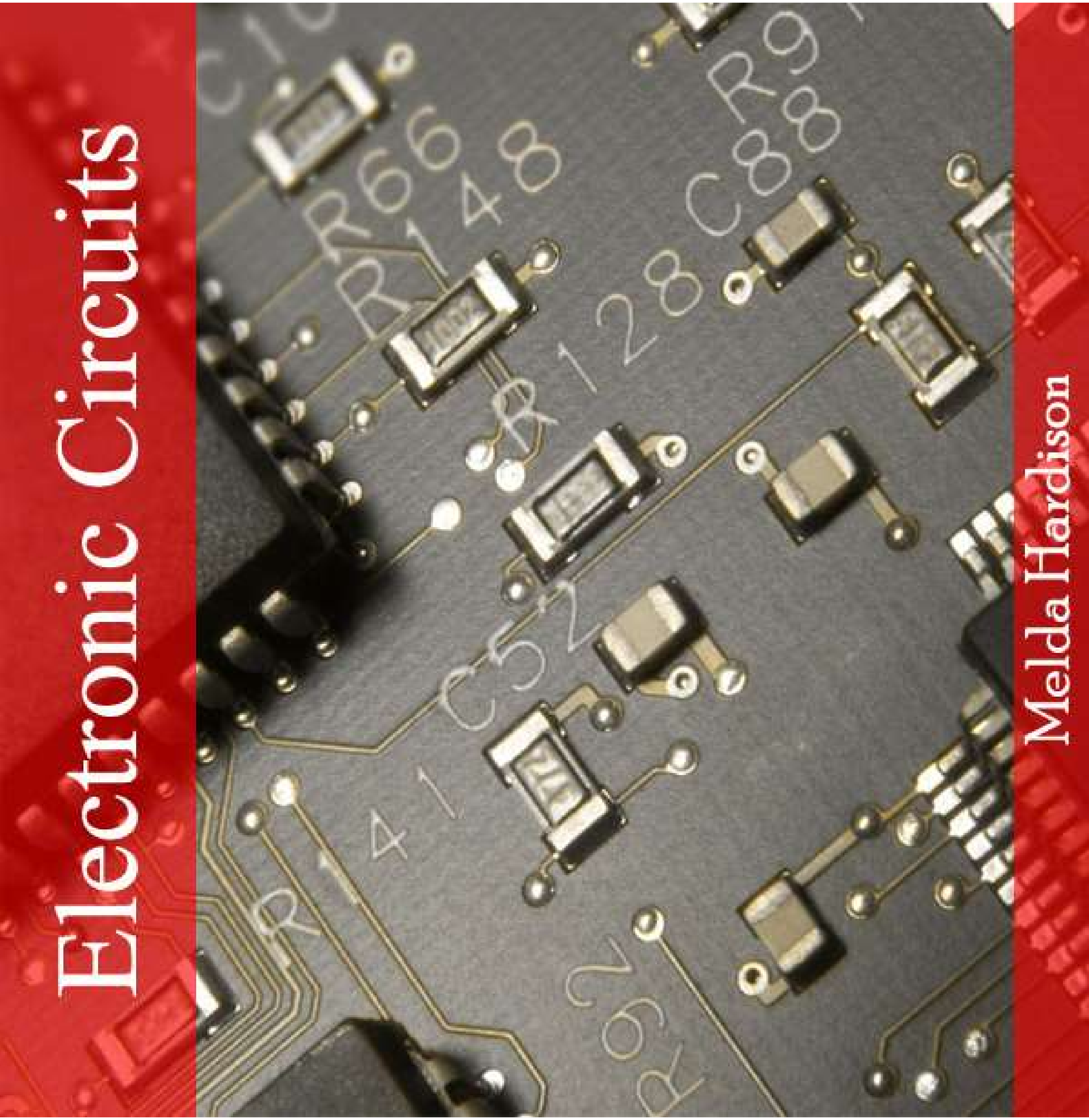

Electronic Circuits
Melda Hardison

First Edition, 2012

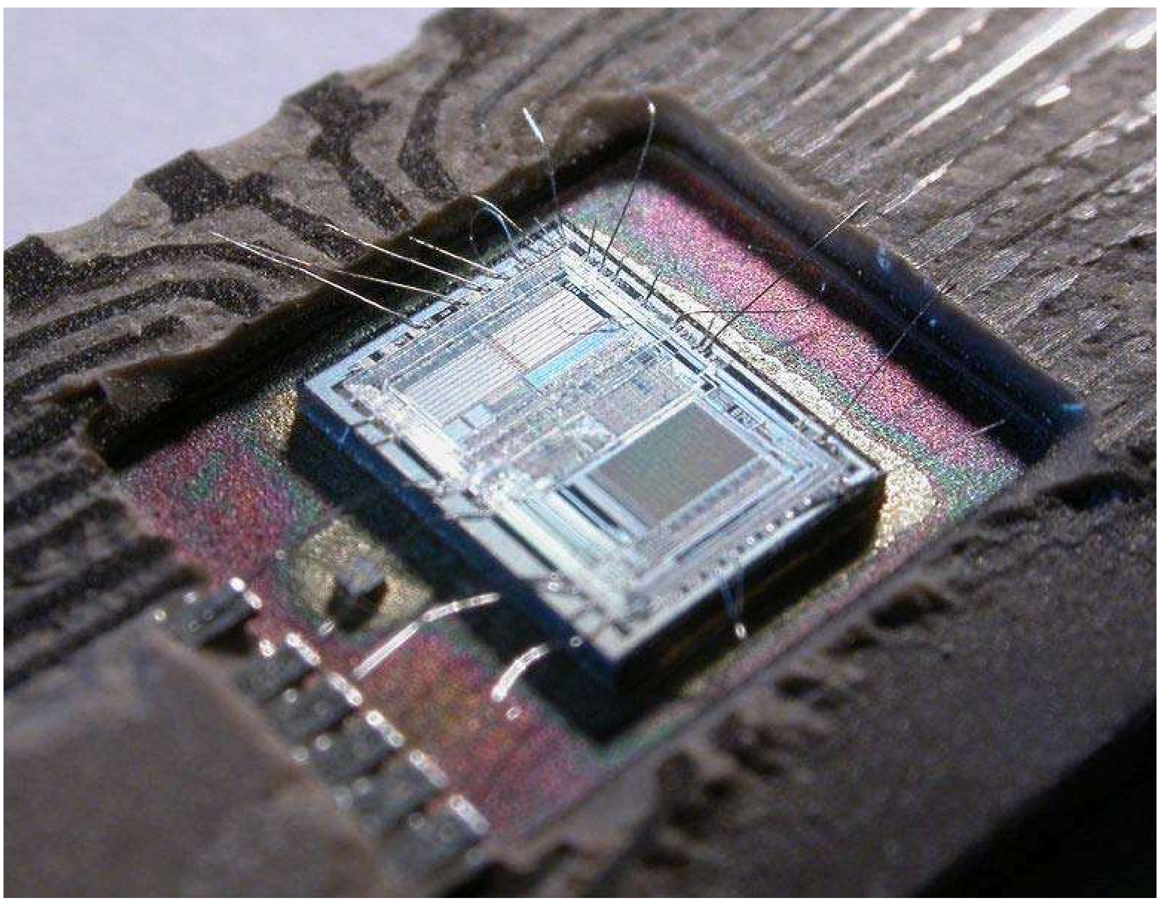
ISBN 978-81-323-2851-3

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

Table of Contents

Chapter- 1

Electronic Circuit



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

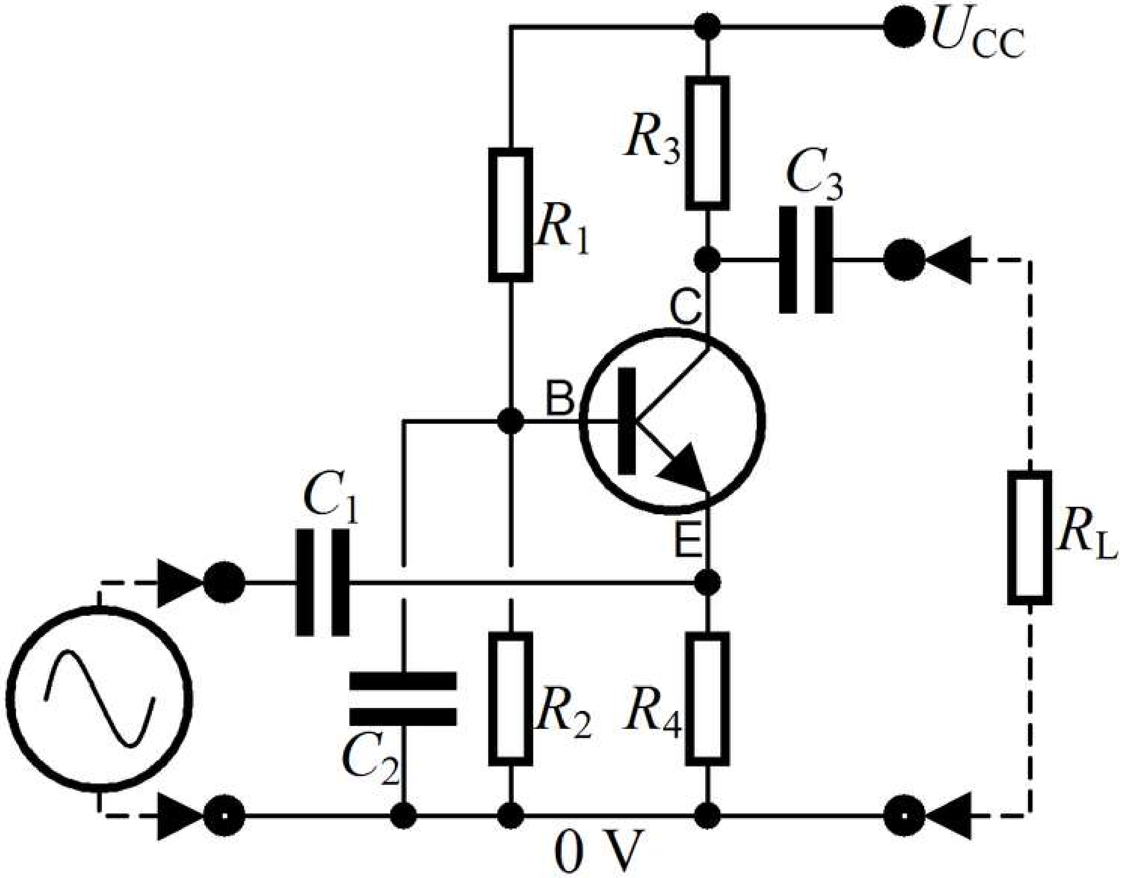
A circuit built on a printed circuit board (PCB).

An **electronic circuit** is composed of individual electronic components, such as resistors, transistors, capacitors, inductors and diodes, connected by conductive wires or traces through which electric current can flow. The combination of components and wires allows various simple and complex operations to be performed: signals can be amplified, computations can be performed, and data can be moved from one place to another. Circuits can be constructed of discrete components connected by individual pieces of wire, but today it is much more common to create interconnections by photolithographic techniques on a laminated substrate (a printed circuit board or PCB) and solder the components to these interconnections to create a finished circuit. In an Integrated Circuit or IC, the components and interconnections are formed on the same substrate, typically a semiconductor such as silicon or (less commonly) gallium arsenide.

Breadboards, perfboards or stripboards are common for testing new designs. They allow the designer to make quick changes to the circuit during development.

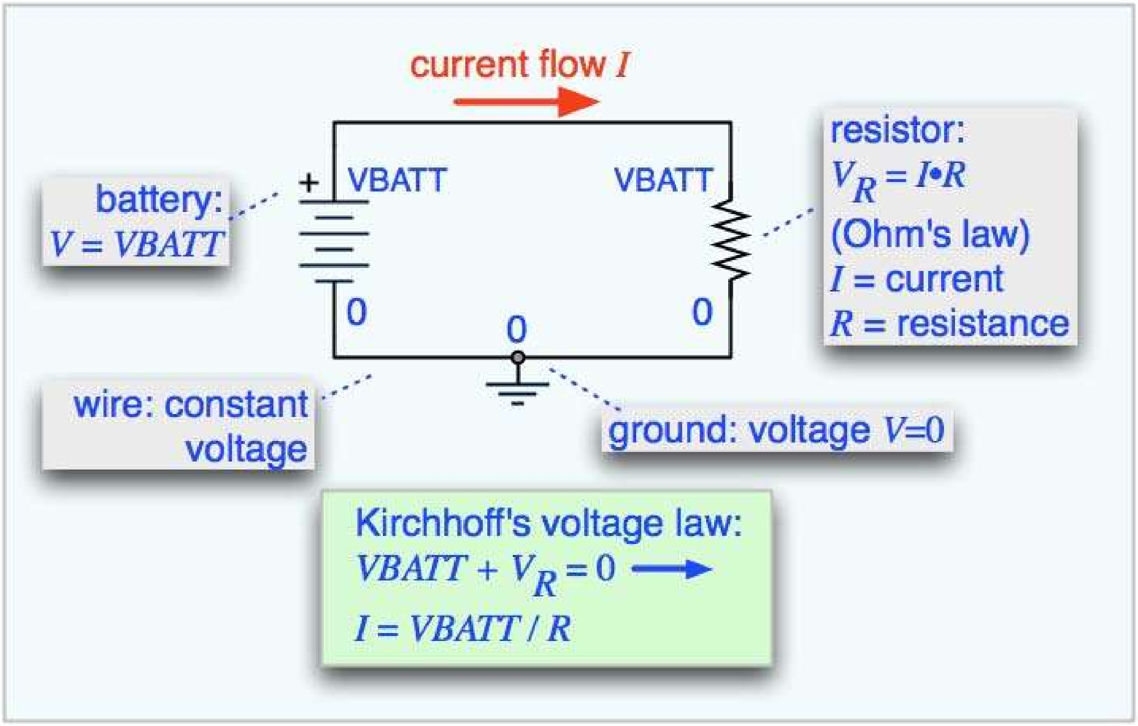
An electronic circuit can usually be categorized as an analog circuit, a digital circuit or a mixed-signal circuit (a combination of analog circuits and digital circuits).

Analog circuits



A circuit diagram representing an analog circuit, in this case a simple amplifier.

Analog electronic circuits are those in which current or voltage may vary continuously with time to correspond to the information being represented. Analog circuitry is constructed from two fundamental building blocks: series and parallel circuits. In a series circuit, the same current passes through a series of components. A string of Christmas lights is a good example of a series circuit: if one goes out, they all do. In a parallel circuit, all the components are connected to the same voltage, and the current divides between the various components according to their resistance.



A simple schematic showing wires, a resistor, and a battery.

The basic components of analog circuits are wires, resistors, capacitors, inductors, diodes, and transistors. (Recently, memristors have been added to the list of available components.) Analog circuits are very commonly represented in schematic diagrams, in which wires are shown as lines, and each component has a unique symbol. Analog circuit analysis employs Kirchhoff's circuit laws: all the currents at a node (a place where wires meet) must add to 0, and the voltage around a closed loop of wires is 0. Wires are usually treated as ideal zero-voltage interconnections; any resistance or reactance is captured by explicitly adding a parasitic element, such as a discrete resistor or inductor. Active components such as transistors are often treated as controlled current or voltage sources: for example, a field-effect transistor can be modeled as a current source from the source to the drain, with the current controlled by the gate-source voltage.

When the circuit size is comparable to a wavelength of the relevant signal frequency, a more sophisticated approach must be used. Wires are treated as transmission lines, with (hopefully) constant characteristic impedance, and the impedances at the start and end determine transmitted and reflected waves on the line. Such considerations typically become important for circuit boards at frequencies above a GHz; integrated circuits are smaller and can be treated as lumped elements for frequencies less than 10 GHz or so.

An alternative model is to take independent power sources and induction as basic electronic units; this allows modeling frequency dependent negative resistors, gyrators, negative impedance converters, and dependent sources as secondary electronic components.

Digital circuits

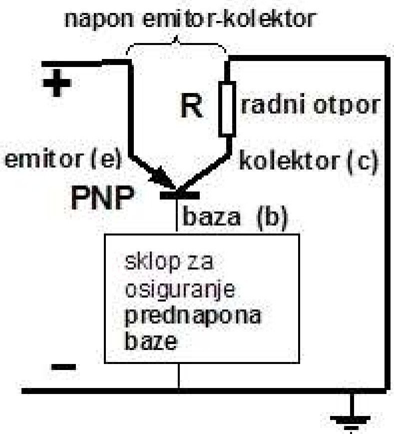
In digital electronic circuits, electric signals take on discrete values, to represent logical and numeric values. These values represent the information that is being processed. In the vast majority of cases, binary encoding is used: one voltage (typically the more positive value) represents a binary '1' and another voltage (usually a value near the ground potential, 0 V) represents a binary '0'. Digital circuits make extensive use of transistors, interconnected to create logic gates that provide the functions of Boolean logic: AND, OR, NOT, and all possible combinations there of. Transistors interconnected so as to provide positive feedback are used as latches and flip flops, circuits that have two or more metastable states, and remain in one of these states until changed by an external input. Digital circuits therefore can provide both logic and memory, enabling them to perform arbitrary computational functions. (Memory based on flip-flops is known as SRAM (static random access memory). Memory based on the storage of charge in a capacitor, DRAM (dynamic random access memory) is also widely used.)

Digital circuits are fundamentally easier to design than analog circuits for the same level of complexity, because each logic gate regenerates the binary signal, so the designer need not account for distortion, gain control, offset voltages, and other concerns faced in an analog design. As a consequence, extremely complex digital circuits, with billions of logic elements integrated on a single silicon chip, can be fabricated at low cost. Such digital integrated circuits are ubiquitous in modern electronic devices, such as calculators, mobile phone handsets, and computers.

Digital circuitry is used to create general purpose computing chips, such as microprocessors, and custom-designed logic circuits, known as Application Specific Integrated Circuits (ASICs). Field Programmable Gate Arrays (FPGAs), chips with logic circuitry whose configuration can be modified after fabrication, are also widely used in prototyping and development.

Mixed-signal circuits

Mixed-signal or hybrid circuits contain elements of both analog and digital circuits. Examples include comparators, timers, PLLs, ADCs (analog-to-digital converters), and DACs (digital-to-analog converters). Most modern radio and communications circuitry uses mixed signal circuits. For example, in a receiver, analog circuitry is used to amplify and frequency-convert signals so that they reach a suitable state to be converted into digital values, after which further signal processing can be performed in the digital domain.



Nekoliko oblika tranzistora i tiristora

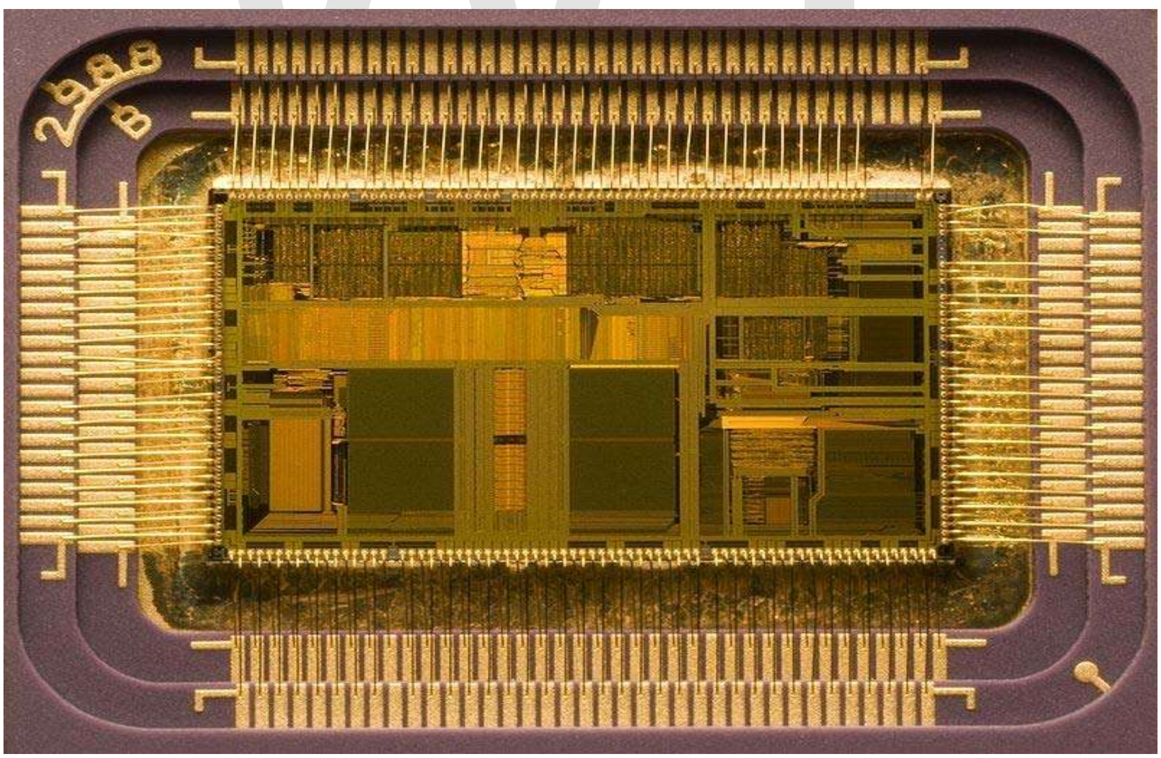
Tranzistor

Chapter- 2

Digital Electronics

Digital electronics represent signals by discrete bands of analog levels, rather than by a continuous range. All levels within a band represent the same signal state. Relatively small changes to the analog signal levels due to manufacturing tolerance, signal attenuation or parasitic noise do not leave the discrete envelope, and as a result are ignored by signal state sensing circuitry.

In most cases the number of these states is two, and they are represented by two voltage bands: one near zero volts and a higher level near the supply voltage, corresponding to the "false" ("0") and "true" ("1") values of the boolean domain respectively.



Intel 80486DX2

Hitachi J100

Digital techniques are useful because it is easier to get an electronic device to switch into one of a number of known states than to accurately reproduce a continuous range of values.

Digital electronic circuits are usually made from large assemblies of logic gates, simple electronic representations of Boolean logic functions.

Advantages

One advantage of digital circuits when compared to analog circuits is that signals represented digitally can be transmitted without degradation due to noise. For example, a continuous audio signal, transmitted as a sequence of 1s and 0s, can be reconstructed without error provided the noise picked up in transmission is not enough to prevent identification of the 1s and 0s. An hour of music can be stored on a compact disc as about 6 billion binary digits.

In a digital system, a more precise representation of a signal can be obtained by using more binary digits to represent it. While this requires more digital circuits to process the signals, each digit is handled by the same kind of hardware. In an analog system, additional resolution requires fundamental improvements in the linearity and noise characteristics of each step of the signal chain.

Computer-controlled digital systems can be controlled by software, allowing new functions to be added without changing hardware. Often this can be done outside of the factory by updating the product's software. So, the product's design errors can be corrected after the product is in a customer's hands.

Information storage can be easier in digital systems than in analog ones. The noise-immunity of digital systems permits data to be stored and retrieved without degradation. In an analog system, noise from aging and wear degrade the information stored. In a digital system, as long as the total noise is below a certain level, the information can be recovered perfectly.

Disadvantages

In some cases, digital circuits use more energy than analog circuits to accomplish the same tasks, thus producing more heat. In portable or battery-powered systems this can limit use of digital systems.

For example, battery-powered cellular telephones often use a low-power analog front-end to amplify and tune in the radio signals from the base station. However, a base station has grid power and can use power-hungry, but very flexible software radios. Such base stations can be easily reprogrammed to process the signals used in new cellular standards.

Digital circuits are sometimes more expensive, especially in small quantities.

Most useful digital systems must translate from continuous analog signals to discrete digital signals. This causes quantization errors. Quantization error can be reduced if the system stores enough digital data to represent the signal to the desired degree of fidelity. The Nyquist-Shannon sampling theorem provides an important guideline as to how much digital data is needed to accurately portray a given analog signal.

In some systems, if a single piece of digital data is lost or misinterpreted, the meaning of large blocks of related data can completely change. Because of the cliff effect, it can be difficult for users to tell if a particular system is right on the edge of failure, or if it can tolerate much more noise before failing.

Digital fragility can be reduced by designing a digital system for robustness. For example, a parity bit or other error management method can be inserted into the signal path. These schemes help the system detect errors, and then either correct the errors, or at least ask for a new copy of the data. In a state-machine, the state transition logic can be designed to catch unused states and trigger a reset sequence or other error recovery routine.

Digital memory and transmission systems can use techniques such as error detection and correction to use additional data to correct any errors in transmission and storage.

On the other hand, some techniques used in digital systems make those systems more vulnerable to single-bit errors. These techniques are acceptable when the underlying bits are reliable enough that such errors are highly unlikely. A single-bit error in audio data stored directly as linear pulse code modulation (such as on a CD-ROM) causes, at worst, a single click. Instead, many people use audio compression to save storage space and download time, even though a single-bit error may corrupt the entire song.

Analog issues in digital circuits

Digital circuits are made from analog components. The design must assure that the analog nature of the components doesn't dominate the desired digital behavior. Digital systems must manage noise and timing margins, parasitic inductances and capacitances, and filter power connections.

Bad designs have intermittent problems such as "glitches", vanishingly-fast pulses that may trigger some logic but not others, "runt pulses" that do not reach valid "threshold" voltages, or unexpected ("undecoded") combinations of logic states.

Additionally, where clocked digital systems interface to analogue systems or systems that are driven from a different clock, the digital system can be subject to metastability where a change to the input violates the set-up time for a digital input latch. This situation will self-resolve, but will take a random time, and while it persists can result in invalid signals being propagated within the digital system for a short time.

Since digital circuits are made from analog components, digital circuits calculate more slowly than low-precision analog circuits that use a similar amount of space and power. However, the digital circuit will calculate more repeatably, because of its high noise immunity. On the other hand, in the high-precision domain (for example, where 14 or more bits of precision are needed), analog circuits require much more power and area than digital equivalents.

Construction

A digital circuit is often constructed from small electronic circuits called logic gates that can be used to create combinational logic. Each logic gate represents a function of boolean logic. A logic gate is an arrangement of electrically controlled switches, better known as transistors.

Each logic symbol is represented by a different shape. The actual set of shapes was introduced in 1984 under IEEE\ANSI standard 91-1984. "The logic symbol given under this standard are being increasingly used now and have even started appearing in the literature published by manufacturers of digital integrated circuits."

The output of a logic gate is an electrical flow or voltage, that can, in turn, control more logic gates.

Logic gates often use the fewest number of transistors in order to reduce their size, power consumption and cost, and increase their reliability.

Integrated circuits are the least expensive way to make logic gates in large volumes. Integrated circuits are usually designed by engineers using electronic design automation software.

Another form of digital circuit is constructed from lookup tables, (many sold as "programmable logic devices", though other kinds of PLDs exist). Lookup tables can perform the same functions as machines based on logic gates, but can be easily reprogrammed without changing the wiring. This means that a designer can often repair design errors without changing the arrangement of wires. Therefore, in small volume products, programmable logic devices are often the preferred solution. They are usually designed by engineers using electronic design automation software.

When the volumes are medium to large, and the logic can be slow, or involves complex algorithms or sequences, often a small microcontroller is programmed to make an embedded system. These are usually programmed by software engineers.

When only one digital circuit is needed, and its design is totally customized, as for a factory production line controller, the conventional solution is a programmable logic controller, or PLC. These are usually programmed by electricians, using ladder logic.

Structure of digital systems

Engineers use many methods to minimize logic functions, in order to reduce the circuit's complexity. When the complexity is less, the circuit also has fewer errors and less electronics, and is therefore less expensive.

The most widely used simplification is a minimization algorithm like the Espresso heuristic logic minimizer within a CAD system, although historically, binary decision diagrams, an automated Quine–McCluskey algorithm, truth tables, Karnaugh Maps, and Boolean algebra have been used.

Representations are crucial to an engineer's design of digital circuits. Some analysis methods only work with particular representations.

The classical way to represent a digital circuit is with an equivalent set of logic gates. Another way, often with the least electronics, is to construct an equivalent system of electronic switches (usually transistors). One of the easiest ways is to simply have a memory containing a truth table. The inputs are fed into the address of the memory, and the data outputs of the memory become the outputs.

For automated analysis, these representations have digital file formats that can be processed by computer programs. Most digital engineers are very careful to select computer programs ("tools") with compatible file formats.

To choose representations, engineers consider types of digital systems. Most digital systems divide into "combinational systems" and "sequential systems." A combinational system always presents the same output when given the same inputs. It is basically a representation of a set of logic functions, as already discussed.

A sequential system is a combinational system with some of the outputs fed back as inputs. This makes the digital machine perform a "sequence" of operations. The simplest sequential system is probably a flip flop, a mechanism that represents a binary digit or "bit".

Sequential systems are often designed as state machines. In this way, engineers can design a system's gross behavior, and even test it in a simulation, without considering all the details of the logic functions.

Sequential systems divide into two further subcategories. "Synchronous" sequential systems change state all at once, when a "clock" signal changes state. "Asynchronous" sequential systems propagate changes whenever inputs change. Synchronous sequential systems are made of well-characterized asynchronous circuits such as flip-flops, that change only when the clock changes, and which have carefully designed timing margins.

The usual way to implement a synchronous sequential state machine is to divide it into a piece of combinational logic and a set of flip flops called a "state register." Each time a clock signal ticks, the state register captures the feedback generated from the previous state of the combinational logic, and feeds it back as an unchanging input to the combinational part of the state machine. The fastest rate of the clock is set by the most time-consuming logic calculation in the combinational logic.

The state register is just a representation of a binary number. If the states in the state machine are numbered (easy to arrange), the logic function is some combinational logic that produces the number of the next state.

In comparison, asynchronous systems are very hard to design because all possible states, in all possible timings must be considered. The usual method is to construct a table of the minimum and maximum time that each such state can exist, and then adjust the circuit to minimize the number of such states, and force the circuit to periodically wait for all of its parts to enter a compatible state (this is called "self-resynchronization"). Without such careful design, it is easy to accidentally produce asynchronous logic that is "unstable", that is, real electronics will have unpredictable results because of the cumulative delays caused by small variations in the values of the electronic components. Certain circuits (such as the synchronizer flip-flops, switch debouncers, arbiters, and the like which allow external unsynchronized signals to enter synchronous logic circuits) are inherently asynchronous in their design and must be analyzed as such.

As of 2005, almost all digital machines are synchronous designs because it is much easier to create and verify a synchronous design—the software currently used to simulate digital machines does not yet handle asynchronous designs. However, asynchronous logic is thought to be superior, if it can be made to work, because its speed is not constrained by an arbitrary clock; instead, it runs at the maximum speed of its logic gates. Building an asynchronous circuit using faster parts makes the circuit faster.

Many digital systems are data flow machines. These are usually designed using synchronous register transfer logic, using hardware description languages such as VHDL or Verilog.

In register transfer logic, binary numbers are stored in groups of flip flops called registers. The outputs of each register are a bundle of wires called a "bus" that carries that number to other calculations. A calculation is simply a piece of combinational logic. Each calculation also has an output bus, and these may be connected to the inputs of several registers. Sometimes a register will have a multiplexer on its input, so that it can store a number from any one of several buses. Alternatively, the outputs of several items may be connected to a bus through buffers that can turn off the output of all of the devices except one. A sequential state machine controls when each register accepts new data from its input.

In the 1980s, some researchers discovered that almost all synchronous register-transfer machines could be converted to asynchronous designs by using first-in-first-out synchronization logic. In this scheme, the digital machine is characterized as a set of data flows. In each step of the flow, an asynchronous "synchronization circuit" determines when the outputs of that step are valid, and presents a signal that says, "grab the data" to the stages that use that stage's inputs. It turns out that just a few relatively simple synchronization circuits are needed.

The most general-purpose register-transfer logic machine is a computer. This is basically an automatic binary abacus. The control unit of a computer is usually designed as a microprogram run by a microsequencer. A microprogram is much like a player-piano roll. Each table entry or "word" of the microprogram commands the state of every bit that controls the computer. The sequencer then counts, and the count addresses the memory or combinational logic machine that contains the microprogram. The bits from the microprogram control the arithmetic logic unit, memory and other parts of the computer, including the microsequencer itself.

In this way, the complex task of designing the controls of a computer is reduced to a simpler task of programming a collection of much simpler logic machines.

Computer architecture is a specialized engineering activity that tries to arrange the registers, calculation logic, buses and other parts of the computer in the best way for some purpose. Computer architects have applied large amounts of ingenuity to computer design to reduce the cost and increase the speed and immunity to programming errors of computers. An increasingly common goal is to reduce the power used in a battery-powered computer system, such as a cell-phone. Many computer architects serve an extended apprenticeship as microprogrammers.

"Specialized computers" are usually a conventional computer with a special-purpose microprogram.

Automated design tools

To save costly engineering effort, much of the effort of designing large logic machines has been automated. The computer programs are called "electronic design automation tools" or just "EDA."

Simple truth table-style descriptions of logic are often optimized with EDA that automatically produces reduced systems of logic gates or smaller lookup tables that still produce the desired outputs. The most common example of this kind of software is the Espresso heuristic logic minimizer.

Most practical algorithms for optimizing large logic systems use algebraic manipulations or binary decision diagrams, and there are promising experiments with genetic algorithms and annealing optimizations.

To automate costly engineering processes, some EDA can take state tables that describe state machines and automatically produce a truth table or a function table for the combinational logic of a state machine. The state table is a piece of text that lists each state, together with the conditions controlling the transitions between them and the belonging output signals.

It is common for the function tables of such computer-generated state-machines to be optimized with logic-minimization software such as Minilog.

Often, real logic systems are designed as a series of sub-projects, which are combined using a "tool flow." The tool flow is usually a "script," a simplified computer language that can invoke the software design tools in the right order.

Tool flows for large logic systems such as microprocessors can be thousands of commands long, and combine the work of hundreds of engineers.

Writing and debugging tool flows is an established engineering specialty in companies that produce digital designs. The tool flow usually terminates in a detailed computer file or set of files that describe how to physically construct the logic. Often it consists of instructions to draw the transistors and wires on an integrated circuit or a printed circuit board.

Parts of tool flows are "debugged" by verifying the outputs of simulated logic against expected inputs. The test tools take computer files with sets of inputs and outputs, and highlight discrepancies between the simulated behavior and the expected behavior.

Once the input data is believed correct, the design itself must still be verified for correctness. Some tool flows verify designs by first producing a design, and then scanning the design to produce compatible input data for the tool flow. If the scanned data matches the input data, then the tool flow has probably not introduced errors.

The functional verification data are usually called "test vectors." The functional test vectors may be preserved and used in the factory to test that newly constructed logic works correctly. However, functional test patterns don't discover common fabrication faults. Production tests are often designed by software tools called "test pattern generators". These generate test vectors by examining the structure of the logic and systematically generating tests for particular faults. This way the fault coverage can closely approach 100%, provided the design is properly made testable.

Once a design exists, and is verified and testable, it often needs to be processed to be manufacturable as well. Modern integrated circuits have features smaller than the wavelength of the light used to expose the photoresist. Manufacturability software adds interference patterns to the exposure masks to eliminate open-circuits, and enhance the masks' resolution and contrast.

Design for testability

"There are several reasons for testing a logic circuit. When the circuit is first developed, it is necessary to verify that the design circuit meets the required functional and timing specifications. When multiple copies of a correctly designed circuit are being manufactured, it is essential to test each copy to ensure that the manufacturing process has not introduced any flaws.

A large logic machine (say, with more than a hundred logical variables) can have an astronomical number of possible states. Obviously, in the factory, testing every state is impractical if testing each state takes a microsecond, and there are more states than the number of microseconds since the universe began. Unfortunately, this ridiculous-sounding case is typical.

Fortunately, large logic machines are almost always designed as assemblies of smaller logic machines. To save time, the smaller sub-machines are isolated by permanently-installed "design for test" circuitry, and are tested independently.

One common test scheme known as "scan design" moves test bits serially (one after another) from external test equipment through one or more serial shift registers known as "scan chains". Serial scans have only one or two wires to carry the data, and minimize the physical size and expense of the infrequently-used test logic.

After all the test data bits are in place, the design is reconfigured to be in "normal mode" and one or more clock pulses are applied, to test for faults (e.g. stuck-at low or stuck-at high) and capture the test result into flip-flops and/or latches in the scan shift register(s). Finally, the result of the test is shifted out to the block boundary and compared against the predicted "good machine" result.

In a board-test environment, serial to parallel testing has been formalized with a standard called "JTAG" (named after the "Joint Test Action Group" that proposed it).

Another common testing scheme provides a test mode that forces some part of the logic machine to enter a "test cycle." The test cycle usually exercises large independent parts of the machine.

Trade-offs

Several numbers determine the practicality of a system of digital logic. Engineers explored numerous electronic devices to get an ideal combination of fanout, speed, low cost and reliability.

The cost of a logic gate is crucial. In the 1930s, the earliest digital logic systems were constructed from telephone relays because these were inexpensive and relatively reliable. After that, engineers always used the cheapest available electronic switches that could still fulfill the requirements.

The earliest integrated circuits were a happy accident. They were constructed not to save money, but to save weight, and permit the Apollo Guidance Computer to control an inertial guidance system for a spacecraft. The first integrated circuit logic gates cost nearly $50 (in 1960 dollars, when an engineer earned $10,000/year). To everyone's surprise, by the time the circuits were mass-produced, they had become the least-expensive method of constructing digital logic. Improvements in this technology have driven all subsequent improvements in cost.

With the rise of integrated circuits, reducing the absolute number of chips used represented another way to save costs. The goal of a designer is not just to make the simplest circuit, but to keep the component count down. Sometimes this results in slightly more complicated designs with respect to the underlying digital logic but nevertheless reduces the number of components, board size, and even power consumption.

For example, in some logic families, NAND gates are the simplest digital gate to build. All other logical operations can be implemented by NAND gates. If a circuit already required a single NAND gate, and a single chip normally carried four NAND gates, then the remaining gates could be used to implement other logical operations like logical and. This could eliminate the need for a separate chip containing those different types of gates.

The "reliability" of a logic gate describes its mean time between failure (MTBF). Digital machines often have millions of logic gates. Also, most digital machines are "optimized" to reduce their cost. The result is that often, the failure of a single logic gate will cause a digital machine to stop working.

Digital machines first became useful when the MTBF for a switch got above a few hundred hours. Even so, many of these machines had complex, well-rehearsed repair procedures, and would be nonfunctional for hours because a tube burned-out, or a moth got stuck in a relay. Modern transistorized integrated circuit logic gates have MTBFs greater than 82 billion hours (8.2×10^{10}) hours, and need them because they have so many logic gates.

Fanout describes how many logic inputs can be controlled by a single logic output without exceeding the current ratings of the gate. The minimum practical fanout is about five. Modern electronic logic using CMOS transistors for switches have fanouts near fifty, and can sometimes go much higher.

The "switching speed" describes how many times per second an inverter (an electronic representation of a "logical not" function) can change from true to false and back. Faster logic can accomplish more operations in less time. Digital logic first became useful when switching speeds got above fifty hertz, because that was faster than a team of humans operating mechanical calculators. Modern electronic digital logic routinely switches at five gigahertz (5×10^9 hertz), and some laboratory systems switch at more than a terahertz (1×10^{12} hertz).

Logic families

Design started with relays. Relay logic was relatively inexpensive and reliable, but slow. Occasionally a mechanical failure would occur. Fanouts were typically about ten, limited by the resistance of the coils and arcing on the contacts from high voltages.

Later, vacuum tubes were used. These were very fast, but generated heat, and were unreliable because the filaments would burn out. Fanouts were typically five to seven, limited by the heating from the tubes' current. In the 1950s, special "computer tubes" were developed with filaments that omitted volatile elements like silicon. These ran for hundreds of thousands of hours.

The first semiconductor logic family was resistor-transistor logic. This was a thousand times more reliable than tubes, ran cooler, and used less power, but had a very low fan-in of three. Diode-transistor logic improved the fanout up to about seven, and reduced the power. Some DTL designs used two power-supplies with alternating layers of NPN and PNP transistors to increase the fanout.

Transistor transistor logic (TTL) was a great improvement over these. In early devices, fanout improved to ten, and later variations reliably achieved twenty. TTL was also fast, with some variations achieving switching times as low as twenty nanoseconds. TTL is still used in some designs.

Emitter coupled logic is very fast but uses a lot of power. It was extensively used for high-performance computers made up of many medium-scale components (such as the Illiac IV).

Modern integrated circuits mostly use variations of CMOS, which is fast, small and low-power. Fanouts of forty or more are possible, with some speed penalty.

Non-electronic logic

It is possible to construct non-electronic digital mechanisms. In principle, any technology capable of representing discrete states and representing logic operations could be used to build mechanical logic. MIT students Erlyne Gee, Edward Hardebeck, Danny Hillis (co-author of The Connection Machine), Margaret Minsky and brothers Barry and Brian Silverman, built two working computers from Tinker toys, string, a brick, and a sharpened pencil. The Tinkertoy computer is in the Boston Museum of Science.

Hydraulic, pneumatic and mechanical versions of logic gates exist and are used in situations where electricity cannot be used. The first two types are considered under the heading of fluidics. One application of fluidic logic is in military hardware that is likely to be exposed to a nuclear electromagnetic pulse (nuclear EMP, or NEMP) that would destroy electrical circuits.

Mechanical logic is frequently used in inexpensive controllers, such as those in washing machines. Famously, the first computer design, by Charles Babbage, was designed to use mechanical logic. Mechanical logic might also be used in very small computers that could be built by nanotechnology.

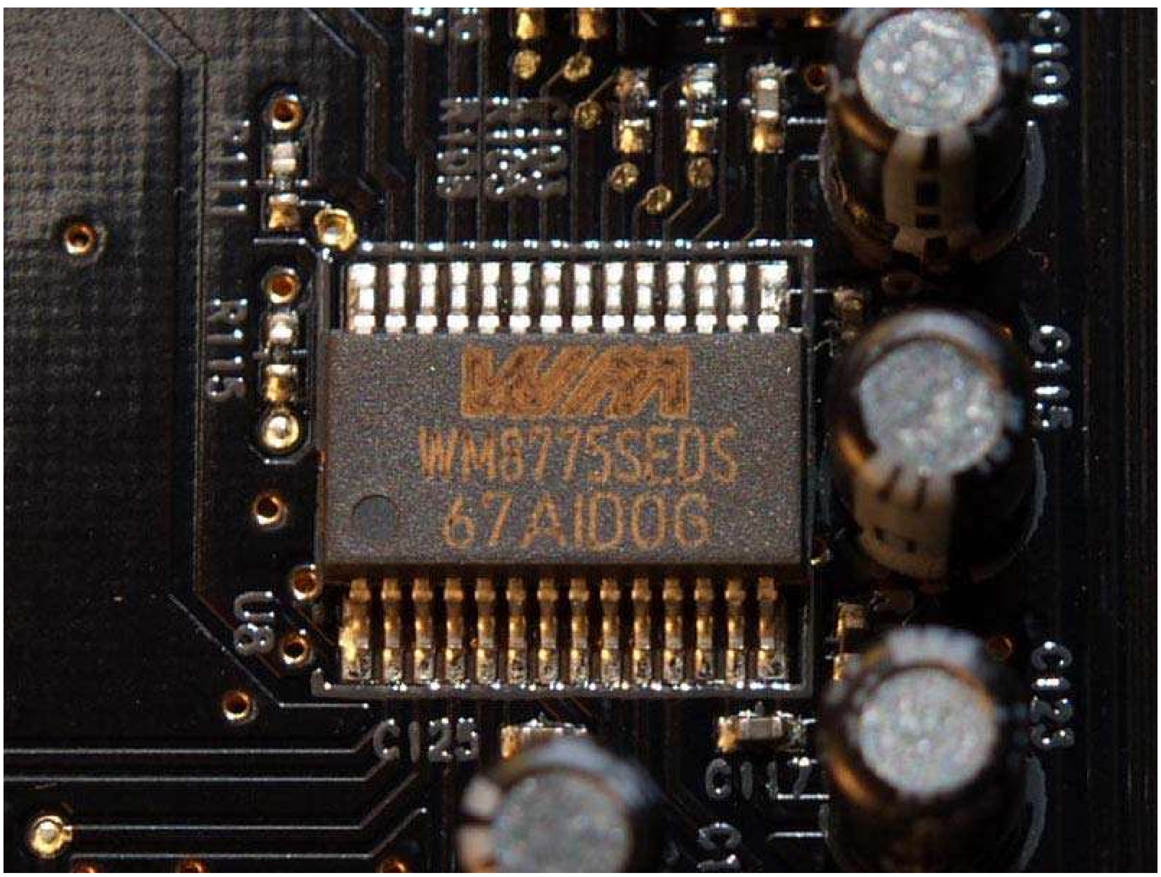
Another example is that if two particular enzymes are required to prevent the construction of a particular protein, this is the equivalent of a biological "NAND" gate.

Recent developments

The discovery of superconductivity has enabled the development of Rapid Single Flux Quantum (RSFQ) circuit technology, which uses Josephson junctions instead of transistors. Most recently, attempts are being made to construct purely optical computing systems capable of processing digital information using nonlinear optical elements.

Chapter- 3

Analog-to-digital Converter



4-channel stereo multiplexed analog-to-digital converter WM8775SEDS made by Wolfson Microelectronics placed on a X-Fi Fatal1ty Pro sound card.

An **analog-to-digital converter** (abbreviated **ADC**, **A/D** or **A to D**) is a device which converts a continuous quantity to a discrete time digital representation. An ADC may also provide an isolated measurement. The reverse operation is performed by a digital-to-analog converter (**DAC**).

Typically, an ADC is an electronic device that converts an input analog voltage or current to a digital number proportional to the magnitude of the voltage or current. However, some non-electronic or only partially electronic devices, such as rotary encoders, can also be considered ADCs.

The digital output may use different coding schemes. Typically the digital output will be a two's complement binary number that is proportional to the input, but there are other possibilities. An encoder, for example, might output a Gray code.

Concepts

Resolution

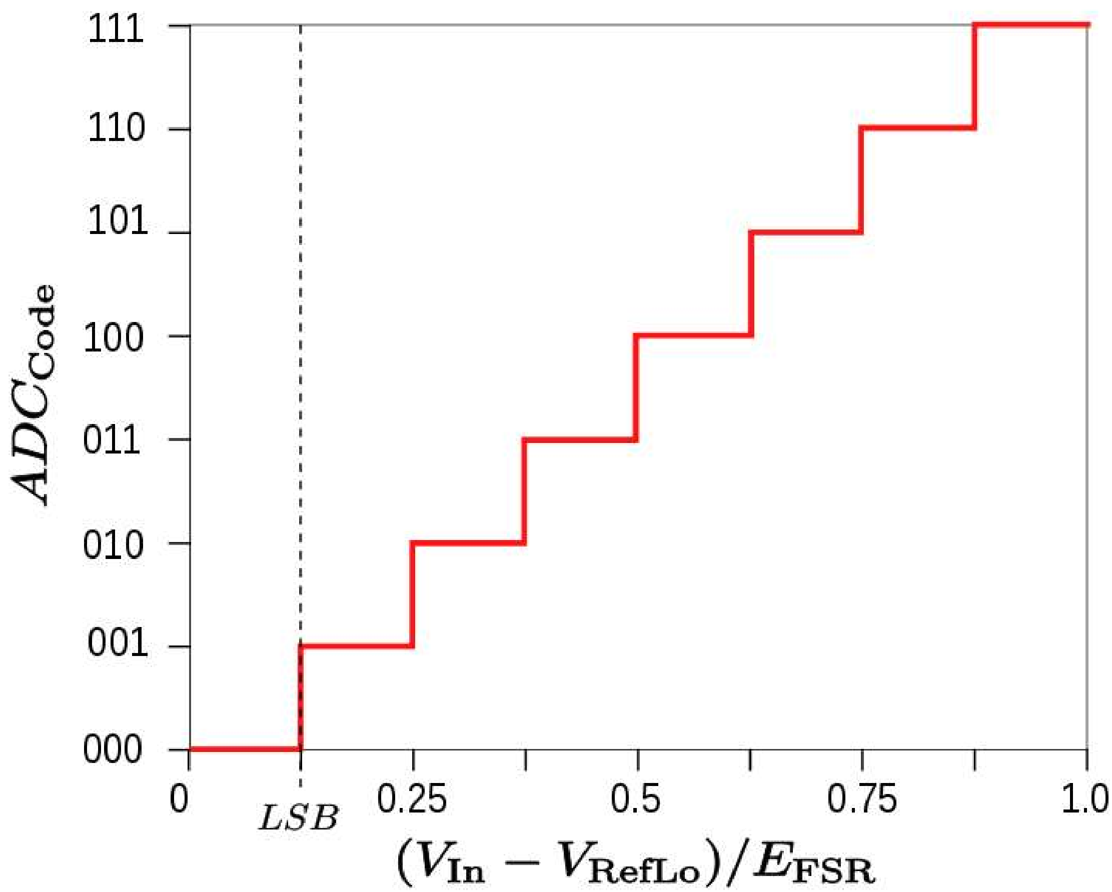


Fig. 1. An 8-level ADC coding scheme.

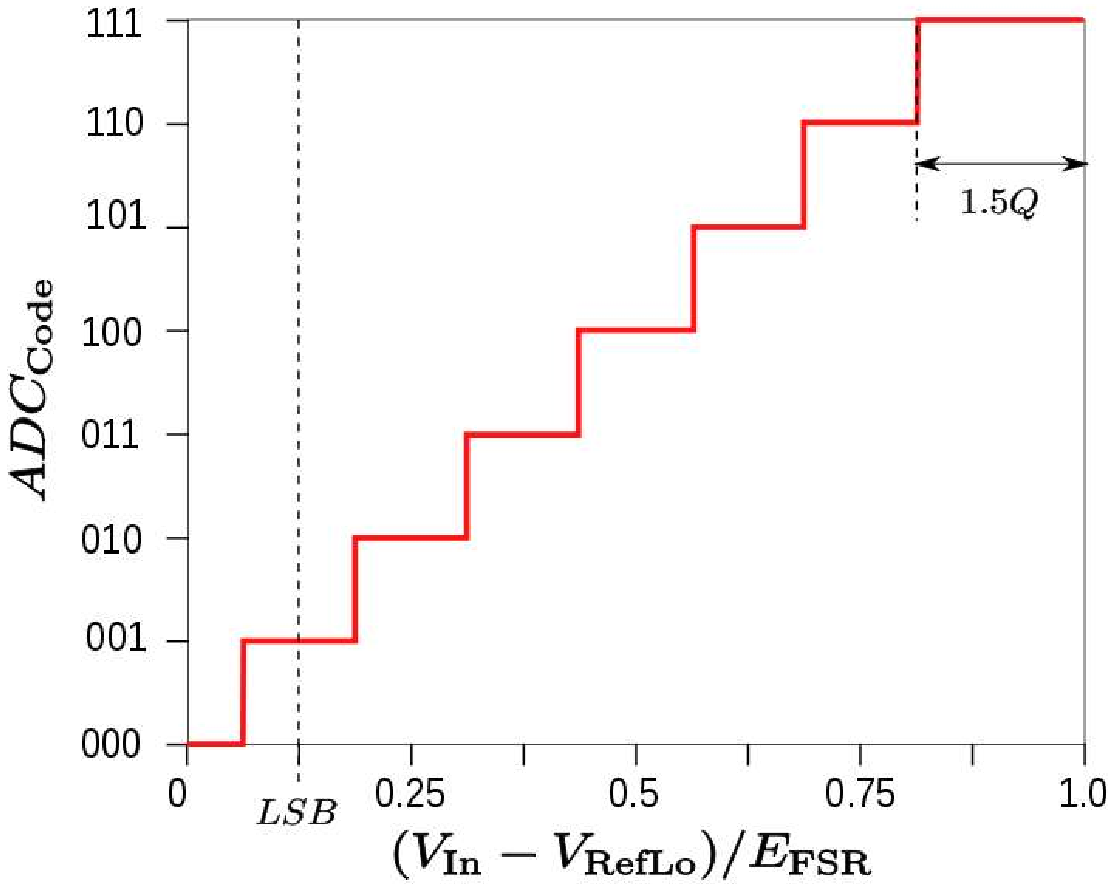


Fig. 2. An 8-level ADC coding scheme. As in figure 1 but with mid-tread coding.

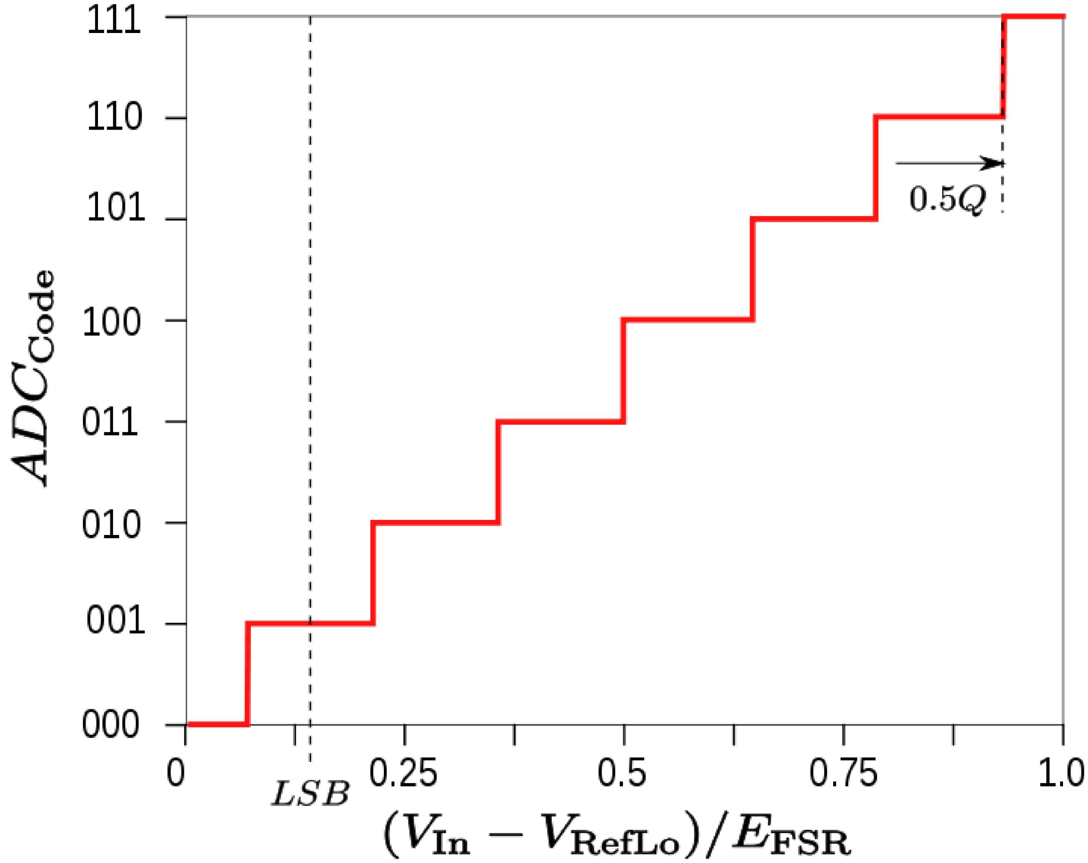


Fig. 3. An 8-level ADC mid-tread coding scheme. As in figure 2 but with equal half-*LSB* intervals at the highest and lowest codes. Note that *LSB* is now slightly larger than in figures 1 and 2.

The resolution of the converter indicates the number of discrete values it can produce over the range of analog values. The values are usually stored electronically in binary form, so the resolution is usually expressed in bits. In consequence, the number of discrete values available, or "levels", is usually a power of two. For example, an ADC with a resolution of 8 bits can encode an analog input to one in 256 different levels, since $2^8 = 256$. The values can represent the ranges from 0 to 255 (i.e. unsigned integer) or from −128 to 127 (i.e. signed integer), depending on the application.

Resolution can also be defined electrically, and expressed in volts. The minimum change in voltage required to guarantee a change in the output code level is called the *LSB* (least significant bit, since this is the voltage represented by a change in the LSB). The resolution *Q* of the ADC is equal to the *LSB* voltage. The voltage resolution of an ADC is equal to its overall voltage measurement range divided by the number of discrete voltage intervals:

$$Q = \frac{E_{\mathrm{FSR}}}{N},$$

where N is the number of voltage intervals and E_{FSR} is the full scale voltage range. E_{FSR} is given by

$$E_{\mathrm{FSR}} = V_{\mathrm{RefHi}} - V_{\mathrm{RefLow}},$$

where V_{RefHi} and V_{RefLow} are the upper and lower extremes, respectively, of the voltages that can be coded.

Normally, the number of voltage intervals is given by

$$N = 2^M,$$

where M is the ADC's resolution in bits.

That is, one voltage interval is assigned per code level. However, figure 3 shows a situation where

$$N = 2^M - 1$$

Some examples:

- Example 1
    - Coding scheme as in figure 1
    - Full scale measurement range = 0 to 10 volts
    - ADC resolution is 12 bits: $2^{12} = 4096$ quantization levels (codes)
    - ADC voltage resolution, $Q = (10\text{ V} - 0\text{ V}) / 4096 = 10\text{ V} / 4096 \approx 0.00244\text{ V} \approx 2.44\text{ mV}$.

- Example 2
    - Coding scheme as in figure 2
    - Full scale measurement range = -10 to +10 volts
    - ADC resolution is 14 bits: $2^{14} = 16384$ quantization levels (codes)
    - ADC voltage resolution is, $Q = (10\text{ V} - (-10\text{ V})) / 16384 = 20\text{ V} / 16384 \approx 0.00122\text{ V} \approx 1.22\text{ mV}$.

- Example 3
    - Coding scheme as in figure 3
    - Full scale measurement range = 0 to 7 volts
    - ADC resolution is 3 bits: $2^3 = 8$ quantization levels (codes)
    - ADC voltage resolution is, $Q = (7\text{ V} - 0\text{ V})/7 = 7\text{ V}/7 = 1\text{ V} = 1000\text{ mV}$

In most ADCs, the smallest output code ("0" in an unsigned system) represents a voltage range which is $0.5Q$, that is, half the ADC voltage resolution (Q). The largest code represents a range of $1.5Q$ as in figure 2 (if this were $0.5Q$ also, the result would be as figure 3). The other $N - 2$ codes are all equal in width and represent the ADC voltage resolution (Q) calculated above. Doing this centers the code on an input voltage that represents the M th division of the input voltage range. This practice is called "mid-tread" operation. This type of ADC can be modeled mathematically as:

$$ADC_{\text{Code}} = \text{round}\left(\left(\frac{2^M}{V_{\text{RefHi}} - V_{\text{RefLow}}}\right) \cdot (V_{\text{In}} - V_{\text{RefLow}})\right)$$

The exception to this convention seems to be the Microchip PIC processor, where all M steps are equal width, as shown in figure 1. This practice is called "Mid-Rise with Offset" operation.

$$ADC_{\text{Code}} = \text{floor}\left(\left(\frac{2^M}{V_{\text{RefHi}} - V_{\text{RefLow}}}\right) \cdot (V_{\text{In}} - V_{\text{RefLow}})\right)$$

In practice, the useful resolution of a converter is limited by the best signal-to-noise ratio (SNR) that can be achieved for a digitized signal. An ADC can resolve a signal to only a certain number of bits of resolution, called the effective number of bits (ENOB). One effective bit of resolution changes the signal-to-noise ratio of the digitized signal by 6 dB, if the resolution is limited by the ADC. If a preamplifier has been used prior to A/D conversion, the noise introduced by the amplifier can be an important contributing factor towards the overall SNR.

Response type

Linear ADCs

Most ADCs are of a type known as linear The term *linear* implies here the range of the input values that map to each output value has a linear relationship with the output value, i.e., that the output value k is used for the range of input values from

$$m(k + b)$$

to

$$m(k + 1 + b),$$

where m and b are constants. Here b is typically 0 or −0.5. When $b = 0$, the ADC is referred to as *mid-rise*, and when $b = -0.5$ it is referred to as *mid-tread*.

Non-linear ADCs

If the probability density function of a signal being digitized is uniform, then the signal-to-noise ratio relative to the quantization noise is the best possible. Because this is often not the case, it is usual to pass the signal through its cumulative distribution function (CDF) before the quantization. This is good because the regions that are more important get quantized with a better resolution. In the dequantization process, the inverse CDF is needed.

This is the same principle behind the companders used in some tape-recorders and other communication systems, and is related to entropy maximization.

For example, a voice signal has a Laplacian distribution. This means that the region around the lowest levels, near 0, carries more information than the regions with higher amplitudes. Because of this, logarithmic ADCs are very common in voice communication systems to increase the dynamic range of the representable values while retaining fine-granular fidelity in the low-amplitude region.

An eight-bit A-law or the μ-law logarithmic ADC covers the wide dynamic range and has a high resolution in the critical low-amplitude region, that would otherwise require a 12-bit linear ADC.

Accuracy

An ADC has several sources of errors. Quantization error and (assuming the ADC is intended to be linear) non-linearity are intrinsic to any analog-to-digital conversion. There is also a so-called *aperture error* which is due to a clock jitter and is revealed when digitizing a time-variant signal (not a constant value).

These errors are measured in a unit called the *LSB*, which is an abbreviation for least significant bit. In the above example of an eight-bit ADC, an error of one LSB is 1/256 of the full signal range, or about 0.4%.

Quantization error

Quantization error (or quantization noise) is the difference between the original signal and the digitized signal. Hence, The magnitude of the quantization error at the sampling instant is between zero and half of one LSB. Quantization error is due to the finite resolution of the digital representation of the signal, and is an unavoidable imperfection in all types of ADCs.

Non-linearity

All ADCs suffer from non-linearity errors caused by their physical imperfections, causing their output to deviate from a linear function (or some other function, in the case of a deliberately non-linear ADC) of their input. These errors can sometimes be mitigated by calibration, or prevented by testing.

Important parameters for linearity are integral non-linearity (INL) and differential non-linearity (DNL). These non-linearities reduce the dynamic range of the signals that can be digitized by the ADC, also reducing the effective resolution of the ADC.

Aperture error

Imagine that we are digitizing a sine wave $x(t) = A\sin(2\pi f_0 t)$. Provided that the actual sampling time *uncertainty* due to the *clock jitter* is Δt, the error caused by this phenomenon can be estimated as $E_{ap} \leq |x'(t)\Delta t| \leq 2A\pi f_0 \Delta t$.

The error is zero for DC, small at low frequencies, but significant when high frequencies have high amplitudes. This effect can be ignored if it is drowned out by the *quantizing error*. Jitter requirements can be calculated using the following formula: $\Delta t < \frac{1}{2^q \pi f_0}$, where q is a number of ADC bits.

ADC resolution in bit	input frequency						
	1 Hz	**44.1 kHz**	**192 kHz**	**1 MHz**	**10 MHz**	**100 MHz**	**1 GHz**
8	1243 µs	28.2 ns	6.48 ns	1.24 ns	124 ps	12.4 ps	1.24 ps
10	311 µs	7.05 ns	1.62 ns	311 ps	31.1 ps	3.11 ps	0.31 ps
12	77.7 µs	1.76 ns	405 ps	77.7 ps	7.77 ps	0.78 ps	0.08 ps
14	19.4 µs	441 ps	101 ps	19.4 ps	1.94 ps	0.19 ps	0.02 ps
16	4.86 µs	110 ps	25.3 ps	4.86 ps	0.49 ps	0.05 ps	–
18	1.21 µs	27.5 ps	6.32 ps	1.21 ps	0.12 ps	–	–
20	304 ns	6.88 ps	1.58 ps	0.16 ps	–	–	–
24	19.0 ns	0.43 ps	0.10 ps	–	–	–	–
32	74.1 ps	–	–	–	–	–	–

This table shows, for example, that it is not worth using a precise 24-bit ADC for sound recording if there is not an *ultra low jitter* clock. One should consider taking this phenomenon into account before choosing an ADC.

Clock jitter is caused by phase noise. The resolution of ADCs with a digitization bandwidth between 1 MHz and 1 GHz is limited by jitter.

When sampling audio signals at 44.1 kHz, the anti-aliasing filter should have eliminated all frequencies above 22 kHz. The input frequency (in this case, 22 kHz), not the ADC clock frequency, is the determining factor with respect to jitter performance.

Sampling rate

The analog signal is continuous in time and it is necessary to convert this to a flow of digital values. It is therefore required to define the rate at which new digital values are sampled from the analog signal. The rate of new values is called the *sampling rate* or *sampling frequency* of the converter.

A continuously varying bandlimited signal can be sampled (that is, the signal values at intervals of time T, the sampling time, are measured and stored) and then the original signal can be *exactly* reproduced from the discrete-time values by an interpolation formula. The accuracy is limited by quantization error. However, this faithful reproduction is only possible if the sampling rate is higher than twice the highest frequency of the signal. This is essentially what is embodied in the Shannon-Nyquist sampling theorem.

Since a practical ADC cannot make an instantaneous conversion, the input value must necessarily be held constant during the time that the converter performs a conversion (called the *conversion time*). An input circuit called a sample and hold performs this task—in most cases by using a capacitor to store the analog voltage at the input, and using an electronic switch or gate to disconnect the capacitor from the input. Many ADC integrated circuits include the sample and hold subsystem internally.

Aliasing

All ADCs work by sampling their input at discrete intervals of time. Their output is therefore an incomplete picture of the behaviour of the input. There is no way of knowing, by looking at the output, what the input was doing between one sampling instant and the next. If the input is known to be changing slowly compared to the sampling rate, then it can be assumed that the value of the signal between two sample instants was somewhere between the two sampled values. If, however, the input signal is changing rapidly compared to the sample rate, then this assumption is not valid.

If the digital values produced by the ADC are, at some later stage in the system, converted back to analog values by a digital to analog converter or DAC, it is desirable that the output of the DAC be a faithful representation of the original signal. If the input signal is changing much faster than the sample rate, then this will not be the case, and spurious signals called *aliases* will be produced at the output of the DAC. The frequency of the aliased signal is the difference between the signal frequency and the sampling rate. For example, a 2 kHz sine wave being sampled at 1.5 kHz would be reconstructed as a 500 Hz sine wave. This problem is called *aliasing*.

To avoid aliasing, the input to an ADC must be low-pass filtered to remove frequencies above half the sampling rate. This filter is called an *anti-aliasing* filter, and is essential for a practical ADC system that is applied to analog signals with higher frequency content.

Although aliasing in most systems is unwanted, it should also be noted that it can be exploited to provide simultaneous down-mixing of a band-limited high frequency signal.

Dither

In A-to-D converters, performance can usually be improved using dither. This is a very small amount of random noise (white noise) which is added to the input before conversion. Its amplitude is set to be twice the value of the least significant bit. Its effect is to cause the state of the LSB to randomly oscillate between 0 and 1 in the presence of very low levels of input, rather than sticking at a fixed value. Rather than the signal simply getting cut off altogether at this low level (which is only being quantized to a resolution of 1 bit), it extends the effective range of signals that the A-to-D converter can convert, at the expense of a slight increase in noise - effectively the quantization error is diffused across a series of noise values which is far less objectionable than a hard cutoff. The result is an accurate representation of the signal over time. A suitable filter at the output of the system can thus recover this small signal variation.

An audio signal of very low level (with respect to the bit depth of the ADC) sampled without dither sounds extremely distorted and unpleasant. Without dither the low level may cause the least significant bit to "stick" at 0 or 1. With dithering, the true level of the audio may be calculated by averaging the actual quantized sample with a series of other samples [the dither] that are recorded over time.

A virtually identical process, also called dither or dithering, is often used when quantizing photographic images to a fewer number of bits per pixel—the image becomes noisier but to the eye looks far more realistic than the quantized image, which otherwise becomes banded. This analogous process may help to visualize the effect of dither on an analogue audio signal that is converted to digital.

Dithering is also used in integrating systems such as electricity meters. Since the values are added together, the dithering produces results that are more exact than the LSB of the analog-to-digital converter.

Note that dither can only increase the resolution of a sampler, it cannot improve the linearity, and thus accuracy does not necessarily improve.

Oversampling

Usually, signals are sampled at the minimum rate required, for economy, with the result that the quantization noise introduced is white noise spread over the whole pass band of the converter. If a signal is sampled at a rate much higher than the Nyquist frequency and then digitally filtered to limit it to the signal bandwidth there are the following advantages:

- digital filters can have better properties (sharper rolloff, phase) than analogue filters, so a sharper anti-aliasing filter can be realised and then the signal can be downsampled giving a better result
- a 20-bit ADC can be made to act as a 24-bit ADC with 256× oversampling
- the signal-to-noise ratio due to quantization noise will be higher than if the whole available band had been used. With this technique, it is possible to obtain an effective resolution larger than that provided by the converter alone
- The improvement in SNR is 3 dB (equivalent to 0.5 bits) per octave of oversampling which is not sufficient for many applications. Therefore,

oversampling is usually coupled with noise shaping. With noise shaping, the improvement is 6L+3 dB per octave where L is the order of loop filter used for noise shaping. e.g. - a 2nd order loop filter will provide an improvement of 15 dB/octave.

Relative speed and precision

The speed of an ADC varies by type. The Wilkinson ADC is limited by the clock rate which is processable by current digital circuits. Currently, frequencies up to 300 MHz are possible. The conversion time is directly proportional to the number of channels. For a successive approximation ADC, the conversion time scales with the logarithm of the number of channels. Thus for a large number of channels, it is possible that the successive approximation ADC is faster than the Wilkinson. However, the time consuming steps in the Wilkinson are digital, while those in the successive approximation are analog. Since analog is inherently slower than digital, as the number of channels increases, the time required also increases. Thus there are competing processes at work. Flash ADCs are certainly the fastest type of the three. The conversion is basically performed in a single parallel step. For an 8-bit unit, conversion takes place in a few tens of nanoseconds.

There is, as expected, somewhat of a trade off between speed and precision. Flash ADCs have drifts and uncertainties associated with the comparator levels, which lead to poor uniformity in channel width. Flash ADCs have a resulting poor linearity. For successive approximation ADCs, poor linearity is also apparent, but less so than for flash ADCs. Here, non-linearity arises from accumulating errors from the subtraction processes. Wilkinson ADCs are the best of the three. These have the best differential non-linearity. The other types require channel smoothing in order to achieve the level of the Wilkinson.

The sliding scale principle

The sliding scale or randomizing method can be employed to greatly improve the channel width uniformity and differential linearity of any type of ADC, but especially flash and successive approximation ADCs. Under normal conditions, a pulse of a particular amplitude is always converted to a certain channel number. The problem lies in that channels are not always of uniform width, and the differential linearity decreases proportionally with the divergence from the average width. The sliding scale principle uses an averaging effect to overcome this phenomenon. A random, but known analog voltage is added to the input pulse. It is then converted to digital form, and the equivalent digital version is subtracted, thus restoring it to its original value. The advantage is that the conversion has taken place at a random point. The statistical distribution of the final channel numbers is decided by a weighted average over a region of the range of the ADC. This in turn desensitizes it to the width of any given channel.

ADC structures

These are the most common ways of implementing an electronic ADC:

- A **direct conversion ADC** or **flash ADC** has a bank of comparators sampling the input signal in parallel, each firing for their decoded voltage range. The comparator bank feeds a logic circuit that generates a code for each voltage range. Direct conversion is very fast, capable of gigahertz sampling rates, but usually has only 8 bits of resolution or fewer, since the number of comparators needed, $2^N - 1$, doubles with each additional bit, requiring a large expensive circuit. ADCs of this type have a large die size, a high input capacitance, high power dissipation, and are prone to produce glitches on the output (by outputting an out-of-sequence code). Scaling to newer submicrometre technologies does not help as the device mismatch is the dominant design limitation. They are often used for video, wideband communications or other fast signals in optical storage.

- A **successive-approximation ADC** uses a comparator to reject ranges of voltages, eventually settling on a final voltage range. Successive approximation works by constantly comparing the input voltage to the output of an internal digital to analog converter (DAC, fed by the current value of the approximation) until the best approximation is achieved. At each step in this process, a binary value of the approximation is stored in a successive approximation register (SAR). The SAR uses a reference voltage (which is the largest signal the ADC is to convert) for comparisons. For example if the input voltage is 60 V and the reference voltage is 100 V, in the 1st clock cycle, 60 V is compared to 50 V (the reference, divided by two. This is the voltage at the output of the internal DAC when the input is a '1' followed by zeros), and the voltage from the comparator is positive (or '1') (because 60 V is greater than 50 V). At this point the first binary digit (MSB) is set to a '1'. In the 2nd clock cycle the input voltage is compared to 75 V (being halfway between 100 and 50 V: This is the output of the internal DAC when its input is '11' followed by zeros) because 60 V is less than 75 V, the comparator output is now negative (or '0'). The second binary digit is therefore set to a '0'. In the 3rd clock cycle, the input voltage is compared with 62.5 V (halfway between 50 V and 75 V: This is the output of the internal DAC when its input is '101' followed by zeros). The output of the comparator is negative or '0' (because 60 V is less than 62.5 V) so the third binary digit is set to a 0. The fourth clock cycle similarly results in the fourth digit being a '1' (60 V is greater than 56.25 V, the DAC output for '1001' followed by zeros). The result of this would be in the binary form 1001. This is also called *bit-weighting conversion*, and is similar to a binary search. The analogue value is rounded to the nearest binary value below, meaning this converter type is mid-rise. Because the approximations are successive (not simultaneous), the conversion takes one clock-cycle for each bit of resolution desired. The clock frequency must be equal to the sampling frequency multiplied by the number of bits of resolution desired. For example, to sample audio at 44.1 kHz with 32 bit resolution, a clock frequency of over 1.4 MHz would be required. ADCs of this type have good resolutions and quite wide ranges. They are more complex than some other designs.

- A **ramp-compare ADC** produces a saw-tooth signal that ramps up or down then quickly returns to zero. When the ramp starts, a timer starts counting. When the ramp voltage matches the input, a comparator fires, and the timer's value is recorded. Timed ramp converters require the least number of transistors. The ramp time is sensitive to temperature because the circuit generating the ramp is often just some simple oscillator. There are two solutions: use a clocked counter driving a DAC and then use the comparator to preserve the counter's value, or calibrate the timed ramp. A special advantage of the ramp-compare system is that comparing a second signal just requires another comparator, and another register to store the voltage value. A very simple (non-linear) ramp-converter can be implemented with a microcontroller and one resistor and capacitor. Vice versa, a filled capacitor can be taken from an integrator, time-to-amplitude converter, phase detector, sample and hold circuit, or peak and hold circuit and discharged. This has the advantage that a slow comparator cannot be disturbed by fast input changes.

- The **Wilkinson ADC** was designed by D. H. Wilkinson in 1950. The Wilkinson ADC is based on the comparison of an input voltage with that produced by a charging capacitor. The capacitor is allowed to charge until its voltage is equal to the amplitude of the input pulse. (A comparator determines when this condition has been reached.) Then, the capacitor is allowed to discharge linearly, which produces a ramp voltage. At the point when the capacitor begins to discharge, a gate pulse is initiated. The gate pulse remains on until the capacitor is completely discharged. Thus the duration of the gate pulse is directly proportional to the amplitude of the input pulse. This gate pulse operates a linear gate which receives pulses from a high-frequency oscillator clock. While the gate is open, a discrete number of clock pulses pass through the linear gate and are counted by the address register. The time the linear gate is open is proportional to the amplitude of the input pulse, thus the number of clock pulses recorded in the address register is proportional also. Alternatively, the charging of the capacitor could be monitored, rather than the discharge.

- An **integrating ADC** (also **dual-slope** or **multi-slope** ADC) applies the unknown input voltage to the input of an integrator and allows the voltage to ramp for a fixed time period (the run-up period). Then a known reference voltage of opposite polarity is applied to the integrator and is allowed to ramp until the integrator output returns to zero (the run-down period). The input voltage is computed as a function of the reference voltage, the constant run-up time period, and the measured run-down time period. The run-down time measurement is usually made in units of the converter's clock, so longer integration times allow for higher resolutions. Likewise, the speed of the converter can be improved by sacrificing resolution. Converters of this type (or variations on the concept) are used in most digital voltmeters for their linearity and flexibility.

- A **delta-encoded ADC** or Counter-ramp has an up-down counter that feeds a digital to analog converter (DAC). The input signal and the DAC both go to a comparator. The comparator controls the counter. The circuit uses negative feedback from the comparator to adjust the counter until the DAC's output is close enough to the input signal. The number is read from the counter. Delta converters have very wide ranges, and high resolution, but the conversion time is dependent on the input signal level, though it will always have a guaranteed worst-case. Delta converters are often very good choices to read real-world signals. Most signals from physical systems do not change abruptly. Some converters combine the delta and successive approximation approaches; this works especially well when high frequencies are known to be small in magnitude.

- A **pipeline ADC** (also called **subranging quantizer**) uses two or more steps of subranging. First, a coarse conversion is done. In a second step, the difference to the input signal is determined with a digital to analog converter (DAC). This difference is then converted finer, and the results are combined in a last step. This can be considered a refinement of the successive approximation ADC wherein the feedback reference signal consists of the interim conversion of a whole range of bits (for example, four bits) rather than just the next-most-significant bit. By combining the merits of the successive approximation and flash ADCs this type is fast, has a high resolution, and only requires a small die size.

- A **Sigma-Delta ADC** (also known as a Delta-Sigma ADC) oversamples the desired signal by a large factor and filters the desired signal band. Generally, a smaller number of bits than required are converted using a Flash ADC after the filter. The resulting signal, along with the error generated by the discrete levels of the Flash, is fed back and subtracted from the input to the filter. This negative feedback has the effect of noise shaping the error due to the Flash so that it does not appear in the desired signal frequencies. A digital filter (decimation filter) follows the ADC which reduces the sampling rate, filters off unwanted noise signal and increases the resolution of the output (sigma-delta modulation, also called delta-sigma modulation).

- A **Time-interleaved ADC** uses M parallel ADCs where each ADC sample data every M:th cycle of the effective sample clock. The result is that the sample rate is increased M times compared to what each individual ADC can manage. In practice, the individual differences between the M ADCs degrade the overall performance reducing the SFDR. However, technologies exist to correct for these time-interleaving mismatch errors.

- An **ADC with intermediate FM stage** first uses a voltage-to-frequency converter to converts the desired signal into an oscillating signal with a frequency proportional to the voltage of the desired signal, and then uses a frequency counter to convert that frequency into a digital count proportional to the desired signal voltage. Longer integration times allow for higher resolutions. Likewise, the speed of the converter can be improved by sacrificing resolution. The two parts of the ADC may be widely separated, with the frequency signal passed through a opto-isolator or transmitted wirelessly. Some such ADCs use sine wave or square wave frequency modulation; others use pulse-frequency modulation.

Such ADCs were once the most popular way to show a digital display of the status of a remote analog sensor.

There can be other ADCs that use a combination of electronics and other technologies:

- A **Time-stretch analog-to-digital converter (TS-ADC)** digitizes a very wide bandwidth analog signal, that cannot be digitized by a conventional electronic ADC, by time-stretching the signal prior to digitization. It commonly uses a photonic preprocessor frontend to time-stretch the signal, which effectively slows the signal down in time and compresses its bandwidth. As a result, an electronic backend ADC, that would have been too slow to capture the original signal, can now capture this slowed down signal. For continuous capture of the signal, the frontend also divides the signal into multiple segments in addition to time-stretching. Each segment is individually digitized by a separate electronic ADC. Finally, a digital signal processor rearranges the samples and removes any distortions added by the frontend to yield the binary data that is the digital representation of the original analog signal.

Commercial analog-to-digital converters

These are usually integrated circuits.

Most converters sample with 6 to 24 bits of resolution, and produce fewer than 1 megasample per second. Thermal noise generated by passive components such as resistors masks the measurement when higher resolution is desired. For audio applications and in room temperatures, such noise is usually a little less than 1 μV (microvolt) of white noise. If the Most Significant Bit corresponds to a standard 2 volts of output signal, this translates to a noise-limited performance that is less than 20~21 bits, and obviates the need for any dithering. Mega- and gigasample per second converters are available, though (Feb 2002). Megasample converters are required in digital video cameras, video capture cards, and TV tuner cards to convert full-speed analog video to digital video files. Commercial converters usually have ±0.5 to ±1.5 LSB error in their output.

In many cases the most expensive part of an integrated circuit is the pins, because they make the package larger, and each pin has to be connected to the integrated circuit's silicon. To save pins, it is common for slow ADCs to send their data one bit at a time over a serial interface to the computer, with the next bit coming out when a clock signal changes state, say from zero to 5 V. This saves quite a few pins on the ADC package, and in many cases, does not make the overall design any more complex (even microprocessors which use memory-mapped I/O only need a few bits of a port to implement a serial bus to an ADC).

Commercial ADCs often have several inputs that feed the same converter, usually through an analog multiplexer. Different models of ADC may include sample and hold circuits, instrumentation amplifiers or differential inputs, where the quantity measured is the difference between two voltages.

Applications

Application to music recording

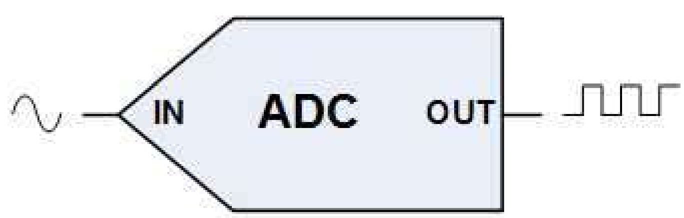
ADCs are integral to current music reproduction technology. Since much music production is done on computers, when an analog recording is used, an ADC is needed to create the PCM data stream that goes onto a compact disc or digital music file.

The current crop of AD converters utilized in music can sample at rates up to 192 kilohertz. High bandwidth headroom allows the use of cheaper or faster anti-aliasing filters of less severe filtering slopes. The proponents of oversampling assert that such shallower anti-aliasing filters produce less deleterious effects on sound quality, exactly because of their gentler slopes. Others prefer entirely filterless AD conversion, arguing that aliasing is less detrimental to sound perception than pre-conversion brickwall filtering. Considerable literature exists on these matters, but commercial considerations often play a significant role. Most high-profile recording studios record in 24-bit/192-176.4 kHz PCM or in DSD formats, and then downsample or decimate the signal for Red-Book CD production (44.1 kHz or at 48 kHz for commonly used for radio/TV broadcast applications).

Digital Signal Processing

AD converters are used virtually everywhere where an analog signal has to be processed, stored, or transported in digital form. Fast video ADCs are used, for example, in TV tuner cards. Slow on-chip 8, 10, 12, or 16 bit ADCs are common in microcontrollers. Very fast ADCs are needed in digital oscilloscopes, and are crucial for new applications like software defined radio.

Electrical Symbol



ELECTRICAL SYMBOL FOR ANALOG TO DIGITAL CONVERTER (ADC)

Chapter- 4

Comparator

In electronics, a **comparator** is a device that compares two voltages or currents and switches its output to indicate which is larger. It is used in Analog-to-digital converter (ADCs).

Input voltage range

The input voltages must stay within the limits specified by the manufacturer. Early integrated comparators, like the LM111 family, and certain high-speed comparators like the LM119 family, require input voltage ranges substantially lower than the power supply voltages (±15 V vs. 36V). *Rail-to-rail* comparators allow any input voltages within the power supply range. When powered from a bipolar (dual rail) supply,

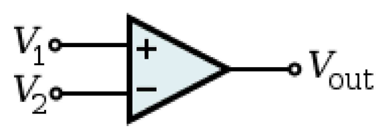
$$V_{S-} \leq V_+, V_- \leq V_{S+}$$

or, when powered from a unipolar TTL/CMOS power supply:

$$0 \leq V_+, V_- \leq V_{cc}$$

Specific rail-to-rail comparators with p-n-p input transistors, like the LM139 family, allow input potential to drop 0.3 Volts *below* the negative supply rail, but do not allow it to rise above the positive rail. Specific ultra-fast comparators, like the LMH7322, allow input signal to swing below the negative rail *and* above the positive rail, although by a narrow margin of only 0.2V. Differential input voltage (the voltage between two inputs) of a modern rail-to-rail comparator is usually limited only by the full swing of power supply.

Op-amp voltage comparator



A simple op-amp comparator

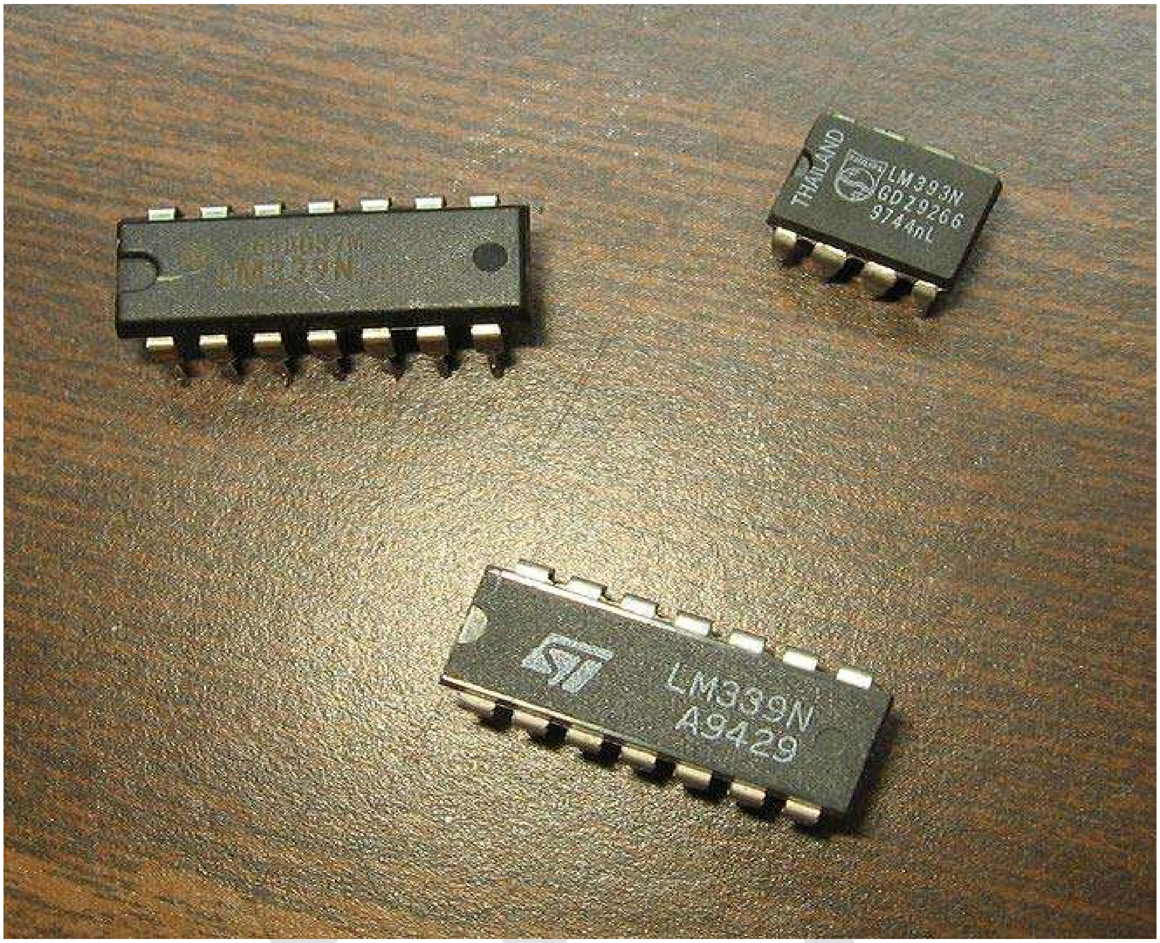
An operational amplifier (op-amp) has a well balanced difference input and a very high gain. This parallels the characteristics of comparators and can be substituted in applications with low-performance requirements.

In theory, a standard op-amp operating in open-loop configuration (without negative feedback) may be used as a low-performance comparator. When the non-inverting input (V+) is at a higher voltage than the inverting input (V-), the high gain of the op-amp causes the output to saturate at the highest positive voltage it can output. When the non-inverting input (V+) drops below the inverting input (V-), the output saturates at the most negative voltage it can output. The op-amp's output voltage is limited by the supply voltage. An op-amp operating in a linear mode with negative feedback, using a balanced, split-voltage power supply, (powered by $\pm V_S$) its transfer function is typically written as: $V_{out} = A_o(V_1 - V_2)$. However, this equation may not be applicable to a comparator circuit which is non-linear and operates open-loop (no negative feedback).

In practice, using an operational amplifier as a comparator presents several disadvantages as compared to using a dedicated comparator:

1. Op-amps are designed to operate in the linear mode with negative feedback. Hence, an op-amp typically has a lengthy recovery time from saturation. Almost all op-amps have an internal compensation capacitor which imposes slew rate limitations for high frequency signals. Consequently an op-amp makes a sloppy comparator with propagation delays that can be as slow as tens of microseconds.
2. Since op-amps do not have any internal hysteresis an external hysteresis network is always necessary for slow moving input signals.
3. The quiescent current specification of an op-amp is valid only when the feedback is active. Some op-amps show an increased quiescent current when the inputs are not equal.
4. A comparator is designed to produce well limited output voltages that easily interface with digital logic. Compatibility with digital logic must be verified while using an op-amp as a comparator.
5. Some multiple-section opamps may exhibit extreme channel-channel interaction when used as comparators.
6. Many opamps have back to back diodes between their inputs. Opamp inputs usually follow each other so this is fine. But comparator inputs are not usually the same. The diodes can cause unexpected current through inputs.

Dedicated voltage comparator chips



Several voltage comparator ICs

A dedicated voltage comparator will generally be faster than a general-purpose operational amplifier pressed into service as a comparator. A dedicated voltage comparator may also contain additional features such as an accurate, internal voltage reference, an adjustable hysteresis and a clock gated input.

A dedicated voltage comparator chip such as LM339 is designed to interface with a digital logic interface (to a TTL or a CMOS). The output is a binary state often used to interface real world signals to digital circuitry. If there is a fixed voltage source from, for example, a DC adjustable device in the signal path, a comparator is just the equivalent of a cascade of amplifiers. When the voltages are nearly equal, the output voltage will not fall into one of the logic levels, thus analog signals will enter the digital domain with unpredictable results. To make this range as small as possible, the amplifier cascade is high gain. The circuit consists of mainly Bipolar transistors except perhaps in the beginning stage which will likely be field effect transistors. For very high frequencies, the input impedance of the stages is low. This reduces the saturation of the slow, large P-N junction bipolar transistors that would otherwise lead to long recovery times. Fast small Schottky diodes, like those found in binary logic designs, improve the performance significantly though the performance still lags that of circuits with amplifiers using analog signals. Slew rate has no meaning for these devices. For applications in flash ADCs the distributed signal across 8 ports matches the voltage and current gain after each amplifier, and resistors then behave as level-shifters.

The LM339 accomplishes this with an open collector output. When the inverting input is at a higher voltage than the non inverting input, the output of the comparator connects to the negative power supply. When the non inverting input is higher than the inverting input, the output is 'floating' (has a very high impedance to ground).

Inputs	Output
$- > +$	Negative
$+ > -$	Floating

With a pull-up resistor and a 0 to +5V power supply, the output takes on the voltages 0 or +5 and can interface with TTL logic:

$$V_{\text{out}} \leq V_{\text{CC}} \text{ when } V_{+} \geq V_{-} \text{ else } 0.$$

Key specifications

While it is easy to understand the basic task of a comparator, that is, comparing two voltages or currents, several parameters must be considered while selecting a suitable comparator:

Speed and power

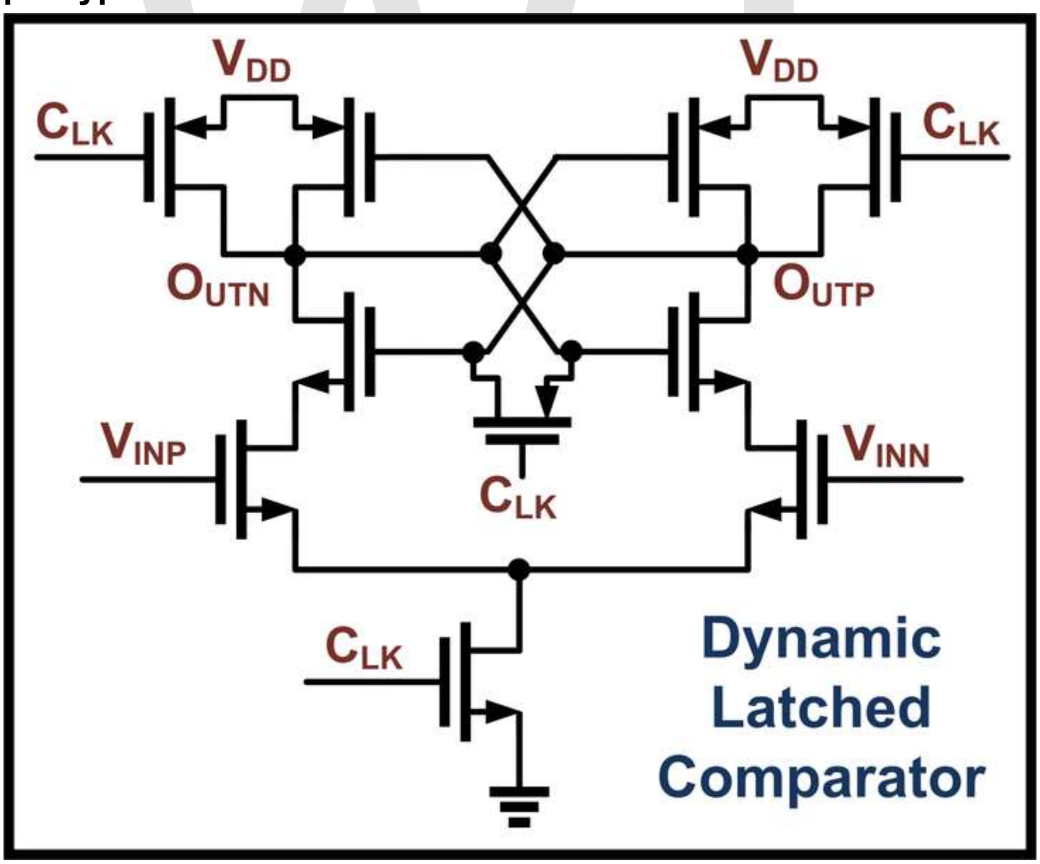
While in general comparators are “fast”, their circuits are not immune to the classic speed-power tradeoff. High speed comparators use transistors with larger aspect ratios and hence also consume more power. Depending on the application, select either a comparator with high speed or one that saves power. For example, nano-powered comparators in space-saving chip-scale packages (UCSP), DFN or SC70 packages such as MAX9027, LTC1540, LPV7215, MAX9060 and MCP6541 are ideal for ultra-low-power, portable applications. Likewise if a comparator is needed to implement a relaxation oscillator circuit to create a high speed clock signal then comparators having few nano seconds of propagation delay may be suitable. ADCMP572 (CML output), LMH7220 (LVDS Output), MAX999 (CMOS output / TTL output), LT1719 (CMOS output / TTL output), MAX9010 (TTL output), and MAX9601 (PECL output) are examples of some good high speed comparators.

Hysteresis

A comparator normally changes its output state when the voltage between its inputs crosses through approximately zero volts. Small voltage fluctuations due to noise, always present on the inputs, can cause undesirable rapid changes between the two output states when the input voltage difference is near zero volts. To prevent this output oscillation, a small hysteresis of a few millivolts is integrated into many modern comparators. For example, the LTC6702, MAX9021 and MAX9031 have internal hysteresis desensitizing them from input noise. In place of one switching point, hysteresis introduces two: one for rising voltages, and one for falling voltages. The difference between the higher-level trip value (VTRIP+) and the lower-level trip value (VTRIP-) equals the hysteresis voltage (VHYST).

If the comparator does not have internal hysteresis or if the input noise is greater than the internal hysteresis then an external hysteresis network can be built using positive feedback from the output to the non-inverting input of the comparator. The resulting Schmitt trigger circuit gives additional noise immunity and a cleaner output signal. Some comparators such as LMP7300, LTC1540, MAX931, MAX971 and ADCMP341 also provide the hysteresis control through a separate hysteresis pin. These comparators make it possible to add a programmable hysteresis without feedback or complicated equations. Using a dedicated hysteresis pin is also convenient if the source impedance is high since the inputs are isolated from the hysteresis network. When hysteresis is added then a comparator cannot resolve signals within the hysteresis band.

Output type



A Low Power CMOS Clocked Comparator

Because comparators have only two output states, their outputs are near zero or near the supply voltage. Bipolar rail-to-rail comparators have a common-emitter output that produces a small voltage drop between the output and each rail. That drop is equal to the collector-to-emitter voltage of a saturated transistor. When output currents are light, output voltages of CMOS rail-to-rail comparators, which rely on a saturated MOSFET, range closer to the rails than their bipolar counterparts.

On the basis of outputs, comparators can also be classified as open drain or push–pull. Comparators with an open-drain output stage use an external pull up resistor to a positive supply that defines the logic high level. Open drain comparators are more suitable for mixed-voltage system design. Since the output is high impedance for logic level high, open drain comparators can also be used to connect multiple comparators on to a single bus. Push pull output does not need a pull up resistor and can also source current unlike an open drain output.

Internal reference

The most frequent application for comparators is the comparison between a voltage and a stable reference. Most comparator manufacturers also offer comparators in which a reference voltage is integrated on to the chip. Combining the reference and comparator in one chip not only saves space, but also draws less supply current than a comparator with an external reference. ICs with wide range of references are available such as MAX9062(200 mV reference), LT6700(400 mV reference), ADCMP350(600mV reference), MAX9025(1.236V reference), MAX9040(2.048V reference), TLV3012(1.24V reference) and TSM109(2.5V reference).

Continuous versus clocked

A continuous comparator will output either a "1" or a "0" any time a high or low signal is applied to its input and will change quickly when the inputs are updated. However, many applications only require comparator outputs at certain instances, such as in A/D converters and memory. By only strobing a comparator at certain intervals, higher accuracy and lower power can be achieved with a clocked (or dynamic) comparator structure, also called a latched comparator. Often latched comparators employ strong positive feedback for a "regeneration phase" when a clock is high, and have a "reset phase" when the clock is low. This is in contrast to a continuous comparator, which can only employ weak positive feedback since there is no reset period.

Applications

Null detectors

A null detector is one that functions to identify when a given value is zero. Comparators can be a type of amplifier distinctively for null comparison measurements. It is the equivalent to a very high gain amplifier with well-balanced inputs and controlled output limits. The circuit compares the two input voltages, determining the larger. The inputs are an unknown voltage and a reference voltage, usually referred to as v_u and v_r. A reference voltage is generally on the non-inverting input (+), while v_u is usually on the inverting input (-). (A circuit diagram would display the inputs according to their sign with respect to the output when a particular input is greater than the other.) The output is either positive or negative, for example +/-12V. In this case, the idea is to detect when there is no difference between in the input voltages. This gives the identity of the unknown voltage since the reference voltage is known.

When using a comparator as a null detector, there are limits as to the accuracy of the zero value measurable. Zero output is given when the magnitude of the difference in the voltages multiplied by the gain of the amplifier is less than the voltage limits. For example, if the gain of the amplifier is 10^6, and the voltage limits are +/-6V, then no output will be given if the difference in the voltages is less than 6μV. One could refer to this as a sort of uncertainty in the measurement.

Zero-crossing detectors

For this type of detector, a comparator detects each time an ac pulse changes polarity. The output of the comparator changes state each time the pulse changes its polarity, that is, the output is HI (high) for a positive pulse and LO (low) for a negative pulse. The comparator also amplifies and squares the input signal.

Relaxation oscillator

A comparator can be used to build a relaxation oscillator. It uses both positive and negative feedback. The positive feedback is a Schmitt trigger configuration. Alone, the trigger is a bistable multivibrator. However, the slow negative feedback added to the trigger by the RC circuit causes the circuit to oscillate automatically. That is, the addition of the RC circuit turns the hysteretic bistable multivibrator into an astable multivibrator.

Level shifter

This circuit requires only a single comparator with an open-drain output as in the LM393, TLV3011 or MAX9028. The circuit provides great flexibility in choosing the voltages to be translated by using a suitable pull up voltage. It also allows the translation of bipolar ±5V logic to unipolar 3V logic by using a comparator like the MAX972.

Analog-to-digital converters

When a comparator performs the function of telling if an input voltage is above or below a given threshold, it is essentially performing a 1-bit quantization. This function is used in nearly all analog to digital converters (such as flash, pipeline, successive approximation, delta-sigma modulation, folding, interpolating, dual-slope and others) in combination with other devices to achieve a multi-bit quantization.

Chapter- 5

Crossbar Switch

In electronics, a **crossbar switch** (also known as **cross-point switch**, **crosspoint switch**, or **matrix switch**) is a switch connecting multiple inputs to multiple outputs in a matrix manner. Originally the term was used literally, for a matrix switch controlled by a grid of crossing metal bars, and later was broadened to matrix switches in general. It is one of the principal switch architectures, together with a memory switch and a crossover switch.

General properties

A crossbar switch is an assembly of individual switches between multiple inputs and multiple outputs. The switches are arranged in a matrix. If the crossbar switch has M inputs and N outputs, then a crossbar has a matrix with M x N cross-points or places where the "bars" cross. At each crosspoint is a switch; when closed, it connects one of M inputs to one of N outputs. A given crossbar is a single layer, non-blocking switch. Collections of crossbars can be used to implement multiple layer and/or blocking switches. A crossbar switching system is also called a co-ordinate switching system.

Applications

Crossbar switches are most famously used in information processing applications such as telephony and packet switching, but they are also used in applications such as mechanical sorting machines with inputs.

The matrix layout of a crossbar switch is also used in some semiconductor memory devices. Here the "bars" are extremely thin metal "wires", and the "switches" are fusible links. The fuses are blown or opened using high voltage and read using low voltage. Such devices are called programmable read-only memory. At the 2008 NSTI Nanotechnology Conference a paper was presented which discussed a nanoscale crossbar implementation of an adding circuit used as an alternative to logic gates for computation.

Furthermore, matrix arrays are fundamental to modern flat-panel displays. Thin-film-transistor LCDs have a transistor at each crosspoint, so they could be considered to include a crossbar switch as part of their structure.

For video switching in home and professional theater applications, a crossbar switch (or a matrix switch, as it is more commonly called in this application) is used to make the output of multiple video appliances available simultaneously to every monitor or every room throughout a building. In a typical installation, all the video sources are located on an equipment rack, and are connected as inputs to the matrix switch.

Where central control of the matrix is practical, a typical rack-mount matrix switch offers front-panel buttons to allow manual connection of inputs to outputs. An example of such a usage might be a sports bar, where numerous programs are displayed simultaneously. In order to accomplish this, a sports bar would ordinarily need to purchase a separate cable or satellite subscription for each display for which independent control is desired. The matrix switch enables the signals to be re-routed on a whim, thus allowing the establishment to purchase only those subscriptions needed to cover the total number of *unique* programs viewed anywhere in the building.

Such switches are used in high-end home theater applications. Video sources typically shared include set-top cable/satellite receivers or DVD changers; the same concept applies to audio as well. The outputs are wired to televisions in individual rooms. The matrix switch is controlled via an Ethernet or RS-232 serial connection by a whole-house automation controller, such as those made by AMX, Crestron, or Control4 - which provides the user interface that enables the user in each room to select which appliance to watch. The actual user interface varies by system brand, and might include a combination of on-screen menus, touch-screens, and handheld remote controls. The system is necessary to enable the user to select the program they wish to watch from the same room they will watch it from, otherwise it would be necessary (and arguably absurd) for them to walk to the equipment rack.

The special crossbar switches used in distributing satellite TV signals are called Multiswitches.

Implementations

Historically, a crossbar switch consisted of metal bars associated with each input and output, together with some means of controlling movable contacts at each cross-point. In the later part of the 20th Century these literal crossbar switches declined and the term came to be used figuratively for rectangular array switches in general. Modern "crossbar switches" are usually implemented with semiconductor technology. An important emerging class of optical crossbars is being implemented with MEMS technology.

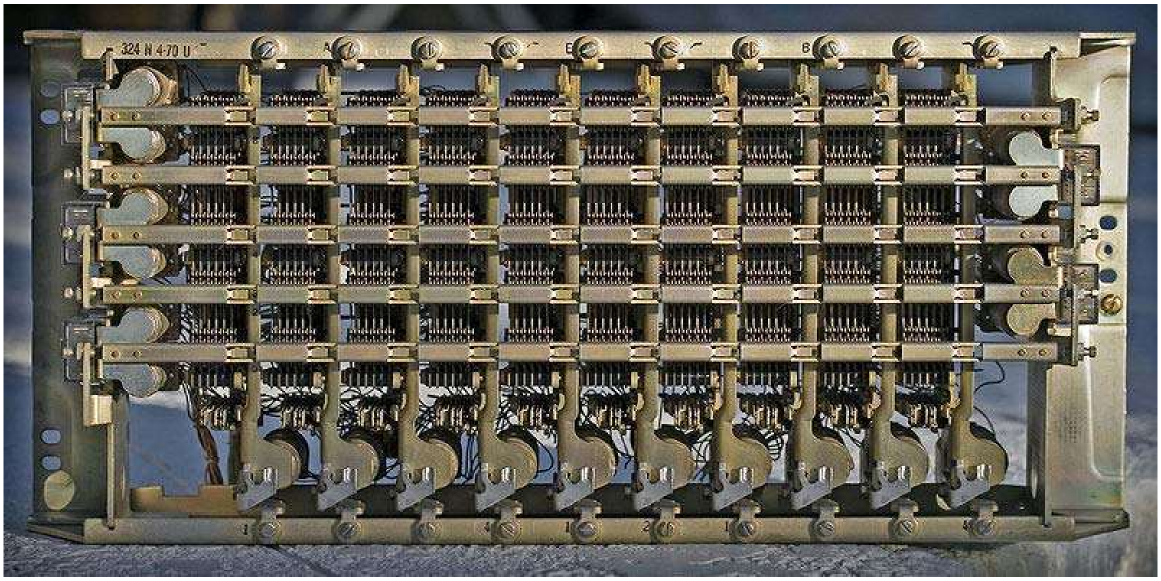
Mechanical

A type of middle 19th Century telegraph exchange consisted of a grid of vertical and horizontal brass bars with a hole at each intersection. The operator inserted a brass pin to connect one telegraph line to another.

Electromechanical/telephony

A telephony crossbar switch is an electromechanical device for switching telephone calls. The first design of what is now called a crossbar switch was Western Electric's "coordinate selector" of 1915. It was little used in America, but the LM Ericsson company used an improved version for rural exchanges in Sweden. To save money on control systems, this system was organized on the stepping switch or selector principle rather than the link principle. The system design used in AT&T's 1XB crossbar exchanges, which entered revenue service from 1938, was developed by Bell Telephone Labs, based on the rediscovered link principle. Delayed by the Second World War, several millions of urban 1XB lines were installed from the 1950s in the United States. Crossbar switching quickly spread to the rest of the world, replacing most earlier designs like the Strowger and Panel systems in larger installations in the U.S. Graduating from entirely electromechanical control on introduction, they were gradually elaborated to have full electronic control and a variety of calling features including short-code and speed-dialing. In the UK the Plessey Company produced a range of crossbar exchanges, but their widespread rollout by the British Post Office began later than in other countries, and then was inhibited by the parallel development of TXE reed relay and electronic exchange systems, so they never achieved a large number of customer connections although they did find some success as tandem switch exchanges.

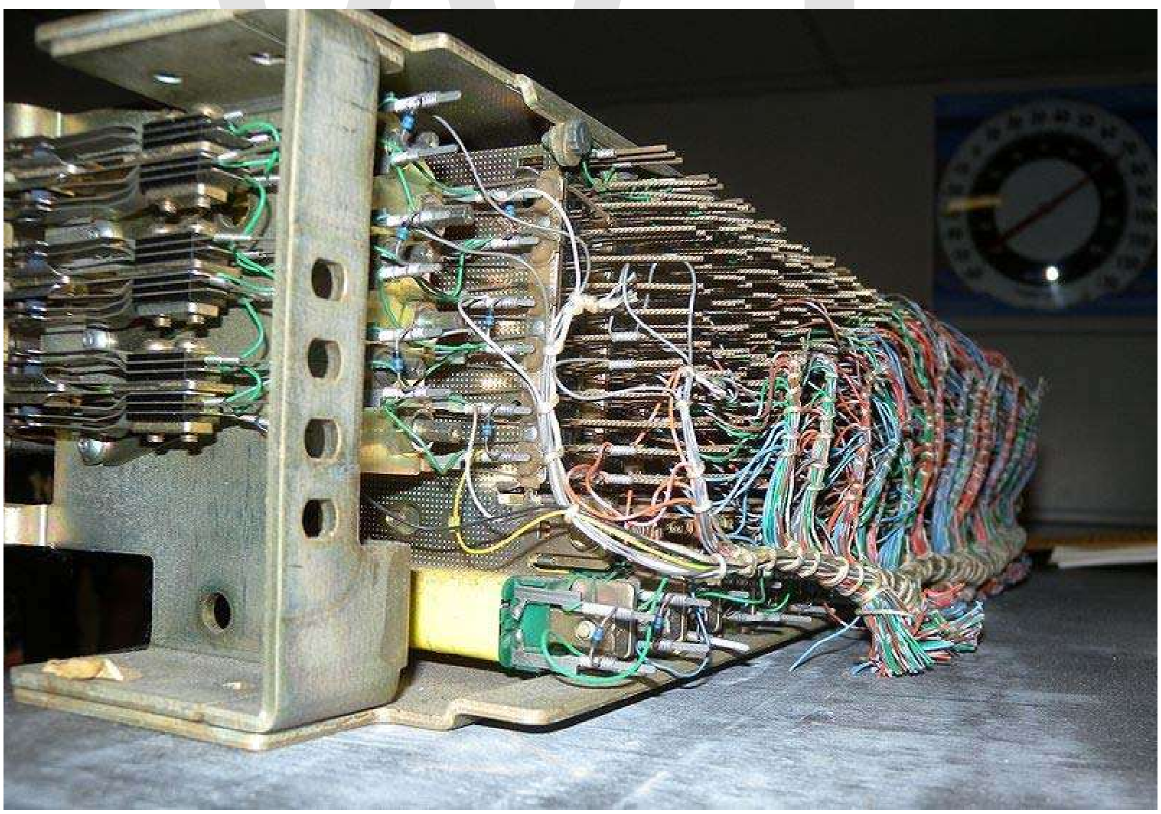
Crossbar switches use switching matrices made from a two-dimensional array of contacts arranged in an x-y format. These switching matrices are operated by a series of horizontal bars arranged over the contacts. Each such "select" bar can be rocked up or down by electromagnets to provide access to two levels of the matrix. A second set of vertical "hold" bars is set at right angles to the first (hence the name, "crossbar") and also operated by electromagnets. The select bars carry spring-loaded wire fingers that enable the hold bars to operate the contacts beneath the bars. When the select and then the hold electromagnets operate in sequence to move the bars, they trap one of the spring fingers to close the contacts beneath the point where two bars cross. This then makes the connection through the switch as part of setting up a calling path through the exchange. Once connected, the select magnet is then released so it can use its other fingers for other connections, while the hold magnet remains energized for the duration of the call to maintain the connection. The crossbar switching interface was referred to as the TXK or TXC switch (*T*elephone e*X*change *C*rossbar) - in the UK.



Western Electric 100 Point six-wire Type B crossbar switch

The Bell System *Type B* crossbar switch of the 1960s was made in the largest quantity. The majority were 200 point switches, with twenty verticals and ten levels of three wires, but this example is a 100 point six wire switch. Each select bar carries ten fingers so any of the ten circuits assigned to the ten verticals can connect to either of two levels. Five select bars, each able to rotate up or down, mean a choice of ten links to the next stage of switching. Each crosspoint in this particular model connected six wires. Note the *Vertical Off-Normal* contacts next to the hold magnets, lined up along the bottom of the switch. They perform logic and memory functions, and the hold bar keeps them in the active position as long as the connection is up. The *Horizontal Off Normals* on the sides of the switch are activated by the horizontal bars when the "butterfly magnets" rotate them. This only happens while the connection is being set up, since the butterflies are only energized then.

Late-model Western Electric crossbar switch



Back of Type C

The majority of Bell System switches were made to connect three wires including the tip and ring of a balanced pair circuit and a sleeve lead for control. Many connected six wires, either for two distinct circuits or for a four wire circuit or other complex connection. The Bell System *Type C* miniature crossbar of the 1970s was similar, but the fingers projected forward from the back and the select bars held paddles to move them. The majority of type C had twelve levels; these were the less common ten level ones. The Northern Electric *Minibar* used in SP1 switch was similar but even smaller. The ITT Pentaconta Multiswitch of the same era had usually 22 verticals, 26 levels, and six to twelve wires. Ericsson crossbar switches sometimes had only five verticals.

Electromechanical/instrumentation

For instrumentation use, James Cunningham, Son and Company made high-speed, very-long-life crossbar switches with physically-small mechanical parts which permitted faster operation than telephone-type crossbar switches. Many of their switches had the mechanical Boolean AND function of telephony crossbar switches, but other models had individual relays (one coil per crosspoint) in matrix arrays, connecting the relay contacts to [x] and [y] buses. These latter types were equivalent to separate relays; there was no logical AND function built in. Cunningham crossbar switches had precious-metal contacts capable of handling millivolt signals.

Telephone exchange

Early crossbar exchanges were divided into an originating side and a terminating side, while the later and prominent Canadian and US SP1 switch and 5XB switch were not. When a user picked up the telephone handset, the resulting line loop operating the user's line relay caused the exchange to connect the user's telephone to an originating sender, which returned the user a dial tone. The sender then recorded the dialed digits and passed them to the originating marker, which selected an outgoing trunk and operated the various crossbar switch stages to connect the calling user to it. The originating marker then passed the trunk call completion requirements (type of pulsing, resistance of the trunk, etc) and the called party's details to the sender and released. The sender then relayed this information to a terminating sender (which could be on either the same or a different exchange). This sender then used a terminating marker to connect the calling user, via the selected incoming trunk, to the called user, and caused the controlling relay set to pass intermittent ring voltage of about 90 VAC at 20 Hz to ring the called user's phone bell, and return ringing tone to the caller.

The crossbar switch itself was simple: exchange design moved all the logical decision-making to the common control elements, which as relay sets were themselves very reliable. The design criterion was to have two hours of "downtime" for service every forty years, which was a huge improvement on earlier electromechanical systems. The exchange design concept lent itself to incremental upgrades, as the control elements could be replaced separately from the call switching elements. The minimum size of a crossbar exchange was comparatively large, but in city areas with a large installed line capacity the whole exchange occupied less space than other exchange technologies of equivalent capacity. For this reason they were also typically the first switches to be replaced with digital systems, which were even smaller and more reliable.

Two principles of using crossbar switches were used. One early method was based on the selector principle, and used the switches as functional replacement for Strowger or stepping switches. Control was distributed to the switches themselves. Call establishment progressed through the exchange stage by stage, as successive digits were dialed. With the selector principle, each switch could only handle its portion of one call at a time. Each moving contact of the array was multipled to corresponding crosspoints on other switches to a selector in the next bank of switches. Thus an exchange with a hundred 10x10 switches in five stages could only have twenty conversations in progress. Distributed control meant there was no common point of failure, but also meant that the setup stage lasted for the ten seconds or so the caller took to dial the required number. In control occupancy terms this comparatively long interval degrades the traffic capacity of a switch.

"Banjo" wiring of a 100 point six wire Type B Bell System switch

Starting with the 1XB switch, the later and more common method was based on the link principle, and used the switches as crosspoints. Each moving contact was multipled to the other contacts on the same level by simpler "banjo" wires, to a link on one of the inputs of a switch in the next stage. The switch could handle its portion of as many calls as it had levels or verticals. Thus an exchange with forty 10x10 switches in four stages could have a hundred conversations in progress. The link principle was more efficient, but required a more complex control system to find idle links through the switching fabric.

This meant common control, as described above: all the digits were recorded, then passed to the common control equipment - the marker - to establish the call at all the separate switch stages simultaneously. A marker-controlled crossbar system had in the marker a highly vulnerable central control; this was invariably protected by having duplicate markers. The great advantage was that the control occupancy on the switches was of the order of one second or less, representing the operate and release lags of the X-then-Y armatures of the switches. The only downside of common control was the need to provide digit recorders enough to deal with the greatest forecast originating traffic level on the exchange.

The Plessey TXK1 or 5005 design used an intermediate form, in which a clear path was marked through the switching fabric by distributed logic, and then closed through all at once.

In some countries, no crossbar exchanges remain in revenue service. However, crossbar exchanges remain in use in countries like Russia, where some massive city telephone networks have not yet been fully upgraded to digital technology. Preserved installations may be seen in museums like The Museum of Communications in Seattle, Washington, and the Science Museum in London.

Changing nomenclature can confuse: in current American terminology a "switch" now frequently refers to a system which is also called a "telephone exchange" (the usual term in English)--that is, a large collection of selectors of some sort within a building. For most of the twentieth century a "Strowger switch" or a "crossbar switch" referred to an individual piece of mechanical equipment making up part of an exchange. Hence the pictures above show a "crossbar switch" using the earlier meaning.

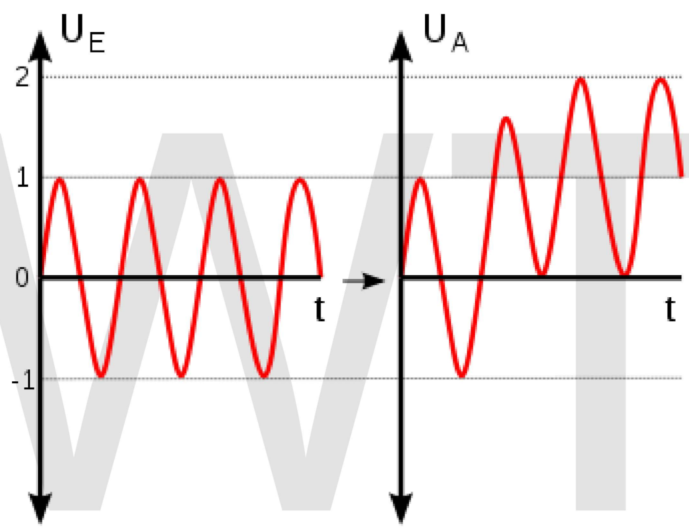
Semiconductor

Semiconductor implementations of crossbar switches typically consist of a set of input amplifiers or retimers connected to a series of metalizations or "bars" within a semiconductor device. A similar set of metalizations or "bars" are connected to output amplifiers or retimers. At each cross-point where the "bars" cross, a pass transistor is implemented which connects the bars. When the pass transistor is enabled, the input is connected to the output.

As computer technologies have improved, crossbar switches have found uses in systems such as the multistage interconnection networks that connect the various processing units in a Uniform Memory Access parallel processor to the array of memory elements.

Chapter- 6

Clamper



Positive unbiased voltage clamping shifts the amplitude of the input waveform so that all parts of it are greater than 0 V

A **clamper** is an electronic circuit that prevents a signal from exceeding a certain defined magnitude by shifting its DC value. The clamper does not restrict the peak-to-peak excursion of the signal, but moves it up or down by a fixed value. A **diode clamp** (a simple, common type) relies on a diode, which conducts electric current in only one direction; resistors and capacitors in the circuit are used to maintain an altered dc level at the clamper output.

General function

A clamping circuit (also known as a clamper) will bind the upper or lower extreme of a waveform to a fixed DC voltage level. These circuits are also known as DC voltage restorers. Clampers can be constructed in both positive and negative polarities. When unbiased, clamping circuits will fix the voltage lower limit (or upper limit, in the case of negative clampers) to 0 Volts. These circuits clamp a peak of a waveform to a specific

DC level compared with a capacitively coupled signal which swings about its average DC level (usually 0 V).

Types

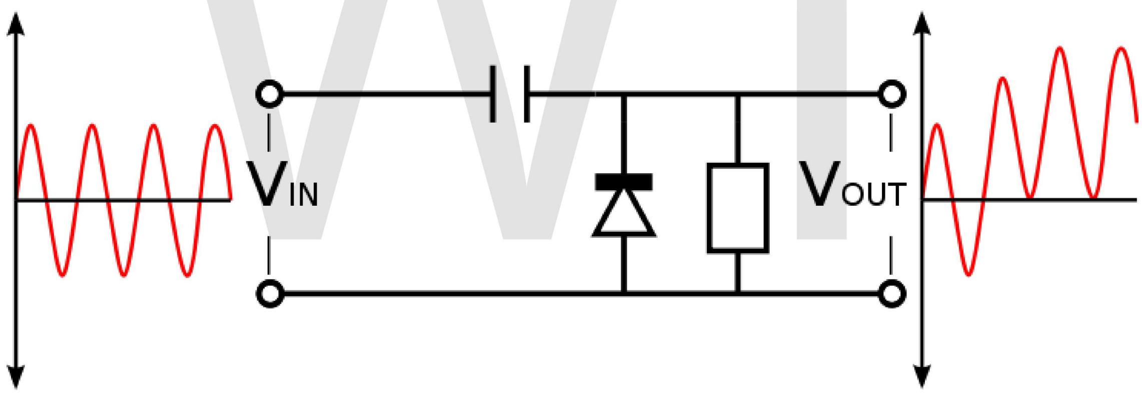
Clamp circuits are categorised by their operation; negative or positive and biased and unbiased. A positive clamp circuit outputs a purely positive waveform from an input signal; it offsets the input signal so that all of the waveform is greater than 0 V. A negative clamp is the opposite of this - this clamp outputs a purely negative waveform from an input signal.

A bias voltage between the diode and ground offsets the output voltage by that amount.

For example, an input signal of peak value 5 V (V_{IN} = 5 V) is applied to a positive clamp with a bias of 3 V (V_{BIAS} = 3 V), the peak output voltage will be

$V_{OUT} = 2V_{IN} + V_{BIAS}$
$V_{OUT} = 2 * 5\ V + 3\ V$
$V_{OUT} = 13\ V$

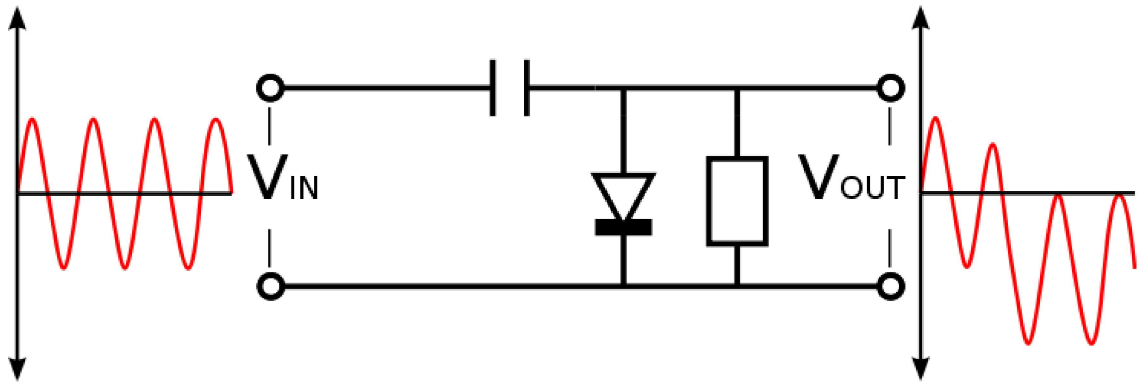
Positive unbiased



A positive unbiased clamp

In the negative cycle of the input AC signal, the diode is forward biased and conducts, charging the capacitor to the peak positive value of V_{IN}. During the positive cycle, the diode is reverse biased and thus does not conduct. The output voltage is therefore equal to the voltage stored in the capacitor plus the input voltage again, so $V_{OUT} = 2V_{IN}$

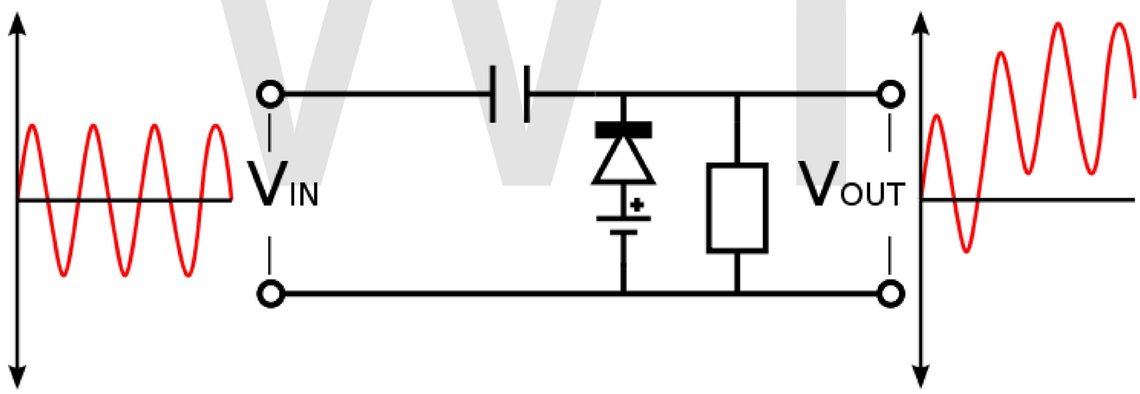
Negative unbiased



A negative unbiased clamp

A negative unbiased clamp is the opposite of the equivalent positive clamp. In the positive cycle of the input AC signal, the diode is forward biased and conducts, charging the capacitor to the peak value of V_{IN}. During the negative cycle, the diode is reverse biased and thus does not conduct. The output voltage is therefore equal to the voltage stored in the capacitor plus the input voltage again, so $V_{OUT} = -2V_{IN}$

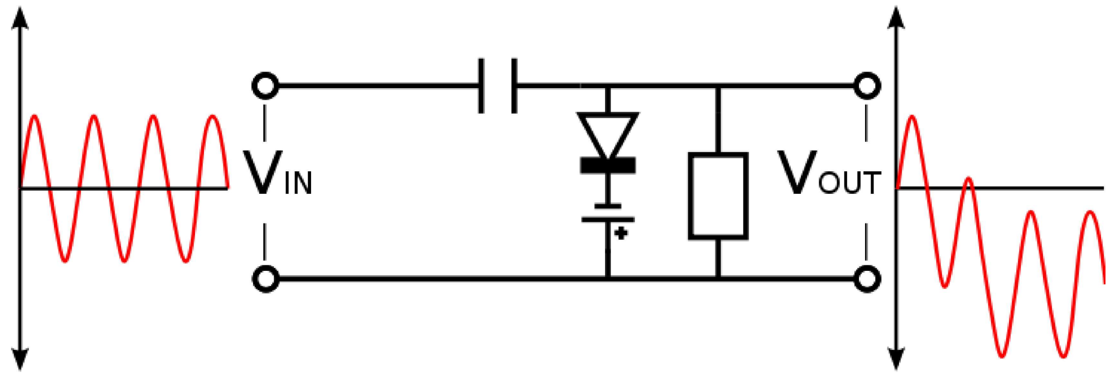
Positive biased



A positive biased clamp

A positive biased voltage clamp is identical to an equivalent unbiased clamp but with the output voltage offset by the bias amount V_{BIAS}. Thus, $V_{OUT} = 2V_{IN} + V_{BIAS}$

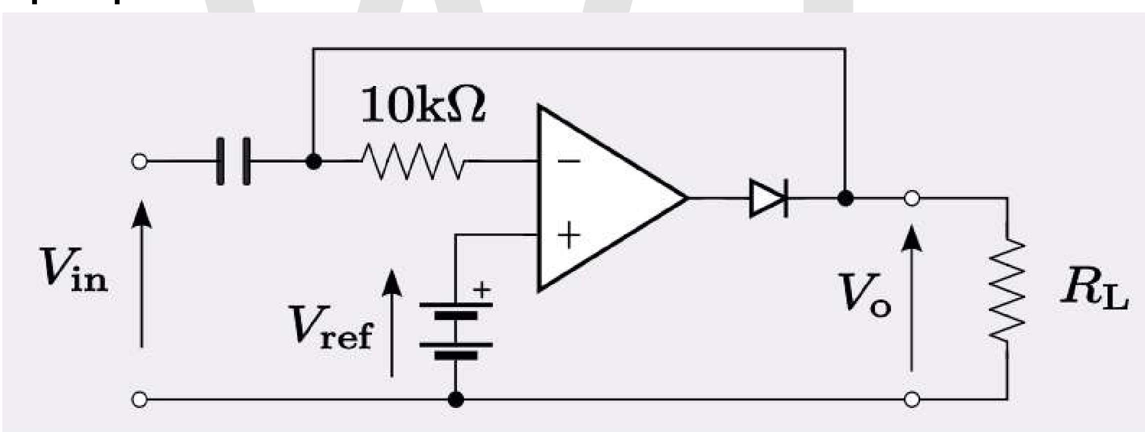
Negative biased



A negative biased clamp

A negative biased voltage clamp is likewise identical to an equivalent unbiased clamp but with the output voltage offset in the negative direction by the bias amount V_{BIAS}. Thus, $V_{OUT} = -2V_{IN} - V_{BIAS}$

Op-amp circuit



Precision op-amp clamp circuit

The figure shows an op-amp clamp circuit with a non-zero reference clamping voltage. The advantage here is that the clamping level is at precisely the reference voltage. There is no need to take into account the forward volt drop of the diode (which is necessary in the preceding simple circuits as this adds to the reference voltage). The effect of the diode volt drop on the circuit output will be divided down by the gain of the amplifier, resulting in an insignificant error.

Clamping for input protection

Clamping can be used to adapt an input signal to a device that cannot make use of or may be damaged by the signal range of the original input.

Principles of operation

The schematic of a clamper reveals that it is a relatively simple device. The two components creating the clamping effect are a capacitor, followed by a diode in parallel with the load. The clamper circuit relies on a change in the capacitor's time constant; this is the result of the diode changing current path with the changing input voltage. The magnitude of R and C are chosen so that $\tau = RC$ is large enough to ensure that the voltage across the capacitor does not discharge significantly during the diode's "Non conducting" interval. During the first negative phase of the AC input voltage, the capacitor in the positive clamper charges rapidly. As V_{in} becomes positive, the capacitor serves as a voltage doubler; since it has stored the equivalent of V_{in} during the negative cycle, it provides nearly that voltage during the positive cycle; this essentially doubles the voltage seen by the load. As V_{in} becomes negative, the capacitor acts as a battery of the same voltage of V_{in}. The voltage source and the capacitor counteract each other, resulting in a net voltage of zero as seen by the load.

Biased versus non-biased

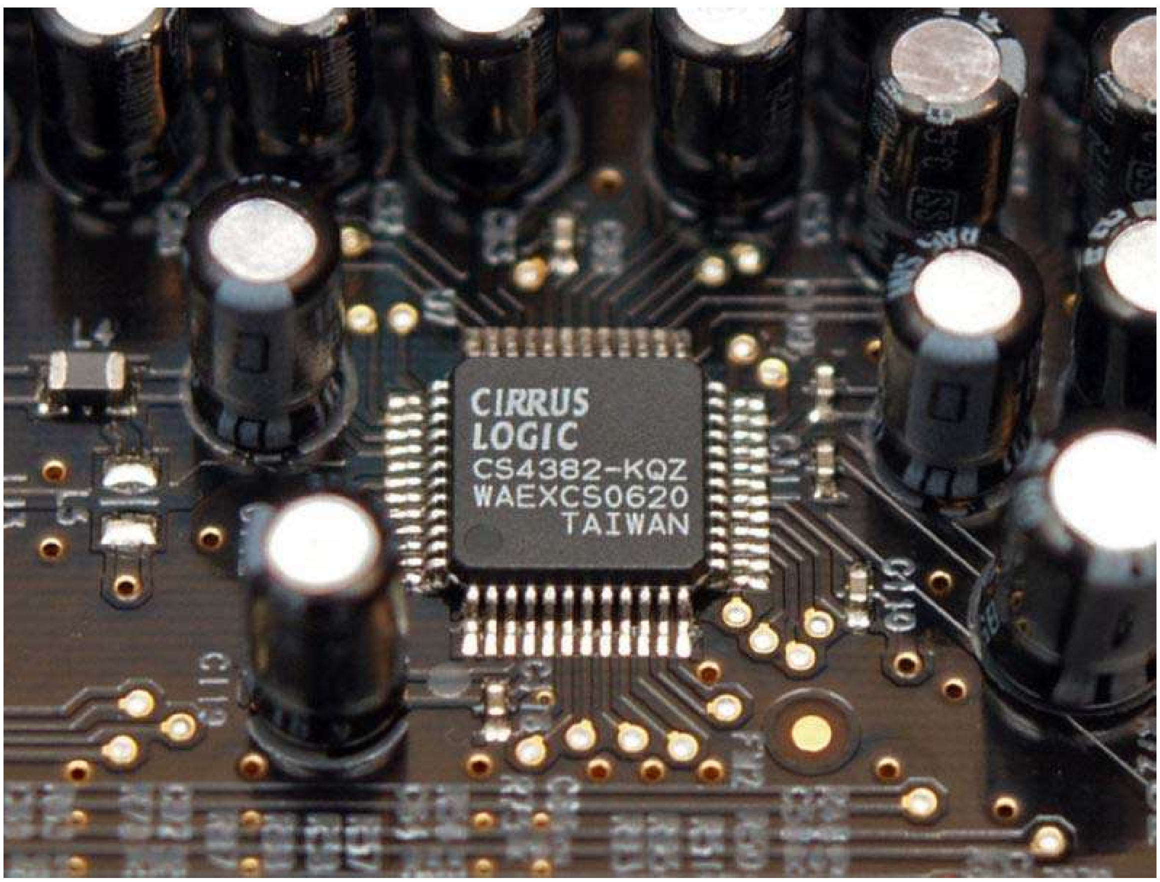
By using a voltage source and resistor, the clamper can be biased to bind the output voltage to a different value. The voltage supplied to the potentiometer will be equal to the offset from zero (assuming an ideal diode) in the case of either a positive or negative clamper (the clamper type will determine the direction of the offset. If a negative voltage is supplied to either positive or negative, the waveform will cross the x-axis and be bound to a value of this magnitude on the opposite side. Zener diodes can also be used in place of a voltage source and potentiometer, hence setting the offset at the Zener voltage.

Examples

One common such clamping circuit is the DC restorer circuit in analog television receiver, which returns the voltage of the signal during the back porch of the line blanking period to 0 V. Since the back porch is required to be at 0 V on transmission, any DC or low frequency hum that has been induced onto the signal can be effectively removed via this method.

Chapter- 7

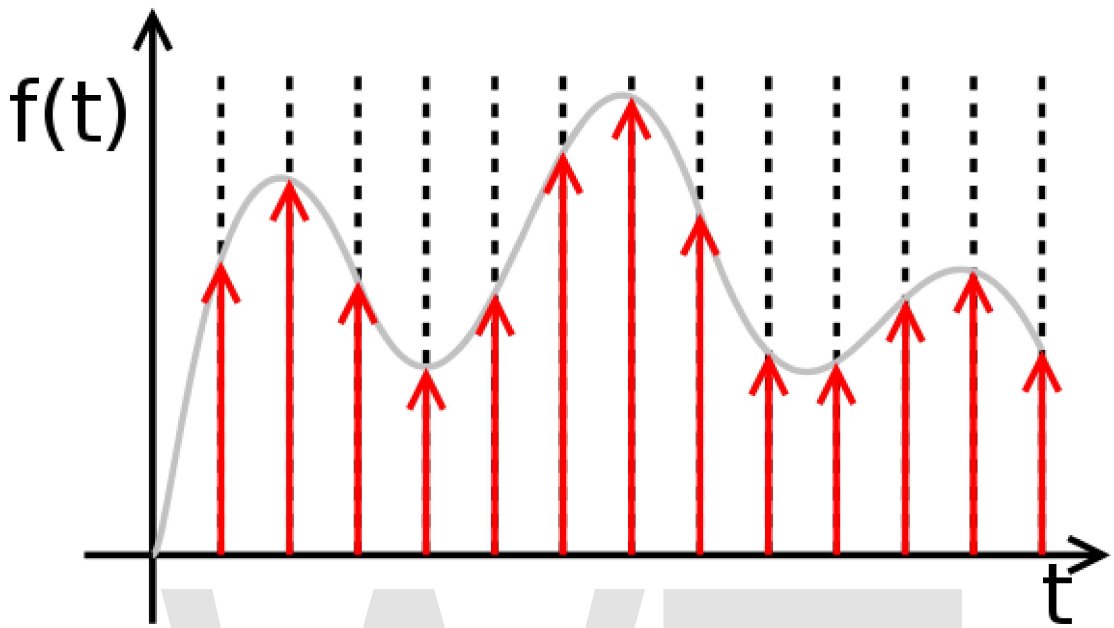
Digital-to-analog Converter



8-channel digital-to-analog converter Cirrus Logic CS4382 as used in a soundcard.

In electronics, a **digital-to-analog converter** (**DAC** or **D-to-A**) is a device that converts a digital (usually binary) code to an analog signal (current, voltage, or electric charge). An analog-to-digital converter (ADC) performs the reverse operation.

Basic ideal operation



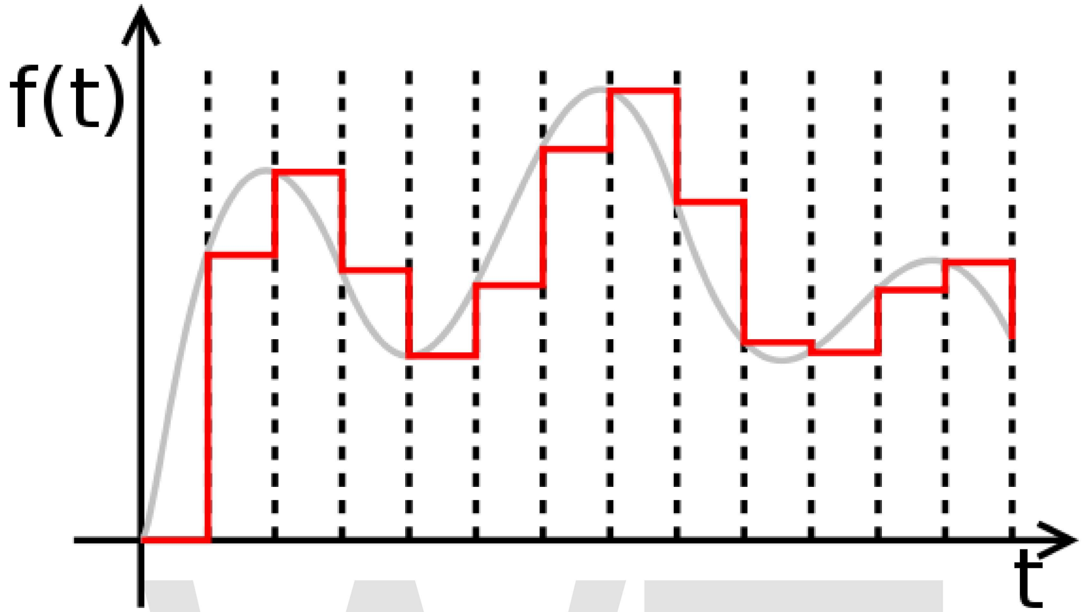
Ideally sampled signal.

A DAC converts an abstract finite-precision number (usually a fixed-point binary number) into a concrete physical quantity (e.g., a voltage or a pressure). In particular, DACs are often used to convert finite-precision time series data to a continually varying physical signal.

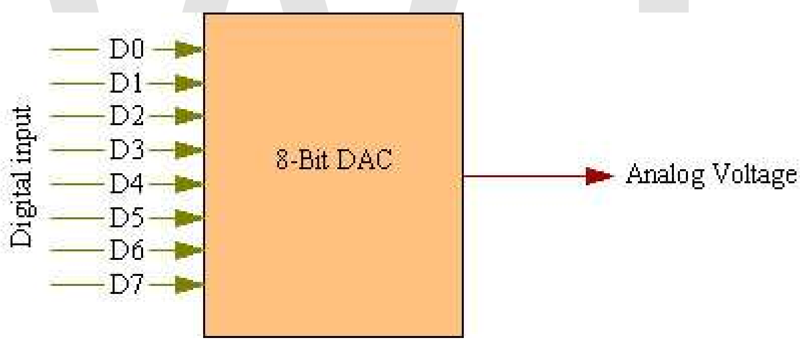
A typical DAC converts the abstract numbers into a concrete sequence of impulses that are then processed by a reconstruction filter using some form of interpolatin to fill in data between the impulses. Other DAC methods (e.g., methods based on Delta-sigma modulation) produce a pulse-density modulated signal that can then be filtered in a similar way to produce a smoothly varying signal.

By the Nyquist–Shannon sampling theorem, sampled data can be reconstructed perfectly provided that its bandwidth meets certain requirements (e.g., a baseband signal with bandwidth less than the Nyquist frequency; BUT requires an infinite number of samples. The finite number used in real life cause other problems especially with the D/A reconstruction of the original signal. However, even with an ideal reconstruction filter, digital sampling introduces quantization error that makes perfect reconstruction practically impossible. Increasing the digital resolution (i.e., increasing the number of bits used in each sample) or introducing sampling dither can reduce this error.

Practical operation



Piecewise constant output of a conventional practical DAC.



A simplified functional diagram of an 8-bit DAC

Instead of impulses, usually the sequence of numbers update the analogue voltage at uniform sampling intervals.

These numbers are written to the DAC, typically with a clock signal that causes each number to be latched in sequence, at which time the DAC output voltage changes rapidly from the previous value to the value represented by the currently latched number. The effect of this is that the output voltage is *held* in time at the current value until the next input number is latched resulting in a piecewise constant or 'staircase' shaped output. This is equivalent to a zero-order hold operation and has an effect on the frequency response of the reconstructed signal.

The fact that DACs output a sequence of piecewise constant values (known as zero-order hold in sample data textbooks) or rectangular pulses causes multiple harmonics above the

Nyquist frequency. Usually, these are removed with a low pass filter acting as a reconstruction filter in applications that require it.

Applications

Audio

Top-loading CD player and external digital-to-analog converter.

Most modern audio signals are stored in digital form (for example MP3s and CDs) and in order to be heard through speakers they must be converted into an analog signal. DACs are therefore found in CD players, digital music players, and PC sound cards.

Specialist standalone DACs can also be found in high-end hi-fi systems. These normally take the digital output of a compatible CD player or dedicated transport and convert the signal into an analog line-level output that can then be fed into an amplifier to drive speakers.

Similar digital-to-analog converters can be found in digital speakers such as USB speakers, and in sound cards.

VOIP (Voice over IP) Phone, Data transmission over the Internet is done digitally so in order for voice to be transmitted it must be converted to digital using an Analog-to-Digital Converter and be converted into analog again using a DAC so the voice it can be heard on the other end.

Video

Video signals from a digital source, such as a computer, must be converted to analog form if they are to be displayed on an analog monitor. As of 2007, analog inputs are more commonly used than digital, but this may change as flat panel displays with DVI and/or HDMI connections become more widespread. A video DAC is, however, incorporated in any digital video player with analog outputs. The DAC is usually integrated with some memory (RAM), which contains conversion tables for gamma correction, contrast and brightness, to make a device called a RAMDAC.

A device that is distantly related to the DAC is the digitally controlled potentiometer, used to control an analog signal digitally.

Mechanical

An unusual application of digital-to-analog conversion was the whiffletree electromechanical digital-to-analog convertor linkage in the IBM Selectric typewriter.

DAC types

The most common types of electronic DACs are:

- The pulse-width modulator, the simplest DAC type. A stable current or voltage is switched into a low-pass analog filter with a duration determined by the digital input code. This technique is often used for electric motor speed control, but has many other applications as well.
- Oversampling DACs or interpolating DACs such as the delta-sigma DAC, use a pulse density conversion technique. The oversampling technique allows for the use of a lower resolution DAC internally. A simple 1-bit DAC is often chosen because the oversampled result is inherently linear. The DAC is driven with a pulse-density modulated signal, created with the use of a low-pass filter, step nonlinearity (the actual 1-bit DAC), and negative feedback loop, in a technique called delta-sigma modulation. This results in an effective high-pass filter acting on the quantization (signal processing) noise, thus steering this noise out of the low frequencies of interest into the megahertz frequencies of little interest, which is called noise shaping. The quantization noise at these high frequencies is removed or greatly attenuated by use of an analog low-pass filter at the output (sometimes a simple RC low-pass circuit is sufficient). Most very high resolution DACs (greater than 16 bits) are of this type due to its high linearity and low cost. Higher oversampling rates can relax the specifications of the output low-pass filter and enable further suppression of quantization noise. Speeds of greater than

100 thousand samples per second (for example, 192 kHz) and resolutions of 24 bits are attainable with delta-sigma DACs. A short comparison with pulse-width modulation shows that a 1-bit DAC with a simple first-order integrator would have to run at 3 THz (which is physically unrealizable) to achieve 24 meaningful bits of resolution, requiring a higher-order low-pass filter in the noise-shaping loop. A single integrator is a low-pass filter with a frequency response inversely proportional to frequency and using one such integrator in the noise-shaping loop is a first order delta-sigma modulator. Multiple higher order topologies (such as MASH) are used to achieve higher degrees of noise-shaping with a stable topology.

- The binary-weighted DAC, which contains one resistor or current source for each bit of the DAC connected to a summing point. These precise voltages or currents sum to the correct output value. This is one of the fastest conversion methods but suffers from poor accuracy because of the high precision required for each individual voltage or current. Such high-precision resistors and current sources are expensive, so this type of converter is usually limited to 8-bit resolution or less.
- The R-2R ladder DAC which is a binary-weighted DAC that uses a repeating cascaded structure of resistor values R and 2R. This improves the precision due to the relative ease of producing equal valued-matched resistors (or current sources). However, wide converters perform slowly due to increasingly large RC-constants for each added R-2R link.
- The thermometer-coded DAC, which contains an equal resistor or current-source segment for each possible value of DAC output. An 8-bit thermometer DAC would have 255 segments, and a 16-bit thermometer DAC would have 65,535 segments. This is perhaps the fastest and highest precision DAC architecture but at the expense of high cost. Conversion speeds of >1 billion samples per second have been reached with this type of DAC.
- Hybrid DACs, which use a combination of the above techniques in a single converter. Most DAC integrated circuits are of this type due to the difficulty of getting low cost, high speed and high precision in one device.
    - The segmented DAC, which combines the thermometer-coded principle for the most significant bits and the binary-weighted principle for the least significant bits. In this way, a compromise is obtained between precision (by the use of the thermometer-coded principle) and number of resistors or current sources (by the use of the binary-weighted principle). The full binary-weighted design means 0% segmentation, the full thermometer-coded design means 100% segmentation.

DAC performance

DACs are very important to system performance. The most important characteristics of these devices are:

- **Resolution**: This is the number of possible output levels the DAC is designed to reproduce. This is usually stated as the number of bits it uses, which is the base two logarithm of the number of levels. For instance a 1 bit DAC is designed to reproduce 2 (2^1) levels while an 8 bit DAC is designed for 256 (2^8) levels. Resolution is related to the **effective number of bits** (ENOB) which is a measurement of the actual resolution attained by the DAC.

- **Maximum sampling frequency**: This is a measurement of the maximum speed at which the DACs circuitry can operate and still produce the correct output. As stated in the Nyquist–Shannon sampling theorem, a signal must be sampled at over twice the frequency of the desired signal. For instance, to reproduce signals in all the audible spectrum, which includes frequencies of up to 20 kHz, it is necessary to use DACs that operate at over 40 kHz. The CD standard samples audio at 44.1 kHz, thus DACs of this frequency are often used. A common frequency in cheap computer sound cards is 48 kHz — many work at only this frequency, offering the use of other sample rates only through (often poor) internal resampling.
- **Monotonicity**: This refers to the ability of a DAC's analog output to move only in the direction that the digital input moves (i.e., if the input increases, the output doesn't dip before asserting the correct output.) This characteristic is very important for DACs used as a low frequency signal source or as a digitally programmable trim element.
- **THD+N**: This is a measurement of the distortion and noise introduced to the signal by the DAC. It is expressed as a percentage of the total power of unwanted harmonic distortion and noise that accompany the desired signal. This is a very important DAC characteristic for dynamic and small signal DAC applications.
- **Dynamic range**: This is a measurement of the difference between the largest and smallest signals the DAC can reproduce expressed in decibels. This is usually related to DAC resolution and noise floor.

Other measurements, such as phase distortion and jitter, can also be very important for some applications.

Bits	Color limit	Frequency	Examples
10	1.024 colors	54 MHz	
12		54 MHz	Sony NS-575p
12	4.096 colors	108 MHz	
12		150 MHz	NeoDigits Helios X5000
12		216 MHz	Philips BDP9000 (Blu-ray)
12		297 MHz	Toshiba HD-XE1
12		216 MHz	Samsung BD-P1200 (Blu-ray)
14	16.384 colors	108 MHz	Pioneer Elite, Black Finish, DV79AVI
14		216 MHz	Marantz DV9600, Sony DVPNS9100ES
16		149 MHz	NeuNeo HVD108

DAC figures of merit

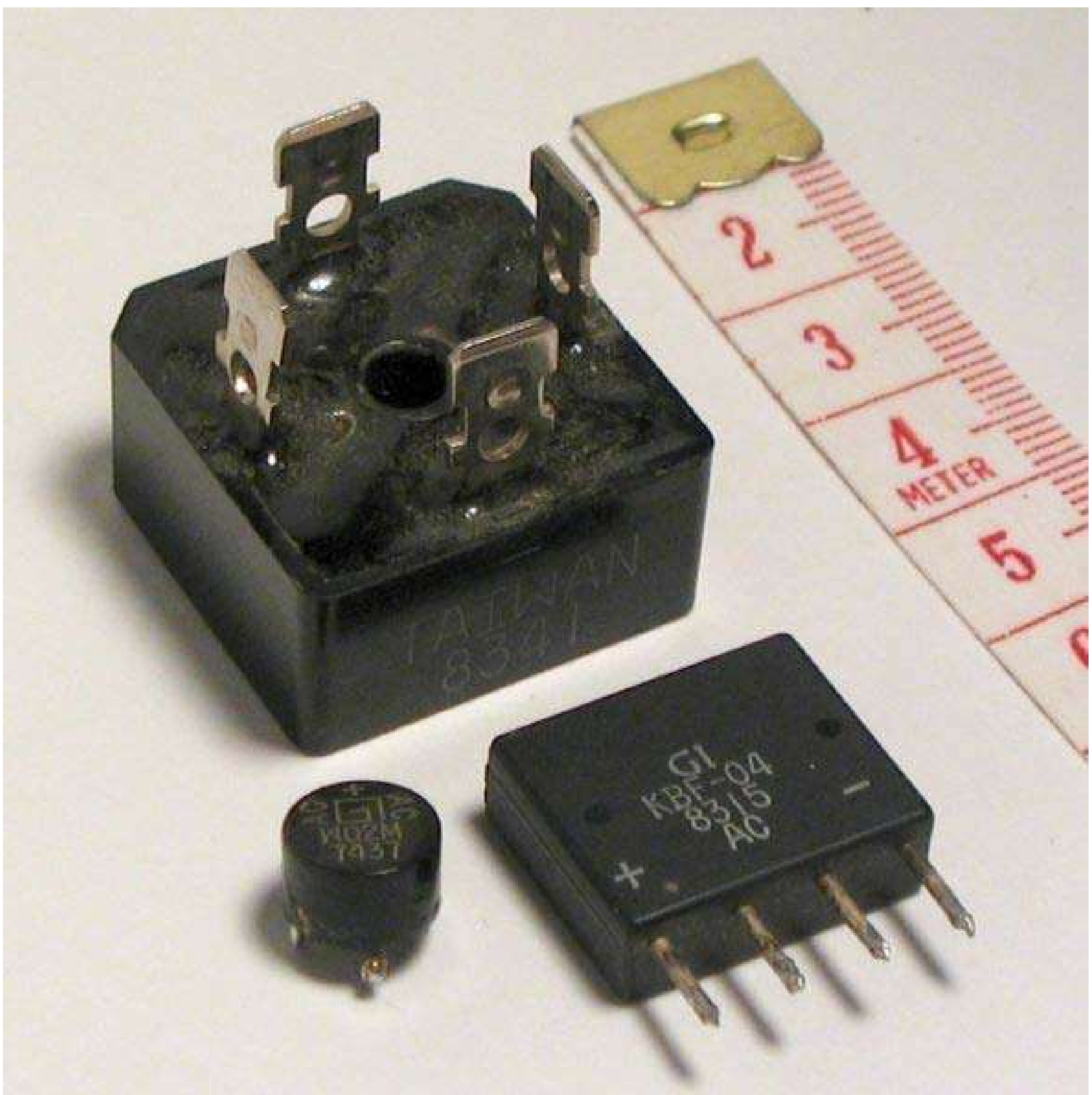
- Static performance:
    - Differential nonlinearity (DNL) shows how much two adjacent code analog values deviate from the ideal 1LSB step
    - Integral nonlinearity (INL) shows how much the DAC transfer characteristic deviates from an ideal one. That is, the ideal characteristic is usually a straight line; INL shows how much the actual voltage at a given code value differs from that line, in LSBs (1LSB steps).
    - Gain
    - Offset
    - Noise is ultimately limited by the thermal noise generated by passive components such as resistors. For audio applications and in room temperatures, such noise is usually a little less than 1 μV (microvolt) of white noise. This limits performance to less than 20~21 bits even in 24-bit DACs.
- Frequency domain performance
    - Spurious-free dynamic range (SFDR) indicates in dB the ratio between the powers of the converted main signal and the greatest undesired spur
    - Signal to noise and distortion ratio (SNDR) indicates in dB the ratio between the powers of the converted main signal and the sum of the noise and the generated harmonic spurs
    - i-th harmonic distortion (HDi) indicates the power of the i-th harmonic of the converted main signal
    - Total harmonic distortion (THD) is the sum of the powers of all HDi
    - If the maximum DNL error is less than 1 LSB, then D/A converter is guaranteed to be monotonic.

However, many monotonic converters may have a maximum DNL greater than 1 LSB.

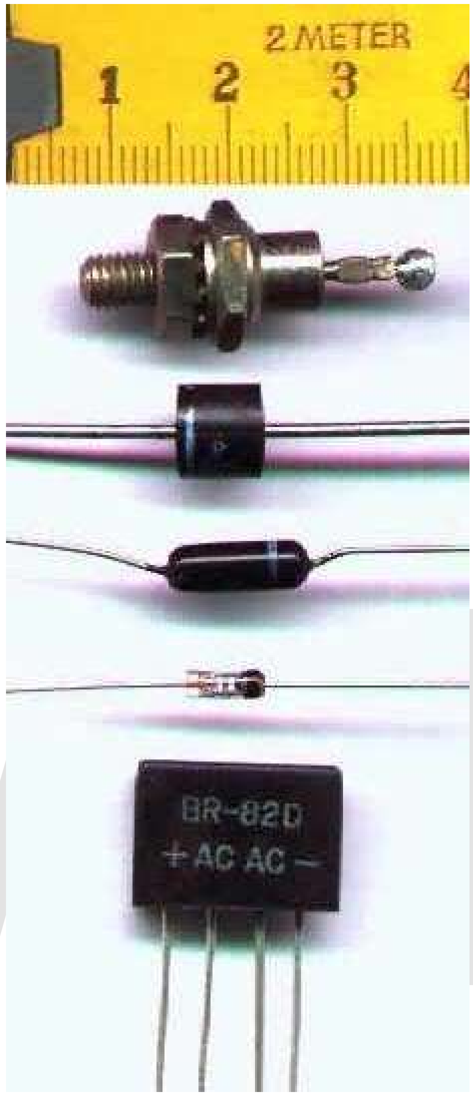
- Time domain performance:
    - Glitch energy
    - Response uncertainty
    - Time nonlinearity (TNL)

Chapter- 8

Diode Bridge



Three bridge rectifiers. The size is generally related to the current handling capability.



Diodes. The one on the bottom is a diode bridge.

A hand made diode bridge. The thick silver bar on the diodes indicates the cathode side of the diode.

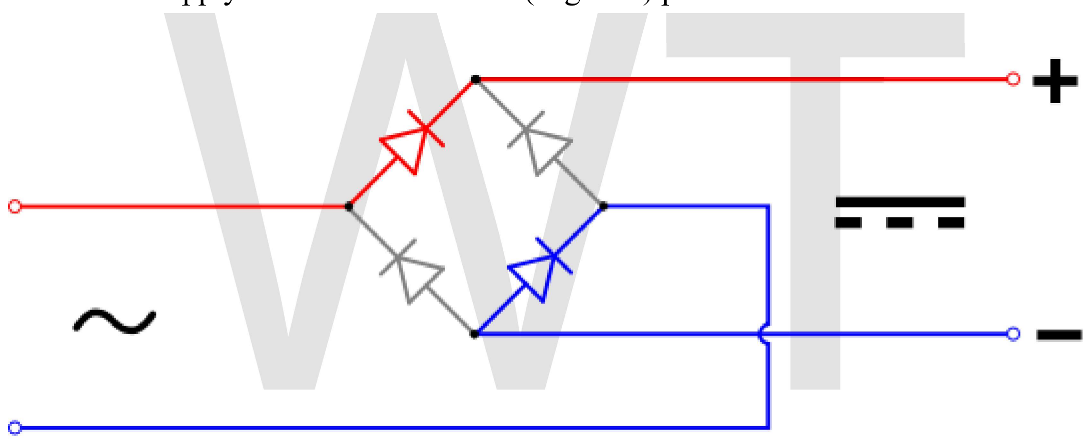
A **diode bridge** is an arrangement of four (or more) diodes in a bridge configuration that provides the same polarity of output for either polarity of input. When used in its most common application, for conversion of an alternating current (AC) input into direct current a (DC) output, it is known as a bridge rectifier. A bridge rectifier provides full-wave rectification from a two-wire AC input, resulting in lower cost and weight as compared to a rectifier with a 3-wire input from a transformer with a center-tapped secondary winding.

The essential feature of a diode bridge is that the polarity of the output is the same regardless of the polarity at the input. The diode bridge circuit is also known as the *Graetz circuit* after its inventor, physicist Leo Graetz.

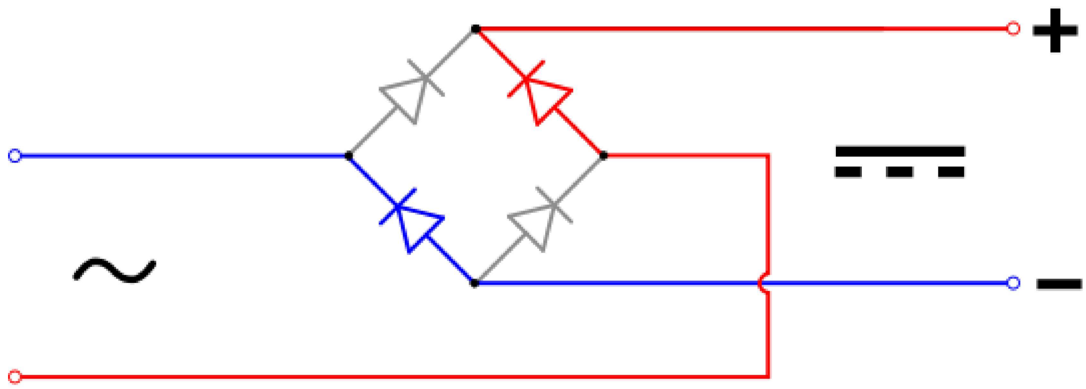
Basic operation

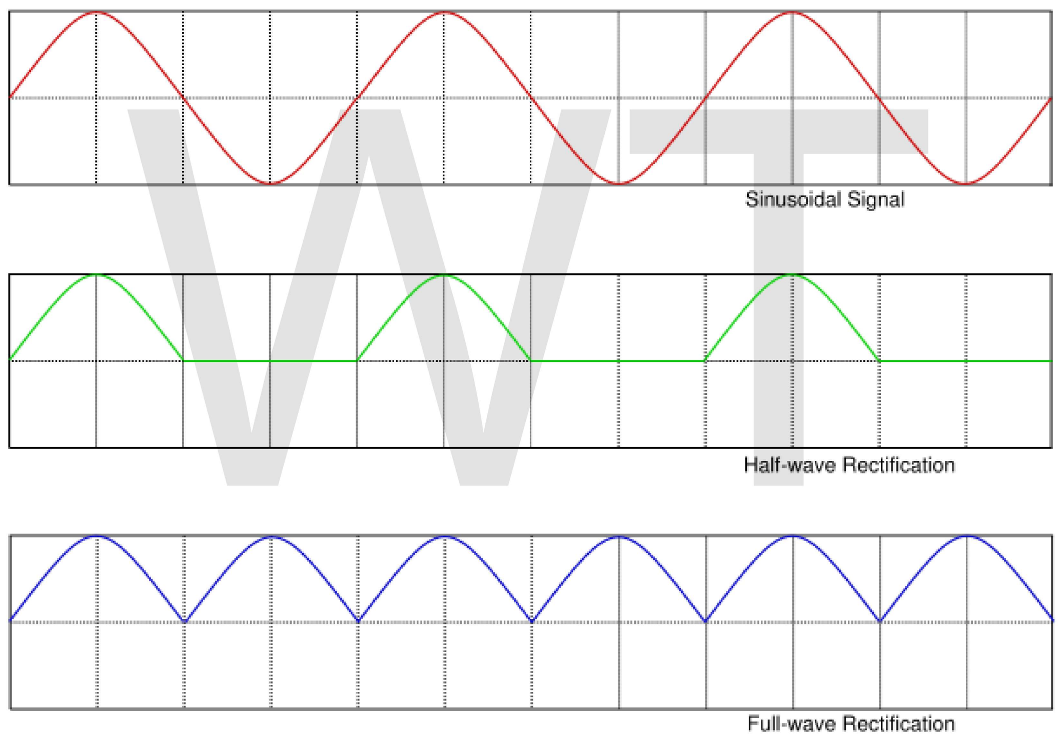
According to the conventional model of current flow originally established by Benjamin Franklin and still followed by most engineers today, current is *assumed* to flow through electrical conductors from the **positive** to the **negative** pole. In actuality, free electrons in a conductor nearly always flow from the **negative** to the **positive** pole. In the vast majority of applications, however, the *actual* direction of current flow is irrelevant. Therefore, in the discussion below the conventional model is retained.

In the diagrams below, when the input connected to the **left** corner of the diamond is **positive**, and the input connected to the **right** corner is **negative**, current flows from the **upper** supply terminal to the right along the **red** (positive) path to the output, and returns to the **lower** supply terminal via the **blue** (negative) path.



When the input connected to the **left** corner is **negative**, and the input connected to the **right** corner is **positive**, current flows from the **upper** supply terminal to the right along the **red** (positive) path to the output, and returns to the **lower** supply terminal via the **blue** (negative) path.





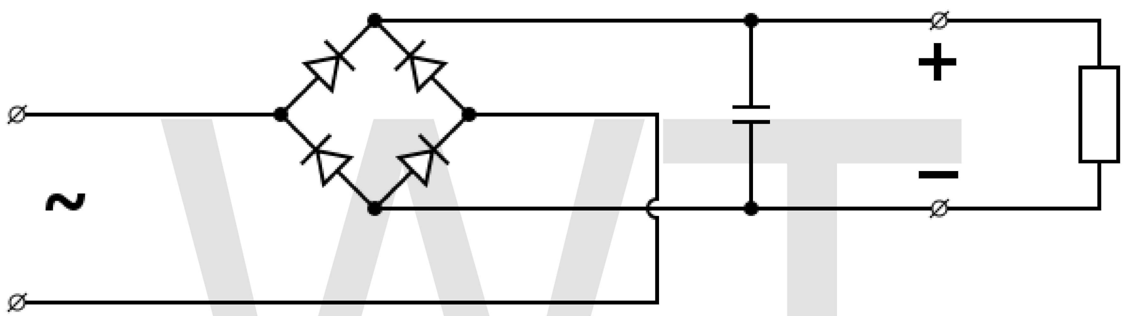
AC, half-wave and full wave rectified signals.

In each case, the upper right output remains positive and lower right output negative. Since this is true whether the input is AC or DC, this circuit not only produces a DC output from an AC input, it can also provide what is sometimes called "reverse polarity protection". That is, it permits normal functioning of DC-powered equipment when batteries have been installed backwards, or when the leads (wires) from a DC power source have been reversed, and protects the equipment from potential damage caused by reverse polarity.

Prior to the availability of integrated circuits, a bridge rectifier was constructed from "discrete components", i.e., separate diodes. Since about 1950, a single four-terminal component containing the four diodes connected in a bridge configuration became a standard commercial component and is now available with various voltage and current ratings.

Output smoothing

For many applications, especially with single phase AC where the full-wave bridge serves to convert an AC input into a DC output, the addition of a capacitor may be desired because the bridge alone supplies an output of fixed polarity but continuously varying or "pulsating" magnitude, an attribute commonly referred to as "ripple".



The function of this capacitor, known as a reservoir capacitor (or smoothing capacitor) is to lessen the variation in (or 'smooth') the rectified AC output voltage waveform from the bridge. One explanation of 'smoothing' is that the capacitor provides a low impedance path to the AC component of the output, reducing the AC voltage across, and AC current through, the resistive load. In less technical terms, any drop in the output voltage and current of the bridge tends to be canceled by loss of charge in the capacitor. This charge flows out as additional current through the load. Thus the change of load current and voltage is reduced relative to what would occur without the capacitor. Increases of voltage correspondingly store excess charge in the capacitor, thus moderating the change in output voltage / current.

The simplified circuit shown has a well-deserved reputation for being dangerous, because, in some applications, the capacitor can retain a *lethal* charge after the AC power source is removed. If supplying a dangerous voltage, a practical circuit should include a reliable way to discharge the capacitor safely. If the normal load cannot be guaranteed to perform this function, perhaps because it can be disconnected, the circuit should include a bleeder resistor connected as close as practical across the capacitor. This resistor should consume a current large enough to discharge the capacitor in a reasonable time, but small enough to minimize unnecessary power waste.

The capacitor and the load resistance have a typical time constant $\tau = RC$ where C and R are the capacitance and load resistance respectively. As long as the load resistor is large enough so that this time constant is much longer than the time of one ripple cycle, the above configuration will produce a smoothed DC voltage across the load.

In some designs, a series resistor at the load side of the capacitor is added. The smoothing can then be improved by adding additional stages of capacitor–resistor pairs, often done only for sub-supplies to critical high-gain circuits that tend to be sensitive to supply voltage noise.

The idealized waveforms shown above are seen for both voltage and current when the load on the bridge is resistive. When the load includes a smoothing capacitor, both the voltage and the current waveforms will be greatly changed. While the voltage is smoothed, as described above, current will flow through the bridge only during the time when the input voltage is greater than the capacitor voltage. For example, if the load draws an average current of n Amps, and the diodes conduct for 10% of the time, the average diode current during conduction must be 10n Amps. This non-sinusoidal current leads to harmonic distortion and a poor power factor in the AC supply.

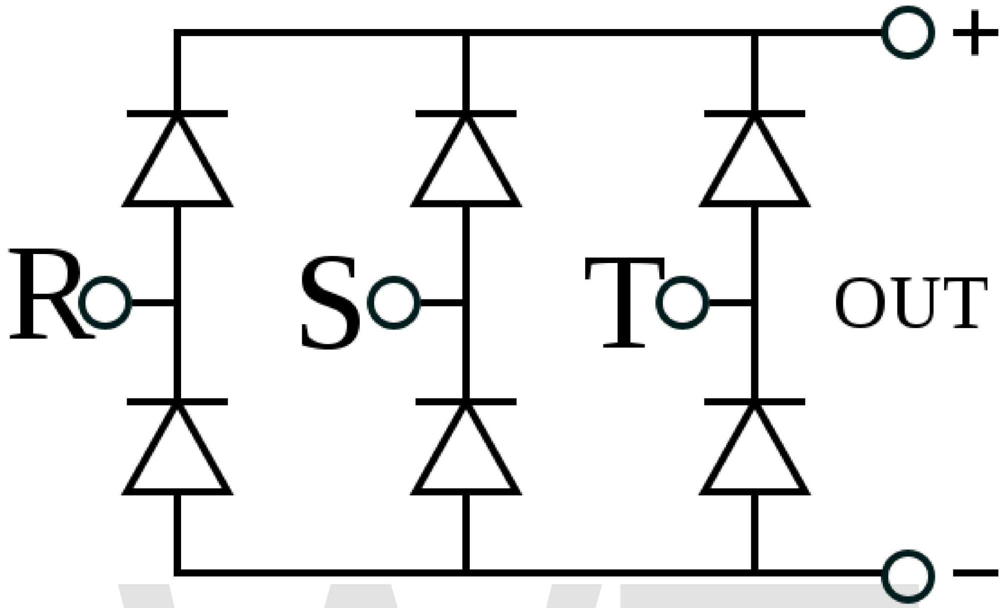
In a practical circuit, when a capacitor is directly connected to the output of a bridge, the bridge diodes must be sized to withstand the current surge that occurs when the power is turned on at the peak of the AC voltage and the capacitor is fully discharged. Sometimes a small series resistor is included before the capacitor to limit this current, though in most applications the power supply transformer's resistance is already sufficient.

Output can also be smoothed using a choke and second capacitor. The choke tends to keep the current (rather than the voltage) more constant. This design is not generally used in modern equipment due to the high cost of an effective choke compared to a resistor and capacitor.

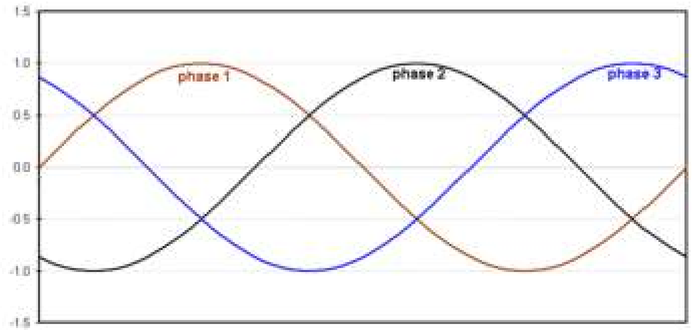
Some early console radios created the speaker's constant field with the current from the high voltage ("B +") power supply, which was then routed to the consuming circuits, (permanent magnets were then too weak for good performance) to create the speaker's constant magnetic field. The speaker field coil thus performed 2 jobs in one: it acted as a choke, filtering the power supply, and it produced the magnetic field to operate the speaker.

Polyphase diode bridges

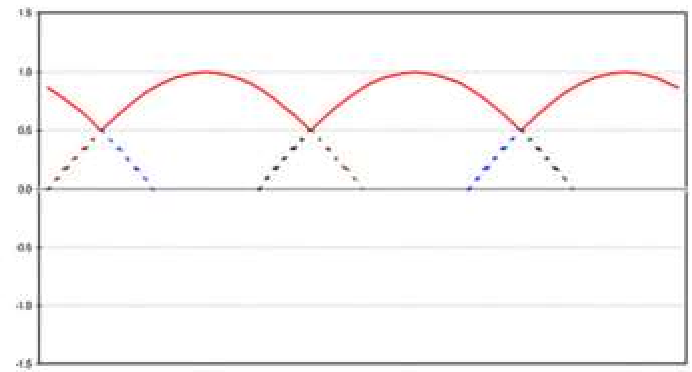
The diode bridge can be generalized to rectify polyphase AC inputs. For example, for a *three*-phase AC input, a half-wave rectifier consists of **three** diodes, but a full-wave *bridge* rectifier consists of **six** diodes.



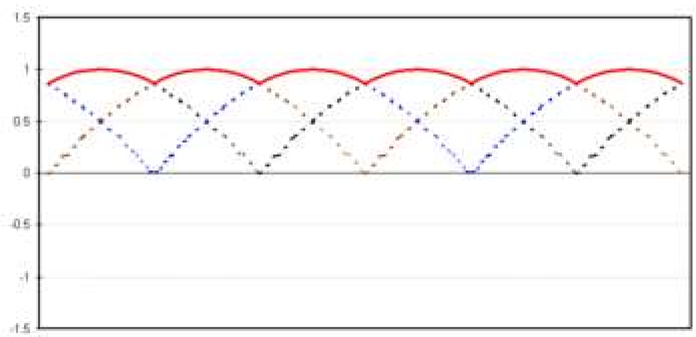
Three phase bridge rectifier.



3-PHASE AC

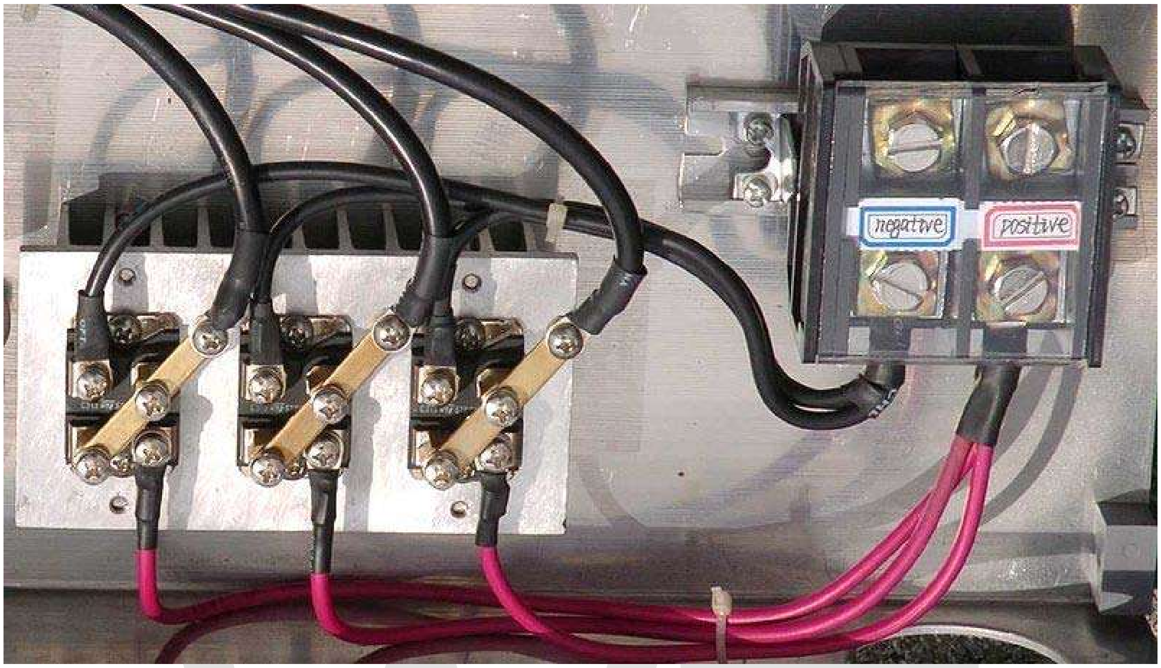


3-PHASE HALF-WAVE RECTIFICATION



3-PHASE FULL WAVE RECTIFICATION

3-phase AC input waveform (top), half-wave rectified waveform (center), and full-wave rectified waveform (bottom).



Three-phase bridge rectifier for a wind turbine.

Chapter- 9

Electrical Ballast

"Choke ballast" (inductor) used in older lighting. This example is from a tanning bed. Requires a lamp starter (below) and capacitor.

Lamp starter, required with some inductor type ballasts. Connects both ends of the lamp together to "preheat" the lamp ends for 1 second before lighting.

An **electrical ballast** (sometimes called **control gear**) is a device intended to limit the amount of current in an electric circuit.

Ballasts vary greatly in complexity. They can be as simple as a series resistor as commonly used with small neon lamps or light-emitting diodes (LEDs). For higher-power installations, too much energy would be wasted in a resistive ballast, so alternatives are used that depend upon the reactance of inductors, capacitors, or both. Finally, ballasts can be as complex as the computerized, remote-controlled electronic ballasts now often used with fluorescent lamps.

Current limiting

Ballasts stabilize the current through an electrical load. These are most often used when an electrical circuit or device presents a negative (differential) resistance to the supply. If such a device were connected to a constant-voltage power supply, it would draw an increasing amount of current until it was destroyed or caused the power supply to fail. To prevent this, a ballast provides a positive resistance or reactance that limits the ultimate current to an appropriate level. In this way, the ballast provides for the proper operation of the negative-resistance device by appearing to be a legitimate, stable resistance in the circuit.

An example of a negative-resistance device is a gas-discharge lamp, where after lamp ignition, increasing arc current reduces the voltage drop.

Ballasts can also be used simply to deliberately reduce the current in an ordinary, positive-resistance circuit.

Prior to the advent of solid-state ignition, automobile ignition systems commonly included a ballast resistor to regulate the voltage applied to the ignition system.

Although LEDs are positive resistance devices, they have insufficient resistance to regulate their current consumption when operated from a voltage controlled source, so ballasts are used to control the current through the LED. Because the power dissipation is minuscule, simple resistor ballasts are normally used.

Resistors

A **ballast resistor** compensates for normal or incidental changes in the physical state of a system. It may be a fixed or variable resistor.

Fixed resistors

For simple, low-powered loads such as a neon lamp or LED, a fixed resistor is commonly used. Because the resistance of the ballast resistor is large it dominates the current in the circuit, even in the face of negative resistance introduced by the neon lamp.

The term also refers to an automobile engine component that lowers the supply voltage to the ignition system after the engine has been started. Because cranking the engine causes a very heavy load on the battery, the system voltage can drop quite low during cranking.

To allow the engine to start, the ignition system must be designed to operate on this lower voltage. But once cranking is completed, the normal operating voltage is regained; this voltage would overload the ignition system. To avoid this problem, a ballast resistor is inserted in series with the supply voltage feeding the ignition system. Occasionally, this ballast resistor will fail and the classic symptom of this failure is that the engine runs while being cranked (while the resistor is bypassed) but stalls immediately when cranking ceases (and the resistor is re-connected in the circuit).

Modern electronic ignition systems do not require a ballast resistor as they are flexible enough to operate on the low cranking voltage or the ordinary operating voltage.

In some old AC/DC receivers (universal sets), the vacuum tube heaters are connected in series. Since the voltage drop across all the filaments in series is sometimes less than the full mains voltage, it was often necessary to get rid of the excess voltage. A ballast resistor was often used for this purpose, as it was cheap and worked with both AC and DC.

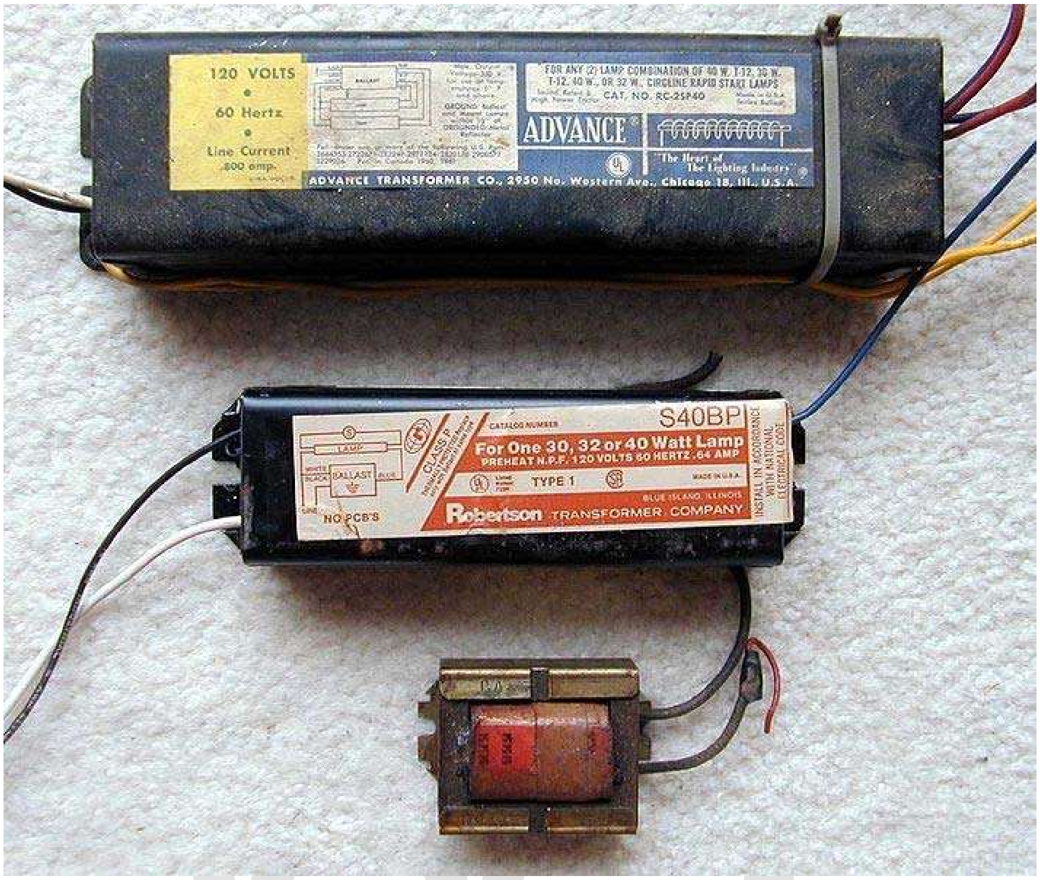
Self-variable resistors

Some ballast resistors have the property of increasing in resistance as current through them increases, and decreasing in resistance as current decreases. Physically, some such devices are often built quite like incandescent lamps. Like the tungsten filament of an ordinary incandescent lamp, if current increases, the ballast resistor gets hotter, its resistance goes up, and its voltage drop increases. If current decreases, the ballast resistor gets colder, its resistance drops, and the voltage drop decreases. Therefore the ballast resistor reduces variations in current, despite variations in applied voltage or changes in the rest of an electric circuit. These devices are sometimes termed barretters.

This property can lead to more precise current control than merely choosing an appropriate fixed resistor. The power lost in the resistive ballast is also reduced because a smaller portion of the overall power is dropped in the ballast compared to what might be required with a fixed resistor.

In times past, household clothes dryers sometimes incorporated a germicidal lamp in series with an ordinary incandescent lamp; the incandescent lamp operated as the ballast for the germicidal lamp. A commonly used light in the home in the 1960s in 220-240V countries was a circleline tube ballasted by an under-run regular mains filament lamp. Self ballasted mercury-vapor lamps incorporate ordinary tungsten filaments within the overall envelope of the lamp to act as the ballast, and it supplements the otherwise lacking red area of the light spectrum produced.

Reactive ballasts



Several typical magnetic ballasts for fluorescent lamps. The top is a high-power factor rapid start series ballast for two 30-40W lamps. The middle is a low power factor preheat ballast for a single 30-40W lamp while the bottom ballast is a simple inductor used with a 15W preheat lamp.

Because of the power that would be lost, resistors are not used as ballasts for lamps of more than about two watts. Instead, a reactance is used. Losses in the ballast due to its resistance and losses in its magnetic core may be significant, on the order of 5 to 25% of the lamp input wattage. Practical lighting design calculations must allow for ballast loss in estimating the running cost of a lighting installation.

An inductor is very common in line-frequency ballasts to provide the proper starting and operating electrical condition to power a fluorescent lamp, neon lamp, or high intensity discharge (HID) lamp. (Because of the use of the inductor, such ballasts are usually called *magnetic ballasts*.) The inductor has two benefits:

1. Its reactance limits the power available to the lamp with only minimal power losses in the inductor
2. The voltage spike produced when current through the inductor is rapidly interrupted is used in some circuits to first strike the arc in the lamp.

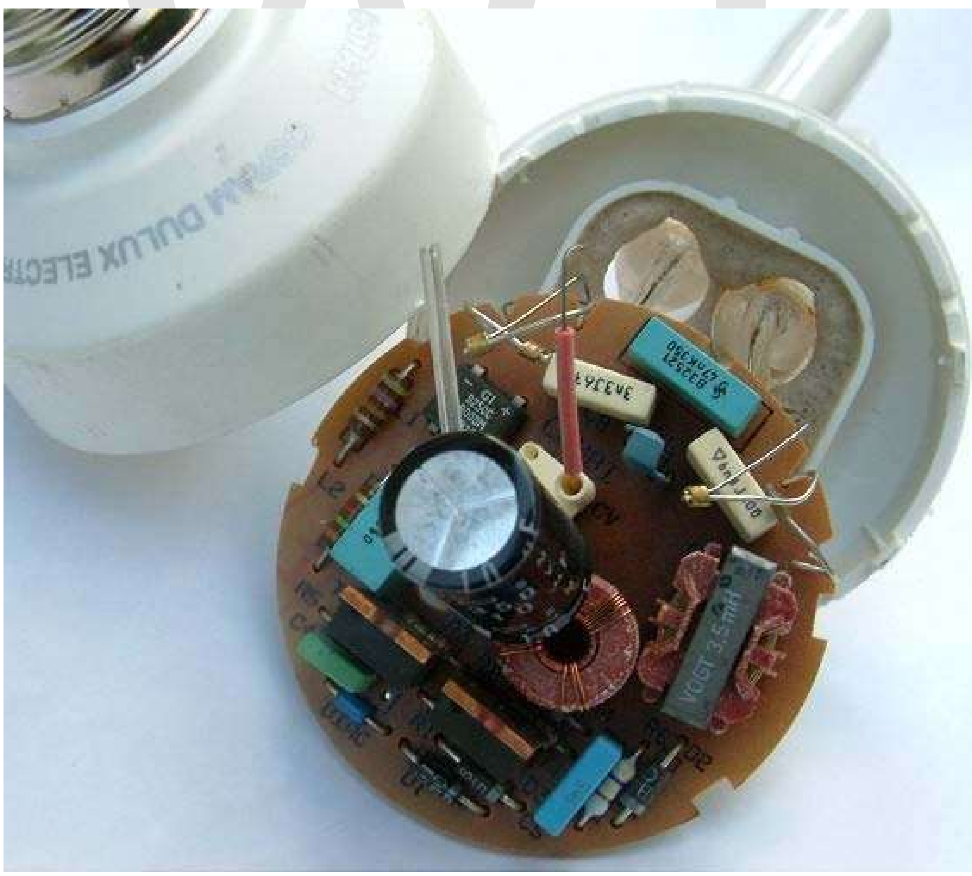
A disadvantage of the inductor is that current is shifted out of phase with the voltage, producing a poor power factor. In more expensive ballasts, a capacitor is often paired with the inductor to correct the power factor. In ballasts that control two or more lamps, line-frequency ballasts commonly use different phase relationships between the multiple lamps. This not only mitigates the flicker of the individual lamps, it also helps maintain a high power factor. These ballasts are often called *lead-lag* ballasts because the current in one lamp leads the mains phase and the current in the other lamp lags the mains phase.

For large lamps, line voltage may not be sufficient to start the lamp, so an autotransformer winding is included in the ballast to step up the voltage. The autotransformer is designed with enough leakage inductance so that the current is appropriately limited.

Because of the large inductors and capacitors that must be used, reactive ballasts operated at line frequency tend to be large and heavy. They commonly also produce acoustic noise (line-frequency hum).

Prior to 1980 in the United States, PCB-based oils were used as an insulating oil in many ballasts to provide cooling and electrical isolation.

Electronic ballasts



Electronic ballast of a compact fluorescent lamp

An **electronic lamp ballast** uses solid state electronic circuitry to provide the proper starting and operating electrical condition to power one or more fluorescent lamps and more recently HID lamps. Electronic ballasts usually change the frequency of the power from the standard mains (e.g., 60 Hz in U.S.) frequency to 20,000 Hz or higher, substantially eliminating the stroboscopic effect of flicker (a product of the line frequency) associated with fluorescent lighting. In addition, because more gas remains ionized in the arc stream, the lamps actually operate at about 9% higher efficacy above approximately 10 kHz. Lamp efficacy increases sharply at about 10 kHz and continues to improve until approximately 20 kHz. Because of the higher efficiency of the ballast itself and the improvement of lamp efficacy by operating at a higher frequency, electronic ballasts offer higher system efficacy for low pressure lamps like the fluorescent lamp. For HID lamps there is no improvement of the lamp efficacy in using higher frequency, but for these lamps the ballast losses are lower at higher frequencies and also the light depreciation is lower meaning more light after a given operating time of say 10 000 hours. Some HID lamp types like the Ceramic discharge metal halide lamp have reduced reliability when operated at high frequencies in the range of 20kHz to 200 kHz and for these lamps a square wave low frequency current drive is mostly used with frequency in the range of 100 to 400 Hz, with the same advantage of lower light depreciation. Electronic ballasts are often based on the SMPS topology, first rectifying the input power and then chopping it at a high frequency. Advanced electronic ballasts may allow dimming via pulse-width modulation or via changing the frequency to a higher value and remote control and monitoring via networks such as LonWorks, DALI, DMX-512, DSI or simple analog control using a 0-10V DC brightness control signal. Recently also systems remotely controlling the dim level via a wireless mesh network have been introduced.

Fluorescent lamp ballasts

Instant start

An instant start ballast starts lamps without heating the cathodes at all by using high voltage (around 600 V). It is the most energy efficient type, but gives the least number of starts from a lamp as emissive oxides are blasted from the cold cathode surfaces each time the lamp is started. This is the best type for installations where lamps are not turned on and off very often.

Rapid start

A rapid start ballast applies voltage and heats the cathodes simultaneously. Provides superior lamp life and more cycle life, but uses slightly more energy as the cathodes in each end of the lamp continue to consume heating power as the lamp operates. A dimming circuit can be used with a dimming ballast, which maintains the heating current while allowing lamp current to be controlled.

Programmed start

A programmed-start ballast is a more advanced version of rapid start. This ballast applies power to the filaments first, then after a short delay to allow the cathodes to preheat, applies voltage to the lamps to strike an arc. This ballast gives the best life and most starts from lamps, and so is preferred for applications with very frequent power cycling such as vision examination rooms and restrooms with a motion detector switch.

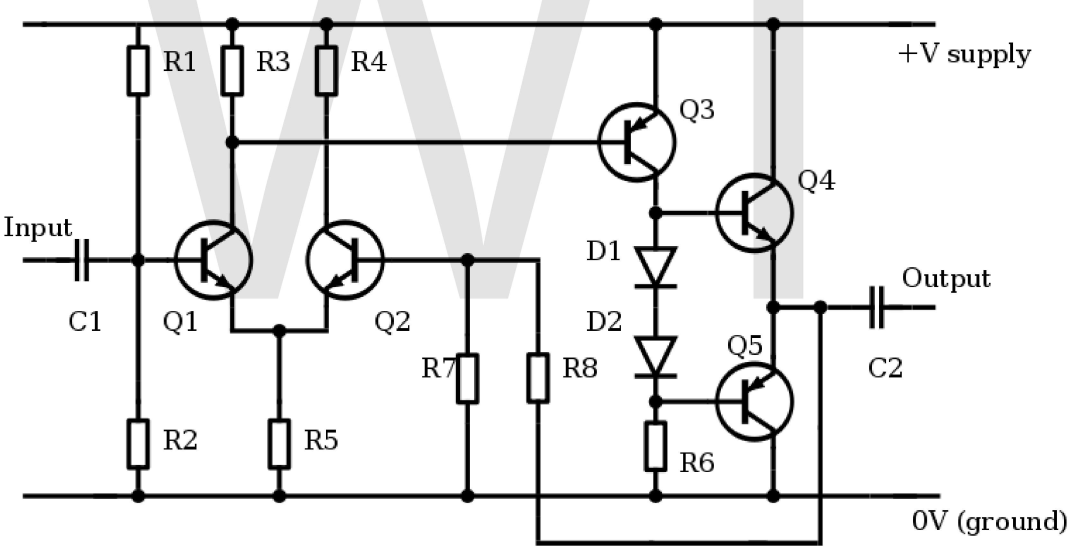
Ballast factor

For a lighting ballast, the *ballast factor* is defined as the light output (in lumens) with a test ballast, compared to the light output with a laboratory reference ballast that operates the lamp at its specified nominal power rating. The ballast factor of practical ballasts must be considered in lighting design; a low ballast factor may save energy, but will produce less light. With fluorescent lamps, an electronic ballast may produce more light than the reference test ballast, which operates the lamp with line frequency current; such electronic ballasts have a ballast factor greater than one.

Chapter- 10

Electronic Amplifier

An **electronic amplifier** is a device for increasing the power of a signal. It does this by taking energy from a power supply and controlling the output to match the input signal shape but with a larger amplitude. In this sense, an amplifier may be considered as modulating the output of the power supply.



A practical amplifier circuit

Types of amplifier

Amplifiers can be specified according to their input and output properties. They have some kind of gain, or multiplication factor relating the magnitude of the output signal to the input signal. The gain may be specified as the ratio of output voltage to input voltage (voltage gain), output power to input power (power gain), or some combination of current, voltage and power. In many cases, with input and output in the same units, gain will be unitless (although often expressed in decibels); for others this is not necessarily so. For example, a transconductance amplifier has a gain with units of conductance (output current per input voltage). The power gain of an amplifier depends on the source and load impedances used as well as its voltage gain; while an RF amplifier may have its impedances optimized for power transfer, audio and instrumentation amplifiers are normally employed with amplifier input and output impedances optimized for least loading and highest quality. So an amplifier that is said to have a gain of 20 dB might have a voltage gain of ten times and an available power gain of much more than 20 dB (100 times power ratio), yet be delivering a much lower power gain if, for example, the input is a 600 ohm microphone and the output is a 47 kilohm power amplifier's input socket.

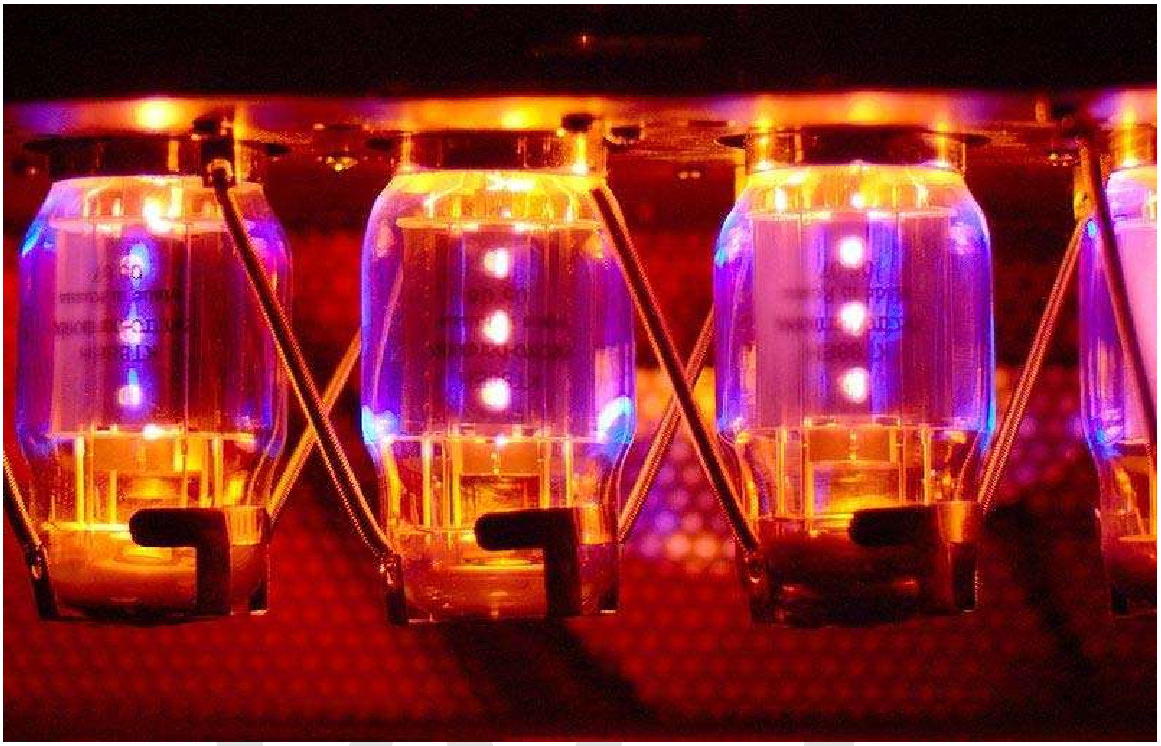
In most cases an amplifier should be linear; that is, the gain should be constant for any combination of input and output signal. If the gain is not constant, e.g., by clipping the output signal at the limits of its capabilities, the output signal will be distorted. There are however cases where variable gain is useful.

There are many types of electronic amplifiers, commonly used in radio and television transmitters and receivers, high-fidelity ("hi-fi") stereo equipment, microcomputers and other electronic digital equipment, and guitar and other instrument amplifiers. Critical components include active devices, such as vacuum tubes or transistors. A brief introduction to the many types of electronic amplifier follows.

Power amplifier

The term "power amplifier" is a relative term with respect to the amount of power delivered to the load and/or sourced by the supply circuit. In general a power amplifier is designated as the last amplifier in a transmission chain (the *output stage*) and is the amplifier stage that typically requires most attention to power efficiency.

Vacuum tube (valve) amplifiers



The glow from four "Electro Harmonix KT88" brand power tubes lights up the inside of a Traynor YBA-200 guitar amplifier

According to Symons, while semiconductor amplifiers have largely displaced valve amplifiers for low power applications, valve amplifiers are much more cost effective in high power applications such as "radar, countermeasures equipment, or communications equipment" (p. 56). Many microwave amplifiers are specially designed valves, such as the klystron, gyrotron, traveling wave tube, and crossed-field amplifier, and these microwave valves provide much greater single-device power output at microwave frequencies than solid-state devices (p. 59).

Valves/tube amplifiers also have niche uses in other areas, such as

- Electric guitar amplification
- in Russian military aircraft, for their EMP tolerance
- niche audio for their sound qualities (recording, and audiophile equipment)

Transistor amplifiers

The essential role of this active element is to magnify an input signal to yield a significantly larger output signal. The amount of magnification (the "forward gain") is determined by the external circuit design as well as the active device.

Many common active devices in transistor amplifiers are bipolar junction transistors (BJTs) and metal oxide semiconductor field-effect transistors (MOSFETs).

Applications are numerous, some common examples are audio amplifiers in a home stereo or PA system, RF high power generation for semiconductor equipment, to RF and Microwave applications such as radio transmitters.

Transistor-based amplifier can be realized using various configurations: for example with a bipolar junction transistor we can realize common base, common collector or common emitter amplifier; using a MOSFET we can realize common gate, common source or common drain amplifier. Each configuration has different characteristic (gain, impedance...).

Operational amplifiers (op-amps)

An operational amplifier is an amplifier circuit with very high open loop gain and differential inputs which employs external feedback for control of its transfer function or gain. Although the term is today commonly applied to integrated circuits, the original operational amplifier design was implemented with valves.

Fully differential amplifiers (FDA)

A fully differential amplifier is a solid state integrated circuit amplifier which employs external feedback for control of its transfer function or gain. It is similar to the operational amplifier but it also has differential output pins.

Video amplifiers

These deal with video signals and have varying bandwidths depending on whether the video signal is for SDTV, EDTV, HDTV 720p or 1080i/p etc.. The specification of the bandwidth itself depends on what kind of filter is used and which point (-1 dB or -3 dB for example) the bandwidth is measured. Certain requirements for step response and overshoot are necessary in order for acceptable TV images to be presented.

Oscilloscope vertical amplifiers

These are used to deal with video signals to drive an oscilloscope display tube and can have bandwidths of about 500 MHz. The specifications on step response, rise time, overshoot and aberrations can make the design of these amplifiers extremely difficult. One of the pioneers in high bandwidth vertical amplifiers was the Tektronix company.

Distributed amplifiers

These use transmission lines to temporally split the signal and amplify each portion separately in order to achieve higher bandwidth than can be obtained from a single amplifying device. The outputs of each stage are combined in the output transmission line. This type of amplifier was commonly used on oscilloscopes as the final vertical amplifier. The transmission lines were often housed inside the display tube glass envelope.

Switched mode amplifiers

These nonlinear amplifiers have much higher efficiencies than linear amps, and are used where the power saving justifies the extra complexity.

Negative resistance devices

Negative resistances can be used as amplifiers, such as the tunnel diode amplifier.

Microwave amplifiers

Travelling wave tube amplifiers

Traveling wave tube amplifiers (TWTAs) are used for high power amplification at low microwave frequencies. They typically can amplify across a broad spectrum of frequencies; however, they are usually not as tunable as klystrons.

Klystrons

Klystrons are vacuum-devices that do not have as wide a bandwidth as TWTAs. They generally are also much heavier than TWTAs, and are therefore ill-suited for light-weight mobile applications. Klystrons are tunable, offering selective output within their specified frequency range.

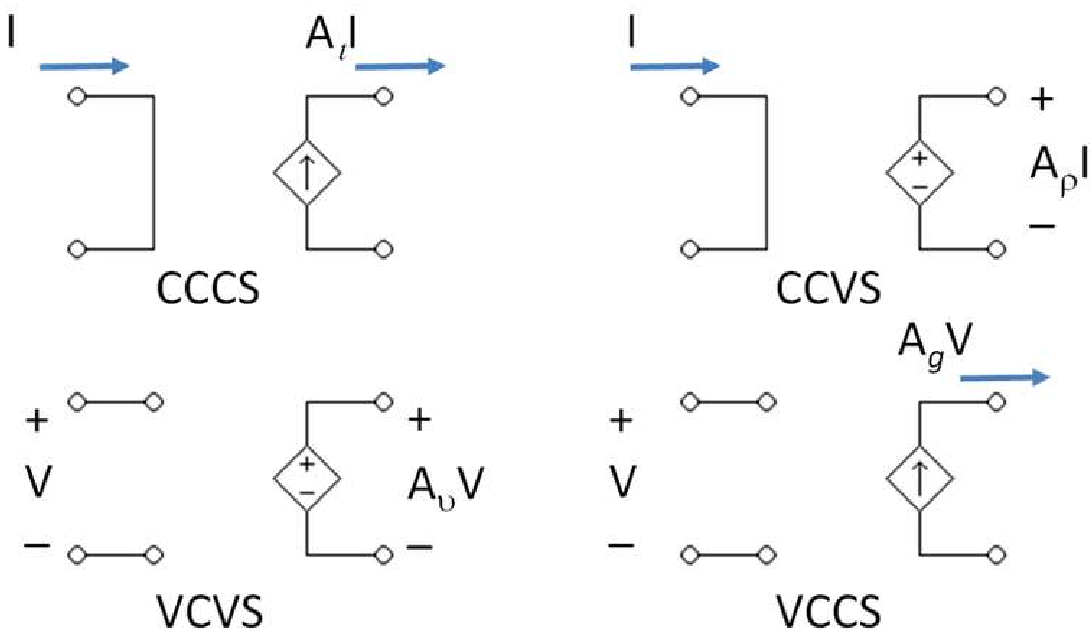
Musical instrument (audio) amplifiers

An audio amplifier is usually used to amplify signals such as music or speech.

Classification of amplifier stages and systems

There are many alternative classifications that address different aspects of amplifier designs, and they all express some particular perspective relating the design parameters to the objectives of the circuit. Amplifier design is always a compromise of numerous factors, such as cost, power consumption, real-world device imperfections, and a multitude of performance specifications. Below are several different approaches to classification:

Input and output variables



The four types of dependent source; control variable on left, output variable on right

Electronic amplifiers use two variables: current and voltage. Either can be used as input, and either as output leading to four types of amplifiers. In idealized form they are represented by each of the four types of dependent source used in linear analysis, as shown in the figure, namely:

Input	Output	Dependent source	Amplifier type
I	**I**	current controlled current source **CCCS**	current amplifier
I	**V**	current controlled voltage source **CCVS**	transresistance amplifier
V	**I**	voltage controlled current source **VCCS**	transconductance amplifier
V	**V**	voltage controlled voltage source **VCVS**	voltage amplifier

Each type of amplifier in its ideal form has an ideal input and output resistance that is the same as that of the corresponding dependent source:

Amplifier type	Dependent source	Input impedance	Output impedance
Current	**CCCS**	**0**	∞
Transresistance	**CCVS**	**0**	**0**
Transconductance	**VCCS**	∞	∞
Voltage	**VCVS**	∞	**0**

In practice the ideal impedances are only approximated. For any particular circuit, a small-signal analysis is often used to find the impedance actually achieved. A small-signal AC test current I_x is applied to the input or output node, all external sources are set to AC zero, and the corresponding alternating voltage V_x across the test current source determines the impedance seen at that node as $R = V_x / I_x$.

Amplifiers designed to attach to a transmission line at input and/or output, especially RF amplifiers, do not fit into this classification approach. Rather than dealing with voltage or current individually, they ideally couple with an input and/or output impedance matched to the transmission line impedance, that is, match *ratios* of voltage to current. Many real RF amplifiers come close to this ideal. Although, for a given appropriate source and load impedance, RF amplifiers can be characterized as amplifying voltage or current, they fundamentally are amplifying power.

Common terminal

One set of classifications for amplifiers is based on which device terminal is common to both the input and the output circuit. In the case of bipolar junction transistors, the three classes are common emitter, common base, and common collector. For field-effect transistors, the corresponding configurations are common source, common gate, and common drain; for triode vacuum devices, *common cathode*, *common grid*, and *common plate*. The output voltage of a common plate amplifier is the same as the input (this arrangement is used as the input presents a high impedance and does not load the signal source, although it does not amplify the voltage), i.e., the output at the cathode follows the input at the grid; consequently it was commonly called a *cathode follower*. By analogy the terms *emitter follower* and *source follower* are sometimes used.

Unilateral or bilateral

When an amplifier has an output that exhibits no feedback to its input side, it is called **unilateral**. The input impedance of a unilateral amplifier is independent of the load, and the output impedance is independent of the signal source impedance.

If feedback connects part of the output back to the input of the amplifier it is called a **bilateral** amplifier. The input impedance of a bilateral amplifier is dependent upon the load, and the output impedance is dependent upon the signal source impedance.

All amplifiers are bilateral to some degree; however they may often be modeled as unilateral under operating conditions where feedback is small enough to neglect for most purposes, simplifying analysis.

Negative feedback is often applied deliberately to tailor amplifier behavior. Some feedback, which may be positive or negative, is unavoidable and often undesirable, introduced, for example, by parasitic elements such as the inherent capacitance between input and output of a device such as a transistor and capacitative coupling due to external wiring. Excessive frequency-dependent positive feedback may cause what is supposed to be an amplifier to become an oscillator.

Linear unilateral and bilateral amplifiers can be represented as two-port networks.

Inverting or non-inverting

Another way to classify amps is the phase relationship of the input signal to the output signal. An **inverting** amplifier produces an output 180 degrees out of phase with the input signal (that is, a polarity inversion or mirror image of the input as seen on an oscilloscope). A **non-inverting** amplifier maintains the phase of the input signal waveforms. An **emitter follower** is a type of non-inverting amplifier, indicating that the signal at the emitter of a transistor is following (that is, matching with unity gain but perhaps an offset) the input signal.

This description can apply to a single stage of an amplifier, or to a complete amplifier system.

Function

Other amplifiers may be classified by their function or output characteristics. These functional descriptions usually apply to complete amplifier systems or sub-systems and rarely to individual stages.

- A **servo amplifier** indicates an integrated feedback loop to actively control the output at some desired level. A **DC servo** indicates use at frequencies down to DC levels, where the rapid fluctuations of an audio or RF signal do not occur. These are often used in mechanical actuators, or devices such as DC motors that must maintain a constant speed or torque. An **AC servo** amp can do this for some ac motors.
- A **linear** amplifier responds to different frequency components independently, and does not generate harmonic distortion or Intermodulation distortion (well, hardly any). A **nonlinear** amplifier does generate distortion (e.g. the output is a current to a lamp that must be either fully on or off, but the input is continuously variable; or the amplifier is used in an analog computer where a special transfer function, such as logarithmic, is desired; or a following tuned circuit will remove the harmonics generated by a non-linear RF amplifier).
- A **wideband** amplifier has a precise amplification factor over a wide range of frequencies, and is often used to boost signals for relay in communications systems. A **narrowband** amp is made to amplify only a specific narrow range of frequencies, to the exclusion of other frequencies.
- An **RF** amplifier refers to an amplifier designed for use in the radio frequency range of the electromagnetic spectrum, and is often used to increase the sensitivity of a receiver or the output power of a transmitter.
- An **audio amplifier** is designed for use in reproducing audio frequencies. This category subdivides into small signal amplification, and power amps which are optimised for driving speakers, sometimes with multiple amps grouped together as separate or bridgeable channels to accommodate different audio reproduction requirements. Frequently used terms within audio amplifiers include:

- preamplifier (preamp), that may include phono or gramophone preamp with equalization for RIAA LP recordings, or tape head preamps with CCIR equalisation filters; they may include filters or tone control circuitry.
    - power amplifier (normally assumed to drive loudspeakers), headphone amplifiers, and public address amplifiers.
    - stereo amplifiers imply two channels of output (left and right), although the term simply means "solid" sound (referring to three-dimensional) - so quadraphonic stereo was used for amplifiers with 4 channels; 5.1 and 7.1 systems refer to Home theatre systems with 5 or 7 normal spacial channels, plus a subwoofer channel (that is not very directional).
- Buffer amplifiers, that may include emitter followers, provide a high impedance input for a device (perhaps another amplifier, or perhaps an energy-hungry load such as lights) that would otherwise draw too much current from the source. Line drivers are a type of buffer intended to feed long or interference-prone interconnect cables, possibly with differential outputs if driving twisted pairs of cables.
- A special type of amplifier is widely used in instruments and for signal processing, among many other varied uses. These are known as **operational amplifiers** or **op-amps**. This is because this type of amplifier is used in circuits that perform mathematical algorithmic functions, or "operations" on input signals to obtain specific types of output signals. A typical modern op-amp has differential inputs (one "inverting", one "non-inverting") and one output. An idealised op-amp has the following characteristics:
    - Infinite input impedance (so as to not load circuitry it is sampling as a control input)
    - Zero output impedance
    - Infinite gain
    - Zero propagation delay

  The performance of an op-amp with these characteristics would be entirely defined by the (usually passive) components forming a negative feedback loop around it, that is, *the amplifier itself has no effect on the output*.
  Today, op-amps are usually provided as integrated circuits, rather than constructed from discrete components. All real-world op-amps fall short of the idealised specification above – but some modern components have remarkable performance and come close in some respects.

Interstage coupling method

Amplifiers are sometimes classified by the coupling method of the signal at the input, output, or between stages. Different types of these include:

Resistive-capacitive (RC) coupled amplifier, using a network of resistors and capacitors
  By design these amplifiers cannot amplify DC signals as the capacitors block the DC component of the input signal. RC-coupled amplifiers were used very often in

circuits with vacuum tubes or discrete transistors. In the days of the integrated circuit a few more transistors on a chip are much cheaper and smaller than a capacitor.

Inductive-capacitive (LC) coupled amplifier, using a network of inductors and capacitors

This kind of amplifier is most often used in selective radio-frequency circuits.

Transformer coupled amplifier, using a transformer to match impedances or to decouple parts of the circuits

Quite often LC-coupled and transformer-coupled amplifiers cannot be distinguished as a transformer is some kind of inductor.

Direct coupled amplifier, using no impedance and bias matching components

This class of amplifier was very uncommon in the vacuum tube days when the anode (output) voltage was at greater than several hundred volts and the grid (input) voltage at a few volts minus. So they were only used if the gain was specified down to DC (e.g., in an oscilloscope). In the context of modern electronics developers are encouraged to use directly coupled amplifiers whenever possible.

Frequency range

Depending on the frequency range and other properties amplifiers are designed according to different principles.

- Frequency ranges down to DC are only used when this property is needed. DC amplification leads to specific complications that are avoided if possible; **DC-blocking** capacitors are added to remove DC and sub-sonic frequencies from audio amplifiers.
- Depending on the frequency range specified different design principles must be used. Up to the MHz range only "discrete" properties need be considered; e.g., a terminal has an input impedance.
- As soon as any connection within the circuit gets longer than perhaps 1% of the wavelength of the highest specified frequency (e.g., at 100 MHz the wavelength is 3 m, so the critical connection length is approx. 3 cm) design properties radically change. For example, a specified length and width of a PCB trace can be used as a selective or impedance-matching entity.
- Above a few 100 MHz, it gets difficult to use discrete elements, especially inductors. In most cases PCB traces of very closely defined shapes are used instead.

The frequency range handled by an amplifier might be specified in terms of bandwidth (normally implying a response that is 3 dB down when the frequency reaches the specified bandwidth), or by specifying a frequency response that is within a certain number of deciBels between a lower and an upper frequency (e.g. "20 Hz to 20 kHz plus or minus 1 dB").

Type of load

- Untuned
    - audio
    - video
- Tuned (RF amps) - used for amplifying a single radio frequency or a band of frequencies

Implementation

Amplifiers are implemented using active elements of different kinds:

- The first active elements were relays. They were for example used in transcontinental telegraph lines: a weak current was used to switch the voltage of a battery to the outgoing line.
- For transmitting audio, carbon microphones were used as the active element. This was used to modulate a radio-frequency source in one of the first AM audio transmissions, by Reginald Fessenden on Dec. 24, 1906.
- In the 1960s, the transistor started to take over. These days, discrete transistors are still used in high-power amplifiers and in specialist audio devices.
- Up to the early 1970s, most amplifiers used vacuum tubes ("valves" in the UK). Today, tubes are used for specialist audio applications such as guitar amplifiers and audiophile amplifiers. Many broadcast transmitters still use vacuum tubes.
- Beginning in the 1970s, more and more transistors were connected on a single chip therefore creating the integrated circuit. Nearly all amplifiers commercially available today are based on integrated circuits.

For exotic purposes, other active elements have been used. For example, in the early days of the communication satellite parametric amplifiers were used. The core circuit was a diode whose capacity was changed by an RF signal created locally. Under certain conditions, this RF signal provided energy that was modulated by the extremely weak satellite signal received at the earth station. The operating principle of a parametric amplifier is somewhat similar to the principle by which children keep their swings in motion: as long as the swing moves you only need to change a parameter of the swinging entity; e.g., you must move your center of gravity up and down. In our case, the capacity of the diode is changed periodically.

Power amplifier classes

Angle of flow or conduction angle

Power amplifier circuits (output stages) are classified as A, B, AB and C for analog designs, and class D and E for switching designs based upon the conduction angle or *angle of flow*, Θ, of the input signal through the (or each) output amplifying device, that is, the portion of the input signal cycle during which the amplifying device conducts. The image of the conduction angle is derived from amplifying a sinusoidal signal. (If the device is always on, $\Theta = 360°$.) The angle of flow is closely related to the amplifier power efficiency. The various classes are introduced below, followed by more detailed discussion under individual headings later on.

Class A

100% of the input signal is used (conduction angle $\Theta = 360°$ or 2π); i.e., the active element remains conducting (works in its "linear" range) all of the time. Where efficiency is not a consideration, most small signal linear amplifiers are designed as class A. Class A amplifiers are typically more linear and less complex than other types, but are very inefficient. This type of amplifier is most commonly used in small-signal stages or for low-power applications (such as driving headphones). Subclass A2 is sometimes used to refer to vacuum tube class A stages where the grid is allowed to be driven slightly positive on signal peaks, resulting in slightly more power than normal class A (A1; where the grid is always negative), but incurring more distortion.

Class B

50% of the input signal is used ($\Theta = 180°$ or π; i.e., the active element works in its linear range half of the time and is more or less turned off for the other half). In most class B, there are two output devices (or sets of output devices), each of which conducts alternately (push–pull) for exactly 180° (or half cycle) of the input signal; selective RF amplifiers can also be implemented using a single active element.

These amplifiers are subject to *crossover distortion* if the transition from one active element to the other is not perfect, as when two complementary transistors (i.e., one PNP, one NPN) are connected as two emitter followers with their base and emitter terminals in common, requiring the base voltage to slew across the region where both devices are turned off.

Class AB

Here the two active elements conduct more than half of the time as a means to reduce the cross-over distortions of class B amplifiers. In the example of the complementary emitter followers a bias network allows for more or less quiescent current thus providing an operating point somewhere between class A and class B. Sometimes a figure is added (e.g., AB_1 or AB_2) for vacuum tube stages where the grid voltage is always negative with respect to the cathode (class AB_1) or may be slightly positive (hence drawing grid current, adding more distortion, but giving slightly higher output power) on signal peaks (class AB_2); another interpretation being higher figures implying a higher quiescent current and therefore more of the properties of class A.

Class C

Less than 50% of the input signal is used (conduction angle $\Theta < 180°$). The advantage is potentially high efficiency, but a disadvantage is high distortion.

Class D

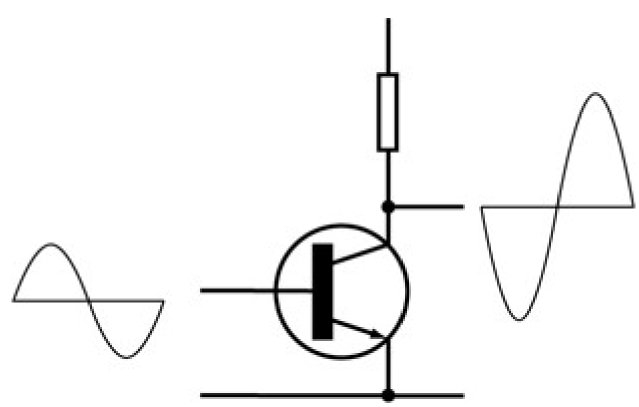
These use switching to achieve a very high power efficiency (more than 90% in modern designs). By allowing each output device to be either fully on or off, losses are minimized. The analog output is created by pulse-width modulation; i.e., the active element is switched on for shorter or longer intervals instead of modifying its resistance. There are more complicated switching schemes like sigma-delta modulation, to improve some performance aspects like lower distortions or better efficiency.

Additional classes

There are several other amplifier classes, although they are mainly variations of the previous classes. For example, class G and class H amplifiers are marked by variation of the supply rails (in discrete steps or in a continuous fashion, respectively) following the input signal. Wasted heat on the output devices can be reduced as excess voltage is kept to a minimum. The amplifier that is fed with these rails itself can be of any class. These kinds of amplifiers are more complex, and are mainly used for specialized applications, such as very high-power units. Also, class E and class F amplifiers are commonly described in literature for radio frequencies applications where efficiency of the traditional classes in are important, yet several aspects not covered elsewhere (e.g.: amplifiers often simply said to have a gain of x dB - so what power gain?) deviate substantially from their ideal values. These classes use harmonic tuning of their output networks to achieve higher efficiency and can be considered a subset of Class C due to their conduction angle characteristics.

The classes can be most easily understood using the diagrams in each section below. For the sake of illustration, a bipolar junction transistor is shown as the amplifying device, but in practice this could be a MOSFET or vacuum tube device. In an analog amplifier (the most common kind), the signal is applied to the input terminal of the device (base, gate or grid), and this causes a proportional output drive current to flow out of the output terminal. The output drive current comes from the power supply.

Class A



Class A amplifier

Amplifying devices operating in class A conduct over the whole of the input cycle such that the output signal is an exact scaled-up replica of the input with no clipping. A *class A amplifier* is distinguished by the *output stage* being biased into class A.

Advantages of class A amplifiers

- Class A designs are simpler than other classes; for example class AB and B designs require two devices (push-pull output) to handle both halves of the waveform; class A can use a single device single-ended.
- The amplifying element is biased so the device is always conducting to some extent, normally implying the quiescent (small-signal) collector current (for transistors; drain current for FETs or anode/plate current for vacuum tubes) is close to the most linear portion of its transconductance curve.
- Because the device is never shut off completely there is no "turn on" time, little problem with charge storage, and generally better high frequency performance and feedback loop stability (and usually fewer high-order harmonics).
- The point at which the device comes closest to being cut off is not close to zero signal, so the problem of crossover distortion associated with class AB and B designs is avoided.

Disadvantage of class A amplifiers

- They are very inefficient; a theoretical maximum of 50% is obtainable with inductive output coupling and only 25% with capacitive coupling, unless Square law output stages are used. In a power amplifier this not only wastes power and limits battery operation, it may place restrictions on the output devices that can be used (for example: ruling out some audio triodes if modern low-efficiency loudspeakers are to be used), and will increase costs. Inefficiency comes not just from the fact that the device is always conducting to some extent (that happens even with class AB, yet its efficiency can be close to that of class B); it is that the standing current is roughly half the maximum output current (although this can be less with Square law output stage), together with the problem that a large part of the power supply voltage is developed across the output device at low signal levels (as with classes AB and B, but unlike output stages such as class D). If high output powers are needed from a class A circuit, the power waste (and the accompanying heat) will become significant. For every watt delivered to the load, the amplifier itself will, *at best*, dissipate another watt. For large powers this means very large and expensive power supplies and heat sinking.

Class A designs have largely been superseded by the more efficient designs for power amplifiers, though they remain popular with some hobbyists, mostly for their simplicity. Also, many audiophiles believe that class A gives the best sound quality (for their absence of crossover distortion and reduced odd-harmonic and high-order harmonic distortion) which provides a small market for expensive **high fidelity** class A amps.

Single-ended and triode class A amplifiers

Some aficionados who prefer class A amplifiers also prefer the use of thermionic valve (or "tube") designs instead of transistors, especially in Single-ended triode output configurations for several claimed reasons:

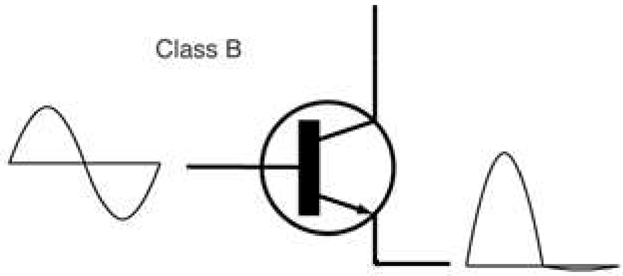
- Single-ended output stages (be they tube or transistor) have an asymmetrical transfer function, meaning that even harmonics in the created distortion tend not to be canceled (as they are in push-pull output stages); by using tubes OR FETs most of the distortion is from the square law transfer characteristic and so second-order, which some consider to be "warmer" and more pleasant.
- For those who prefer low distortion figures, the use of tubes with class A (generating little odd-harmonic distortion, as mentioned above) together with symmetrical circuits (such as push-pull output stages, or balanced low-level stages) results in the cancellation of most of the even distortion harmonics, hence the removal of most of the distortion.
- Though good amplifier design can reduce harmonic distortion patterns to almost nothing, distortion is essential to the sound of electric guitar amplifiers, for example, and is held by recording engineers to offer more flattering microphones and to enhance "clinical-sounding" digital technology.
- Historically, valve amplifiers often used a class A power amplifier simply because valves are large and expensive; many class A designs use only a single device.

Transistors are much cheaper, and so more elaborate designs that give greater efficiency but use more parts are still cost-effective. A classic application for a pair of class A devices is the long-tailed pair, which is exceptionally linear, and forms the basis of many more complex circuits, including many audio amplifiers and almost all op-amps. Class A amplifiers are often used in output stages of high quality op-amps (although the accuracy of the bias in low cost op-amps such as the **741** may result in class A or class AB or class B, varying from device to device or with temperature). They are sometimes used as medium-power, low-efficiency, and high-cost audio amplifiers. The power consumption is unrelated to the output power. At idle (no input), the power consumption is essentially the same as at high output volume. The result is low efficiency and high heat dissipation.

Class B and AB

Class B or AB push–pull circuits are the most common design type found in audio power amplifiers. Class AB is widely considered a good compromise for audio amplifiers, since much of the time the music is quiet enough that the signal stays in the "class A" region, where it is amplified with good fidelity, and by definition if passing out of this region, is large enough that the distortion products typical of class B are relatively small. The crossover distortion can be reduced further by using negative feedback. Class B and AB amplifiers are sometimes used for RF linear amplifiers as well. Class B amplifiers are also favored in battery-operated devices, such as transistor radios.

Class B



Class B amplifier

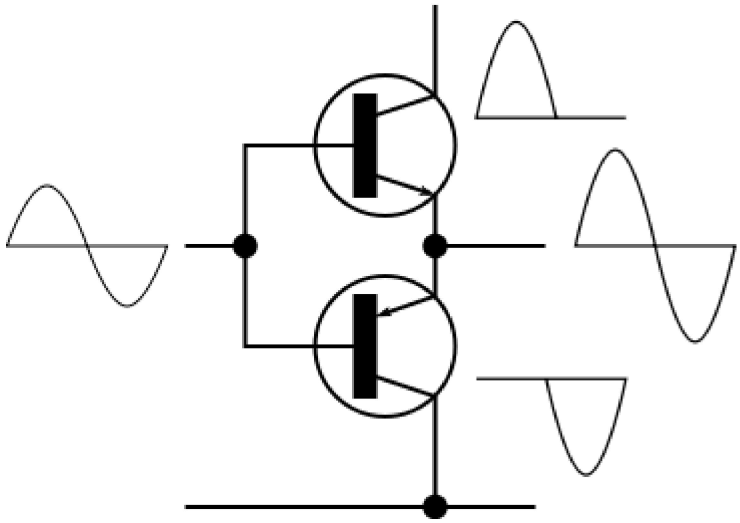
Class B amplifiers only amplify half of the input wave cycle, thus creating a large amount of distortion, but their efficiency is greatly improved and is much better than class A. Class B has a maximum theoretical efficiency of 78.5% (i.e., π/4). This is because the amplifying element is switched off altogether half of the time, and so cannot dissipate power. A single class B element is rarely found in practice, though it has been used for driving the loudspeaker in the early IBM Personal Computers with beeps, and it can be used in RF power amplifier where the distortion levels are less important. However, class C is more commonly used for this.

A practical circuit using class B elements is the push-pull stage, such as the very simplified complementary pair arrangement shown below. Here, complementary or quasi-complementary devices are each used for amplifying the opposite halves of the input signal, which is then recombined at the output. This arrangement gives excellent efficiency, but can suffer from the drawback that there is a small mismatch in the cross-over region - at the "joins" between the two halves of the signal, as one output device has to take over supplying power exactly as the other finishes. This is called crossover distortion. An improvement is to bias the devices so they are not completely off when they're not in use. This approach is called *class AB* operation.

Digital class B

A limited power output class B amplifier with a single-ended supply rail of 5±0.5 V.

Class AB

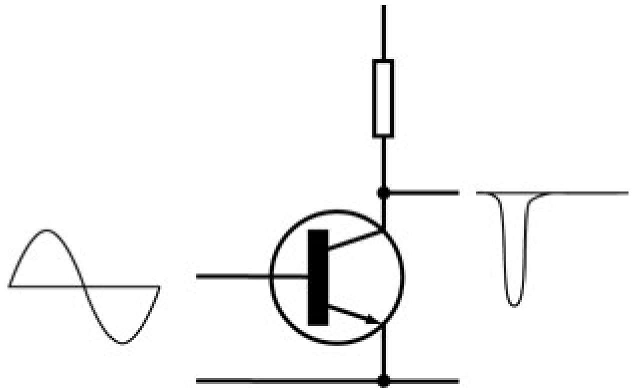


Class B push–pull amplifier

In class AB operation, each device operates the same way as in class B over half the waveform, but also conducts a small amount on the other half. As a result, the region where both devices simultaneously are nearly off (the "dead zone") is reduced. The result is that when the waveforms from the two devices are combined, the crossover is greatly minimised or eliminated altogether. The exact choice of **quiescent current**, the standing current through both devices when there is no signal, makes a large difference to the level of distortion (and to the risk of thermal runaway, that may damage the devices); often the bias voltage applied to set this quiescent current has to be adjusted with the temperature of the output transistors (for example in the circuit at the beginning the diodes would be mounted physically close to the output transistors, and chosen to have a matched temperature coefficient). Another approach (often used as well as thermally-tracking bias voltages) is to include small value resistors in series with the emitters.

Class AB sacrifices some efficiency over class B in favor of linearity, thus is less efficient (below 78.5% for full-amplitude sinewaves in transistor amplifiers, typically; much less is common in class AB vacuum tube amplifiers). It is typically much more efficient than class A.

Class C



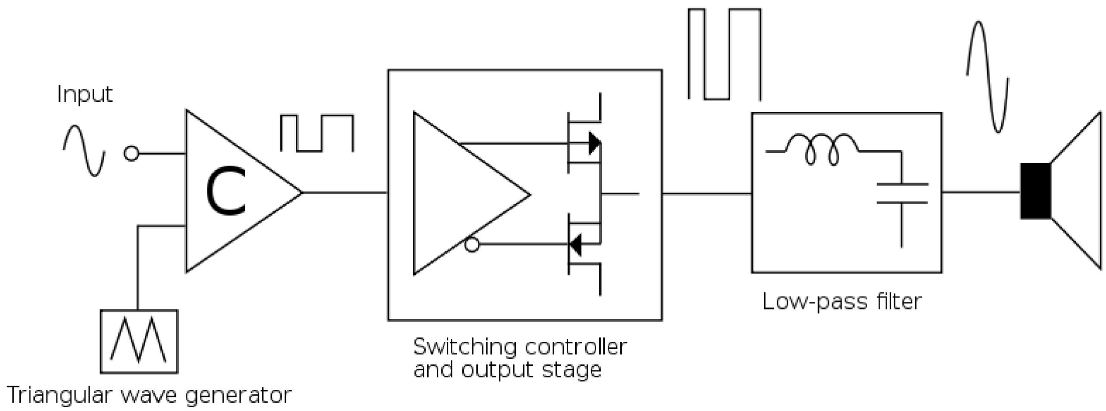
Class C amplifier

Class C amplifiers conduct less than 50% of the input signal and the distortion at the output is high, but high efficiencies (up to 90%) are possible. Some applications (for example, megaphones) can tolerate the distortion. A much more common application for class C amplifiers is in RF transmitters, where the distortion can be vastly reduced by using tuned loads on the amplifier stage. The input signal is used to roughly switch the amplifying device on and off, which causes pulses of current to flow through a tuned circuit.

The class C amplifier has two modes of operation: tuned and untuned. The diagram shows a waveform from a simple class C circuit without the tuned load. This is called untuned operation, and the analysis of the waveforms shows the massive distortion that appears in the signal. When the proper load (e.g., a pure inductive-capacitive filter) is used, two things happen. The first is that the output's bias level is clamped, so that the output variation is centered at one-half of the supply voltage. This is why tuned operation is sometimes called a *clamper*. This action of elevating bias level allows the waveform to be restored to its proper shape, allowing a complete waveform to be re-established despite having only a one-polarity supply. This is directly related to the second phenomenon: the waveform on the center frequency becomes much less distorted. The distortion that is present is dependent upon the bandwidth of the tuned load, with the center frequency seeing very little distortion, but greater attenuation the farther from the tuned frequency that the signal gets.

The tuned circuit will only resonate at particular frequencies, and so the unwanted frequencies are dramatically suppressed, and the wanted full signal (sine wave) will be extracted by the tuned load (e.g., a high-quality bell will ring at a particular frequency when it is hit periodically with a hammer). Provided the transmitter is not required to operate over a very wide band of frequencies, this arrangement works extremely well. Other residual harmonics can be removed using a filter.

Class D



Block diagram of a basic switching or PWM (class D) amplifier.

Boss Audio class D mono car audio amplifier with a low pass filter for powering subwoofers

In the class D amplifier the input signal is converted to a sequence of higher voltage output pulses. The averaged-over-time power values of these pulses are directly proportional to the instantaneous amplitude of the input signal. The frequency of the output pulses is typically ten or more times the highest frequency in the input signal to be amplified. The output pulses contain inaccurate spectral components (that is, the pulse frequency and its harmonics) which must be removed by a low-pass passive filter. The resulting filtered signal is then an amplified replica of the input.

These amplifiers use pulse width modulation, pulse density modulation (sometimes referred to as pulse frequency modulation) or more advanced form of modulation such as Delta-sigma modulation (for example, in the Analog Devices AD1990 class D audio power amplifier). Output stages such as those used in pulse generators are examples of class D amplifiers. The term *class D* is usually applied to devices intended to reproduce signals with a bandwidth well below the switching frequency.

Class D amplifiers can be controlled by either analog or digital circuits. The digital control introduces additional distortion called *quantization error* caused by its conversion of the input signal to a digital value.

The main advantage of a class D amplifier is power efficiency. Because the output pulses have a fixed amplitude, the switching elements (usually MOSFETs, but valves and bipolar transistors were once used) are switched either completely on or completely off, rather than operated in linear mode. A MOSFET operates with the lowest resistance when fully-on and thus has the lowest power dissipation when in that condition, except when fully off. When operated in a linear mode the MOSFET has variable amounts of resistance that vary linearly with the input voltage and the resistance is something other than the minimum possible, therefore more electrical energy is dissipated as heat. Compared to class A/B operation, class D's lower losses permit the use of a smaller heat sink for the MOSFETS while also reducing the amount of AC power supply power required. Thus, class D amplifiers do not need as large or as heavy power supply transformers or heatsinks, so they are smaller and more compact in size than an equivalent class AB amplifier.

Class D amplifiers have been widely used to control motors, and almost exclusively for small DC motors, but they are now also used as audio amplifiers, with some extra circuitry to allow analogue to be converted to a much higher frequency pulse width modulated signal. The relative difficulty of achieving good audio quality means that nearly all are used in applications where quality is not a factor, such as modestly-priced bookshelf audio systems and "DVD-receivers" in mid-price home theater systems.

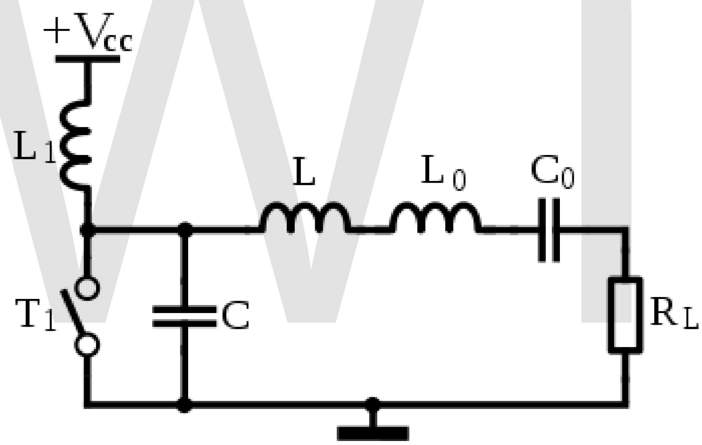
High quality class D audio amplifiers have now appeared in the market and these revised designs have been said to rival good traditional AB amplifiers in terms of quality. Before these higher quality designs existed an earlier use of class D amplifiers and prolific area of application was high-powered, subwoofer amplifiers in cars. Because subwoofers are generally limited to a bandwidth of no higher than 150 Hz, the switching speed for the amplifier does not have to be as high as for a full range amplifier. Class D amplifiers for driving subwoofers are relatively inexpensive, in comparison to class AB amplifiers.

The letter *D* used to designate this amplifier class is simply the next letter after *C*, and does not stand for *digital*. Class D and class E amplifiers are sometimes mistakenly described as "digital" because the output waveform superficially resembles a pulse-train of digital symbols, but a class D amplifier merely converts an input waveform into a continuously pulse-width modulated (square wave) analog signal. (A digital waveform would be pulse-code modulated.)

Additional classes

Class E

The class E/F amplifier is a highly efficient switching power amplifier, typically used at such high frequencies that the switching time becomes comparable to the duty time. As said in the class D amplifier, the transistor is connected via a serial LC circuit to the load, and connected via a large L (inductor) to the supply voltage. The supply voltage is connected to ground via a large capacitor to prevent any RF signals leaking into the supply. The class E amplifier adds a C (capacitor) between the transistor and ground and uses a defined L_1 to connect to the supply voltage.



Class E amplifier

The following description ignores DC, which can be added easily afterwards. The above mentioned C and L are in effect a parallel LC circuit to ground. When the transistor is on, it pushes through the serial LC circuit into the load and some current begins to flow to the parallel LC circuit to ground. Then the serial LC circuit swings back and compensates the current into the parallel LC circuit. At this point the current through the transistor is zero and it is switched off. Both LC circuits are now filled with energy in C and L_0. The whole circuit performs a damped oscillation. The damping by the load has been adjusted so that some time later the energy from the Ls is gone into the load, but the energy in both C_0 peaks at the original value to in turn restore the original voltage so that the voltage across the transistor is zero again and it can be switched on.

With load, frequency, and duty cycle (0.5) as given parameters and the constraint that the voltage is not only restored, but peaks at the original voltage, the four parameters (L, L_0, C and C_0) are determined. The class E amplifier takes the finite on resistance into account and tries to make the current touch the bottom at zero. This means that the voltage and the current at the transistor are symmetric with respect to time. The Fourier transform allows an elegant formulation to generate the complicated LC networks and says that the first harmonic is passed into the load, all even harmonics are shorted and all higher odd harmonics are open.

Class E uses a significant amount of second-harmonic voltage. The second harmonic can be used to reduce the overlap with edges with finite sharpness. For this to work, energy on the second harmonic has to flow from the load into the transistor, and no source for this is visible in the circuit diagram. In reality, the impedance is mostly reactive and the only reason for it is that class E is a class F amplifier with a much simplified load network and thus has to deal with imperfections.

In many amateur simulations of class E amplifiers, sharp current edges are assumed nullifying the very motivation for class E and measurements near the transit frequency of the transistors show very symmetric curves, which look much similar to class F simulations.

The class E amplifier was invented in 1972 by Nathan O. Sokal and Alan D. Sokal, and details were first published in 1975. Some earlier reports on this operating class have been published in Russian.

Class F

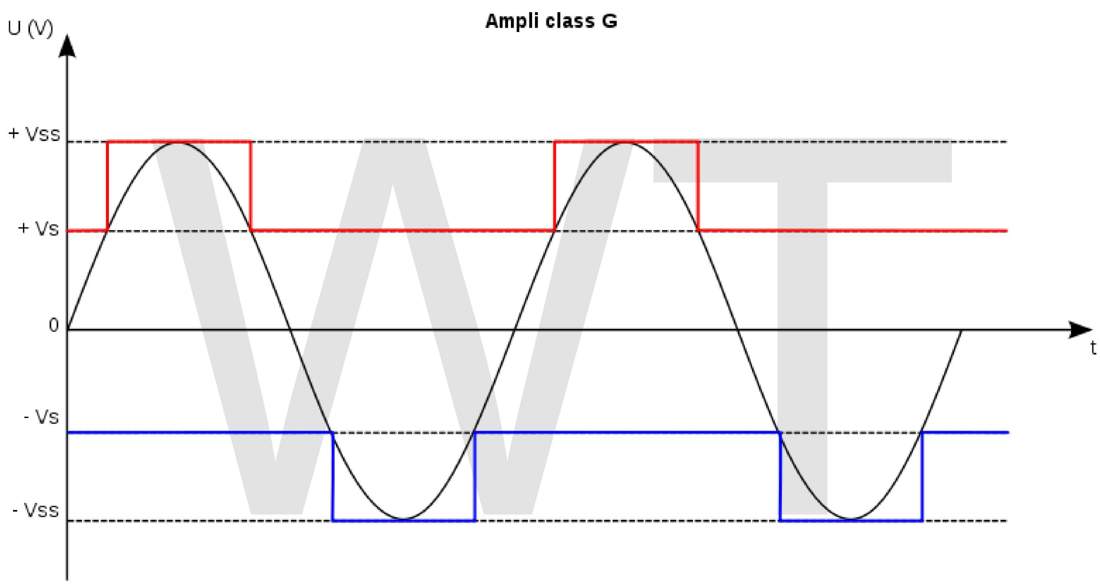
In push-pull amplifiers and in CMOS, the even harmonics of both transistors just cancel. Experiment shows that a square wave can be generated by those amplifiers and theory shows that square waves do consist of odd harmonics only. In a class D amplifier, the output filter blocks all harmonics; i.e., the harmonics see an open load. So even small currents in the harmonics suffice to generate a voltage square wave. The current is in phase with the voltage applied to the filter, but the voltage across the transistors is out of phase. Therefore, there is a minimal overlap between current through the transistors and voltage across the transistors. The sharper the edges, the lower the overlap.

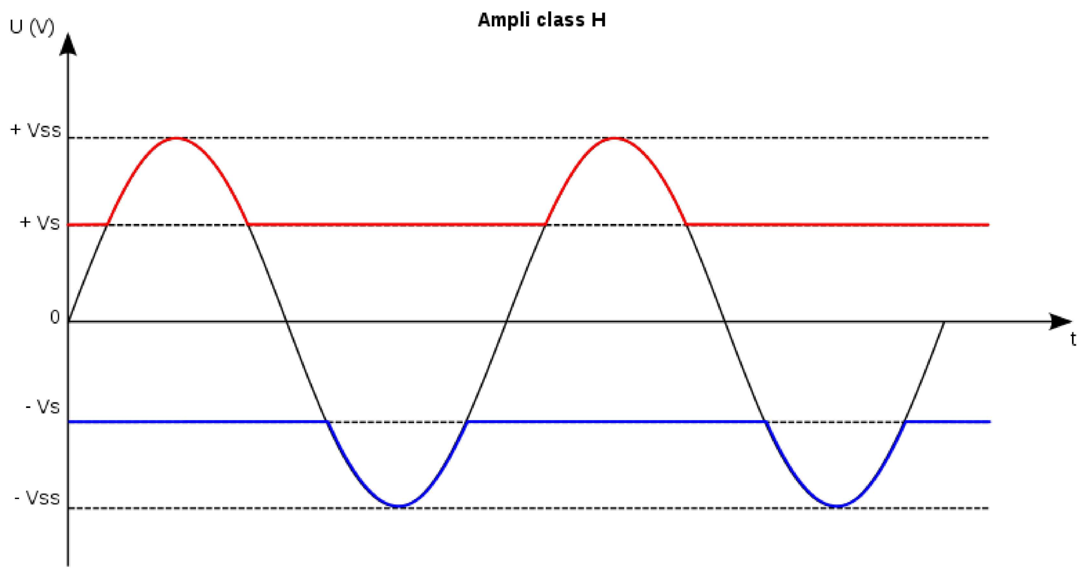
While class D sees the transistors and the load as two separate modules, class F admits imperfections like the parasitics of the transistor and tries to optimise the global system to have a high impedance at the harmonics. Of course there has to be a finite voltage across the transistor to push the current across the on-state resistance. Because the combined current through both transistors is mostly in the first harmonic, it looks like a sine. That means that in the middle of the square the maximum of current has to flow, so it may make sense to have a dip in the square or in other words to allow some overswing of the voltage square wave. A class F load network by definition has to transmit below a cutoff frequency and reflect above.

Any frequency lying below the cutoff and having its second harmonic above the cutoff can be amplified, that is an octave bandwidth. On the other hand, an inductive-capacitive series circuit with a large inductance and a tunable capacitance may be simpler to implement. By reducing the duty cycle below 0.5, the output amplitude can be modulated. The voltage square waveform will degrade, but any overheating is compensated by the lower overall power flowing. Any load mismatch behind the filter can only act on the first harmonic current waveform, clearly only a purely resistive load makes sense, then the lower the resistance, the higher the current.

Class F can be driven by sine or by a square wave, for a sine the input can be tuned by an inductor to increase gain. If class F is implemented with a single transistor, the filter is complicated to short the even harmonics. All previous designs use sharp edges to minimise the overlap.

Classes G and H





There are a variety of amplifier designs that enhance class AB output stages with more efficient techniques to achieve greater efficiencies with low distortion. These designs are common in large audio amplifiers since the heatsinks and power transformers would be prohibitively large (and costly) without the efficiency increases. The terms "class G" and "class H" are used interchangeably to refer to different designs, varying in definition from one manufacturer or paper to another.

Class G amplifiers (which use "rail switching" to decrease power consumption and increase efficiency) are more efficient than class AB amplifiers. These amplifiers provide several power rails at different voltages and switch between them as the signal output approaches each level. Thus, the amplifier increases efficiency by reducing the wasted power at the output transistors. Class G amplifiers are more efficient than class AB but less efficient when compared to class D, without the negative EMI effects of class D.

Class H amplifiers take the idea of class G one step further creating an infinitely variable supply rail. This is done by modulating the supply rails so that the rails are only a few volts larger than the output signal at any given time. The output stage operates at its maximum efficiency all the time. Switched-mode power supplies can be used to create the tracking rails. Significant efficiency gains can be achieved but with the drawback of more complicated supply design and reduced THD performance.

The voltage signal shown is thus a larger version of the input, but has been changed in sign (inverted) by the amplification. Other arrangements of amplifying device are possible, but that given (that is, common emitter, common source or common cathode) is the easiest to understand and employ in practice. If the amplifying element is linear, then the output will be faithful copy of the input, only larger and inverted. In practice, transistors are not linear, and the output will only approximate the input. Non-linearity from any of several sources is the origin of distortion within an amplifier. Which class of amplifier (A, B, AB or C) depends on how the amplifying device is biased — in the diagrams the bias circuits are omitted for clarity.

Any real amplifier is an imperfect realization of an ideal amplifier. One important limitation of a real amplifier is that the output it can generate is ultimately limited by the power available from the power supply. An amplifier will saturate and clip the output if the input signal becomes too large for the amplifier to reproduce or if operational limits for a device are exceeded.

For additional information on class H: Efficiency Class H

Doherty amplifiers

A hybrid configuration receiving new attention is the Doherty amplifier, invented in 1934 by William H. Doherty for Bell Laboratories (whose sister company, Western Electric, was then an important manufacturer of radio transmitters). The Doherty amplifier consists of a class B *primary* or *carrier* stage in parallel with a class C *auxiliary* or *peak* stage. The input signal is split to drive the two amplifiers and a combining network sums the two output signals. Phase shifting networks are employed in the inputs and the outputs. During periods of low signal level, the class B amplifier efficiently operates on the signal and the class C amplifier is cutoff and consumes little power. During periods of high signal level, the class B amplifier delivers its maximum power and the class C amplifier delivers up to its maximum power. The efficiency of previous AM transmitter designs was proportional to modulation but, with average modulation typically around 20%, transmitters were limited to less than 50% efficiency. In Doherty's design, even with zero modulation, a transmitter could achieve at least 60% efficiency.

As a successor to Western Electric for broadcast transmitters, the Doherty concept was considerably refined by Continental Electronics Manufacturing Company of Dallas, TX. Perhaps, the ultimate refinement was the screen-grid modulation scheme invented by Joseph B. Sainton. The Sainton amplifier consists of a class C primary or carrier stage in parallel with a class C auxiliary or peak stage. The stages are split and combined through 90-degree phase shifting networks as in the Doherty amplifier. The unmodulated radio frequency carrier is applied to the control grids of both tubes. Carrier modulation is applied to the screen grids of both tubes. The bias point of the carrier and peak tubes is different, and is established such that the peak tube is cutoff when modulation is absent (and the amplifier is producing rated unmodulated carrier power) whereas both tubes contribute twice the rated carrier power during 100% modulation (as four times the carrier power is required to achieve 100% modulation). As both tubes operate in class C, a significant improvement in efficiency is thereby achieved in the final stage. In addition, as the tetrode carrier and peak tubes require very little drive power, a significant improvement in efficiency within the driver stage is achieved as well (317C, et al.). The released version of the Sainton amplifier employs a cathode-follower modulator, not a push-pull modulator. Previous Continental Electronics designs, by James O. Weldon and others, retained most of the characteristics of the Doherty amplifier but added screen-grid modulation of the driver (317B, et al.).

The Doherty amplifier remains in use in very-high-power AM transmitters, but for lower-power AM transmitters, vacuum-tube amplifiers in general were eclipsed in the 1980s by arrays of solid-state amplifiers, which could be switched on and off with much finer granularity in response to the requirements of the input audio. However, interest in the Doherty configuration has been revived by cellular-telephone and wireless-Internet applications where the sum of several constant-envelope users creates an aggregate AM result. The main challenge of the Doherty amplifier for digital transmission modes is in aligning the two stages and getting the class-C amplifier to turn on and off very quickly.

Recently, Doherty amplifiers have found widespread use in cellular base station transmitters for GHz frequencies. Implementations for transmitters in mobile devices have also been demonstrated.

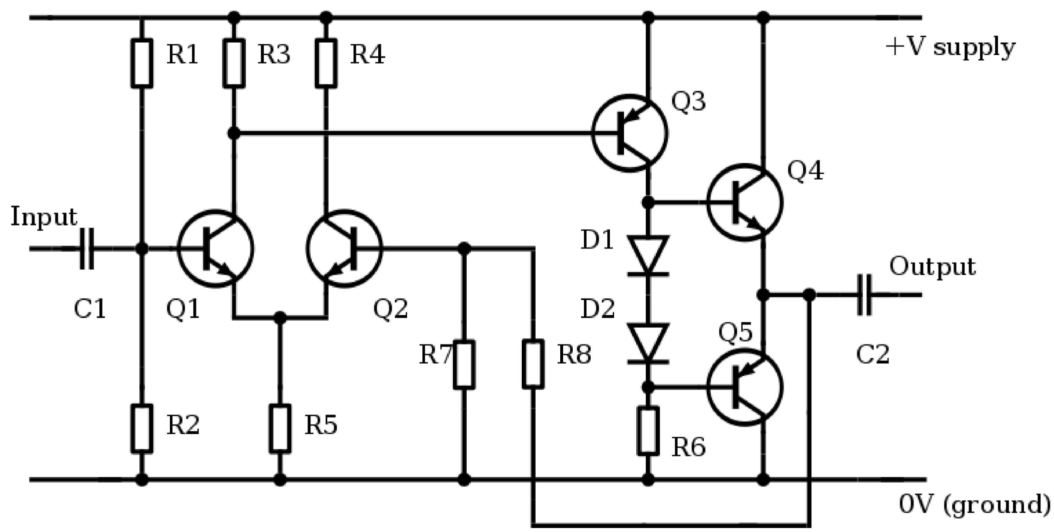
Special classes

Several audio amplifier manufacturers have started "inventing" new classes as a way to differentiate themselves. These class names usually do not reflect any revolutionary amplification technique, and are used mostly for marketing purposes. This can easily be determined by the fact that the class name is trademarked or copyrighted. For example, Crown's K and I-Tech Series as well as several other models utilise Crown's patented class I (or BCA) technology. Lab.gruppen use a form of class D amplifier called class TD or tracked class D which tracks the waveform to more accurately amplify it without the drawbacks of traditional class D amplifiers.

"Class T" was a trademark of TriPath company which manufactures audio amplifier ICs. This new class T is a revision of the common class D amplifier, but with changes to ensure fidelity over the full audio spectrum, unlike traditional class D designs. It operates at different frequencies depending on the power output, with values ranging from as low as 200 kHz to 1.2 MHz, using a proprietary modulator. Tripath ceased operations in 2007, its patents acquired by Cirrus Logic for their Mixed-Signal Audio division. Some Kenwood Recorder use Class W amplifier

"Class Z" is a trademark of Zetex Semiconductors (now part of Diodes Inc. of Dallas, TX) and is a direct-digital-feedback technology. Zetex-patented circuits are being utilised in the latest power amplifiers by NAD Electronics of Canada.

Amplifier circuit



The practical amplifier circuit to the right could be the basis for a moderate-power audio amplifier. It features a typical (though substantially simplified) design as found in modern amplifiers, with a class AB push–pull output stage, and uses some overall negative feedback. Bipolar transistors are shown, but this design would also be realizable with FETs or valves.

The input signal is coupled through capacitor C1 to the base of transistor Q1. The capacitor allows the AC signal to pass, but blocks the DC bias voltage established by resistors R1 and R2 so that any preceding circuit is not affected by it. Q1 and Q2 form a differential amplifier (an amplifier that multiplies the difference between two inputs by some constant), in an arrangement known as a long-tailed pair. This arrangement is used to conveniently allow the use of negative feedback, which is fed from the output to Q2 via R7 and R8.

The negative feedback into the difference amplifier allows the amplifier to compare the input to the actual output. The amplified signal from Q1 is directly fed to the second stage, Q3, which is a common emitter stage that provides further amplification of the signal and the DC bias for the output stages, Q4 and Q5. R6 provides the load for Q3 (A better design would probably use some form of active load here, such as a constant-current sink). So far, all of the amplifier is operating in class A. The output pair are arranged in class AB push–pull, also called a complementary pair. They provide the majority of the current amplification (while consuming low quiescent current) and directly drive the load, connected via DC-blocking capacitor C2. The diodes D1 and D2 provide a small amount of constant voltage bias for the output pair, just biasing them into the conducting state so that crossover distortion is minimized. That is, the diodes push the output stage firmly into class-AB mode (assuming that the base-emitter drop of the output transistors is reduced by heat dissipation).

This design is simple, but a good basis for a practical design because it automatically stabilises its operating point, since feedback internally operates from DC up through the audio range and beyond. Further circuit elements would probably be found in a real design that would roll off the frequency response above the needed range to prevent the possibility of unwanted oscillation. Also, the use of fixed diode bias as shown here can cause problems if the diodes are not both electrically and thermally matched to the output transistors — if the output transistors turn on too much, they can easily overheat and destroy themselves, as the full current from the power supply is not limited at this stage.

A common solution to help stabilise the output devices is to include some emitter resistors, typically an ohm or so. Calculating the values of the circuit's resistors and capacitors is done based on the components employed and the intended use of the amp.

Notes on implementation

Real world amplifiers are imperfect.

- One consequence is that the power supply itself may influence the output, and must itself be considered when designing the amplifier
- The amplifier circuit has an "open loop" performance, that can be described as various parameters (gain, slew rate, output impedance, distortion, bandwidth, signal to noise ratio, etc.)
- Many modern amplifiers use negative feedback techniques to hold the gain at the desired value.

Different methods of supplying power result in many different methods of bias. Bias is a technique by which the active devices are set up to operate in a particular regime, or by which the DC component of the output signal is set to the midpoint between the maximum voltages available from the power supply. Most amplifiers use several devices at each stage; they are typically matched in specifications except for polarity. Matched inverted polarity devices are called complementary pairs. Class A amplifiers generally use only one device, unless the power supply is set to provide both positive and negative voltages, in which case a dual device symmetrical design may be used. Class C amplifiers, by definition, use a single polarity supply.

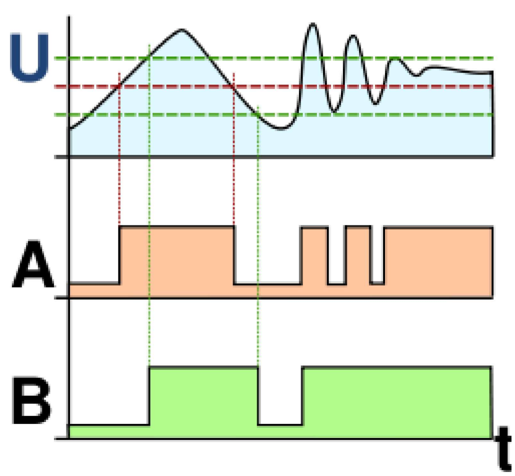
Amplifiers often have multiple stages in cascade to increase gain. Each stage of these designs may be a different type of amp to suit the needs of that stage. For instance, the first stage might be a class A stage, feeding a class AB push–pull second stage, which then drives a class G final output stage, taking advantage of the strengths of each type, while minimizing their weaknesses.

Chapter- 11

Schmitt Trigger

In electronics, **Schmitt trigger** is a generic name of *threshold circuits* with positive feedback having a loop gain > 1. The circuit is named "trigger" because the output retains its value until the input changes sufficiently to trigger a change: in the non-inverting configuration, when the input is higher than a certain chosen threshold, the output is high; when the input is below a different (lower) chosen threshold, the output is low; when the input is between the two, the output retains its value. This dual threshold action is called *hysteresis* and implies that the Schmitt trigger possess memory and can act as a bistable circuit (latch). There is a close relation between the two kinds of circuits that actually are the same: a Schmitt trigger can be converted into a latch and v.v., a latch can be converted into a Schmitt trigger.

Schmitt trigger devices are typically used in open loop configurations for noise immunity and closed loop negative feedback configurations to implement bistable regulators, triangle/square wave generators, etc.



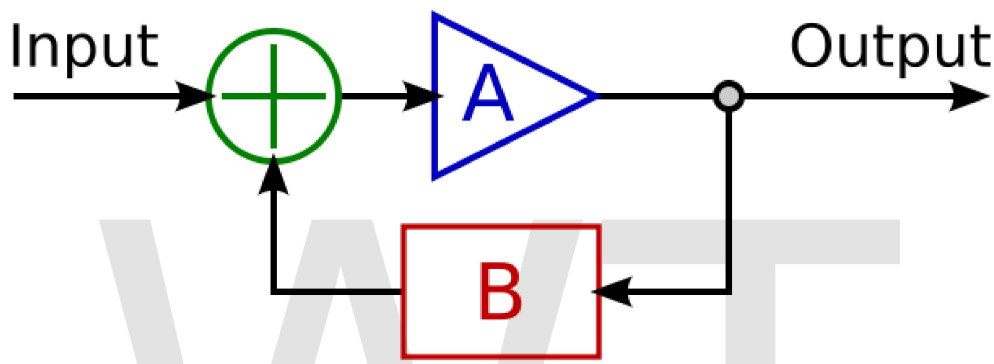
The effect of using a Schmitt trigger (B) instead of a comparator (A).

Invention

The Schmitt trigger was invented by US scientist Otto H. Schmitt in 1934 while he was still a graduate student, later described in his doctoral dissertation (1937) as a "thermionic trigger". It was a direct result of Schmitt's study of the neural impulse propagation in squid nerves.

Implementation

Fundamental idea



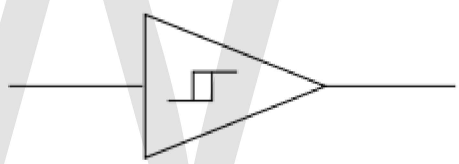
Schmitt trigger is a system with an avalanche-like positive feedback (B < 1; B.A > 1), in which the output helps the input

Circuits with hysteresis are based on the fundamental positive feedback idea: *any active circuit can be made to behave as a Schmitt trigger by applying a positive feedback so that the loop gain is more than one*. The positive feedback is introduced by adding a part of the output voltage to the input voltage; so, these circuits contain an *attenuator* (the B box in the figure on the right) and a *summer* (the circle with "+" inside) in addition to an amplifier acting as a comparator. There are three specific techniques for implementing this general idea. The first two of them are dual versions (series and parallel) of the general positive feedback system. In these configurations, the output voltage increases the effective difference input voltage of the comparator by *decreasing the threshold* or by *increasing the circuit input voltage*; the threshold and memory properties are incorporated in one element. In the third technique, the threshold and memory properties are separated.

Dynamic threshold (series feedback): *when the input voltage crosses the threshold in some direction the very circuit changes its own threshold to the opposite direction*. For this purpose, it subtracts a part of its output voltage from the threshold (it is equal to adding voltage to the input voltage). Thus the output affects the threshold and does not impact on the input voltage. These circuits are implemented by a differential amplifier with *series positive feedback* where the input is connected to the inverting input and the output - to the non-inverting input. In this arrangement, attenuation and summation are separated: a voltage divider acts as an attenuator and the loop acts as a simple series voltage summer. Examples: the classic transistor emitter-coupled Schmitt trigger, op-amp inverting Schmitt trigger, etc.

Modified input voltage (parallel feedback): *when the input voltage crosses the threshold in some direction the circuit changes the very input voltage in the same direction* (now it adds a part of its output voltage directly to the input voltage). Thus the output "helps" the input voltage and does not affect the threshold. These circuits can be implemented by a single-ended non-inverting amplifier with *parallel positive feedback* where the input and the output sources are connected through resistors to the input. The two resistors form a weighted parallel summer incorporating both the attenuation and summation. Examples: the less familiar collector-base coupled Schmitt trigger, op-amp non-inverting Schmitt trigger, etc.

Two different unidirectional thresholds *are assigned in this case to two separate open-loop comparators (without hysteresis) driving an RS trigger (2-input memory cell)*. The trigger is toggled high when the input voltage crosses down to up the high threshold and low when the input voltage crosses up to down the low threshold. Again, there is a positive feedback but now it is concentrated only in the memory cell. Example: 555 timer, switch debounce circuit.



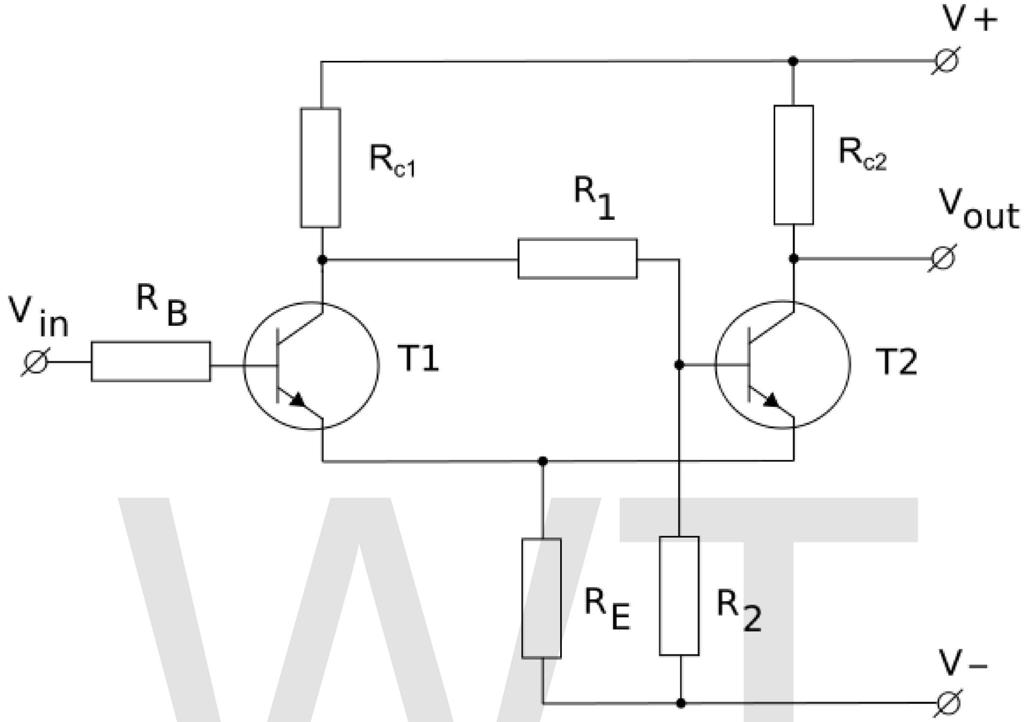
The symbol of Schmitt trigger

Some circuits and elements exhibiting negative resistance can also act as Schmitt triggers: negative impedance converters (NIC), neon lamps, tunnel diodes (e.g., a diode with an "N"-shaped current–voltage characteristic in the first quadrant), etc. In the last case, an oscillating input will cause the diode to move from one rising leg of the "N" to the other and back again as the input crosses the rising and falling switching thresholds.

The symbol for Schmitt triggers in circuit diagrams is a triangle with a symbol inside representing its ideal hysteresis curve.

Transistor Schmitt triggers

Classic emitter-coupled circuit



Schmitt trigger implemented by two emitter-coupled transistor stages

The original Schmitt trigger is based on the dynamic threshold idea that is implemented by a voltage divider with a switchable upper leg (the collector resistors R_{c1} and R_{c2}) and a steady lower leg (R_E). T1 acts as a comparator with a differential input (T1 base-emitter junction) consisting of an inverting (T1 base) and a non-inverting (T1 emitter) inputs. The input voltage is applied to the inverting input; the output voltage of the voltage divider is applied to the non-inverting input thus determining its threshold. The comparator output drives the second common collector stage T2 (an *emitter follower*) through the voltage follower R_1-R_2. The emitter-coupled transistors T1 and T2 actually compose an electronic double throw switch that switches over the upper legs of the voltage divider and changes the threshold in a different (to the input voltage) direction.

This configuration can be considered as a differential amplifier with series positive feedback between its non-inverting input (T2 base) and output (T1 collector) that forces the transition process. There is also a smaller negative feedback introduced by the emitter resistor R_E. To make the positive feedback dominate over the negative one and to obtain a hysteresis, the proportion between the two collector resistors is chosen $R_{c1} > R_{c2}$. Thus less current flows through and less voltage drop is across R_E when T1 is switched on than in the case when T2 is switched on. As a result, the circuit has two different thresholds in regard to ground (V_- in the picture).

Operation

Initial state. For NPN transistors as shown, imagine the input voltage is below the shared emitter voltage (high threshold for concreteness) so that T1 base-emitter junction is backward-biased and T1 does not conduct. T2 base voltage is determined by the mentioned divider so that T2 is conducting and the trigger output is in the low state. The two resistors R_{c2} and R_E form another voltage divider that determines the high threshold. Neglecting V_{BE}, the high threshold value is approximately

$$V_{HT} = \frac{R_E}{R_E + R_{c2}} V_+ .$$

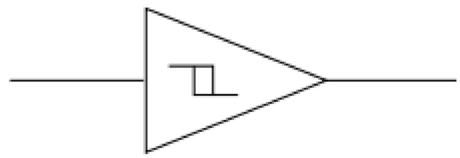
The output voltage is low but well above the ground. It is approximately equal to the high threshold and may not be low enough to be a logical zero for next digital circuits. This may require additional shifting circuit following the trigger circuit.

Crossing up the high threshold. When the input voltage (T1 base voltage) rises slightly above the voltage across the emitter resistor R_E (the high threshold), T1 begins conducting. Its collector voltage goes down and T2 begins going cut-off, because the voltage divider now provides lower T2 base voltage. The common emitter voltage follows this change and goes down thus making T1 conduct more. The current begins steering from the right leg of the circuit to the left one. Although T1 is more conducting, it passes less current through R_E (since $R_{c1} > R_{c2}$); the emitter voltage continues dropping and the effective T1 base-emitter voltage continuously increases. This avalanche-like process continues until T1 becomes completely turned on (saturated) and T2 turned off. The trigger is transitioned to the high state and the output (T2 collector) voltage is close to V+. Now, the two resistors R_{c1} and R_E form a voltage divider that determines the low threshold. Its value is approximately

$$V_{LT} = \frac{R_E}{R_E + R_{c1}} V_+ .$$

Crossing down the low threshold. With the trigger now in the high state, if the input voltage lowers enough (below the low threshold), T1 begins cutting-off. Its collector current reduces; as a result, the shared emitter voltage lowers slightly and T1 collector voltage rises significantly. R_1-R_2 voltage divider conveys this change to T2 base voltage and it begins conducting. The voltage across R_E rises, further reducing the T1 base-emitter potential in the same avalanche-like manner, and T1 ceases to conduct. T2 becomes completely turned-on (saturated) and the output voltage becomes low again.

Variations

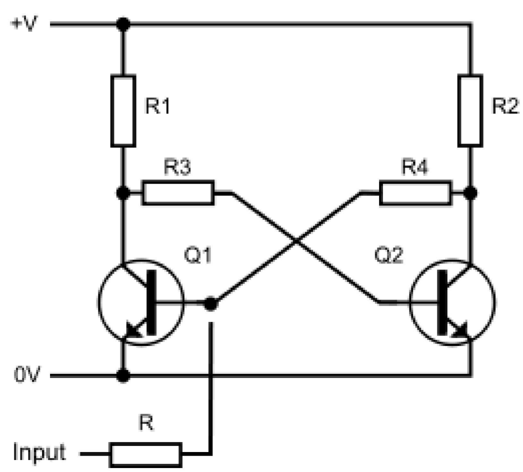


The symbol of inverting Schmitt trigger

Non-inverting circuit. The classic non-inverting Schmitt trigger can be turned into an inverting trigger by taking V_{out} from the emitters instead from T2 collector. In this configuration, the output voltage is equal to the dynamic threshold (the shared emitter voltage) and both the output levels stay away from the supply rails. Another disadvantage is that the load changes the thresholds; so, it has to be high enough. The base resistor R_B is obligatory to prevent the impact of the input voltage through T1 base-emitter junction on the emitter voltage.

Direct-coupled circuit. To simplify the circuit, the voltage divider R_1-R_2 can be omitted connecting T1 collector directly to T2 base. The base resistor R_B can be omitted as well so that the input voltage source drives directly T1 base. In this case, the common emitter voltage and T1 collector voltage are not suitable for outputs. Only T2 collector should be used as an output since, when the input voltage exceedes the high threshold and T1 saturates, its base-emitter junction is forward-biased and transfers the input voltage variations directly to the emitters. As a result, the common emitter voltage and T1 collector voltage follow the input voltage. This situation is typical for over-driven transistor differential amplifiers and ECL gates.

Collector-base coupled circuit



BJT bistable collector-base coupled circuit can be converted to a Schmitt trigger by connecting an additional base resistor to some of the bases

Like every latch, the fundamental collector-base coupled bistable circuit possesses a hysteresis. So, it can be converted to a Schmitt trigger by connecting an additional base resistor R to some of the inputs (Q1 base in the figure). The two resistors R and R_4 form a parallel voltage summer (the circle in the block diagram above) that sums output (Q2 collector) voltage and the input voltage, and drives the single-ended transistor "comparator" Q1. When the base voltage crosses the threshold ($V_{BE0} \infty 0.65$ V) in some direction, a part of Q2 collector voltage is added in the same direction to the input voltage. Thus the output modifies the input voltage by means of parallel positive feedback and does not affect the threshold (the base-emitter voltage).

Comparison between emitter- and collector-coupled circuit

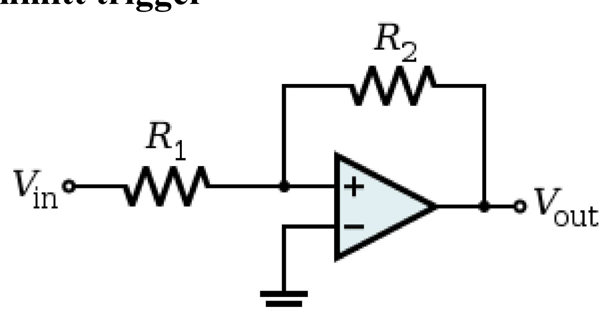
The emitter-coupled version has the advantage that the input transistor is backward-biased when the input voltage is quite below the high threshold; so, the transistor is surely cut-off. It was important when germanium transistors were used for implementing the circuit and this advantage has determined its popularity. The input base resistor can be omitted since the emitter resistor limits the current when the input base-emitter junction is forward-biased.

The emitter-coupled Schmitt trigger has not low enough level at output *logical zero* and needs an additional output shifting circuit. The collector-coupled trigger has extremely low (almost zero) output level at output *logical zero*.

Op-amp implementations

Schmitt triggers are commonly implemented using an operational amplifier or the more dedicated comparator. An open-loop op-amp and comparator may be considered as an analog-digital device having analog inputs and a digital output that extracts the sign of the voltage difference between its two inputs. The positive feedback is applied by adding a part of the output voltage to the input voltage in series or parallel manner. Due to the extremely high op-amp gain, the loop gain is also high enough and provides the avalanche-like process.

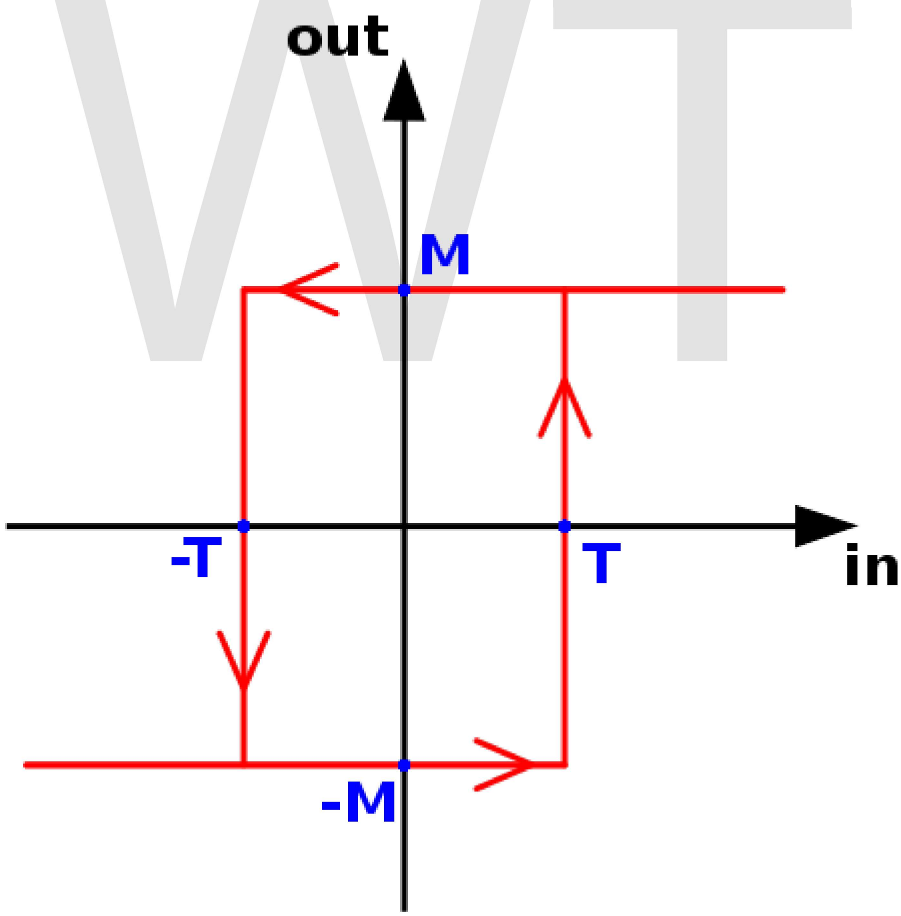
Non-inverting Schmitt trigger



Schmitt trigger implemented by a non-inverting comparator

In this circuit, the two resistors R_1 and R_2 form a parallel voltage summer. It adds a part of the output voltage to the input voltage thus "helping" it during and after switching that occurs when the resulting voltage is near the ground. This *parallel positive feedback* creates the needed hysteresis that is controlled by the proportion between the resistances of R_1 and R_2. The output of the parallel voltage summer is single-ended (it produces voltage in respect to ground); so, the circuit does not need an amplifier with a differential input. Since the conventional op-amps usually have a differential input, the inverting input is grounded (not used).

The output voltage always has the same sign as the *op-amp input voltage* but it not always has the same sign as the *circuit input voltage* (the signs of the two input voltages can differ). When the circuit input voltage is above the high threshold or below the low threshold, the output voltage has the same sign as the *circuit input voltage* (the circuit is non-inverting). It acts like a comparator that switches at a different point depending on whether the output of the comparator is high or low. When the circuit input voltage is between the thresholds, the output voltage is undefined; it depends on the last state (the circuit behaves as an elementary latch).

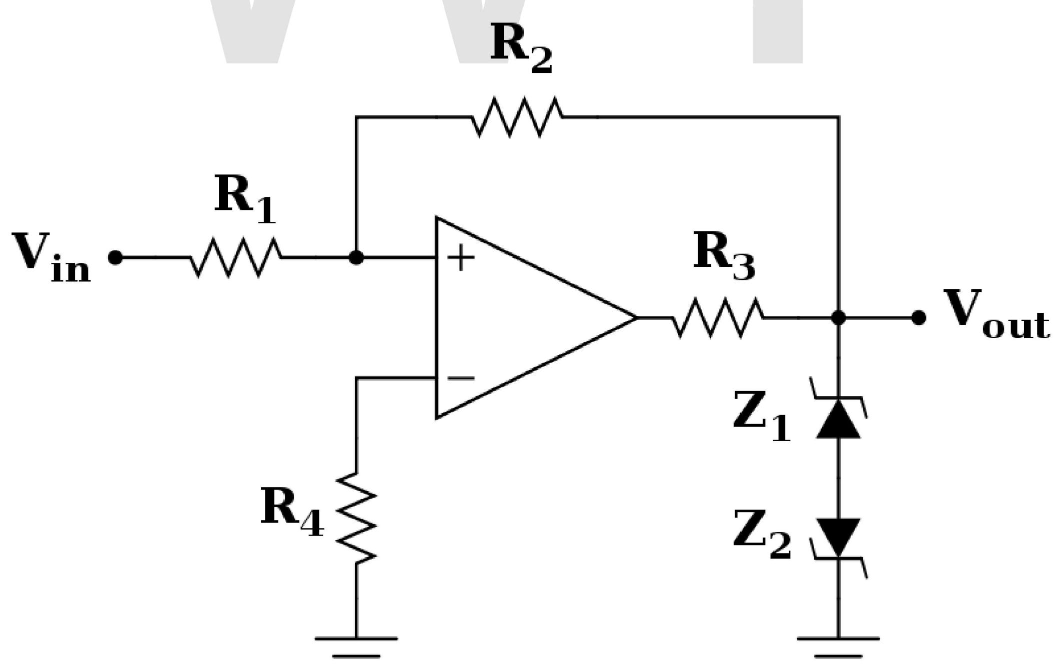


Typical hysteresis curve (Non-inverting) (which matches the curve shown on a Schmitt trigger symbol)

For instance, if the Schmitt trigger is currently in the high state, the output will be at the positive power supply rail ($+V_S$). The output voltage V_+ of the resistive summer can be found by applying the superposition theorem:

$$V_+ = \frac{R_2}{R_1 + R_2} \cdot V_{\text{in}} + \frac{R_1}{R_1 + R_2} \cdot V_{\text{s}}$$

The comparator will switch when $V_+=0$. Then $R_2 \cdot V_{\text{in}} = -R_1 \cdot V_{\text{s}}$(the same result can be obtained by applying the current conservation principle). So V_{in} must drop below $-\frac{R_1}{R_2}V_s$ to get the output to switch. Once the comparator output has switched to $-V_S$, the threshold becomes $+\frac{R_1}{R_2}V_s$ to switch back to high. So this circuit creates a switching band centered around zero, with trigger levels $\pm\frac{R_1}{R_2}V_s$ (it can be shifted to the left or the right by applying a bias voltage to the inverting input). The input voltage must rise above the top of the band, and then below the bottom of the band, for the output to switch on (plus) and then back off (minus). If R_1 is zero or R_2 is infinity (i.e., an open circuit), the band collapses to zero width, and it behaves as a standard comparator. The transfer characteristic is shown in the picture on the right. The value of the threshold T is given by $\frac{R_1}{R_2}V_s$ and the maximum value of the output M is the power supply rail.

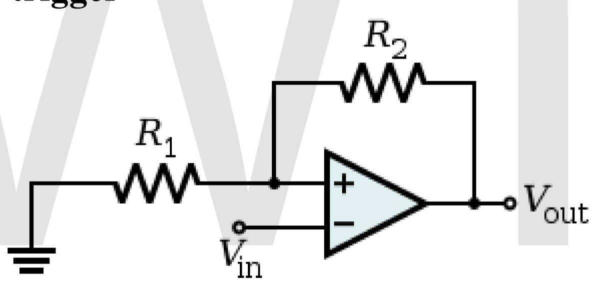


A practical Schmitt trigger configuration with precise thresholds

A unique property of circuits with parallel positive feedback is the impact on the input source. In circuits with negative parallel feedback (e.g., an inverting amplifier), the virtual ground at the inverting input separates the input source from the op-amp output. Here there is no permanent virtual ground (there is only an instant one during the transition) and the steady op-amp output voltage is applied through R_1 - R_2 network to the input source. It is interesting fact that the op-amp output passes an opposite current through the input source (it injects current into the source when the input voltage is positive and it draws current from the source when it is negative).

A practical Schmitt trigger with precise thresholds is shown in the figure on the left. The transfer characteristic has exactly the same shape of the previous basic configuration, and the threshold values are the same as well. On the other hand, in the previous case, the output voltage was depending on the power supply, while now it is defined by the Zener diodes (which could also be replaced with a single double-anode Zener diode). In this configuration, the output levels can be modified by appropriate choice of Zener diode, and these levels are resistant to power supply fluctuations (i.e., they increase the PSRR of the comparator). The resistor R_3 is there to limit the current through the diodes, and the resistor R_4 minimizes the input voltage offset caused by the comparator's input leakage currents.

Inverting Schmitt trigger



Schmitt trigger implemented by an inverting comparator

In the inverting version, the attenuation and summation are separated. The two resistors R_1 and R_2 act only as a "pure" attenuator (voltage divider). The input loop acts as a simple series voltage summer that adds a part of the output voltage in series to the circuit input voltage. This *series positive feedback* creates the needed hysteresis that is controlled by the proportion between the resistances of R_1 and the whole resistance (R_1 and R_2). The effective voltage applied to the op-amp input is floating; so, the op-amp must have a differential input.

The circuit is named *inverting* since the output voltage always has an opposite sign to the input voltage when it is out of the hysteresis cycle (when the input voltage is above the high threshold or below the low threshold). However, if the input voltage is within the hysteresis cycle (between the high and low thresholds), the circuit can be inverting as well as non-inverting. The output voltage is undefined; it depends on the last state and the circuit behaves as an elementary latch.

To compare the two versions, the circuit operation will be considered at the same conditions as above. If the Schmitt trigger is currently in the high state, the output will be at the positive power supply rail ($+V_S$). The output voltage V_+ of the voltage divider is:

$$V_+ = \frac{R_1}{R_1 + R_2} \cdot V_s$$

The comparator will switch when $V_{in} = V_+$. So V_{in} must exceed above this voltage to get the output to switch. Once the comparator output has switched to $-V_S$, the threshold becomes $-\frac{R_1}{R_1 + R_2} V_s$ to switch back to high. So this circuit creates a switching band centered around zero, with trigger levels $\pm\frac{R_1}{R_1 + R_2} V_s$ (it can be shifted to the left or the right by connecting R_1 to bias voltage). The input voltage must rise above the top of the band, and then below the bottom of the band, for the output to switch off (minus) and then back on (plus). If R_1 is infinity or R_2 is zero (i.e., an short circuit), the band collapses to zero width, and it behaves as a standard comparator.

In contrast with the parallel version, this circuit does not impact on the input source since the source is separated from the voltage divider output by the high op-amp input differential impedance.

Applications

Schmitt triggers are typically used in open loop configurations for noise immunity and closed loop configurations to implement function generators.

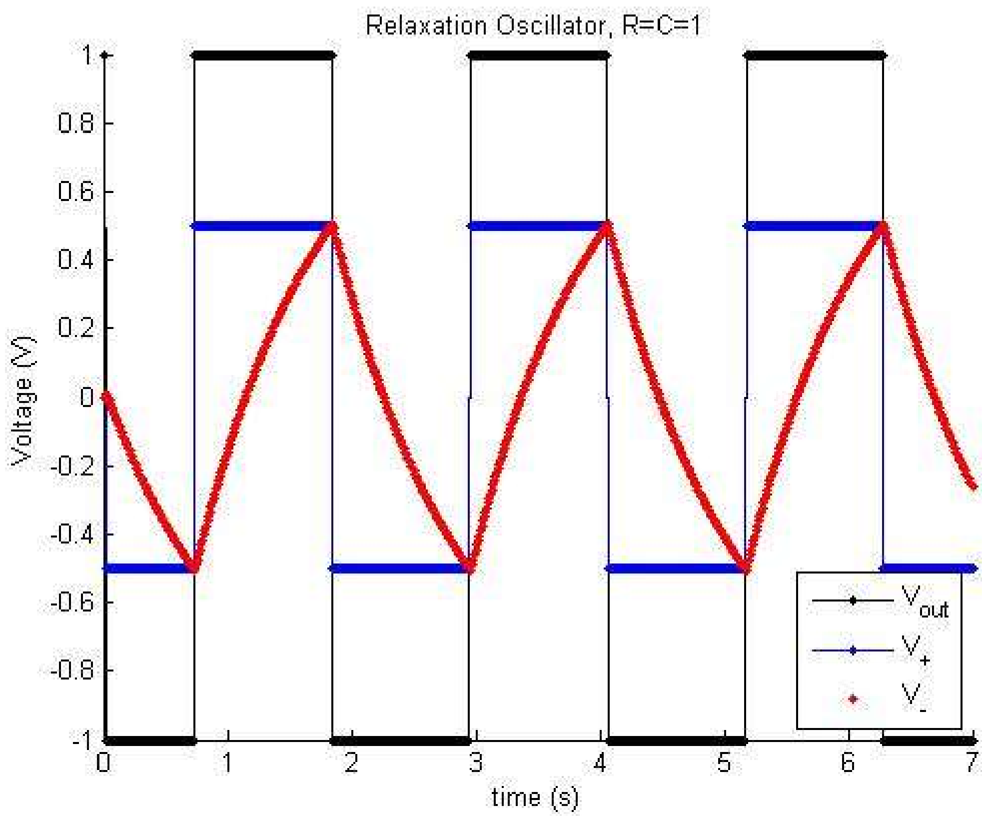
Noise immunity

One application of a Schmitt trigger is to increase the noise immunity in a circuit with only a single input threshold. With only one input threshold, a noisy input signal near that threshold could cause the output to switch rapidly back and forth from noise alone. A noisy Schmitt Trigger input signal near one threshold can cause only one switch in output value, after which it would have to move beyond the other threshold in order to cause another switch.

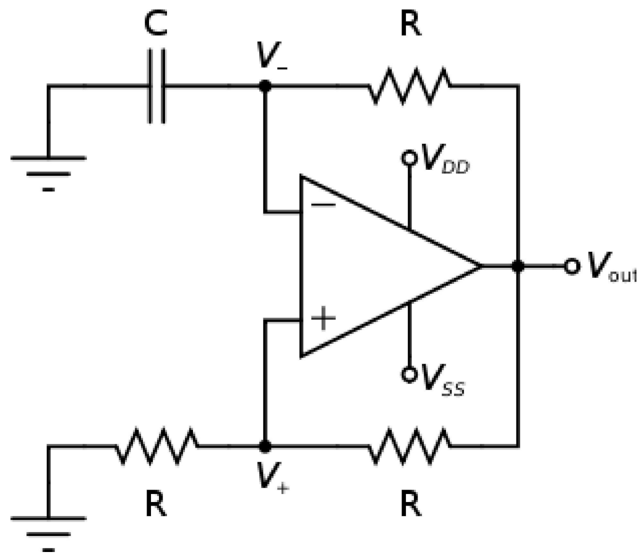
For example, in Fairchild Semiconductor's QSE15x family of infrared photosensors, an amplified infrared photodiode generates an electric signal that switches frequently between its absolute lowest value and its absolute highest value. This signal is then low-pass filtered to form a smooth signal that rises and falls corresponding to the relative amount of time the switching signal is on and off. That filtered output passes to the input of a Schmitt trigger. The net effect is that the output of the Schmitt trigger only passes from low to high after a received infrared signal excites the photodiode for longer than some known delay, and once the Schmitt trigger is high, it only moves low after the infrared signal ceases to excite the photodiode for longer than a similar known delay. Whereas the photodiode is prone to spurious switching due to noise from the environment, the delay added by the filter and Schmitt trigger ensures that the output only switches when there is certainly an input stimulating the device.

As discussed in the example above, the Fairchild Semiconductor QSE15x family of photosensors use a Schmitt trigger internally for noise immunity. Schmitt triggers are common in many switching circuits for similar reasons (e.g., for switch debouncing). The extended content below contains a list of IC including input Schmitt triggers.

Use as an oscillator



Output and capacitor waveforms for comparator-based relaxation oscillator



A comparator-based implementation of a relaxation oscillator

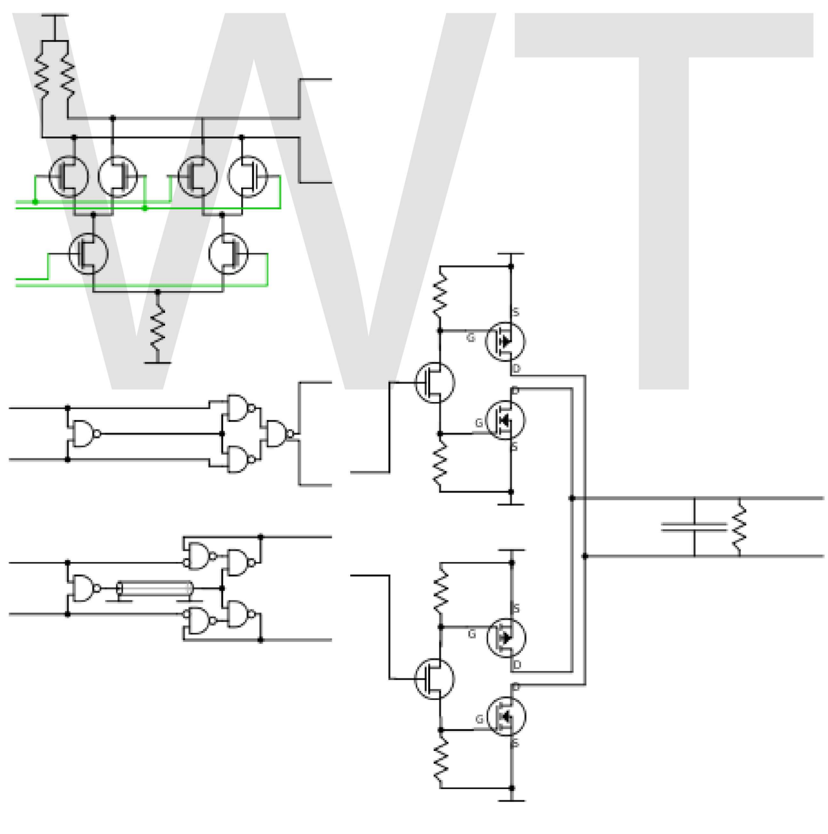
A Schmitt trigger is a bistable multivibrator, and it can be used to implement another type of multivibrator, the relaxation oscillator. This is achieved by connecting a single RC integrating circuit between the output and the input of an inverting Schmitt trigger. The output will be a continuous square wave whose frequency depends on the values of R and C, and the threshold points of the Schmitt trigger. Since multiple Schmitt trigger circuits can be provided by a single integrated circuit (e.g. the 4000 series CMOS device type 40106 contains 6 of them), a spare section of the IC can be quickly pressed into service as a simple and reliable oscillator with only two external components.

Here, a comparator-based Schmitt trigger is used in its inverting configuration. Additionally, slow negative feedback is added with an integrating RC network. The result, which is shown on the right, is that the output automatically oscillates from V_{SS} to V_{DD} as the capacitor charges from one Schmitt trigger threshold to the other.

Chapter- 12

Phase Detector and Nodal Analysis

Phase detector



Four phase detectors. Signal flow is from left to right. In the upper left is a Gilbert cell, which works well for sine waves and square waves, but less well for pulses. In the case of square waves it acts as an XOR gate, which can also be made from NAND gates. On the middle left are two phase detectors: adding feedback and removing one NAND gate produces a time frequency detector. The delay line avoids a dead band. On the right is a charge pump with a filter at its output.

A **phase detector** or **phase comparator** is a frequency mixer, analog multiplier or logic circuit that generates a voltage signal which represents the difference in phase between two signal inputs. It is an essential element of the phase-locked loop (PLL).

Detecting phase differences is very important in many applications, such as motor control, radar and telecommunication systems, servo mechanisms, and demodulators.

Types

Phase detectors for phase-locked loop circuits may be classified in two types. A Type I detector is designed to be driven by analog signals or square-wave digital signals and produces an output pulse at the difference frequency. The Type I detector always produces an output waveform, which must be filtered to control the phase-locked loop voltage-controlled oscillator (VCO). A type II detector is sensitive only to the relative timing of the edges of the input and reference pulses, and produces a constant output proportional to phase difference when both signals are at the same frequency. This output will tend not to produce ripple in the control voltage of the VCO.

Analog phase detector

The phase detector needs to compute the phase difference of its two input signals. Let α be the phase of the first input and β be the phase of the second. The actual input signals to the phase detector, however, are not α and β, but rather sinusoids such as sin(α) and cos(β). In general, computing the phase difference would involve computing the arcsine and arccosine of each normalized input (to get an ever increasing phase) and doing a subtraction. Such an analog calculation is difficult. Fortunately, the calculation can be simplified by using some approximations.

Assume that the phase differences will be small (much less than 1 radian, for example). The small-angle approximation for the sine function and the sine angle addition formula yield:

$$\alpha - \beta \approx \sin(\alpha - \beta) = \sin\alpha\cos\beta - \sin\beta\cos\alpha$$

The expression suggests a quadrature phase detector can be made by summing the outputs of two multipliers. The quadrature signals may be formed with phase shift networks. Two common implementations for multipliers are the **double balanced diode mixer** (diode ring) and the **four-quadrant multiplier** (Gilbert cell).

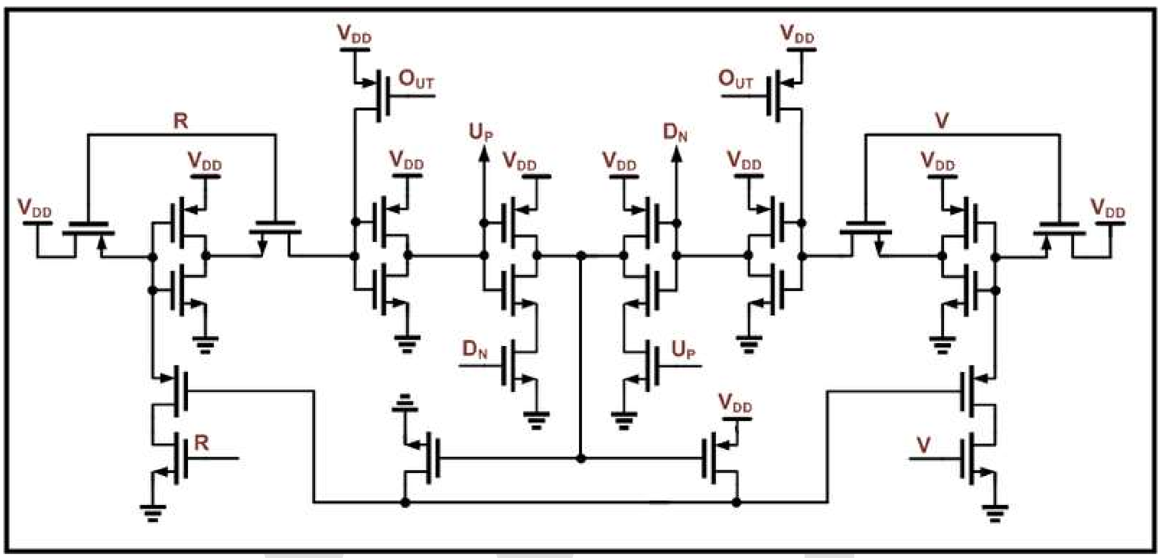
Instead of using two multipliers, a more common phase detector uses a single multiplier and a different trigonometric identity:

$$\sin\alpha\cos\beta = \frac{\sin(\alpha - \beta)}{2} + \frac{\sin(\alpha + \beta)}{2} \approx \frac{\alpha - \beta}{2} + \frac{\sin(\alpha + \beta)}{2}$$

The first term provides the desired phase difference. The second term is a sinusoid at twice the reference frequency, so it can be filtered out.

A mixer-based detector (e.g., a Schottky diode-based double-balanced mixer) provides "the ultimate in phase noise floor performance" and "in system sensitivity." since it does not create finite pulse widths at the phase detector output. Another advantage of a mixer-based PD is its relative simplicity. Both the quadrature and simple multiplier phase detectors have an output that depends on the input amplitudes as well as the phase difference. In practice, the input amplitudes are normalized.

Digital phase detector



An Example CMOS Digital Phase Frequency Detector (Transistor Variety). Inputs are R and V while the outputs UP and DN feed to a charge pump.

A phase detector suitable for square wave signals can be made from an exclusive-OR (XOR) logic gate. When the two signals being compared are completely in-phase, the XOR gate's output will have a constant level of zero. When the two signals differ in phase by 1°, the XOR gate's output will be high for 1/180th of each cycle — the fraction of a cycle during which the two signals differ in value. When the signals differ by 180° — that is, one signal is high when the other is low, and vice versa — the XOR gate's output remains high throughout each cycle.

The XOR detector compares well to the analog mixer in that it locks near a 90° phase difference and has a square-wave output at twice the reference frequency. The square-wave changes duty-cycle in proportion to the phase difference resulting. Applying the XOR gate's output to a low-pass filter results in an analog voltage that is proportional to the phase difference between the two signals. It requires inputs that are symmetrical square waves, or nearly so. The remainder of its characteristics are very similar to the analog mixer for capture range, lock time, reference spurious and low-pass filter requirements.

Digital phase detectors can also be based on a sample and hold circuit, a charge pump, or a logic circuit consisting of flip-flops. When a phase detector that's based on logic gates is used in a PLL, it can quickly force the VCO to synchronize with an input signal, even when the frequency of the input signal differs substantially from the initial frequency of the VCO. Such phase detectors also have other desirable properties, such as better accuracy when there are only small phase differences between the two signals being compared. This is because a digital phase detector has a nearly infinite pull-in range in comparison to an XOR detector.

Phase-frequency detector

A phase-frequency detector is an asynchronous sequential logic circuit originally made of four flip-flops (e.g., the phase-frequency detectors found in both the RCA CD4046 and the motorola MC4344 ICs introduced in the 1970s). The logic determines which of the two signals has a zero-crossing earlier or more often. When used in a PLL application, lock can be achieved even when it is off frequency and is known as a *Phase Frequency Detector*. Such a detector has the advantage of producing an output even when the two signals being compared differ not only in phase but in frequency. A phase frequency detector prevents a "false lock" condition in PLL applications, in which the PLL synchronizes with the wrong phase of the input signal or with the wrong frequency (e.g., a harmonic of the input signal).

A **bang-bang** charge pump phase detector supplies current pulses with fixed total charge, either positive or negative, to the capacitor acting as an integrator. A phase detector for a bang-bang charge pump must always have a **dead band** where the phases of inputs are close enough that the detector fires either both or neither of the charge pumps, for no total effect. Bang-bang phase detectors are simple, but are associated with significant minimum peak-to-peak jitter, because of drift within the dead band.

In 1976 it was shown that by using a three-state phase detector configuration (using only two flip-flops) instead of the original RCA/Motorola twelve-state configurations, this problem could be elegantly overcome. For other types of phase-frequency detectors other, though possibly less-elegant, solutions exist to the dead zone phenomenon. Other solutions are necessary since the three-state phase-frequency detector does not work for certain applications involving randomized signal degradation, which can be found on the inputs to some signal regeneration systems (e.g., clock recovery designs).

A **proportional** phase detector employs a charge pump that supplies charge amounts in proportion to the phase error detected. Some have dead bands and some do not. Specifically, some designs produce both "up" and "down" control pulses even when the phase difference is zero. These pulses are small, nominally the same duration, and cause the charge pump to produce equal-charge positive and negative current pulses when the phase is perfectly matched. Phase detectors with this kind of control system don't exhibit a dead band and typically have lower minimum peak-to-peak jitter when used in PLLs.

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.

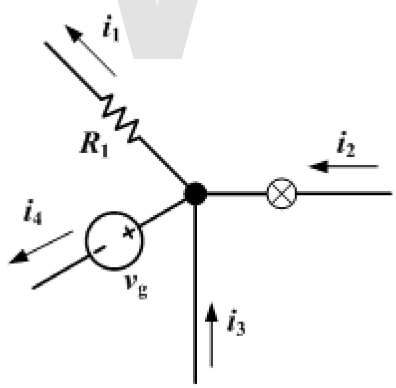
Electronic phase detector

Some signal processing techniques such as those used in radar may require both the amplitude and the phase of a signal, to recover all the information encoded in that signal. One technique is to feed an amplitude-limited signal into one port of a product detector and a reference signal into the other port; the output of the detector will represent the phase difference between the signals. If the signal is different in frequency from the reference, the detector output will be periodic at the difference frequency.

Optical phase detectors

Phase detectors are also known in optics as interferometers. For pulsed (amplitude modulated) light, it is said to measure the phase between the carriers. It is also possible to measure the delay between the envelopes of two short optical pulses by means of cross correlation in a nonlinear crystal. And it is possible to measure the phase between the envelope and the carrier of an optical pulse, by sending a pulse into an nonlinear crystal. There the spectrum gets wider and at the edges the shape depends significantly on the phase.

Nodal analysis



Kirchhoff's current law is the basis of nodal analysis.

In electric circuits analysis, **nodal analysis**, **node-voltage analysis**, or the **branch current method** is a method of determining the voltage (potential difference) between "nodes" (points where elements or branches connect) in an electrical circuit in terms of the branch currents.

In analyzing a circuit using Kirchhoff's circuit laws, one can either do nodal analysis using Kirchhoff's current law (KCL) or mesh analysis using Kirchhoff's voltage law

(KVL). Nodal analysis writes an equation at each electrical node, requiring that the branch currents incident at a node must sum to zero. The branch currents are written in terms of the circuit node voltages. As a consequence, each branch constitutive relation must give current as a function of voltage; an admittance representation. For instance, for a resistor, $I_{branch} = V_{branch} * G$, where G (=1/R) is the admittance (conductance) of the resistor.

Nodal analysis is possible when all the circuit elements' branch constitutive relations have an admittance representation. Nodal analysis produces a compact set of equations for the network, which can be solved by hand if small, or can be quickly solved using linear algebra by computer. Because of the compact system of equations, many circuit simulation programs (e.g. SPICE) use nodal analysis as a basis. When elements do not have admittance representations, a more general extension of nodal analysis, modified nodal analysis, can be used.

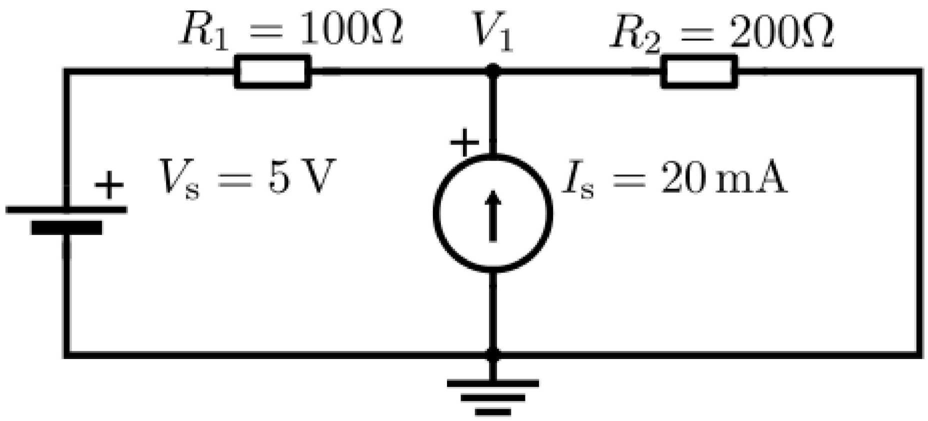
While simple examples of nodal analysis focus on linear elements, more complex nonlinear networks can also be solved with nodal analysis by using Newton's method to turn the nonlinear problem into a sequence of linear problems.

Method

1. Note all connected wire segments in the circuit. These are the *nodes* of nodal analysis.
2. Select one node as the ground reference. The choice does not affect the result and is just a matter of convention. Choosing the node with most connections can simplify the analysis.
3. Assign a variable for each node whose voltage is unknown. If the voltage is already known, it is not necessary to assign a variable.
4. For each unknown voltage, form an equation based on Kirchhoff's current law. Basically, add together all currents leaving from the node and mark the sum equal to zero.
5. If there are voltage sources between two unknown voltages, join the two nodes as a supernode. The currents of the two nodes are combined in a single equation, and a new equation for the voltages is formed.
6. Solve the system of simultaneous equations for each unknown voltage.

Examples

Basic case



Basic example circuit with one unknown voltage, V_1.

The only unknown voltage in this circuit is V_1. There are three connections to this node and consequently three currents to consider. The direction of the currents in calculations is chosen to be away from the node.

1. Current through resistor R_1: $(V_1 - V_S) / R_1$
2. Current through resistor R_2: V_1 / R_2
3. Current through current source I_S: $-I_S$

With Kirchhoff's current law, we get:

$$\frac{V_1 - V_S}{R_1} + \frac{V_1}{R_2} - I_S = 0$$

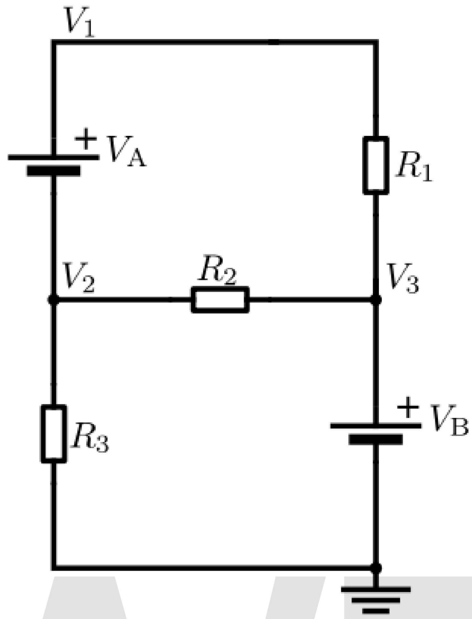
This equation can be solved in respect to V_1:

$$V_1 = \left(\frac{V_S}{R1} + I_S\right) : \left(\frac{1}{R_1} + \frac{1}{R_2}\right)$$

Finally, the unknown voltage can be solved by substituting numerical values for the symbols. Any unknown currents are easy to calculate after all the voltages in the circuit are known.

$$V_1 = \left(\frac{5\ \mathrm{V}}{100\,\Omega} + 20\ \mathrm{mA}\right) : \left(\frac{1}{100\,\Omega} + \frac{1}{200\,\Omega}\right) \approx 4.667\ \mathrm{V}$$

Supernodes



In this circuit, V_A is between two unknown voltages, and is therefore a supernode.

In this circuit, we initially have two unknown voltages, V_1 and V_2. The voltage at V_3 is already known to be V_B because the other terminal of the voltage source is at ground potential.

The current going through voltage source V_A cannot be directly calculated. Therefore we can not write the current equations for either V_1 or V_2. However, we know that the same current leaving node V_2 must enter node V_1. Even though the nodes can not be individually solved, we know that the combined current of these two nodes is zero. This combining of the two nodes is called the supernode technique, and it requires one additional equation: $V_1 = V_2 + V_A$.

The complete set of equations for this circuit is:

$$\begin{cases} \frac{V_1 - V_B}{R_1} + \frac{V_2 - V_B}{R_2} + \frac{V_2}{R_3} = 0 \\ V_1 = V_2 + V_A \end{cases}$$

By substituting V_1 to the first equation and solving in respect to V2, we get:

$$V_2 = \frac{(R_1 + R_2)R_3 V_B - R_2 R_3 V_A}{(R_1 + R_2)R_3 + R_1 R_2}$$