit) SIMM on a typical PC motherboard. Just like the 80386, the 32-bit address bus of the 80486 enabled up to 4 gigabytes of memory to be directly addressed using a flat memory model with 32-bit linear addresses in protected mode. Just as with the 80386, the ability to use memory directly without segmentation helped performance in compliant operating systems and applications.

Models

There are several suffixes and variants including:

- **i486DX**: The original chip (without any clock doubling).

- **i486DX-S**: SL Enhanced 486DX
- **Intel RapidCAD**: a specially packaged Intel 486DX and a dummy floating point unit (FPU) designed as pin-compatible replacements for an Intel 80386 processor and 80387 FPU.
- **i486SX**: an i486DX with the FPU part disabled or missing. Early variants were parts with disabled (defective) FPUs, later versions had the FPU removed from the die to reduce area and hence cost.
- **i486SX-S**: SL Enhanced 486SX
- **i486DX2**: the internal processor clock runs at twice the clock rate of the external bus clock.
- **i486SX2**: i486DX2 with the FPU disabled.
- **i486SL**: low power version of the i486DX, reduced VCore, SMM (System Management Mode), stop clock, and power saving features - mainly for use in portable computers.
- **i486SL-NM**: i486SL based on i486SX
- **i487SX**: i486DX with a slightly different pinout sold as an FPU upgrade to i486SX systems; it was widely documented that an i487SX when installed completely disabled the existing i486SX on the motherboard, replacing it.
- **i486 OverDrive**: i486SX, i486SX2, i486DX2 or i486DX4. Marked as upgrade processors, some models had different pinouts or voltage handling abilities from 'standard' chips of the same speed stepping. Fitted to a coprocessor or "OverDrive" socket on the motherboard, worked the same as the i487SX.
- **i486DX4**: designed to run at triple clock rate (not quadruple as often believed; the DX3, which was meant to run at 2.5x the clock speed, was never released). DX4 models that featured write-back cache were identified by an "&EW" laser etched into their top surface, while the write-through models were identified by "&E".
- **i486GX**: Embedded Ultra-Low power CPU with all features of the i486SX and 16 Bit external data bus. This CPU is for embedded battery-operated and hand-held applications.



The specified maximum internal clock frequency (on Intel's versions) ranged from 16 to 100 MHz. The 16 MHz i486SX model was used by Dell Computers.

One of the few 486 models specified for a 50 MHz bus (486DX-50) initially had overheating problems and was moved to the 0.8 micrometre fabrication process. However, problems continued when the 486DX-50 was installed in local bus systems due to the high bus speed, making it rather unpopular with mainstream consumers as local bus video was considered a requirement at the time, though it remained popular with users of EISA systems. The 486DX-50 was soon eclipsed by the clock-doubled i486DX2 which instead ran the CPU logic at twice the external bus speed.

Other makers of 486-like CPUs



STMicroelectronics It's ST ST486DX2-40



UMC Green CPU U5SX

486 compatible processors have been produced by other companies such as IBM, Texas Instruments, AMD, Cyrix, UMC, and SGS Thompson. Some were clones (identical at the microarchitectural level), others were clean-room implementations of the Intel instruction-set. (IBM's multiple source requirement is one of the reasons behind its x86-manufacturing since the 80286.) The 486 was, however, covered by many of Intel's patents covering new R&D as well as that of the prior 80386. Intel and IBM have broad cross-licenses of these patents, and AMD was granted rights to the relevant patents in the 1995 settlement of a lawsuit between the companies.

AMD produced several clones of the 486 using a 40 MHz bus (486DX-40, 486DX/2-80, and 486DX/4-120) which had no equivalent available from Intel, as well as a part

specified for 90 MHz, using a 30 MHz external clock, that was sold only to OEMs. The fastest running 486 CPU, the Am5x86, ran at 133 MHz and was released by AMD in 1995. 150 MHz and 160 MHz parts were planned but never officially released.



Cyrix made a variety of 486-compatible processors, positioned at the cost-sensitive desktop and low-power (laptop) markets. Unlike AMD's 486 clones, the Cyrix processors were the result of clean-room reverse-engineering. Cyrix's early offerings included the 486DLC and 486SLC, two hybrid chips which plugged into 386DX or SX sockets respectively, and offered 1 KB of cache (versus 8 KB for the then-current Intel/AMD parts). Cyrix also made "real" 486 processors, which plugged into the i486's socket and offered 2 or 8 KB of cache. Clock-for-clock, the Cyrix-made chips were generally slower than their Intel/AMD equivalents, though later products with 8 KB caches were more competitive, if late to market.

Competitive alternatives

The Motorola 68040 (best known for its use in the Macintosh Quadra series), while not compatible with the 486, was often positioned as the 486's equivalent in features and performance. While the 68040 outperformed the 486 significantly on a clock for clock basis, the 486 had the ability to be clocked significantly faster without suffering from overheating problems. For a time Apple attempted to compete with Intel's clock doubling 486DX2 systems by publicizing doubled clock rates for its '040-based Macintosh Performa systems, despite the lack of any clock doubling.

Motherboards and buses

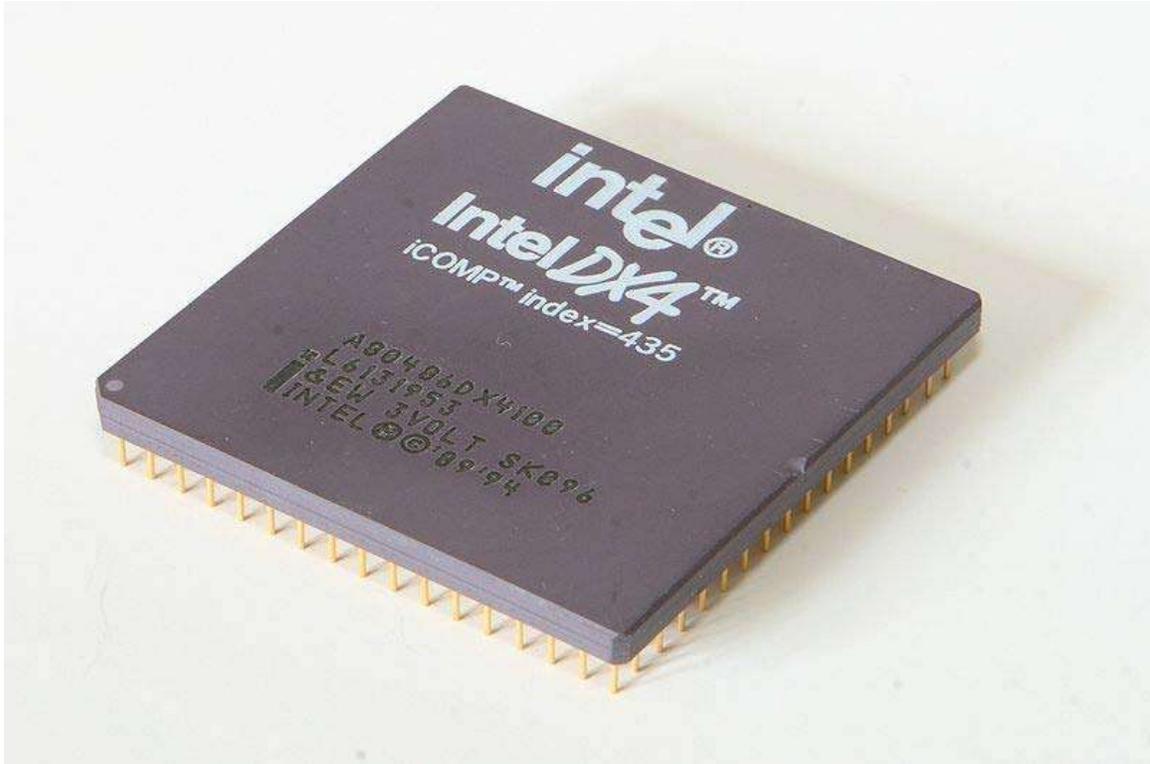
Early 486 machines were equipped with several ISA slots (using an emulated PC/AT-bus) and sometimes one or two 8-bit-only slots (compatible with the PC/XT-bus). Many motherboards enabled overclocking of these up from the default 6 or 8 MHz to perhaps 16.5 or 20 MHz (half the i486 bus clock) in a number of steps, often from within the BIOS setup. Especially older peripheral cards normally worked well at such speeds as they often used standard MSI chips instead of slower (at the time) custom VLSI designs. This could give significant performance gains (such as for old video cards moved from a 386 or 286 computer, for example). However, operation beyond 8 or 10 MHz could sometimes lead to stability problems, at least in systems equipped with SCSI or sound cards.

Some motherboards came equipped with a 32-bit bus called EISA that was backward compatible with the ISA-standard. EISA offered a number of attractive features such as increased bandwidth, extended addressing, IRQ sharing, and card configuration through software (rather than through jumpers, DIP switches, etc.) However, EISA cards were expensive and therefore mostly employed in servers and workstations. Consumer desktops often used the simpler but faster VESA Local Bus, unfortunately somewhat prone to electrical and timing-based instability; typical consumer desktops had ISA slots combined with a single VLB slot for a video card. VLB was gradually replaced by PCI during the final years of the 80486 period. Few Pentium class motherboards had VLB support as VLB was based directly on the i486 bus; it was no trivial matter adapting it to the quite different P5 Pentium-bus. ISA persisted through the P5 Pentium generation and was not completely displaced by PCI until the Pentium III era.



Late 486 boards were normally equipped with both PCI- and ISA-slots, and sometimes a single VLB slot as well. In this configuration VLB or PCI throughput suffered depending on how buses were bridged. The VLB slot in these systems was usually only fully compatible with video cards (quite fitting as "VESA" stands for *Video Electronics Standards Association*); VLB-IDE, multi I/O, or SCSI cards could have problems on motherboards with PCI slots. The VL-Bus operated at the same clock speed as the i486-bus (basically *being* a local 486-bus) while the PCI bus also usually depended on the i486 clock but sometimes had a divider setting available via the BIOS. This could be set to 1/1 or 1/2, sometimes even 2/3 (for 50 MHz CPU clocks). Some motherboards limited the PCI clock to the specified maximum of 33 MHz and certain network cards depended on this frequency for correct bit-rates. The ISA clock was typically generated by a divider of the CPU/VLB/PCI clock (as implied above).

One of the earliest complete systems to use the 80486 chip was the Apricot VX FT, produced by United Kingdom hardware manufacturer Apricot Computers. Even overseas in the United States it drew attention as "The World's First 486" in a popular September 1989 issue of *Byte* magazine (shown right).



Later 486 boards also supported Plug-And-Play, a specification designed by Microsoft to make component installation easier for consumers that began as a part of Windows 95.

Obsolescence

The AMD Am5x86 and Cyrix Cx5x86 were the last 486 processors that were often used in late generation 486 motherboards with PCI slots and 72-pin SIMMs that are designed to be able to run Windows 95, and also often used as upgrades for older 486 motherboards. While the Cyrix Cx5x86 faded quite quickly when the Cyrix 6x86 took over, the AMD Am5x86 was important during the time when the AMD K5 was delayed.

In the general purpose desktop computer role, the 486s were used as budget machines for people who could not afford the latest computers, until around 2001, when Windows 95 support ended and Windows 98, ME, 2000, and XP required much more powerful computers to perform well. In fact, the Windows XP kernel could not run on a 486 at all due to the use of new instruction set extensions introduced with the P5 Pentium processor family.

Although the 486 became obsolete for personal computer applications by the mid-1990s, Intel had continued production for use in embedded systems. In May 2006 Intel announced that production of the 80486 would stop at the end of September 2007.

Chapter 8

Frequency Multiplier

In electronics, a **frequency multiplier** is an electronic circuit that generates an output signal whose output frequency is a harmonic of its input frequency. Frequency multipliers consist of a nonlinear circuit that distorts the input signal and consequently generates harmonics of the input signal. A subsequent bandpass filter selects the desired harmonic frequency and removes the unwanted fundamental and other harmonics from the output.

Frequency multipliers are often used in frequency synthesizers and communications circuits. It can be more economic to develop a lower frequency signal with lower power and less expensive devices, and then use a frequency multiplier chain to generate an output frequency in the microwave or millimeter wave range. Some modulation schemes, such as frequency modulation, survive the nonlinear distortion without ill effect (but schemes such as amplitude modulation do not).

Frequency multiplication is also used in nonlinear optics. The nonlinear distortion in crystals can be used to generate harmonics of laser light.

Theory

A pure sinewave at frequency f has no harmonics. If it goes through a linear amplifier, the result continues to be pure (but may acquire a phase shift).

If the sinewave is run through a stateless nonlinear circuit (transcribing function), the resulting distortion creates harmonics. The distorted signal can be described by a Fourier series in f .

$$x(t) = \sum_{k=-\infty}^{\infty} c_k e^{i2\pi k f t}.$$

The nonzero c_k represent the generated harmonics. The Fourier coefficients are given by integrating over the fundamental period T :

$$c_k = \frac{1}{2\pi} \int_0^T x(t) e^{-i2\pi k t/T} dt$$

These harmonics can be selected by a bandpass filter.

The power in the distorted signal is spread across all the resulting harmonics. An ideal halfwave rectifier, for example, has all nonzero coefficients. An approximate circuit could use a diode.

From a conversion efficiency standpoint, the nonlinear circuit should maximize the coefficient for the desired harmonic and minimize the others. Consequently, the transcribing function is often specially chosen. Easy choices are to use an even function to generate even harmonics or an odd function to for odd harmonics. A full wave rectifier, for example, is good for making a doubler. On the other hand, a tripler may over drive an amplifier to symmetrically distort the positive and negative peaks.

YIG multipliers often want to select an arbitrary harmonic, so they use a stateful distortion circuit that converts the input sine wave into an approximate impulse train. The ideal (but impractical) impulse train generates an infinite number of (weak) harmonics. In practice, an impulse train generated by a monostable circuit will have many usable harmonics. YIG multipliers using step recovery diodes may, for example, take an input frequency of 1 to 2 GHz and produce outputs up to 18 GHz. Sometimes the frequency multiplier circuit will adjust the width of the impulses to improve conversion efficiency for a specific harmonic.

Circuits

Spark generator

Before amplifiers, frequency multipliers were the way to generate radio frequencies.

- Spark gap transmitter
- Arc converter

Diode

Clipping circuits. Full wave bridge doubler.

Class C amplifier and multiplier

Efficiently generating power becomes more important at high power levels. Linear Class A amplifiers are at best 25 percent efficient. Push-pull Class B amplifiers are at best 50 percent efficient. The basic problem is the amplifying element is dissipating power. Switching Class C amplifiers are nonlinear, but they can be better than 50 percent efficient because an ideal switch does not dissipate any power.

A clever design can use the nonlinear Class C amplifier for both gain and as a frequency multiplier.

Step recovery diode

Generating a large number of useful harmonics requires a fast nonlinear device.

Step recovery diodes.

Varactor diode

Resistive loaded varactors. Regenerative varactors. Penfield.

Frequency multipliers have much in common with frequency mixers, and some of the same nonlinear devices are used for both: transistors operated in Class C and diodes. In transmitting circuits many of the amplifying devices (vacuum tubes or transistors) operate nonlinearly and create harmonics, so an amplifier stage can be made a multiplier by tuning the tuned circuit at the output to a multiple of the input frequency. Usually the power (gain) produced by the nonlinear device drops off rapidly at the higher harmonics, so most frequency multipliers just double or triple the frequency, and multiplication by higher factors is accomplished by cascading doubler and tripler stages.

Previous

Frequency multipliers use circuits tuned to a harmonic of the input frequency. Non-linear elements such as diodes may be added to enhance the production of harmonic frequencies. Since the power in the harmonics declines rapidly, usually a frequency multiplier is tuned to only a small multiple (twice, three times, or five times) of the input frequency. Usually amplifiers are inserted in a chain of frequency multipliers to ensure adequate signal level at the final frequency.

Since the tuned circuits have a limited bandwidth, if the base frequency is changed significantly (more than one percent or so), the multiplier stages may have to be adjusted; this can take significant time if there are many stages.

PLLs with frequency dividers

In **digital electronics**, frequency multipliers are often used along with frequency dividers and phase-locked loops to generate any desired frequency from an external reference frequency. The frequency multiplication is carried out in the phase-locked loop's feedback loop, by using a frequency divider on the output of the voltage controlled oscillator (VCO). This **divided-down output** is fed-back to the input comparator and compared to the reference frequency. Since the divided down frequency is smaller than the reference frequency, the comparator generates a voltage signal to the VCO, telling it to increase the output frequency. It continues to do this via the feedback loop, raising the VCO output frequency, until the divided-down frequency from the VCO output is equal

to the reference frequency. At this point the comparator stabilizes and generates no more signals to the VCO, or only minor changes to maintain stability. The output frequency from the VCO will be stable at the input reference frequency multiplied by the value of the feedback divider.

A PLL with a frequency divider in its feedback loop acts as a frequency multiplier and is a type of frequency synthesizer.

Integer-N synthesizer

In a configuration with an integer-N divider, its VCO's output frequency is N times its reference, or input, frequency.

Fractional-N synthesizer

Periodic changes in the integer value of an integer-N frequency divider will effectively result in a multiplier with both whole number and fractional component. Such a multiplier is called a fractional-N synthesizer after its fractional component. Fractional-N synthesizers provide an effective means of achieving fine frequency resolution with lower values of N, allowing loop architectures with tens of thousands of times less phase noise than alternative designs with lower reference frequencies and higher integer N values. They also allow a faster settling time because of their higher reference frequencies, allowing wider closed and open loop bandwidths.

Delta sigma synthesizer

A delta sigma synthesizer adds a randomization to programmable-N frequency divider of the fractional-N synthesizer. This is done to shrink sidebands created by periodic changes of an integer-N frequency divider.

Chapter 9

Frequency Synthesizer and Current Loop

Frequency synthesizer

A **frequency synthesizer** is an electronic system for generating any of a range of frequencies from a single fixed timebase or oscillator. They are found in many modern devices, including radio receivers, mobile telephones, radiotelephones, walkie-talkies, CB radios, satellite receivers, GPS systems, etc. A frequency synthesizer can combine frequency multiplication, frequency division, and frequency mixing (the frequency mixing process generates sum and difference frequencies) operations to produce the desired output signal.

Types

Three types of synthesizer can be distinguished. The first and second type are routinely found as stand-alone architecture: **Direct Analog Synthesis** (also called a **mix-filter-divide** architecture as found in the 1960s HP 5100A) and by comparison the more modern **Direct Digital Synthesizer** (DDS) (Table-Look-Up). The third type are routinely used as communication system IC building-blocks: indirect digital (PLL) synthesizers including integer-N and fractional-N.

Digiphase Synthesizer

It is in some ways similar to a DDS, but it has architectural differences. One of its big advantages is to allow a much finer resolution than other types of synthesizers with a given reference frequency.

History

Although frequency as the inverse of a wave period is a relatively recent idea, the origins of frequency synthesis can be found in the much older concept of angular velocity. The wheel trains of timekeeping devices have gear ratio relationships that were well-studied at least as far back as the time of Christian Huygens, who died in 1695.

Prior to widespread use of synthesizers, radio and television receivers relied on manual tuning of a local oscillator, such as with the turret tuner commonly used in television receivers prior to the 1980s. Variations in temperature and aging of components caused

frequency drift. Automatic frequency control (AFC) solves some of the drift problem, but manual retuning was often necessary. Since transmitter frequencies are well known and very stable, an accurate means of generating fixed, stable frequencies would solve the problem.

A simple and effective solution employs the use of many stable resonators or oscillators, one for each tuning frequency. Quartz crystals offer good stability and are often used for this purpose. This "brute force" technique is practical when only a handful of frequencies are required, but quickly becomes costly and impractical in many applications. For example, the FM radio band in many countries supports 100 individual frequencies from about 88 MHz to 108 MHz. Cable television can support even more frequencies or channels over a much wider band. A large number of crystals increases cost and requires greater space.

Many coherent and incoherent techniques have been devised over the years. Some approaches include phase locked loops, double mix, triple mix, harmonic, double mix divide, and direct digital synthesis (DDS). The choice of approach depends on several factors, such as cost, complexity, frequency step size, switching rate, phase noise, and spurious output.

Coherent techniques generate frequencies derived from a single, stable master oscillator. In most applications, crystal oscillator are common, but other resonators and frequency sources can be used. Incoherent techniques derive frequencies from a set of several stable oscillators. The vast majority of synthesizers in commercial applications use coherent techniques due to simplicity and low cost.

Synthesizers used in commercial radio receivers are largely based on phase-locked loops or PLLs. Many types of frequency synthesiser are available as integrated circuits, reducing cost and size. High end receivers and electronic test equipment use more sophisticated techniques, often in combination.

System analysis and design

A well-thought-out *design procedure* is considered to be the first significant step to a successful synthesizer project. In the system design of a frequency synthesizer, states Manassewitsch, there are as many "best" design procedures as there are experienced synthesizer designers. System analysis of a frequency synthesizer involves output frequency range (or frequency bandwidth or tuning range), frequency increments (or resolution or frequency tuning), frequency stability (or phase stability, compare spurious outputs), phase noise performance (e.g., spectral purity), switching time (compare settling time and rise time), and size, power consumption, and cost. James A. Crawford says that these are mutually contradictory requirements

Trial-and-error superseded by calculation and control theory

The trial and error method was once the work-horse for designers of frequency synthesizers.

This began to change with the works of Floyd M. Gardner (his 1966 *Phaselock techniques*) and Venceslav F. Kroupa (his 1973 *Frequency Synthesis*). Manassewitsch calls this the Brute-force approach. Techniques and formulae have been provided by Dean Banerjee.

Gearbox approach

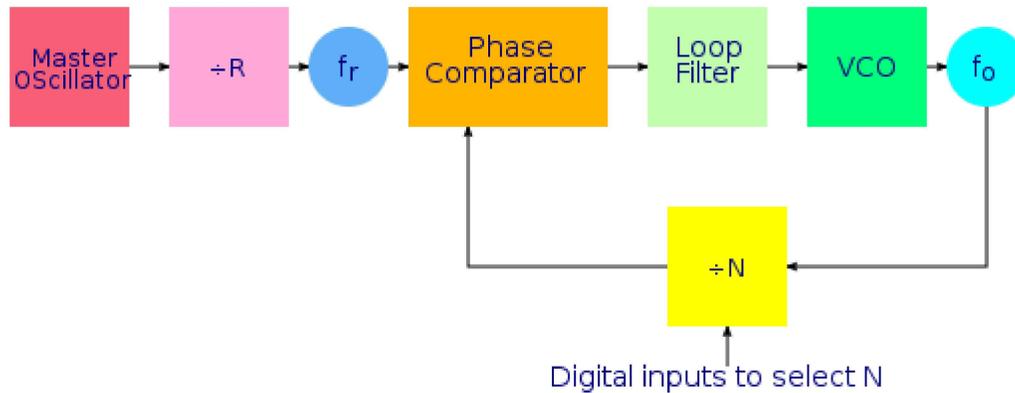
Surprisingly sophisticated mathematical techniques analogous to mechanical gear ratio relationships can be employed in frequency synthesis when the frequency synthesis factor is composed of multiplicative integers in the numerator and denominator. This method allows for effective planning of distribution and suppression of spectral spurs.

Modulo-N approach

Variable frequency synthesizers including DDS are routinely designed using this method.

Principle of PLL synthesizers

A phase locked loop is a feedback control system. It compares the phases of two input signals and produces an error signal that is proportional to the difference between their phases. The error signal is then low pass filtered and used to drive a voltage-controlled oscillator (VCO) which creates an output frequency. The output frequency is fed through a frequency divider back to the input of the system, producing a negative feedback loop. If the output frequency drifts, the phase error signal will increase, driving the frequency in the opposite direction so as to reduce the error. Thus the output is *locked* to the frequency at the other input. This other input is called the **reference** and is usually derived from a crystal oscillator, which is very stable in frequency. The block diagram below shows the basic elements and arrangement of a PLL based frequency synthesizer.



The key to the ability of a frequency synthesizer to generate multiple frequencies is the divider placed between the output and the feedback input. This is usually in the form of a digital counter, with the output signal acting as a clock signal. The counter is preset to some initial count value, and counts down at each cycle of the clock signal. When it reaches zero, the counter output changes state and the count value is reloaded. This circuit is straightforward to implement using flip-flops, and because it is digital in nature, is very easy to interface to other digital components or a microprocessor. This allows the frequency output by the synthesizer to be easily controlled by a digital system.

Example

Suppose the reference signal is 100 kHz, and the divider can be preset to any value between 1 and 100. The error signal produced by the comparator will only be zero when the output of the divider is also 100 kHz. For this to be the case, the VCO must run at a frequency which is 100 kHz x the divider count value. Thus it will produce an output of 100 kHz for a count of 1, 200 kHz for a count of 2, 1 MHz for a count of 10 and so on. Note that only whole multiples of the reference frequency can be obtained with the simplest integer N dividers. Fractional N dividers are readily available .

Practical considerations

In practice this type of frequency synthesiser cannot operate over a very wide range of frequencies, because the comparator will have a limited bandwidth and may suffer from aliasing problems. This would lead to false locking situations, or an inability to lock at all. In addition, it is hard to make a high frequency VCO that operates over a very wide range. This is due to several factors, but the primary restriction is the limited capacitance range of varactor diodes. However, in most systems where a synthesiser is used, we are not after a huge range, but rather a finite number over some defined range, such as a number of radio channels in a specific band.

Many radio applications require frequencies that are higher than can be directly input to the digital counter. To overcome this, the entire counter could be constructed using high-speed logic such as ECL, or more commonly, using a fast initial division stage called a

prescaler which reduces the frequency to a manageable level. Since the prescaler is part of the overall division ratio, a fixed prescaler can cause problems designing a system with narrow channel spacings - typically encountered in radio applications. This can be overcome using a dual-modulus prescaler.

Further practical aspects concern the amount of time the system can switch from channel to channel, time to lock when first switched on, and how much noise there is in the output. All of these are a function of the *loop filter* of the system, which is a low-pass filter placed between the output of the frequency comparator and the input of the VCO. Usually the output of a frequency comparator is in the form of short error pulses, but the input of the VCO must be a smooth noise-free DC voltage. (Any noise on this signal naturally causes frequency modulation of the VCO.). Heavy filtering will make the VCO slow to respond to changes, causing drift and slow response time, but light filtering will produce noise and other problems with harmonics. Thus the design of the filter is critical to the performance of the system and in fact the main area that a designer will concentrate on when building a synthesiser system.

WWT

Current loop

A **current loop** describes two different electrical signalling schemes.

Digital



RS-232 / Current loop converter

For digital serial communications, a current loop is a communication interface that uses current instead of voltage for signaling. Current loops can be used over moderately long distances (tens of kilometres), and can be interfaced with optically isolated links.

Long before the RS-232 standard, current loops were used to send digital data in serial form for teletypewriters. More than two teletypes could be connected on a single circuit allowing a simple form of networking. Older teletypes used a 60 mA current loop. Later machines, such as the ASR33 teletypewriter, operated on a lower 20 mA current level and most early minicomputers featured a 20 mA current loop interface, with an RS-232 port generally available as a more expensive option. The original IBM PC serial port card had provisions for a 20 mA current loop. A digital current loop uses the absence of current for high (space or break), and the presence of current in the loop for low (mark).

The maximum resistance for a current loop is limited by the available voltage. Current loop interfaces usually use voltages much higher than those found on an RS-232 interface, and cannot be interconnected with voltage-type inputs without some form of level translator circuit.

MIDI (Musical Instrument Digital Interface) is a digital current loop interface.

Analog

Analog current loops are used where a device must be monitored or controlled remotely over a pair of conductors. Only one current level can be present at any time.

Given its analog nature, current loops are easier to understand and debug than more complicated digital fieldbuses, requiring only a handheld digital multimeter in most situations. Using fieldbuses and solving related problems usually requires much more education and understanding than required by simple current loop systems.

Additional digital communication to the device can be added to current loop using HART Protocol. Digital process buses such as FOUNDATION Fieldbus and Profibus may replace analog current loops.

Process-control use

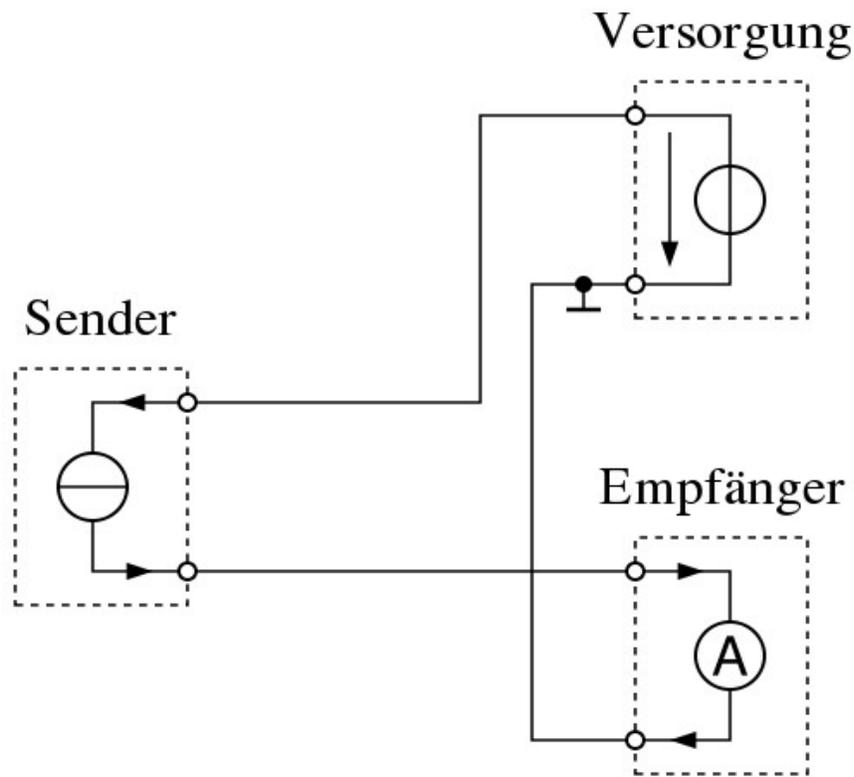
For industrial process control instruments, analog **4–20 mA** and 10–50 mA current loops are commonly used for analog signaling, with 4 mA representing the lowest end of the range and 20 mA the highest. The key advantages of the current loop are that the accuracy of the signal is not affected by voltage drop in the interconnecting wiring, and that the loop can supply operating power to the device. Even if there is significant electrical resistance in the line, the current loop transmitter will maintain the proper current, up to its maximum voltage capability. The *live-zero* represented by 4 mA allows the receiving instrument to detect some failures of the loop, and also allows transmitter devices to be powered by the same current loop (called *two-wire* transmitters). Such instruments are used to measure pressure, temperature, flow, pH or other process variables. A current loop can also be used to control a valve positioner or other output actuator. An analog current loop can be converted to a voltage input with a precision resistor. Since input terminals of instruments may have one side of the current loop input tied to the chassis ground (earth), analog isolators may be required when connecting several instruments in series.

Depending on the source of current for the loop, devices may be classified as *active* (supplying power) or *passive* (relying on loop power). For example, a chart recorder may provide loop power to a pressure transmitter. The pressure transmitter modulates the current on the loop to send the signal to the strip chart recorder, but does not in itself supply power to the loop and so is passive. (A *4-wire* instrument has a power supply input separate from the current loop.) Another loop may contain two passive chart

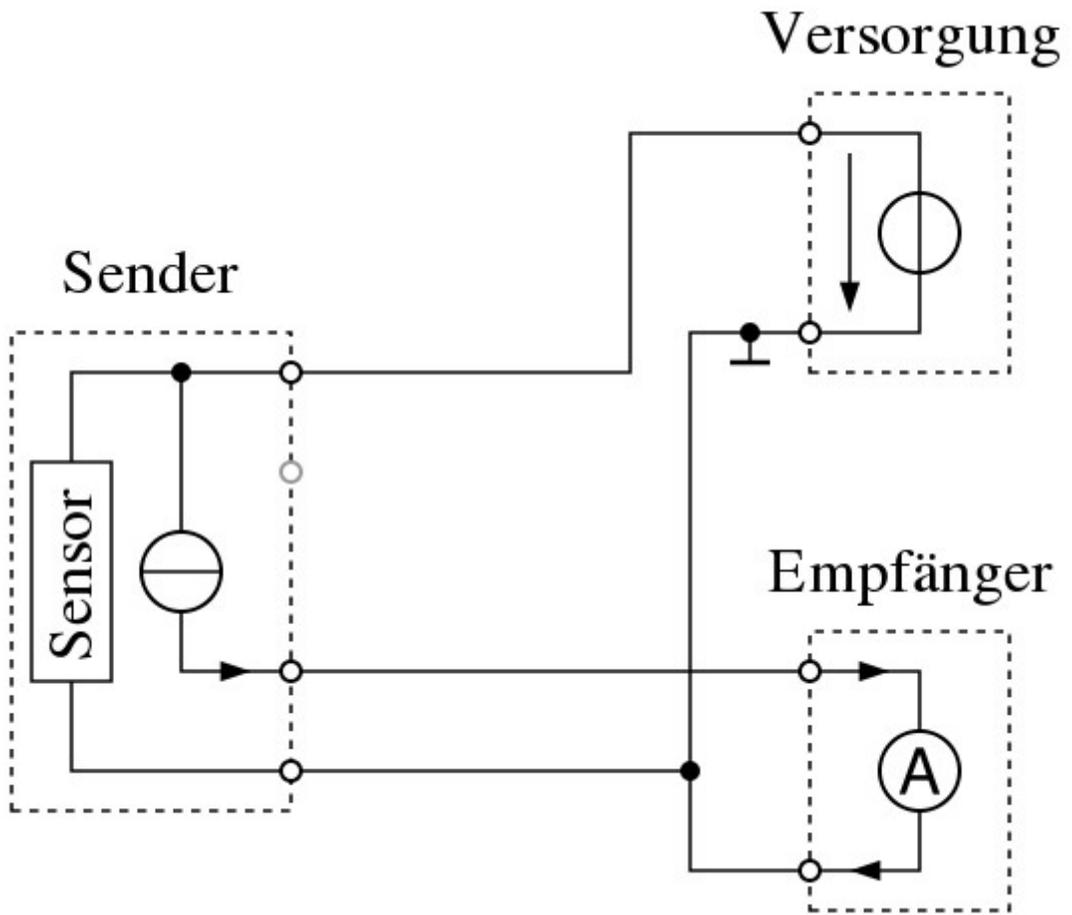
recorders, a passive pressure transmitter, and a 24 V battery. (The battery is the active device).

Panel mount displays and chart recorders are commonly termed 'indicator devices' or 'process monitors'. Several passive indicator devices may be connected in series, but a loop must have only one transmitter device and only one power source (active device).

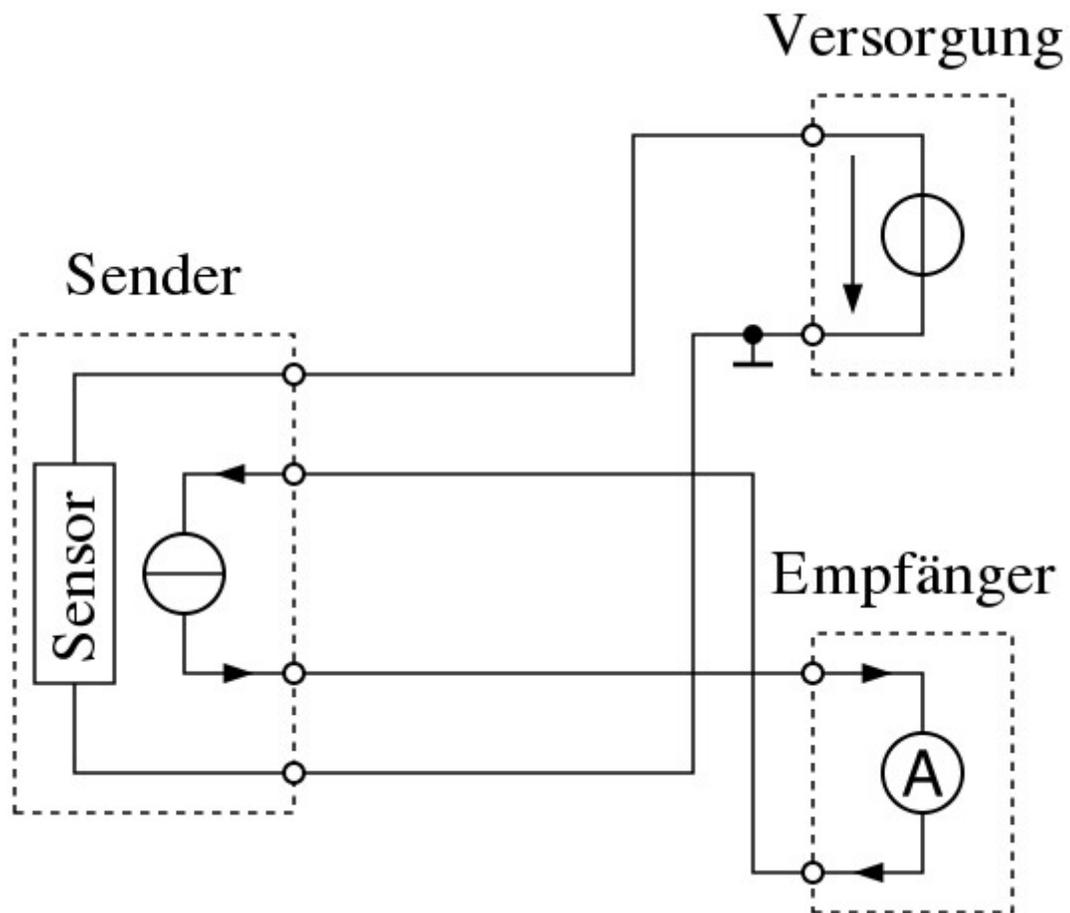
The relationship between current value and process variable measurement is set by calibration, which assigns different ranges of engineering units to the span between 4 and 20 mA. The mapping between engineering units and current can be inverted, so that 4 mA represents the maximum and 20 mA the minimum.



Typ 2



Typ 3



Typ 4

Long circuits

Analog current loops were occasionally carried between buildings by dry pairs in telephone cables leased from the local telephone company. 4–20 mA loops were more common in the days of analog telephony. These circuits require end-to-end direct current (DC) continuity. DC continuity is not available over a microwave radio, optical fiber, or a multiplexed telephone circuit connection.

Basic DC circuit theory shows that the current is the same all along the line. It was common to see 4–20 mA circuits that had loop lengths in miles or circuits working over telephone cable pairs that were longer than ten thousand feet end-to-end. There are still legacy systems in place using this technology. In Bell System circuits, voltages up to 125 VDC were employed.

Discrete control

Discrete control functions can be represented by discrete levels of current sent over a loop. This would allow multiple control functions to be operated over a single pair of wires. Currents required for a specific function vary from one application or manufacturer to another. There is no specific current that is tied to a single meaning. It is almost universal that 0 mA indicates the circuit has failed. In the case of a fire alarm, 6 mA could be normal, 15 mA could mean a fire has been detected, and 0 mA would produce a trouble indication, telling the monitoring site the alarm circuit had failed. Some devices, such as two-way radio remote control consoles, can reverse the polarity of currents and can multiplex audio onto a DC current.

These devices can be employed for any remote control need a designer might imagine. For example, a current loop could actuate an evacuation siren or command synchronized traffic signals.

Two-way radio use



A Motorola T-1300 series remote control is built in a telephone housing. The dial is replaced with a speaker and volume control. This remote control uses a two-wire circuit to control a base station.

Current loop circuits are one possible way used to control radio base stations at distant sites. The two-way radio industry calls this type of remote control **DC remote**. This name

comes from the need for DC circuit continuity between the control point and the radio base station. The purpose current loop remote control is to save the cost of extra pairs of wires between the operating point and the radio transceiver. Some equipment, such as the Motorola MSF-5000 base station, uses currents below 4 mA for some functions. An alternative type, the Tone remote, is more complex but requires only an audio path between control point and base station. The patent does not describe this tone remote but confirms the use of the phrase to describe this system of signaling.

For example, a taxi dispatch base station might be physically located on the rooftop of an eight-story building. The taxi company office might be in the basement of a different building nearby. The office would have a remote control unit that would operate the taxi company base station over a current loop circuit. The circuit would normally be over a telephone line or similar wiring. Control function currents come from the remote control console at the dispatch office end of a circuit. In two-way radio use, an idle circuit would normally have no current present.

In two-way radio use, radio manufacturers use different currents for specific functions. Polarities are changed to get more possible functions over a single circuit. For example, imagine one possible scheme where the presence of these currents cause the base station to change state:

- no current means *receive on channel 1*, (the default).
- +6 mA might mean *transmit on channel 1*
- -6 mA might mean *stay in receive mode but switch to channel 2*. So long as the -6 mA current were present, the remote base station would continue to receive on channel 2.
- -12 mA might command the base station to *transmit on channel 2*.

Note that this circuit is polarity-sensitive. If a telephone company cable splicer accidentally reversed the conductors, selecting channel 2 would lock the transmitter on.

Each current level could close a set of contacts, or operate solid-state logic, at the other end of the circuit. That contact closure caused a change of state on the controlled device. Some remote control equipment could have options set to allow compatibility between manufacturers. That is, a base station that was configured to transmit with a +18 mA current could have options changed to (instead) make it transmit when +6 mA was present.

In two-way radio use, AC signals were also present on the circuit pair. If the base station were idle, receive audio would be sent over the line from the base station to the dispatch office. In the presence of a transmit command current, the remote control console would send audio to be transmitted. The voice of the user in the dispatch office would be superimposed over the DC current that caused the transmitter to operate.

Chapter 10

Leased Line

A **leased line** is a service contract between a provider and a customer, whereby the provider agrees to deliver a symmetric telecommunications line connecting two or more locations in exchange for a monthly rent (hence the term lease). It is sometimes known as a 'Private Circuit' or 'Data Line' in the UK or as CDN (Circuito Diretto Numerico) in Italy. Unlike traditional PSTN lines it does not have a telephone number, each side of the line being permanently connected to the other. Leased lines can be used for telephone, data or Internet services. Some are ringdown services, and some connect two PBXes.

Typically, leased lines are used by businesses to connect geographically distant offices. Unlike dial-up connections, a leased line is always active. The fee for the connection is a fixed monthly rate. The primary factors affecting the monthly fee are distance between end points and the speed of the circuit. Because the connection doesn't carry anybody else's communications, the carrier can assure a given level of quality.

An internet leased line is a premium internet connectivity product, delivered over fiber normally, which is dedicated and provides uncontended, symmetrical speeds, Full Duplex. It is also known as an ethernet leased line, DIA line, data circuit or private circuit.

For example, a T-1 channel can be leased, and provides a maximum transmission speed of 1.544 Mbps. The user can divide the connection into different lines for multiplexing data and voice communication, or use the channel for one high speed data circuit. Increasingly, leased lines are being used by companies, and even individuals, for Internet access because they afford faster data transfer rates and are cost-effective for heavy users of the Internet.

History

Leased lines services (or private line services) became digital in the 1970s with the conversion of the Bell backbone network from analog to digital circuits . This conversion allowed AT&T to offer Dataphone Digital Services (later re-branded digital data services) that started the deployment of ISDN and T1 lines to customer premises to connect .

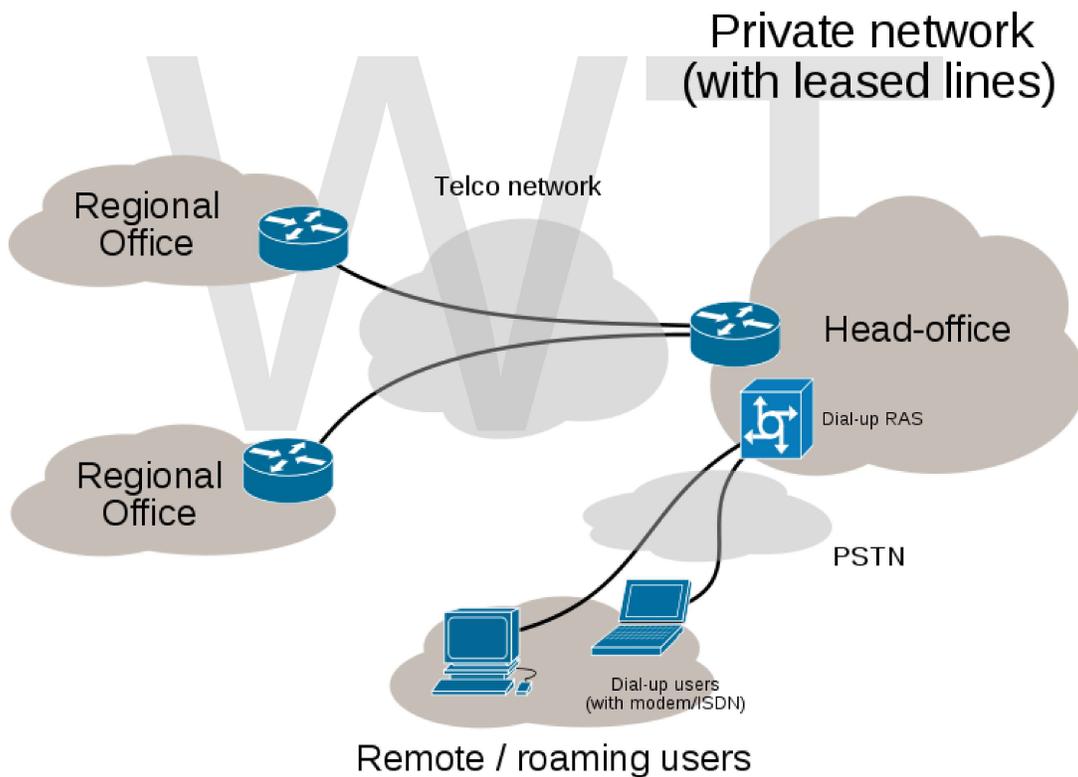
Leased lines were used to connect mainframe computers with terminals and remote sites, via IBM Systems Network Architecture (created in 1974) or DECnet (created in 1975).

With the extension of digital services in the 1980s leased lines were used to connect customer premises to Frame Relay or ATM networks. Access data rates increased from the original T1 option up to T3 circuits.

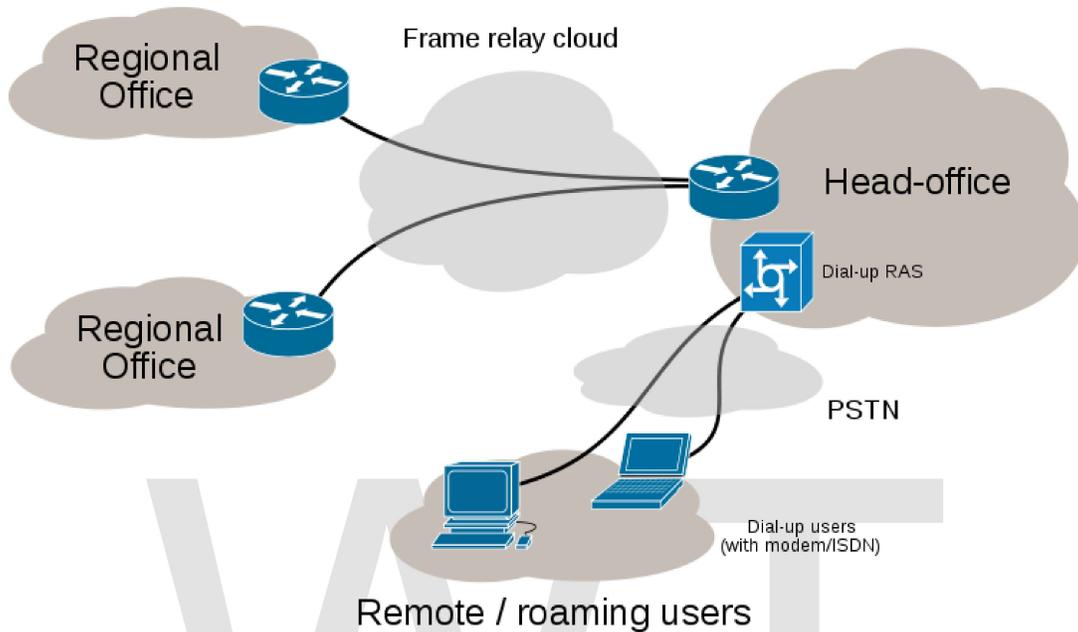
In the 1990s with the advances of the Internet, leased lines were also used to connect customer premises to ISP Point of Presence's whilst the following decade saw a convergence of the aforementioned services (frame relay, ATM, Internet for businesses) with the MPLS integrated offerings.

Access data rates also evolved dramatically to speeds of up to 10Gb/s in the early 21st century with the Internet boom and increased offering in long-haul optical networks or Metropolitan Area Networks.

Applications



Frame-relay network



Leased lines are used to build up private networks, private telephone networks (by interconnecting PBX's) or access the internet or a partner network (extranet).

Here's is a review of the leased line applications in Network designs over time:

Site to site data connectivity

Terminating a leased line with two routers can extend network capabilities across sites. Leased lines were first used in the 1970s by enterprise with proprietary protocols such as IBM System Network Architecture and Digital Equipment DECnet, and with TCP/IP in University and Research networks before the Internet became widely available. Note that other Layer 3 protocols were used such as Novell IPX on enterprise networks until TCP/IP became ubiquitous in the 2000s. Today, point to point data circuits are typically provisioned as either TDM, Ethernet, or Layer 3 MPLS.

Site to site PBX connectivity

Terminating a leased line with two PBX allowed customers to by-pass PSTN for inter-site telephony. This allowed the customers to manage their own dial plan (and to use short extensions for internal telephone number) as well as to make significant savings if enough voice traffic was carried across the line (specially when the savings on the telephone bill exceeded the fixed cost of the leased line).

Site to network connectivity

As demand grew on data network telcos started to build more advanced network using packet switching on top of their infrastructure. Thus number of telecommunication companies added ATM, Frame-relay or ISDN offerings to their services portfolio. Leased lines were used to connect the customer site to the telco network access point.

International Private Lease Circuit

An IPLC is an International Private Leased Circuit that functions as a point-to-point private line. IPLCs are usually Time-division multiplexing (TDM) circuits that utilize the same circuit amongst many customers. The nature of TDM requires the use of a CSU/DSU and a router. Usually the router will include the CSU/DSU.

Then came the Internet (in the mid-1990s) and since the most common application for leased line is to connect a customer to its ISP Point of presence. With the changes that Internet brought in the networking world other technologies were developed to propose alternative to Frame-relay or ATM networks such as VPN's (hardware and software) and MPLS networks (that are in effect an upgrade to TCP/IP of existing ATM/Frame-relay infrastructures).

Availability

In the United Kingdom

In the U.K., leased lines are available at speeds from 64Kb/s increasing in 64Kb/s increments to 2.048Mb/s over a channelised E1 tail circuit and at speeds between 2.048Mb/s to 34.368Mb/s via channelised E3 tail circuits. The NTE will terminate the circuit and provide the requested presentation most frequently X.21 however higher speed interfaces are available such as G.703 or 10baseT. Some ISPs however use the term more loosely, defining a leased line as “any dedicated bandwidth service delivered over a leased fibre connection”.

In the United States

In the U.S., low-speed leased lines (56 kbit/s and below) are usually provided using analog modems. Higher-speed leased lines are usually presented using **FT1 (Fractional T1)**: a **T1** bearer circuit with 1 to 24, 56k or 64k timeslots. Customers must manage their own network termination equipment—Channel Service Unit and Data Service Unit (CSU/DSU).

In Hong Kong

In Hong Kong, leased lines are usually available at speeds of 64k, 128k, 256k, 512k, T1 (channelized or not) or E1 (less common). Whatever the speed, telcos usually provide the CSU/DSU and present to the customer on V.35 interface.

In India

In India, leased lines are available at speeds of 64k, 128k, 256k, 512k, 1mbps, 2mbps, 4mbps, 8mbps, 16mbps T1 (1.544mbps) or E1 (2.048mbps). Customers are connected either through OFC, telephone lines ADSL, or through Wifi. Customers would have to manage their own network termination equipment, namely the Channel service unit and Data service unit. All service providers give a 99% uptime guarantee.

In Italy

In Italy, leased lines are available at speeds of 64k (terminated by DCE2 or DCE2plus modem) or multiple of 64k from 128k up to framed or unframed E1 (DCE3 modem) in digital form (PDH service, known as CDN, Circuito Diretto Numerico). Local TELCOs also may provide CDA (Circuito Diretto Analogico), that are plain copper dry pair between two buildings, without any line termination: in the past (pre-2002) a full analog base band was provided, giving an option to customer to deploy xDSL technology between sites: nowadays everything is limited at 4 kHz of bearer channel, so the service is just a POTS connection without any setup channel.

For many purposes, leased lines are gradually being replaced by DSL and metro Ethernet.

Leased line alternatives

Leased lines are more expensive than alternative connectivity services including (ADSL, SDSL, etc.) because they are reserved exclusively to the leaseholder. Some internet service providers have therefore developed alternative products that aim to deliver leased-line type services (Carrier Ethernet-based, zero contention, guaranteed availability), with more moderate bandwidth, over the standard UK national broadband network. While a leased line is full-duplex, most leased line alternatives provide only half-duplex or in many cases asymmetrical service.

Chapter 11

Balanced Line and Balanced Circuit

Balanced line

In telecommunications and professional audio, a **balanced line** or **balanced signal pair** is a transmission line consisting of two conductors of the same type, each of which have equal impedances along their lengths and equal impedances to ground and to other circuits. The chief advantage of the balanced line format is good rejection of external noise. Common forms of balanced line are twin-lead, used for radio frequency signals and twisted pair, used for lower frequencies. They are to be contrasted to unbalanced lines, such as coaxial cable, which is designed to have its return conductor connected to ground, or circuits whose return conductor actually is ground. Balanced and unbalanced circuits can be interconnected using a transformer called a balun.

Circuits driving balanced lines must themselves be balanced to maintain the benefits of balance. This may be achieved by differential signaling, transformer coupling or by merely balancing the impedance in each conductor.

Lines carrying symmetrical signals (those with equal but opposite voltages to ground on each leg) are often referred to as balanced, but this is an entirely different meaning. The two conditions are not related.

Explanation

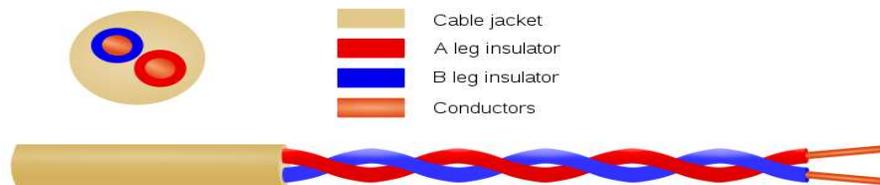


Fig. 1. Balanced line in twisted pair format. This line is intended for use with 2-wire circuits.

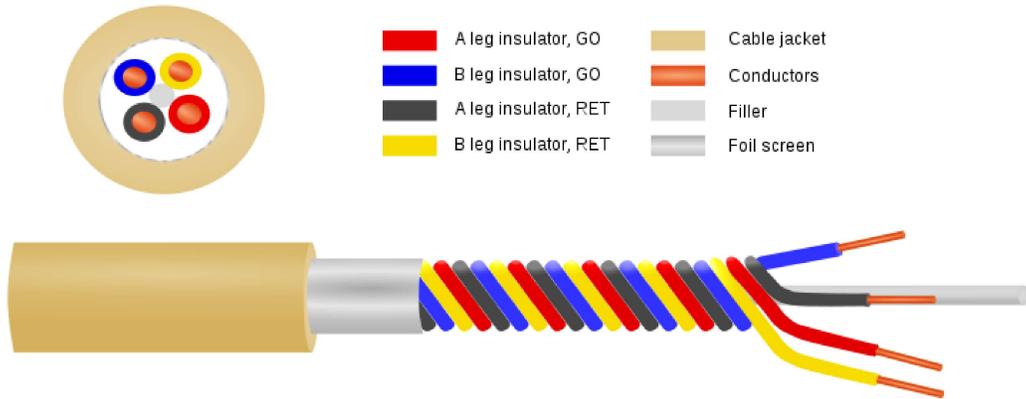


Fig. 2. Balanced line in star quad format. This line is intended for use with 4-wire circuits or two 2-wire circuits. It is also used with microphone signals in professional audio.

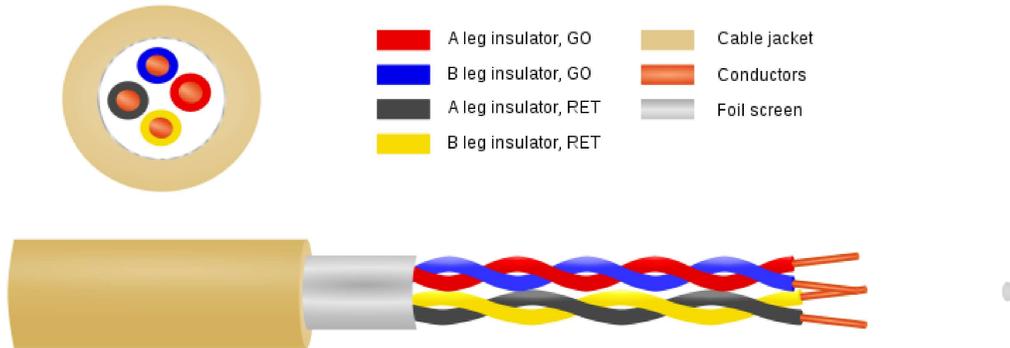


Fig. 3. Balanced line in DM quad format. This line is intended for use with 4-wire circuits or two 2-wire circuits.



Fig. 4. Balanced line in twin lead format. This line is intended for use with RF circuits, particularly antennae.

Transmission of a signal over a balanced line reduces the influence of noise or interference due to external stray electric fields. Any external signal sources tend to induce only a common mode signal on the line and the balanced impedances to ground minimizes differential pickup due to stray electric fields. The conductors are sometimes twisted together to ensure that each conductor is equally exposed to any external magnetic fields that could induce unwanted noise.

Some balanced lines also have electromagnetic shielding to reduce the amount of noise introduced.

A balanced line allows a differential receiver to reduce the noise on a connection by rejecting common-mode interference. The lines have the same impedance to ground, so the interfering fields or currents induce the same voltage in both wires. Since the receiver responds only to the difference between the wires, it is not influenced by the induced noise voltage. If twisted pair becomes unbalanced, for example due to insulation failure, noise will be induced. Examples of twisted pairs include Cat-3 Ethernet cables or telephone wires.

Compared to unbalanced circuits, balanced lines reduce the amount of noise per distance, allowing a longer cable run to be practical. This is because electromagnetic interference will affect both signals the same way. Similarities between the two signals are automatically removed at the end of the transmission path when one signal is subtracted from the other.

Telephone systems

The first application for balanced lines was for telephone lines. Interference that was of little consequence on a telegraph system (which is in essence digital) could be very disturbing for a telephone user. The initial format was to take two single-wire unbalanced telegraph lines and use them as a pair. This proved insufficient, however, with the growth of electric power transmission which tended to use the same routes. A telephone line running alongside a power line for many miles will inevitably have more interference induced in one leg than the other since one of them will be nearer to the power line. This

issue was addressed by swapping the positions of the two legs every few hundred yards with a cross-over, thus ensuring that both legs had equal interference induced and allowing common-mode rejection to do its work. As the telephone system grew, it became preferable to use cable rather than open wires to save space, and also to avoid poor performance during bad weather. The format used for balanced telephone cables was twisted pair, however, this did not become widespread until repeater amplifiers became available. On an unamplified line cable could only manage a maximum distance of 30 km. Open wires, on the other hand, with their lower capacitance had been used for enormous distances - the longest was the 1500 km from New York to Chicago built in 1893. Loading coils were used to improve the distance achievable with cable but the problem was not finally overcome until amplifiers started to be installed in 1912. Twisted pair balanced lines are still widely used for the telephone subscribers local end.

Telephone trunk lines, and especially frequency division multiplexing carrier systems, are usually 4-wire circuits rather than 2-wire circuits (or at least they were before fibre-optic became widespread) and require a different kind of cable. This format requires the conductors to be arranged in two pairs, one pair for the sending (go) signal and the other for the return signal. The greatest source of interference on this kind of transmission is usually the crosstalk between the go and return circuits themselves. The most common cable format is star quad, where the diagonally opposite conductors form the pairs. This geometry gives maximum common mode rejection between the two pairs. An alternative format is DM quad which consists of two twisted pairs with the twisting at different pitches.

Audio systems

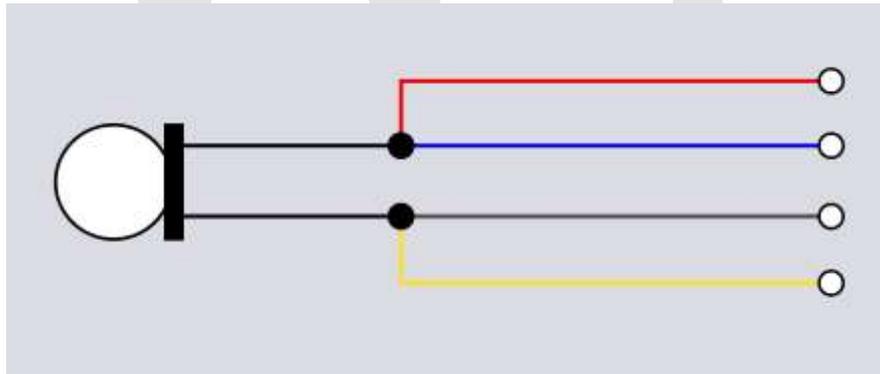


Fig. 5. Microphones connected to star quad cable join together diametrically opposite conductors to maintain balance. This is different from the usage on 4-wire circuits. The colours in this diagram correspond with the colouring in figure 2.

An example of balanced lines is the connection of microphones to a mixer in professional systems. Classically, both dynamic and condenser microphones used transformers to provide a differential-mode signal. While transformers are still used in the large majority of modern dynamic microphones, more recent condenser microphones are more likely to use electronic drive circuitry. Each leg, irrespective of any signal, should have an

identical impedance to ground. Pair cable (or a pair-derivative such as star quad) is used to maintain the balanced impedances and close twisting of the cores ensures that any interference is common to both conductors. Providing that the receiving end (usually a mixing console) does not disturb the line balance, and is able to ignore common-mode (noise) signals, and can extract differential ones, then the system will have excellent immunity to induced interference.

Typical professional audio sources, such as microphones, have three-pin XLR connectors. One is the shield or chassis ground, while the other two are signal connections. These signal wires carry two copies of the same signal, but with opposite polarity. (They are often termed "hot" and "cold," and the AES14-1992(r2004) Standard [and EIA Standard RS-297-A] suggest that the pin that carries the positive signal that results from a positive air pressure on a transducer will be deemed 'hot'. Pin 2 has been designated as the 'hot' pin, and that designation serves useful for keeping a consistent polarity in the rest of the system.) Since these conductors travel the same path from source to destination, the assumption is that any interference is induced upon both conductors equally. The appliance receiving the signals compares the difference between the two signals (often with disregard to electrical ground) allowing the appliance to ignore any induced electrical noise. Any induced noise would be present in equal amounts and in identical polarity on each of the balanced signal conductors, so the two signals' difference from each other would be unchanged. The successful rejection of induced noise from the desired signal depends in part on the balanced signal conductors receiving the same amount and type of interference. This typically leads to twisted, braided, or co-jacketed cables for use in balanced signal transmission.

Balanced and differential

Most explanations of balanced lines assume symmetrical (antiphase) signals but this is an unfortunate confusion - signal symmetry and balanced lines are quite independent of each other. Essential in a balanced line is matched impedances in the driver, line and receiver. These conditions assure that external noise affects each leg of the differential line equally and thus appears as a common mode signal that is removed by the receiver. There are balanced drive circuits that have excellent common-mode impedance matching between "legs" but do *not* provide symmetrical signals. Symmetrical differential signals exist to prevent interference with *other* circuits - the electromagnetic fields are canceled out by the equal and opposite currents. But they are not necessary for interference rejection *from* other circuits.

Baluns

To convert a signal from balanced to unbalanced requires a balun. For example, baluns can be used to send line level audio or E-carrier level 1 signals over coaxial cable (which is unbalanced) through 300 feet (91 m) of Category 5 cable by using a pair of baluns at each end of the CAT5 run. The balun takes the unbalanced signal, and creates an inverted copy of that signal. It then sends these 2 signals across the CAT5 cable as a balanced

signal. Upon reception at the other end, the balun takes the difference of the two signals, thus removing any noise picked up along the way and recreating the unbalanced signal.

A once common application of a radio frequency balun was found at the antenna terminals of a television receiver. Typically a 300-ohm balanced twin lead antenna input could only be connected to a coaxial cable from a cable TV system through a balun.

Characteristic Impedance

The characteristic impedance Z_0 of a transmission line is an important parameter at higher frequencies of operation. For a parallel 2-wire transmission line,

$$Z_0 = \frac{1}{\pi} \sqrt{\frac{\mu}{\epsilon}} \ln \left(\frac{l}{R} + \sqrt{\left(\frac{l}{R}\right)^2 - 1} \right),$$

where l is half the distance between the wire centres, R is the wire radius and μ , ϵ are respectively the permeability and permittivity of the surrounding medium. A commonly used approximation that is valid when the wire separation is much larger than the wire radius and in the absence of magnetic materials is

$$Z_0 = \frac{120}{\sqrt{\epsilon_r}} \ln \left(\frac{2l}{R} \right),$$

where ϵ_r is the relative permittivity of the surrounding medium.

Electric power lines

In electric power transmission, the three conductors used for three-phase power transmission are referred to as a balanced line since the instantaneous sum of the three line voltages is nominally zero. However, *balance* in this field is referring to the symmetry of the source and load: it has nothing to do with the impedance balance of the line itself, the sense of the meaning in telecommunications.

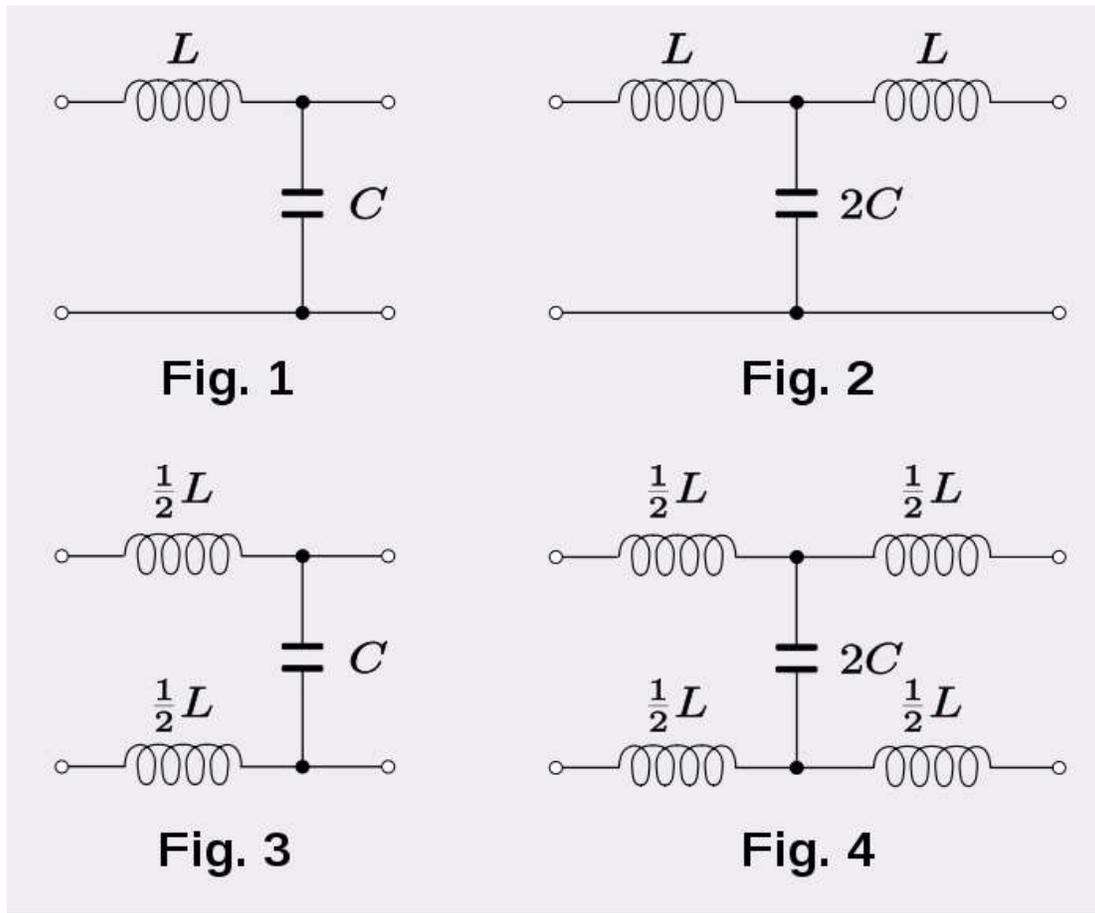
For the transmission of single-phase electric power as used for railway electrification systems, two conductors are used to carry in-phase and out-of-phase voltages such that the line is balanced.

Balanced circuit

A **balanced circuit** is circuitry for use with a balanced line or the balanced line itself. Balanced lines are a common method of transmitting many types of electrical communication signals between two points on two wires. In a balanced line the two signal lines are of a matched impedance to help ensure that interference induced in the line is common-mode and can be removed at the receiving end by circuitry with good common-mode rejection. To maintain the balance, circuit blocks which interface to the line, or are connected in the line, must also be balanced.

Balanced lines work because the interfering noise from the surrounding environment is induced into both wires equally. By measuring the difference between the two wires at the receiving end, the original signal is recovered while the noise is cancelled. Any inequality in the noise induced in each wire is an imbalance and will result in the noise not being fully cancelled. One requirement for balance is that both wires are an equal distance from the noise source. This is often achieved by placing the wires as close together as possible and twisting them together. Another requirement is that the impedance to ground (or to whichever reference point is being used by the difference detector) is the same for both conductors at all points along the length of the line. If one wire has a higher impedance to ground it will tend to have a higher noise induced, destroying the balance.

Balance and symmetry



Examples circuits using a low-pass filter to demonstrate. **Fig. 1.** Unbalanced, asymmetrical circuit. **Fig. 2.** Unbalanced, symmetrical circuit. **Fig. 3.** Balanced, asymmetrical circuit. **Fig. 4.** Balanced, symmetrical circuit.

A balanced circuit will normally show a symmetry of its components about a horizontal line mid-way between the two conductors (example in figure 3). This is different from what is normally meant by a symmetrical circuit which is a circuit showing symmetry of its components about a vertical line at its mid-point. An example of a symmetrical circuit is shown in figure 2. Circuits designed for use with balanced lines will often be designed to be both balanced and symmetrical as shown in figure 4. The advantages of symmetry are that the same impedance is presented at both ports and that the circuit has the same effect on signals travelling in both directions on the line.

Balance and symmetry are usually associated with reflected horizontal and vertical physical symmetry respectively as shown in figures 1 to 4. However, physical symmetry is not a necessary requirement for these conditions. It is only necessary that the electrical impedances are symmetrical. It is possible to design circuits that are not physically symmetrical but which have equivalent impedances which are symmetrical.

Balanced signals and balanced circuits

A balanced signal is one where the voltages on each wire are symmetrical with respect to ground (or some other reference). That is, the signals are inverted with respect to each other. A balanced circuit is a circuit where the two sides have identical transmission characteristics in all respects. A balanced line is a line in which the two wires will carry balanced currents (that is, equal and opposite currents) when balanced (symmetrical) voltages are applied. The condition for balance of lines and circuits will be met, in the case of passive circuitry, if the impedances are balanced. The line and circuit remain balanced, and the benefits of common-mode noise rejection continue to apply, whether or not the applied signal is itself balanced (symmetrical), always provided that the generator producing that signal maintains the impedance balance of the line.

Driving and receiving circuits

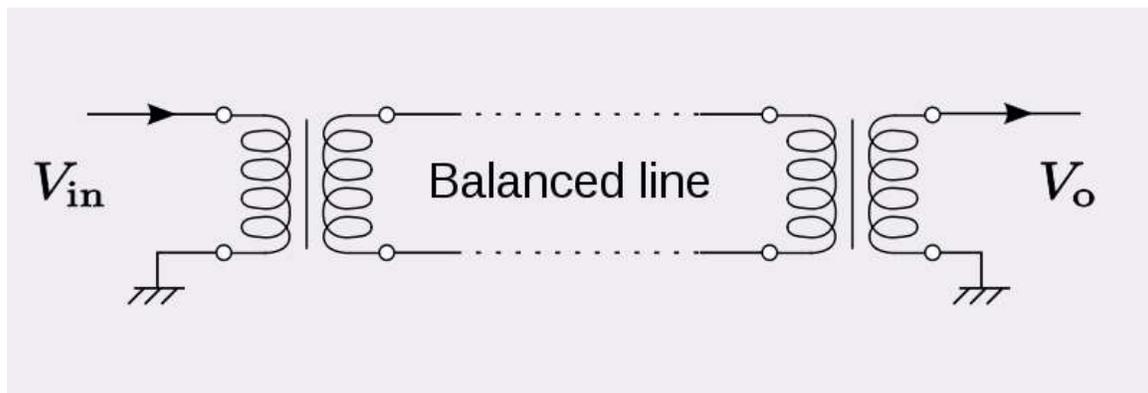


Fig. 5. Balanced line connected by transformers

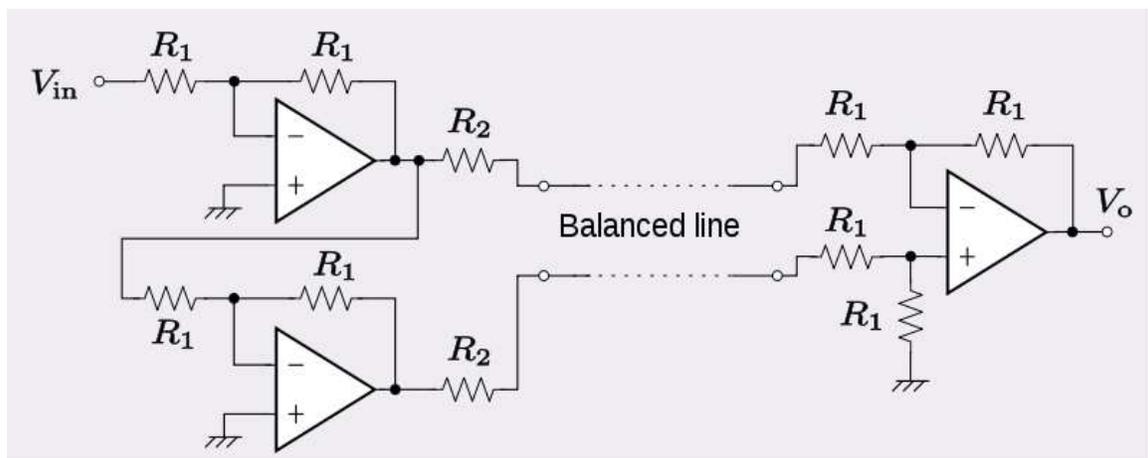


Fig. 6. Balanced line connected to electronically balanced circuitry

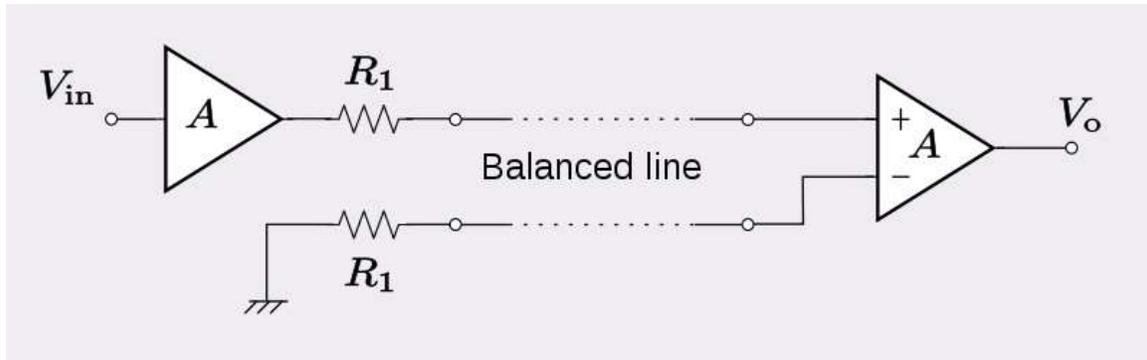


Fig. 7. Balanced line connected to an unbalanced signal, but with balanced impedances.

There are a number of ways that a balanced line can be driven and the signal detected. In all methods, for the continued benefit of good noise immunity, it is essential that the driving and receiving circuit maintain the impedance balance of the line. It is also essential that the receiving circuit detects only differential signals and rejects common-mode signals. It is not essential (although it is often the case) that the transmitted signal is balanced, that is, symmetrical about ground.

Transformer balance

The conceptually simplest way to connect to a balanced line is through transformers at each end shown in figure 5. Transformers were the original method of making such connections in telephony, and before the advent of active circuitry were the only way. In the telephony application they are known as repeating coils. Transformers have the additional advantage of completely isolating (or "floating") the line from earth and earth loop currents, which are an undesirable possibility with other methods, are completely eliminated.

The side of the transformer facing the line, in a good quality design, will have the winding laid in two parts (often with a centre tap provided) which are carefully balanced to maintain the line balance. Line side and equipment side windings are more useful concepts than the more usual primary and secondary windings when discussing these kinds of transformers. At the sending end the line side winding is the secondary, but at the receiving end the line side winding is the primary. When discussing a two-wire circuit primary and secondary cease to have any meaning at all, since signals are flowing in both directions at once.

The equipment side winding of the transformer does not need to be so carefully balanced. In fact, One leg of the equipment side can be earthed without effecting the balance on the line as shown in figure 5. With transformers the sending and receiving circuitry can be entirely unbalanced with the transformer providing the balancing.

Electronic balance

Electronic balance, or active balance, is achieved using differential amplifiers at each end of the line. An op-amp implementation of this is shown in figure 6, other circuitry is possible. Unlike transformer balance, there is no isolation of the circuitry from the line. Each of the two wires is driven by an op amp circuit which are identical except that one is inverting and one is non-inverting. Each one produces an unbalanced signal individually but together they drive the line with a symmetrical balanced signal. Because the currents in the two lines are equal and opposite, this has the further advantage that radiated signals cancel each other except in the near field of the conductors, thereby reducing cross-talk onto other conductors.

While it is not possible to create an isolated drive with op-amp circuitry alone, it is possible to create a floating output. This is important if one leg of the line might become grounded or connected to some other voltage reference. Grounding one leg of the line in the circuit of figure 6 will result in the line voltage being halved since only one op-amp is now providing signal. To achieve a floating output additional feedback paths are required between the two op-amps resulting in a more complex circuit than figure 6, but still avoiding the expense of a transformer. A floating op-amp output can only float within the limits of the op-amp's supply rails. An isolated output can be achieved without transformers with the addition of opto-isolators.

Impedance balance

As noted above, it is possible to drive a balanced line with an unbalanced signal and still maintain the line balance. This is represented in outline in figure 7. Amplifier *A* is assumed to be an ideal (that is, zero output impedance) unbalanced output amp. This is connected through a resistor to one leg of the line. The other leg is connected through another resistor of the same value. The impedance to ground of both legs is the same and the line remains balanced. The receiving amplifier still rejects any common-mode noise as it has a differential input. On the other hand the line signal is not balanced. The voltages at the input to the two legs, V_+ and V_- are given by;

$$V_+ = V_{in} \frac{Z_{in} + R_1}{Z_{in} + 2R_1}$$
$$V_- = V_{in} \frac{R_1}{Z_{in} + 2R_1}$$

Where Z_{in} is the input impedance of the line. These are clearly not symmetrical since V_- is much smaller than V_+ . They are not even opposite polarities.

Balanced to unbalanced conversion

A circuit that has the specific purpose of converting between balanced and unbalanced formats is called a balun. A balun could be a transformer with one leg earthed on the

unbalanced side as described in the transformer balance section above. Other circuits are possible such as autotransformers or active circuits.

Connectors

Common connectors used with balanced circuits include RJ-11 (telephone instruments), RJ-45 (broadband data) and XLR (professional audio) connectors. Also 1/4" tip ring sleeve (TRS) connectors were once widely used on manual switchboards and other telephone infrastructure. TRS connectors are now more commonly seen in miniature sizes (2.5 and 3.5 mm) being used for unbalanced stereo audio; however, some professional audio equipment such as mixing consoles still commonly use balanced "line-level" connections with 1/4" TRS jacks.



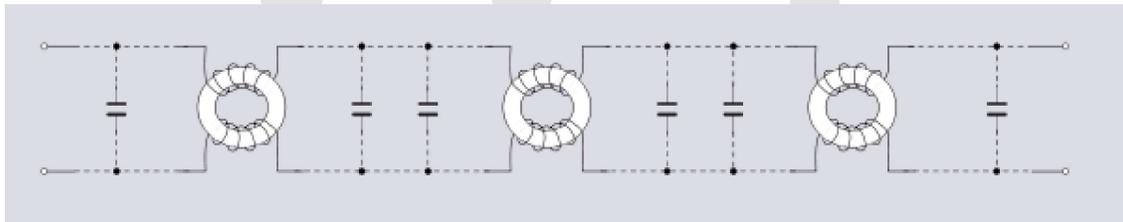
Chapter 12

Loading Coil

In electronics, a **loading coil** or **load coil** is a coil (inductor) that does not provide coupling to any other circuit, but is inserted in a circuit to increase its inductance. The need was discovered by Oliver Heaviside in studying the disappointing slow speed of the Transatlantic telegraph cable. He concluded additional inductance was required to prevent amplitude and time delay distortion of the transmitted signal. The mathematical condition for distortionless transmission is known as the Heaviside condition. Previous telegraph lines were overland or shorter, hence had less delay and the need for extra inductance was not so great. Submarine communications cables are particularly subject to the problem, but early 20th century ones using balanced pairs were often continuously loaded by iron tape rather than discretely by load coils.

Loading coils are archaically known as **Pupin coils** after Mihajlo Pupin (especially when used for the Heaviside condition), and the process of inserting them is sometimes called *pupinization*.

Applications



Schematic of a balanced loaded line. The capacitors are shown connected with dotted lines to indicate that the capacitance is actually distributed along the line rather than the discrete elements shown. The windings of the loading coil are wound such that the magnetic flux induced in the core is in the same direction for both windings.

Voice circuits

A common application of loading coils is to improve the voice-frequency amplitude response characteristics of the twisted balanced pairs in a telephone cable.

Loading coils inserted periodically in series with a pair of wires reduce the attenuation at the higher voice frequencies up to the cutoff frequency of the low-pass filter formed by

the inductance of the coils (plus the distributed inductance of the wires) and the distributed capacitance between the wires. Above the cutoff frequency, attenuation increases rapidly. The shorter the interval between the coils, the higher the cut-off frequency.

It should be emphasised that the cutoff effect is an artifact of using lumped inductors. With loading methods using continuous distributed inductance there is no cutoff.

Without loading coils, the line response is dominated by the resistance and capacitance of the line with the attenuation gently increasing with frequency. With loading coils of exactly the right inductance, neither capacitance nor inductance dominate: the response is flat, waveforms are undistorted and the characteristic impedance is resistive up to the cutoff frequency. The coincidental formation of an audio frequency filter is also beneficial in that noise is reduced.

DSL

When loading coils are in place, signal attenuation remains low for signals within the passband of the transmission line but increases rapidly for frequencies above the audio cutoff frequency. Thus, if the pair is subsequently reused to support applications that require higher frequencies (such as analog or digital carrier systems or DSL), any loading coils that were present on the line must be removed or replaced with one which is transparent to DSL. Using coils with parallel capacitors will form a filter with the topology of an m-derived filter and a band of frequencies above the cut-off will also be passed.

If they are not removed, as when the subscriber is an extended distance (e.g. over 4 miles) from the Central Office, DSL can not be supported. This sometimes happens in dense, growing areas (subject to frequent national numbering scheme repartitioning) such as Southern California in the late 1990s and early 21st century.

Carrier systems

American early and middle 20th Century telephone cables had load coils at intervals of a mile (1.61 km), usually in coil cases holding many. The coils must be removed to pass high frequencies, but the coil cases provided convenient places for repeaters for digital T-carrier systems, which could carry 1.5 Mbit/s across that distance. Due to narrower streets and higher cost of copper, European cables had thinner wires and needed closer intervals. Intervals of a kilometer allowed European systems to carry 2 Mbit/s.

Radio antennae

A (mobile) radio antenna, shorter than a quarter wavelength for practical reasons, presents capacitive reactance to a transmission line. This can be canceled by inserting an equal and opposite (inductive) reactance in series, by means of a loading coil typically at

the base or center of the antenna. Consequently the antenna presents a resistance (desirable) to the transmission line.

Campbell equation

The Campbell equation is a relationship due to George Ashley Campbell for predicting the propagation constant of a loaded line. It is stated as;

$$\cosh(\gamma'd) = \cosh(\gamma d) + \frac{Z}{2Z_0} \sinh(\gamma d)$$

where,

γ is the propagation constant of the unloaded line

γ' is the propagation constant of the loaded line

d is the interval between coils on the loaded line

Z is the impedance of a loading coil and

Z_0 is the characteristic impedance of the unloaded line.

A more engineer friendly rule of thumb is that the approximate requirement for spacing loading coils is ten coils per wavelength of the maximum frequency being transmitted. This approximation can be arrived at by treating the loaded line as a constant k filter and applying image filter theory to it. From basic image filter theory the angular cutoff frequency and the characteristic impedance of a low-pass constant k filter are given by;

$$\omega_c = \frac{1}{\sqrt{L_{\frac{1}{2}} C_{\frac{1}{2}}}} \quad \text{and,} \quad Z_0 = \sqrt{\frac{L_{\frac{1}{2}}}{C_{\frac{1}{2}}}}$$

where $L_{\frac{1}{2}}$ and $C_{\frac{1}{2}}$ are the half section element values.

From these basic equations the necessary loading coil inductance and coil spacing can be found;

$$L = \frac{Z_0}{\omega_c} \quad \text{and,} \quad d = \frac{2}{\omega_c Z_0 C}$$

where C is the capacitance per unit length of the line.

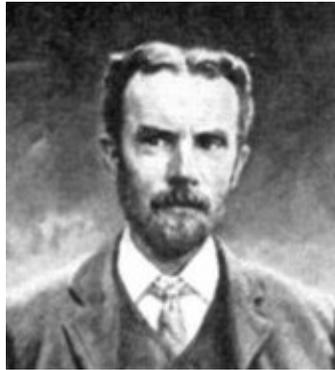
Expressing this in terms of number of coils per cutoff wavelength yields;

$$\frac{\lambda_c}{d} = \pi v Z_0 C$$

where v is the velocity of propagation of the cable in question.

History

Oliver Heaviside



Oliver Heaviside

The origin of the loading coil can be found in the work of Oliver Heaviside on the theory of transmission lines. Heaviside (1881) represented the line as a network of infinitesimally small circuit elements. By applying his operational calculus to the analysis of this network he discovered (1887) what has become known as the Heaviside condition. This is the condition that must be fulfilled in order for a transmission down a line to be free from distortion. The Heaviside condition is that the line series impedance, Z , must be proportional to the line shunt admittance, Y , at all frequencies. In terms of the primary line coefficients this is the condition;

$$\frac{R}{G} = \frac{L}{C}$$

where;

R is the series resistance of the line per unit length

L is the series self-inductance of the line per unit length

G is the shunt leakage conductance of the line insulator per unit length

C is the shunt capacitance between the line conductors per unit length

Heaviside was aware that this condition was not met in the practical telegraph cables in use in his day. In general, a real cable would have,

$$\frac{R}{G} \gg \frac{L}{C}$$

This is mainly due to the low value of leakage through the cable insulator, which is even more pronounced in modern cables which have better insulators than in Heaviside's day. In order to meet the condition, the choices are therefore to try and increase G or L or to decrease R or C . Decreasing R requires larger conductors. Copper was already in use in

telegraph cables and this is the very best conductor available short of using silver. Decreasing R means using more copper and a more expensive cable. Decreasing C would also mean a larger cable (although not necessarily more copper). Increasing G is highly undesirable, while it would reduce distortion, it would at the same time increase the signal loss. Heaviside considered, but rejected, this possibility which left him with the strategy of increasing L as the way to reduce distortion.

Heaviside immediately (1887) proposed several methods of increasing the inductance, including spacing the conductors further apart and loading the insulator with iron dust. Finally, Heaviside made the proposal (1893) to use discrete inductors at intervals along the line. However, he never succeeded in persuading the British GPO to take up the idea. Brittain attributes this to Heaviside's failure to provide engineering details on the size and spacing of the coils for particular cable parameters. Heaviside's eccentric character and setting himself apart from the establishment may also have played a part in their ignoring of him.

John Stone

John S. Stone worked for the American Telephone & Telegraph Company (AT&T) and was the first to attempt to apply Heaviside's ideas to real telecommunications. Stone's idea (1896) was to use a bimetallic iron-copper cable which he had patented. This cable of Stone's would increase the line inductance due to the iron content and had the potential to meet the Heaviside condition. However, Stone left the company in 1899 and the idea was never implemented.

George Campbell

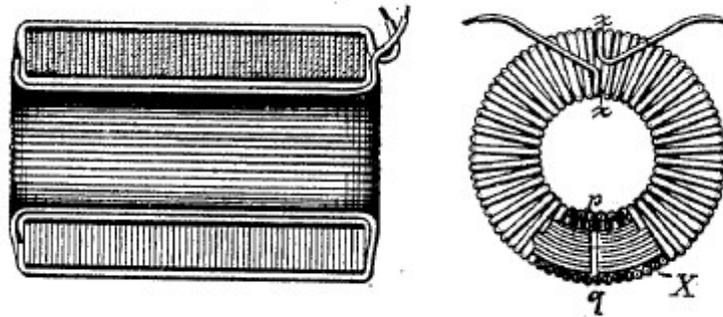
George Campbell was another AT&T engineer working for them in their Boston facility. Campbell was tasked with continuing the investigation into Stone's bimetallic cable, but soon abandoned this in favour of the loading coils idea. This was an independent discovery, Campbell being aware of Heaviside's work in discovering the Heaviside condition, but apparently not aware of Heaviside's suggestion of using loading coils to force a line to meet it. The motivation for the change of direction was Campbell's limited budget.

Campbell was struggling to set up a practical demonstration over a real telephone route with the budget he had been allocated. After considering that his artificial line simulators used lumped components rather than the distributed quantities found in a real line, he wondered if he could not insert the inductance with lumped components instead of using Stone's distributed line. When his calculations showed that the manholes on telephone routes were sufficiently close together to be able to insert the loading coils without the expense of either having to dig up the route or lay in new cables he changed to this new plan. The very first demonstration of loading coils on a telephone cable was on a 46-mile length of the so-called Pittsburgh cable (the test was actually in Boston, the cable had previously been used for testing in Pittsburgh) on September 6, 1899 carried out by

Campbell himself and his assistant. The first telephone cable using loaded lines put into public service was between Jamaica Plain and West Newton in Boston on May 18, 1900.

Campbell's work on loading coils provided the theoretical basis for his subsequent work on filters which proved to be so important for frequency-division multiplexing. The cut-off phenomena of loading coils, an undesirable side-effect, can be exploited to produce a desirable filter frequency response.

Michael Pupin



Pupin's design of loading coil

Michael Pupin, inventor and Serbian immigrant to the USA, also played a part in the story of loading coils. Pupin filed a rival patent to the one of Campbell's. This patent of Pupin's dates from 1899. There is an earlier patent (1894, filed December 1893) which is sometimes cited as Pupin's loading coil patent but is, in fact, something different. The confusion is easy to understand, Pupin himself claims that he first thought of the idea of loading coils while climbing a mountain in 1894, although there is nothing from him published at that time.

Pupin's 1894 patent "loads" the line with capacitors rather than inductors, a scheme that has been criticised as being theoretically flawed and never put into practice. To add to the confusion, one variant of the capacitor scheme proposed by Pupin does indeed have coils. However, these are not intended to compensate the line in any way. They are there merely to restore DC continuity to the line so that it may be tested with regular equipment. Pupin states that the inductance is to be so large that it will block all AC signals above 50 Hz. Consequently, only the capacitor is adding any significant impedance to the line and "the coils will not exercise any material influence on the results before noted".

Legal battle

Heaviside never patented his idea; indeed, he took no commercial advantage of any of his work. Despite the legal disputes surrounding this invention, it is unquestionable that

Campbell was the first to actually construct a telephone circuit using loading coils. There also can be little doubt that Heaviside was the first to publish and many would dispute Pupin's priority.

AT&T fought a legal battle with Pupin over his claim. Pupin was first to patent but Campbell had already conducted practical demonstrations before Pupin had even filed his patent (December 1899). Campbell's delay in filing was due to the slow internal machinations of AT&T.

However, AT&T foolishly deleted from Campbell's proposed patent application all the tables and graphs detailing the exact value of inductance that would be required before the patent was submitted. Since Pupin's patent contained a (less accurate) formula, AT&T was open to claims of incomplete disclosure. Fearing that there was a risk that the battle would end with the invention being declared unpatentable due to Heaviside's prior publication, they decided to desist from the challenge and buy an option on Pupin's patent for a yearly fee so that AT&T would control both patents. By January 1901 Pupin had been paid \$200,000 and by 1917, when the AT&T monopoly ended and payments ceased, he had received a total of \$455,000.

Benefit to AT&T

The invention was of enormous value to AT&T. Telephone cables could now be used to twice the distance previously possible, or alternatively, a cable of half the previous quality (and cost) could be used over the same distance. When considering whether to allow Campbell to go ahead with the demonstration, their engineers had estimated that they stood to save \$700,000 in new installation costs in New York and New Jersey alone. It has been estimated that AT&T saved \$100 million in the first quarter of the 20th century. Heaviside, who began it all, came away with nothing. He was offered a token payment but would not accept, wanting the credit for his work. He remarked ironically that if his prior publication had been admitted it would "interfere . . . with the flow of dollars in the proper direction . . ."

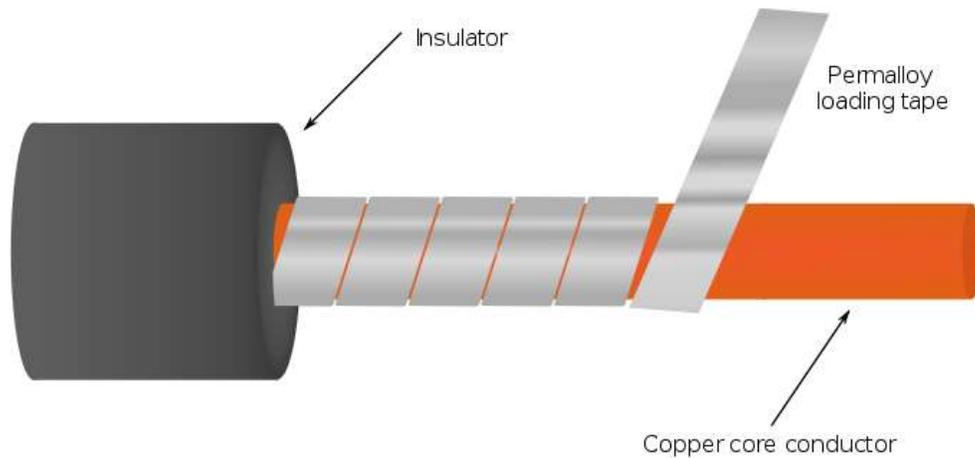
Krarup cable

Loading coils were not without their problems. For submarine cables where they were of most benefit, they were difficult to lay. The cable needed to be heavier and both this and the discontinuities in the profile where the coils occurred caused stresses in the cable during laying. Without great care, the cable might part and would be enormously expensive, possibly impossible, to fix. A second problem was that the material science of the time had difficulties sealing the joint between coil and cable against ingress of seawater. When this occurred, of course, the cable was ruined.

A Danish engineer, Carl Emil Krarup, invented a form of continuously loaded cable which solved these problems and the cable is named for him. Krarup cable has iron wires continuously wound around the central copper conductor with adjacent turns in contact with each other. This cable was the first use of continuous loading on any

telecommunication cable. In 1902 Krarup both wrote his paper on this subject and saw the installation of the first cable between Helsingør (Denmark) and Helsingborg (Sweden).

Permalloy cable

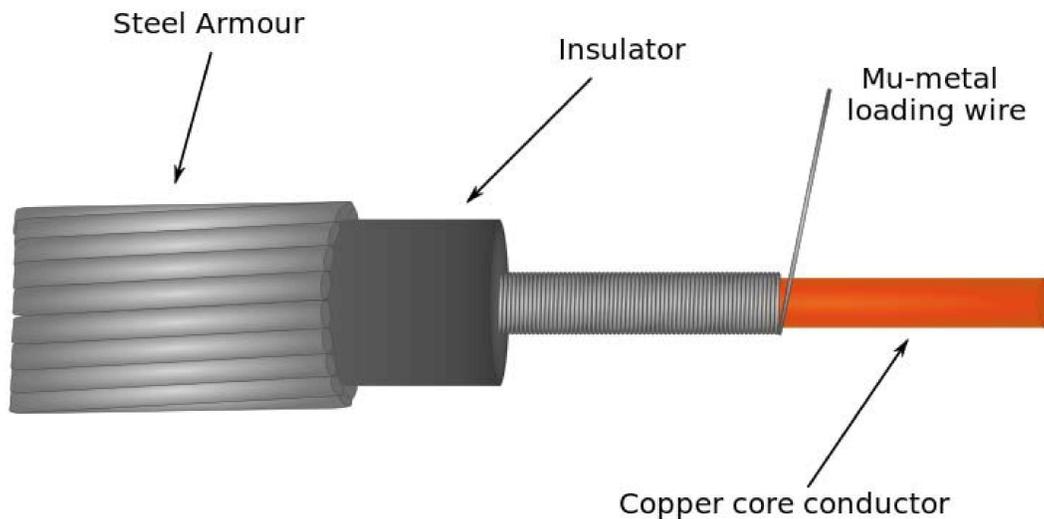


Permalloy cable construction

Even though Krarup cable added inductance to the line, it did not add enough to meet the Heaviside condition. AT&T searched for a better material with higher magnetic permeability. In 1914 Gustav Elmen discovered permalloy, a magnetic nickel-iron annealed alloy. Oliver E. Buckley, along with his colleagues at Bell Labs, H. D. Arnold and Elmen, c.1915 proposed a method of constructing submarine cable using permalloy tape wrapped around the copper conductors. This construction greatly improved the performance of the cable.

The cable was tested in a trial in Bermuda in 1923. The first permalloy cable to be put into service was between New York and Horta (Azores) in September 1924.

Mu-metal cable



Mu-metal cable construction

Mu-metal has similar magnetic properties to permalloy but the addition of copper to the alloy increases the ductility and allows the metal to be drawn into wire. Mu-metal cable is easier to construct than permalloy cable, the mu-metal being wound around the core copper conductor in much the same way as the iron wire in Krarup cable. A further advantage with mu-metal cable is that the construction lends itself to a variable loading profile whereby the loading is tapered towards the ends.

Mu-metal was invented (1923) by The Telegraph Construction and Maintenance Company Ltd., London, who made the cable, initially, for the Western Union Telegraph Co. Western Union were in competition with AT&T and the Western Electric Company who were using permalloy (the patent for permalloy was held by Western Electric).

Current practice

Loaded cable is no longer a useful technology for submarine communication cables, having first been superseded by co-axial cable using electrically powered in-line repeaters and then by fibre-optic cable. Manufacture of loaded cable declined in the 1930s and was then superseded by other technologies post-war. Loading coils can still be found in some telephone landlines today but new installations would use more modern technology.

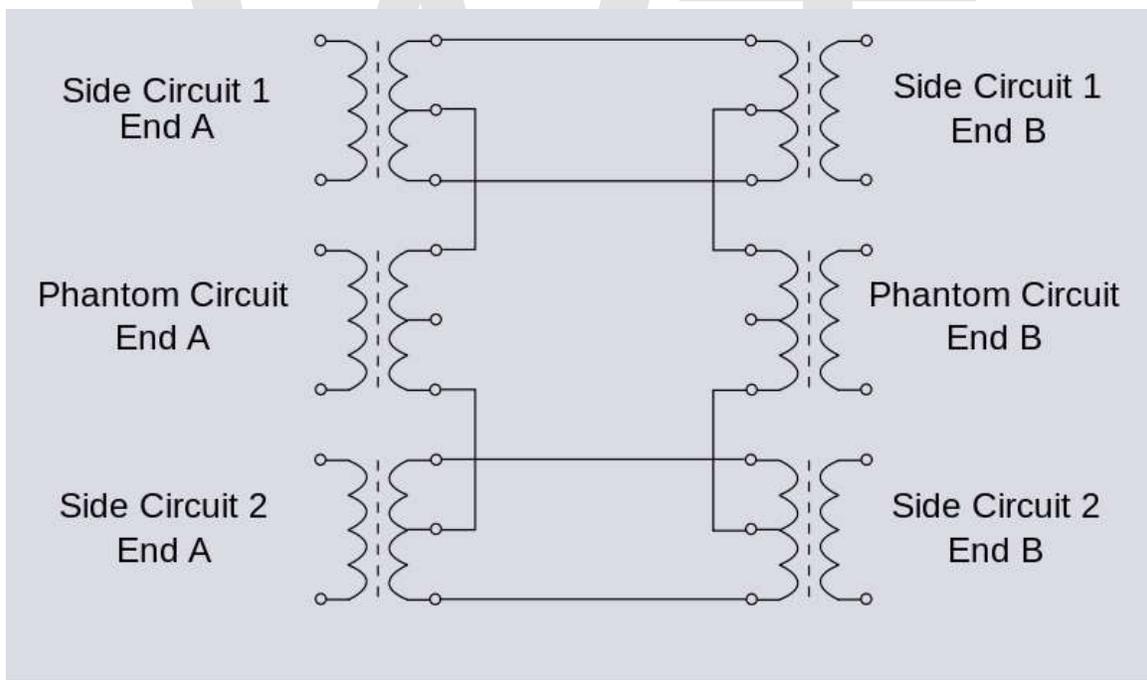
Chapter 13

Phantom Circuit and Frequency Mixer

Phantom circuit

In telecommunication and electrical engineering, a **phantom circuit** is an electrical circuit derived from suitably arranged wires with one or more conductive paths being a circuit in itself and at the same time acting as one conductor of another circuit.

Phantom group



Phantom circuit derived from two subscriber circuits

A **phantom group** is composed of three circuits that are derived from two single-channel circuits to form a *phantom circuit*. Here the phantom circuit is a third circuit derived from two suitably arranged pairs of wires, called side circuits, with each pair of wires being a circuit in itself and at the same time acting as one conductor of the third circuit. The "side circuits" within phantom circuits can be coupled to their respective voltage drops by center-tapped transformers, usually called "repeating coils". The center taps are on the

line side of the side circuits. Current from the phantom circuit is split evenly by the center taps. This cancels crosstalk from the phantom circuit to the side circuits.

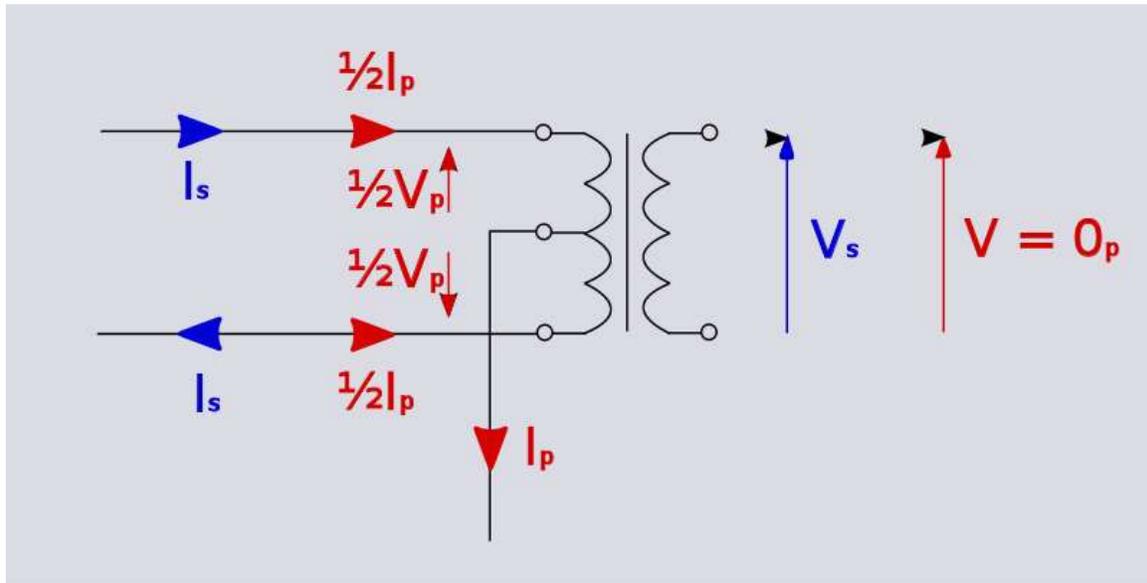


Diagram showing how the phantom currents (red) cancel in the transformer. Side circuit currents (blue) do not cancel and are transmitted through the transformer.

Phantom working increased the number of circuits on long distance routes in the early 20th century without putting up more wires. Phantoming declined with the adoption of carrier systems.

It is theoretically possible to create a phantom circuit from two other phantom circuits and so on up in a pyramid with a maximum $2n-1$ circuits being derived from n original circuits. However, more than one level of phantoming is usually impractical. Isolation between the phantom circuit and the side circuits relies on accurate balance of the line and transformers. Imperfect balance results in crosstalk between the phantom and side circuits and this effect accumulates as each level of phantoms is added. Even small levels of crosstalk are unacceptable on analogue telecommunications circuits since speech crosstalk is still intelligible down to quite low levels.

Phantom microphone powering

Recording and broadcast studios commonly use phantom powering as a means to provide power to microphones. Power may be needed either for a device that requires power such as a pre-amp on an electret microphone or because the microphone is a type that intrinsically requires powering such as a condenser microphone. Since the microphone has only one pair of wires the return path for the power has to be provided elsewhere. This is usually done via the microphone cable screen.

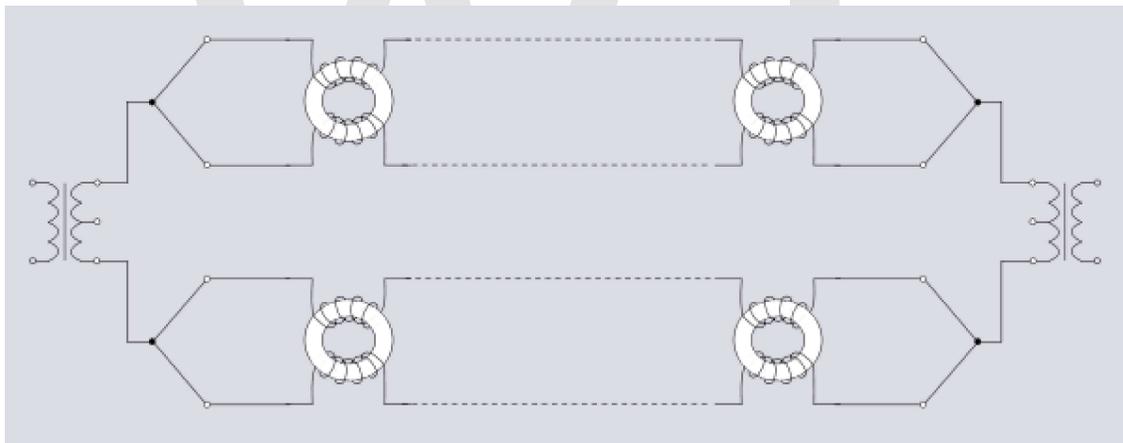
DC phantom

Simple DC signalling can be achieved on a telecommunications line in a similar way to phantom powering of microphones. A switch connected to the transformer centre-tap at one end of the line can operate a similarly connected relay at the other end. The return path is through the ground connection. This arrangement can be used for remotely controlling equipment.

Carrier circuit phantoms

From the 1950s to around the 1980s, using phantoms on star-quad trunk carrier circuits was a popular method of deriving a high quality broadcast audio circuit. The multiplexed FDM telecommunications carrier system usually did not use the baseband of the cable because it was inconvenient to separate low frequencies with filters. On the other hand, a one-way audio phantom could be formed from the two pairs (go and return signals) making up the star-quad cable.

Unloaded phantom



Unloaded phantom configuration. The windings of the loading coil are wound such that the magnetic flux induced in the core is normally in the same direction for both windings. However, in the phantom configuration the flux cancels.

Unloaded phantom is a phantom configuration of loaded lines (a circuit fitted with loading coils). The idea here is not to create additional circuits. Rather, the purpose is to cancel or greatly reduce the effect of the loading coils fitted to a line. The reason for doing this is that loaded lines have a definite cut-off frequency and it may be desired to equalise the line to a frequency which is higher than this, for example to make a circuit suitable for use by a broadcaster. Ideally, the loading would be removed or reduced for a permanent connection, but this is not feasible for temporary arrangements such as a requirement for outside broadcast. Instead, two circuits in a phantom configuration can be

used to greatly reduce the inductance being inserted by the loading coils, and hence the loading effect.

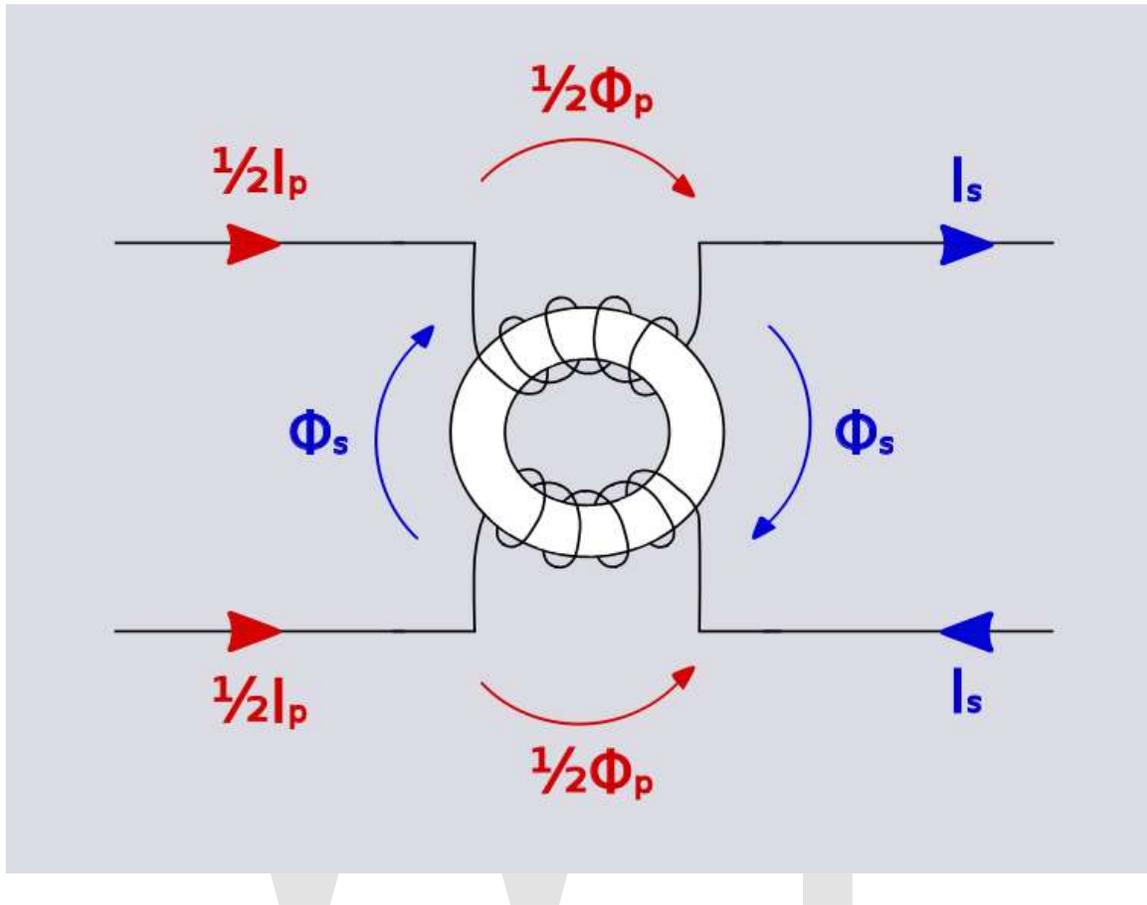


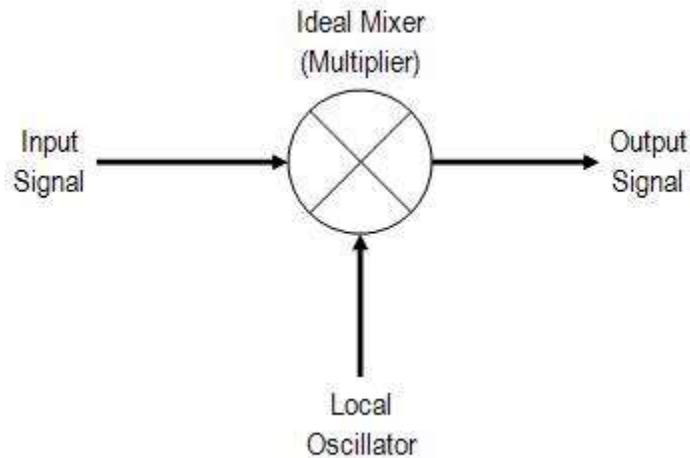
Diagram showing how the flux due to the phantom currents (red) is cancelled in the load coil. Flux due to normal line currents (blue) is additive.

It works because the loading coils used on balanced lines have two windings, one for each leg of the circuit. They are both wound on a common core and the windings are so arranged that the magnetic flux induced by both of them is in the same direction. Both windings induce an emf in each other as well as their own self-induction. This effect greatly increases the inductance of the coil and hence its loading effectiveness. By contrast, when the circuit is in the phantom configuration the currents in the two wires of each pair are in the same direction and the magnetic flux is being cancelled. This has precisely the opposite effect and the inductance is greatly reduced.

This configuration is most commonly used on the two pairs of a star-quad cable. It is not so successful with other pairs of wires. The difference in the path of the two pairs can easily destroy the balance and results in crosstalk and interference.

This configuration can also be called "bunched pairs". However, "bunched pairs" can also refer to the straightforward connection of two lines in parallel which is not a phantom circuit and will not reduce the loading.

Frequency mixer



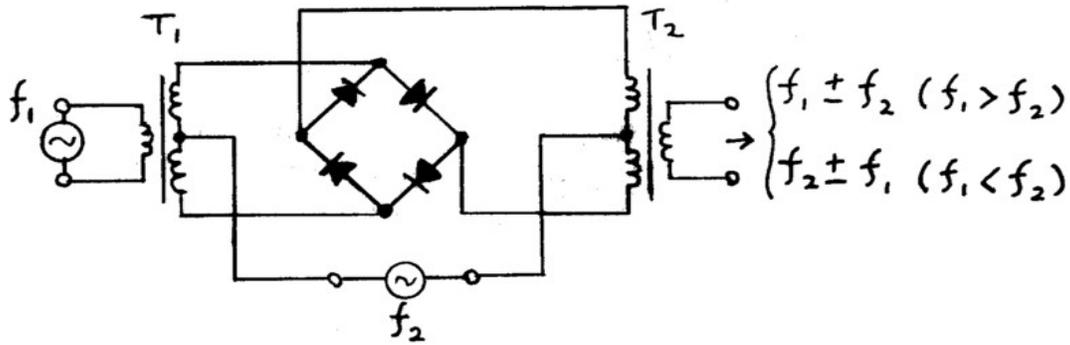
Frequency Mixer Symbol

In electronics a **mixer** or **frequency mixer** is a nonlinear electrical circuit that creates new frequencies from two signals applied to it. In its most common application, two signals at frequencies f_1 and f_2 are applied to a mixer, and it produces new signals at the sum $f_1 + f_2$ and difference $f_1 - f_2$ of the original frequencies. Other frequency components may also be produced in a practical frequency mixer.

Mixers are widely used to shift signals from one frequency range to another, a process known as heterodyning, for convenience in transmission or further signal processing. For example, a key component of a superheterodyne receiver is a mixer used to move received signals to a common intermediate frequency. Frequency mixers are also used to modulate a carrier frequency in radio transmitters.

Types

Passive mixers use one or more diodes and rely on the non-linear relation between voltage and current to provide the multiplying element. In a passive mixer, the desired output signal is always of lower power than the input signals. Active mixers can increase the strength of the product signal. Active types improve isolation between the ports, but may have higher noise and more power consumption and be less tolerant of overload. Mixers may be built of discrete components, may be part of integrated circuits, or can be delivered as hybrid modules. Mixers may also be classified by their topology. Unbalanced mixers allow some of the input signal power to pass through to the output. A single-balanced mixer is arranged so that the local oscillator (or RF) signal port cancels. A doubly-balanced mixer has symmetrical paths for both inputs, and will have no output if either input signal is not present.



Schematic diagram of a double-balanced passive diode mixer. There is no output unless both f_1 and f_2 inputs are present.

Selection of a mixer type is a trade off for a particular application. Mixer circuits are characterized by conversion gain, and noise figure. Balanced and double-balanced designs allow less of the input signals to feed through to the output.

Nonlinear electronic components that are used as mixers include diodes, transistors biased near cutoff, and at lower frequencies, analog multipliers. Ferromagnetic-core inductors driven into saturation have also been used. In nonlinear optics, crystals with nonlinear characteristics are used to mix two frequencies of laser light to create optical heterodynes.

Diode

A diode can be used to create a simple mixer. The importance of the diode is that it is non-linear (or non-Ohmic), which means its response (current) is not proportional to its input (voltage). The diode therefore does not reproduce the frequencies of its driving voltage in the current through it, which allows the desired frequency manipulation. Certain other non-linear devices could be utilized similarly.

The current I through an ideal diode as a function of the voltage V across it is given by

$$I = I_S \left(e^{\frac{qV_D}{nkT}} - 1 \right)$$

where what is important is that V appears in e 's exponent. The exponential can be expanded as

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

and can be approximated for small x (that is, small voltages) by the first few terms of that series:

$$e^x - 1 \approx x + \frac{x^2}{2}$$

Suppose that the sum of the two input signals $v_1 + v_2$ is applied to a diode, and that an output voltage is generated that is proportional to the current through the diode (perhaps by providing the voltage that is present across a resistor in series with the diode). Then, disregarding the constants in the diode equation, the output voltage will have the form

$$v_o = (v_1 + v_2) + \frac{1}{2}(v_1 + v_2)^2 + \dots$$

The first term on the right is the original two signals, as expected, followed by the square of the sum, which can be rewritten as $(v_1 + v_2)^2 = v_1^2 + 2v_1v_2 + v_2^2$, where the multiplied signal is obvious. The ellipsis represents all the higher powers of the sum which we assume to be negligible for small signals.

Switching

Another form of mixer operates by switching, with the smaller input signal being passed inverted or uninverted according to the phase of the local oscillator (LO). This would be typical of the normal operating mode of a packaged double balanced mixer module such as an SBL-1, with the local oscillator drive considerably higher than the signal amplitude.

The aim of a switching mixer is to achieve linear operation over the signal level, and hard switching driven by the local oscillator. Mathematically the switching mixer is not much different from a multiplying mixer, just because instead of the LO sine wave term we would use the signum function. In the frequency domain the switching mixer operation leads to the usual sum and difference frequencies, but also to further terms e.g. $+3*f_{LO}$, $+5*f_{LO}$, etc. The advantage of a switching mixer is that it can achieve - with the same effort - a lower noise figure (NF) and larger conversion gain. This come because the switching diodes or transistors act either like a low resistor (switch closed) or large resistor (switch open) and in both cases only minimum noise is added. From the circuit perspective many multiplying mixers can be used as switching mixers, just by increasing the LO amplitude. So RF engineers simply talk about mixers, and mean switching mixers.

Applications

The mixer circuit can be used not only to shift the frequency of an input signal as in a receiver, but also as a product detector, modulator, phase detector or frequency multiplier. For example a communications receiver might contain two mixer stages for conversion of the input signal to an intermediate frequency, and another mixer employed as a detector for demodulation of the signal.

Chapter 14

Phase-locked Loop

A **phase-locked loop** or **phase lock loop** (PLL) is a control system that tries to generate an output signal whose phase is related to the phase of the input "reference" signal. It is an electronic circuit consisting of a variable frequency oscillator and a phase detector. This circuit compares the phase of the input signal with the phase of the signal derived from its output oscillator and adjusts the frequency of its oscillator to keep the phases matched. The signal from the phase detector is used to control the oscillator in a feedback loop.

Frequency is the derivative of phase. Keeping the input and output phase in lock step implies keeping the input and output frequencies in lock step. Consequently, a phase-locked loop can track an input frequency, or it can generate a frequency that is a multiple of the input frequency. The former property is used for demodulation, and the latter property is used for indirect frequency synthesis.

Phase-locked loops are widely used in radio, telecommunications, computers and other electronic applications. They may generate stable frequencies, recover a signal from a noisy communication channel, or distribute clock timing pulses in digital logic designs such as microprocessors. Since a single integrated circuit can provide a complete phase-locked-loop building block, the technique is widely used in modern electronic devices, with output frequencies from a fraction of a hertz up to many gigahertz.

Practical analogies

Automobile race analogy

For a practical idea of what is going on, consider an auto race. There are many cars, and each of them wants to go around the track as fast as possible. Each lap corresponds to a complete cycle, and each car will complete dozens of laps per hour. The number of laps per hour (a speed) is a frequency, but the number of laps (a distance) corresponds to a phase. At one instant, car 3 may have gone 37.23 laps.

During most of the race, each car is on its own and is trying to beat every other car on the course. However, if there is an accident, a pace car comes out to set a safe speed. None of the race cars is permitted to pass the pace car (or the race cars in front of them), but each of the race cars wants to stay as close to the pace car as it can. While it is on the track, the pace car is a reference, and the race cars become phase-locked loops. Each driver will

measure the phase difference (a distance in laps) between him and the pace car. If the driver is far away, he will increase his engine speed to close the gap. If he's too close to the pace car, he will slow down. The result is all the race cars lock on to the phase of the pace car. The cars travel around the track in a tight group that is a small fraction of a lap.

Clock analogy

Phase can be proportional to time, so a phase difference can be a time difference. Clocks are, with varying degrees of accuracy, phase-locked (time-locked) to a master clock.

Left on its own, each clock will mark time at slightly different rates. A wall clock, for example, might be fast by a few seconds per hour compared to the reference clock at NIST. Over time, that time difference would become substantial.

To keep his clock in synch, each week the owner compares the time on his wall clock to a more accurate clock (a phase comparison), and he resets his clock. Left alone, the wall clock will continue to diverge from the reference clock at the same few seconds per hour rate.

Some clocks have a timing adjustment (a fast-slow control). When the owner compared his wall clock's time to the reference time, he noticed that his clock was too fast. Consequently, he could turn the timing adjust a small amount to make the clock run a little slower. If things work out right, his clock will be more accurate. Over a series of weekly adjustments, the wall clock's notion of a second would agree with the reference time (within the wall clock's stability).

An early mechanical version of a phase-locked loop was used in 1921 in the Short-Synchronome clock.

History

Automatic synchronization of electronic oscillators was described in 1923. Earliest research towards what became known as the phase-locked loop goes back to 1932, when British researchers developed an alternative to Edwin Armstrong's superheterodyne receiver, the Homodyne or direct-conversion receiver. In the homodyne or synchrodyne system, a local oscillator was tuned to the desired input frequency and multiplied with the input signal. The resulting output signal included the original modulation information. The intent was to develop an alternative receiver circuit that required fewer tuned circuits than the superheterodyne receiver. Since the local oscillator would rapidly drift in frequency, an automatic correction signal was applied to the oscillator, maintaining it in the same phase and frequency as the desired signal. The technique was described in 1932, in a paper by Henri de Bellescize, in the French journal *L'Onde Électrique*.

In analog television receivers since at least the late 1930s, phase-locked-loop horizontal and vertical sweep circuits are locked to synchronization pulses in the broadcast signal.

When Signetics introduced a line of monolithic integrated circuits that were complete phase-locked loop systems on a chip in 1969, applications for the technique multiplied. A few years later RCA introduced the "CD4046" CMOS Micropower Phase-Locked Loop, which became a popular integrated circuit.

Structure and function

Phase-locked loop mechanisms may be implemented as either analog or digital circuits. Both implementations use the same basic structure. Both analog and digital PLL circuits include four basic elements:

- Phase detector,
- low-pass filter
- Variable frequency oscillator, and
- feedback path (which may include a frequency divider).

Variations

There are several variations of PLLs. Some terms that are used are analog phase-locked loop (APLL) also referred to as a linear phase-locked loop (LPLL), digital phase-locked loop (DPLL), all digital phase-locked loop (ADPLL), and software phase-locked loop (SPLL).

Analog or Linear PLL (LPLL)

Phase detector is an analog multiplier. Loop filter is active or passive. Uses a Voltage-controlled oscillator (VCO).

Digital PLL (DPLL)

An analog PLL with a digital phase detector (such as XOR, edge-trigger JK, phase frequency detector). May have digital divider in the loop.

All digital PLL (ADPLL)

Phase detector, filter and oscillator are digital. Uses a numerically-controlled oscillator (NCO).

Software PLL (SPLL)

Functional blocks are implemented by software rather than specialized hardware.

Performance parameters

- Type and order
- Lock range: The frequency range the PLL is able to stay locked. Mainly defined by the VCO range.
- Capture range: The frequency range the PLL is able to lock-in, starting from unlocked condition. This range is usually smaller than the lock range and will depend e.g. on phase detector.
- Loop bandwidth: Defining the speed of the control loop.
- Transient response: Like overshoot and settling time to a certain accuracy (like 50ppm).

- Steady-state errors: Like remaining phase or timing error
- Output spectrum purity: Like sidebands generated from a certain VCO tuning voltage ripple.
- Phase-noise: Defined by noise energy in a certain frequency band (like 10kHz offset from carrier). Highly dependent on VCO phase-noise, PLL bandwidth, etc.
- General parameters: Such as power consumption, supply voltage range, output amplitude, etc.

Applications

Phase-locked loops are widely used for synchronization purposes; in space communications for coherent demodulation and threshold extension, bit synchronization, and symbol synchronization. Phase-locked loops can also be used to demodulate frequency-modulated signals. In radio transmitters, a PLL is used to synthesize new frequencies which are a multiple of a reference frequency, with the same stability as the reference frequency.

Other applications include:

- Demodulation of both FM and AM signals
- Recovery of small signals that otherwise would be lost in noise (lock-in amplifier)
- Recovery of clock timing information from a data stream such as from a disk drive
- Clock multipliers in microprocessors that allow internal processor elements to run faster than external connections, while maintaining precise timing relationships
- DTMF decoders, modems, and other tone decoders, for remote control and telecommunications

Clock recovery

Some data streams, especially high-speed serial data streams (such as the raw stream of data from the magnetic head of a disk drive), are sent without an accompanying clock. The receiver generates a clock from an approximate frequency reference, and then phase-aligns to the transitions in the data stream with a PLL. This process is referred to as clock recovery. In order for this scheme to work, the data stream must have a transition frequently enough to correct any drift in the PLL's oscillator. Typically, some sort of redundant encoding is used; 8B10B is very common.

Deskewing

If a clock is sent in parallel with data, that clock can be used to sample the data. Because the clock must be received and amplified before it can drive the flip-flops which sample the data, there will be a finite, and process-, temperature-, and voltage-dependent delay between the detected clock edge and the received data window. This delay limits the frequency at which data can be sent. One way of eliminating this delay is to include a deskew PLL on the receive side, so that the clock at each data flip-flop is phase-matched

to the received clock. In that type of application, a special form of a PLL called a delay-locked loop (DLL) is frequently used.

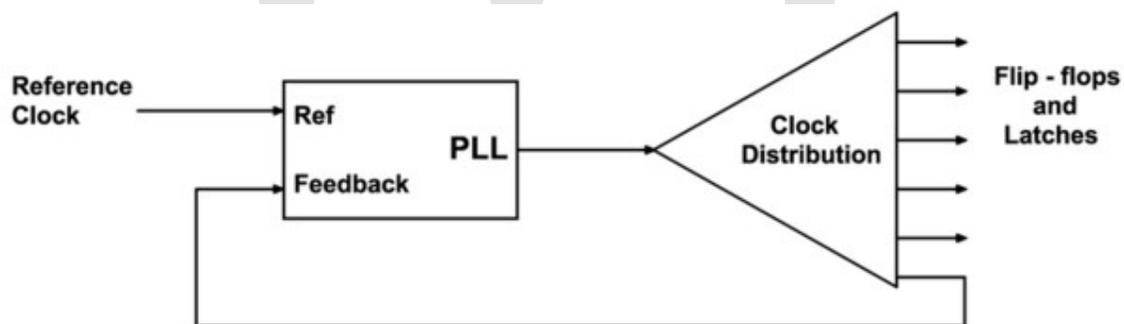
Clock generation

Many electronic systems include processors of various sorts that operate at hundreds of megahertz. Typically, the clocks supplied to these processors come from clock generator PLLs, which multiply a lower-frequency reference clock (usually 50 or 100 MHz) up to the operating frequency of the processor. The multiplication factor can be quite large in cases where the operating frequency is multiple gigahertz and the reference crystal is just tens or hundreds of megahertz.

Spread spectrum

All electronic systems emit some unwanted radio frequency energy. Various regulatory agencies (such as the FCC in the United States) put limits on the emitted energy and any interference caused by it. The emitted noise generally appears at sharp spectral peaks (usually at the operating frequency of the device, and a few harmonics). A system designer can use a spread-spectrum PLL to reduce interference with high-Q receivers by spreading the energy over a larger portion of the spectrum. For example, by changing the operating frequency up and down by a small amount (about 1%), a device running at hundreds of megahertz can spread its interference evenly over a few megahertz of spectrum, which drastically reduces the amount of noise seen on broadcast FM radio channels, which have a bandwidth of several tens of kilohertz.

Clock distribution



Typically, the reference clock enters the chip and drives a phase locked loop (PLL), which then drives the system's clock distribution. The clock distribution is usually balanced so that the clock arrives at every endpoint simultaneously. One of those endpoints is the PLL's feedback input. The function of the PLL is to compare the distributed clock to the incoming reference clock, and vary the phase and frequency of its output until the reference and feedback clocks are phase and frequency matched.

PLLs are ubiquitous—they tune clocks in systems several feet across, as well as clocks in small portions of individual chips. Sometimes the reference clock may not actually be a

pure clock at all, but rather a data stream with enough transitions that the PLL is able to recover a regular clock from that stream. Sometimes the reference clock is the same frequency as the clock driven through the clock distribution, other times the distributed clock may be some rational multiple of the reference.

Jitter and noise reduction

One desirable property of all PLLs is that the reference and feedback clock edges be brought into very close alignment. The average difference in time between the phases of the two signals when the PLL has achieved lock is called the **static phase offset** (also called the **steady-state phase error**). The variance between these phases is called **tracking jitter**. Ideally, the static phase offset should be zero, and the tracking jitter should be as low as possible.

Phase noise is another type of jitter observed in PLLs, and is caused by the oscillator itself and by elements used in the oscillator's frequency control circuit. Some technologies are known to perform better than others in this regard. The best digital PLLs are constructed with emitter-coupled logic (ECL) elements, at the expense of high power consumption. To keep phase noise low in PLL circuits, it is best to avoid saturating logic families such as transistor-transistor logic (TTL) or CMOS.

Another desirable property of all PLLs is that the phase and frequency of the generated clock be unaffected by rapid changes in the voltages of the power and ground supply lines, as well as the voltage of the substrate on which the PLL circuits are fabricated. This is called substrate and supply noise rejection. The higher the noise rejection, the better.

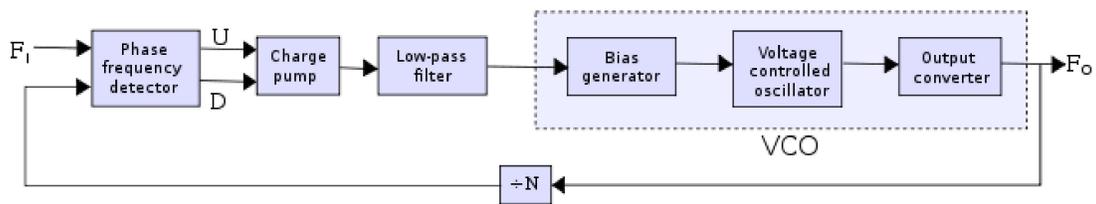
To further improve the phase noise of the output, an injection locked oscillator can be employed following the VCO in the PLL.

Frequency Synthesis

In digital wireless communication systems (GSM, CDMA etc.), PLLs are used to provide the local oscillator for up-conversion during transmission and down-conversion during reception. In most cellular handsets this function has been largely integrated into a single integrated circuit to reduce the cost and size of the handset. However, due to the high performance required of base station terminals, the transmission and reception circuits are built with discrete components to achieve the levels of performance required. GSM local oscillator modules are typically built with a frequency synthesizer integrated circuit and discrete resonator VCOs.

Frequency synthesizer manufacturers include Analog Devices, National Semiconductor and Texas Instruments. VCO manufacturers include Sirenza, Z-Communications, Inc. (Z-COMM).

Phase-locked loop block diagram



Digital phase-locked loop block diagram

A phase detector compares two input signals and produces an error signal which is proportional to their phase difference. The error signal is then low-pass filtered and used to drive a VCO which creates an output phase. The output is fed through an optional divider back to the input of the system, producing a negative feedback loop. If the output phase drifts, the error signal will increase, driving the VCO phase in the opposite direction so as to reduce the error. Thus the output phase is locked to the phase at the other input. This input is called the reference.

Analog phase locked loops are generally built with an analog phase detector, low pass filter and VCO placed in a negative feedback configuration. A digital phase locked loop uses a digital phase detector; it may also have a divider in the feedback path or in the reference path, or both, in order to make the PLL's output signal frequency a rational multiple of the reference frequency. A non-integer multiple of the reference frequency can also be created by replacing the simple divide-by-N counter in the feedback path with a programmable pulse swallowing counter. This technique is usually referred to as a fractional-N synthesizer or fractional-N PLL.

The oscillator generates a periodic output signal. Assume that initially the oscillator is at nearly the same frequency as the reference signal. If the phase from the oscillator falls behind that of the reference, the phase detector changes the control voltage of the oscillator so that it speeds up. Likewise, if the phase creeps ahead of the reference, the phase detector changes the control voltage to slow down the oscillator. Since initially the oscillator may be far from the reference frequency, practical phase detectors may also respond to frequency differences, so as to increase the lock-in range of allowable inputs.

Depending on the application, either the output of the controlled oscillator, or the control signal to the oscillator, provides the useful output of the PLL system.

Elements

Phase detector

The two inputs of the phase detector are the reference input and the feedback from the VCO. The PD output controls the VCO such that the phase difference between the two inputs is held constant, making it a negative feedback system. There are several types of phase detectors in the two main categories of analog and digital.

Different types of phase detectors have different performance characteristics.

For instance, the frequency mixer produces harmonics that adds complexity in applications where spectral purity of the VCO signal is important. The resulting unwanted (spurious) sidebands, also called "reference spurs" can dominate the filter requirements and reduce the capture range and lock time well below the requirements. In these applications the more complex digital phase detectors are used which do not have as severe a reference spur component on their output. Also, when in lock, the steady-state phase difference at the inputs using this type of phase detector is near 90 degrees. The actual difference is determined by the DC loop gain.

A **bang-bang** charge pump phase detector must always have a **dead band** where the phases of inputs are close enough that the detector detects no phase error. For this reason, bang-bang phase detectors are associated with significant minimum peak-to-peak jitter, because of drift within the dead band. However these types, having outputs consisting of very narrow pulses at lock, are very useful for applications requiring very low VCO spurious outputs. The narrow pulses contain very little energy and are easy to filter out of the VCO control voltage. This results in low VCO control line ripple and therefore low FM sidebands on the VCO.

In PLL applications it is frequently required to know when the loop is out of lock. The more complex digital phase-frequency detectors usually have an output that allows a reliable indication of an out of lock condition.

Filter

The block commonly called the PLL loop filter (usually a low pass filter) generally has two distinct functions.

The primary function is to determine loop dynamics, also called stability. This is how the loop responds to disturbances, such as changes in the reference frequency, changes of the feedback divider, or at startup. Common considerations are the range over which the loop can achieve lock (pull-in range, lock range or capture range), how fast the loop achieves lock (lock time, lock-up time or settling time) and damping behavior. Depending on the application, this may require one or more of the following: a simple proportion (gain or attenuation), an integral (low pass filter) and/or derivative (high pass filter). Loop parameters commonly examined for this are the loop's gain margin and phase margin.

Common concepts in control theory including the PID controller are used to design this function.

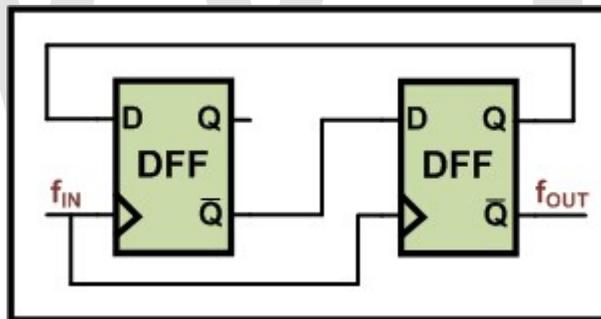
The second common consideration is limiting the amount of reference frequency energy (ripple) appearing at the phase detector output that is then applied to the VCO control input. This frequency modulates the VCO and produces FM sidebands commonly called "reference spurious". The low pass characteristic of this block can be used to attenuate this energy, but at times a band reject "notch" may also be useful.

The design of this block can be dominated by either of these considerations, or can be a complex process juggling the interactions of the two. Typical trade-offs are: increasing the bandwidth usually degrades the stability or too much damping for better stability will reduce the speed and increase settling time. Often also the phase-noise is affected.

Oscillator

All phase-locked loops employ an oscillator element with variable frequency capability. This can be an analog VCO either driven by analog circuitry in the case of an APLL or driven digitally through the use of a digital-to-analog converter as is the case for some DPLL designs. Pure digital oscillators such as a numerically-controlled oscillator are used in ADPLLs.

Feedback path and optional divider



An Example Digital Divider (by 4) for use in the Feedback Path of a Multiplying PLL

PLLs may include a divider between the oscillator and the feedback input to the phase detector to produce a frequency synthesizer. A programmable divider is particularly useful in radio transmitter applications, since a large number of transmit frequencies can be produced from a single stable, accurate, but expensive, quartz crystal-controlled reference oscillator.

Some PLLs also include a divider between the reference clock and the reference input to the phase detector. If the divider in the feedback path divides by N and the reference input divider divides by M , it allows the PLL to multiply the reference frequency by N / M . It might seem simpler to just feed the PLL a lower frequency, but in some cases the

reference frequency may be constrained by other issues, and then the reference divider is useful.

Frequency multiplication in a sense can also be attained by locking the PLL to the 'N'th harmonic of the signal.

It should also be noted that the feedback is not limited to a frequency divider. This element can be other elements such as a frequency multiplier, or a mixer. The multiplier will make the VCO output a sub-multiple (rather than a multiple) of the reference frequency. A mixer can translate the VCO frequency by a fixed offset. It may also be a combination of these. An example being a divider following a mixer; this allows the divider to operate at a much lower frequency than the VCO without a loss in loop gain.

Modeling

Time domain model

The equations governing a phase-locked loop with an analog multiplier as the phase detector may be derived as follows. Let the input to the phase detector be $x_c(t)$ and the output of the VCO is $x_r(t)$ with frequency $\omega_r(t)$, then the output of the phase detector $x_m(t)$ is given by

$$x_m(t) = x_c(t) \cdot x_r(t)$$

the VCO frequency may be written as a function of the VCO input $y(t)$ as

$$\omega_r(t) = \omega_f + g_v y(t)$$

where g_v is the *sensitivity* of the VCO and is expressed in Hz / V.

Hence the VCO output takes the form

$$x_r(t) = A_r \cos\left(\int_0^t \omega_r(\tau) d\tau\right) = A_r \cos(\omega_f t + \varphi(t))$$

where

$$\varphi(t) = \int_0^t g_v y(\tau) d\tau$$

The loop filter receives this signal as input and produces an output

$$x_f(t) = F_{\text{filter}}(x_m(t))$$

where F_{Filter} is the operator representing the loop filter transformation.

When the loop is closed, the output from the loop filter becomes the input to the VCO thus

$$y(t) = x_f(t) = F_{\text{filter}}(x_m(t))$$

We can deduce how the PLL reacts to a sinusoidal input signal:

$$x_c(t) = A_c \sin(\omega_c t).$$

The output of the phase detector then is:

$$x_m(t) = A_c \sin(\omega_c t) A_r \cos(\omega_f t + \varphi(t)).$$

This can be rewritten into sum and difference components using trigonometric identities:

$$x_m(t) = \frac{A_c A_f}{2} \sin(\omega_c t - \omega_f t - \varphi(t)) + \frac{A_c A_f}{2} \sin(\omega_c t + \omega_f t + \varphi(t))$$

As an approximation to the behaviour of the loop filter we may consider only the difference frequency being passed with no phase change, which enables us to derive a small-signal model of the phase-locked loop. If we can make $\omega_f \approx \omega_c$, then the $\sin(\cdot)$ can be approximated by its argument resulting in:

$y(t) = x_f(t) \simeq -A_c A_f \varphi(t) / 2$. The phase-locked loop is said to be *locked* if this is the case.

Linearized phase domain model

Phase locked loops can also be analyzed as control systems by applying the Laplace transform. The loop response can be written as:

$$\frac{\theta_o}{\theta_i} = \frac{K_p K_v F(s)}{s + K_p K_v F(s)}$$

Where

- θ_o is the output phase in radians
- θ_i is the input phase in radians
- K_p is the phase detector gain in volts per radian
- K_v is the VCO gain in radians per volt-second
- $F(s)$ is the loop filter transfer function (dimensionless)

The loop characteristics can be controlled by inserting different types of loop filters. The simplest filter is a one-pole RC circuit. The loop transfer function in this case is:

$$F(s) = \frac{1}{1 + sRC}$$

The loop response becomes:

$$\frac{\theta_o}{\theta_i} = \frac{\frac{K_p K_v}{RC}}{s^2 + \frac{s}{RC} + \frac{K_p K_v}{RC}}$$

This is the form of a classic harmonic oscillator. The denominator can be related to that of a second order system:

$$s^2 + 2s\zeta\omega_n + \omega_n^2$$

Where

- ζ is the damping factor
- ω_n is the natural frequency of the loop

For the one-pole RC filter,

$$\omega_n = \sqrt{\frac{K_p K_v}{RC}}$$

$$\zeta = \frac{1}{2\sqrt{K_p K_v RC}}$$

The loop natural frequency is a measure of the response time of the loop, and the damping factor is a measure of the overshoot and ringing. Ideally, the natural frequency should be high and the damping factor should be near 0.707 (critical damping). With a single pole filter, it is not possible to control the loop frequency and damping factor independently. For the case of critical damping,

$$RC = \frac{1}{2K_p K_v}$$

$$\omega_c = K_p K_v \sqrt{2}$$

A slightly more effective filter, the lag-lead filter includes one pole and one zero. This can be realized with two resistors and one capacitor. The transfer function for this filter is

$$F(s) = \frac{1 + sCR_2}{1 + sC(R_1 + R_2)}$$

This filter has two time constants

$$\begin{aligned}\tau_1 &= C(R_1 + R_2) \\ \tau_2 &= CR_2\end{aligned}$$

Substituting above yields the following natural frequency and damping factor

$$\begin{aligned}\omega_n &= \sqrt{\frac{K_p K_v}{\tau_1}} \\ \zeta &= \frac{1}{2\omega_n \tau_1} + \frac{\omega_n \tau_2}{2}\end{aligned}$$

The loop filter components can be calculated independently for a given natural frequency and damping factor

$$\begin{aligned}\tau_1 &= \frac{K_p K_v}{\omega_n^2} \\ \tau_2 &= \frac{2\zeta}{\omega_n} - \frac{1}{K_p K_v}\end{aligned}$$

Real world loop filter design can be much more complex e.g. using higher order filters to reduce various types or source of phase noise.