Chapter 20

Vacuum Tubes - The First Active Devices

Active devices refer to those things that can electronically ***control*** electron flow. The first active devices were vacuum tubes, followed by the ***transistor junction.*** This chapter begins the fist active devices and their development into increasingly more complex device that made all electronics possible.

In the Beginning: the Fleming Valve

The first such devices were ***diodes*** (of a sort), double-electrode structures that would allow electron flow to go only one way from a heated filament to a *plate.* Although technically they were half-wave *rectifiers,* they worked slightly better than a variety of diode-like devices loosely based on semiconducting junctions as *detectors* of broadband damped-wave on-off-keyed radio signals.[1]. A *new* device back around the year 1900 was the *galena crystal*, a rather crude point-contact solid-state diode.[2]

English academician John Ambrose Fleming experimentally developed the first vacuum tube diode about 1888 while investigating the *Edison Effect* mentioned in 1883 literature. The first ***Fleming valve*** has been described as an Edison incandescent lamp with a metal cylinder around it. The glowing filament acted to emit electrons, some of which would go to the cylinder, later called a ***plate.*** Electrons would be repulsed by the high temperature of the filament so electron flow would always be from filament to plate. The English used ***valve*** as a descriptor for this new device since it was, in effect, a one-way *valve* for electrons. Americans adopted the name ***tube*** probably because the first valves formed in the USA were made from glass tubing.

The key to *valve* or *tube* operation was the ***vacuum*** inside it. Electrons would literally move around in a vacuum whereas they would be physically inhibited by another substance such as air. The *Fleming Valve* was tried out in early Marconi receivers and it worked better than the other strange devices used before that. While this Fleming Valve worked better, the 'radio receivers' of early days were little more than ***tuned*** *half-wave rectifiers.* That would change about 1906 when Lee De Forest began some experiments of his own.

---

[1] A variety of such semiconducting junctions existed before 1900, nearly all of them quite inefficient but better than nothing at all. There was no theory to back up their applications and all inventions in radio of those early days was empirical.

[2] That gave birth to numerous new names and catch-phrases, such as *cats whisker*, a holder for a needle that was manually positioned on an exposed surface of the crystal held in a molded lead base. Once that location was found, the user was said to have found the *sweet spot.* Crude receivers were, and still are, called *Crystal Sets.*

The First True Active Device

De Forest introduced a third electrode in the space between filament and plate, a literal wire ***grid***. When this *grid* was at a potential (or voltage) more negative than the filament, few electrons would flow into it. If the relative grid voltage was less negative, more electrons would flow from the filament *cathode* through the grid to the plate or *anode.* Voltage on the grid (relative to the filament) ***could control the electron flow from filament to plate.*** Not only that, the grid, the third electrode, would draw very little electron current when negatively biased. It was an ideal amplifier!

Lee De Forest called his new invention an *audion,* supposedly a contraction of two other words, in his January 1906 patent. In a later improvement patent application, De Forest called it a *De Forest* ***Triode***, possibly to get away from the early name pertaining to audio frequencies. This was a forerunner to more tube descriptions having multiple grids: Tetrode with two grids, Pentode with three grids, even Pentagrid having five grids.[3]

In all circuits that followed, vacuum tube plates would always be connected to a ***positive*** supply voltage relative to the filament.[4] In very old radio receivers powered by batteries, the filaments were lit by an ***A*** battery of low voltage, a ***B*** battery for high voltage plate supply, and, perhaps, a ***C*** battery for low current demand biasing. Since the plate supply is always positive, the high voltage battery was colloquially called ***B+***, spoken as ***B-Plus***, in common use in electronics labs as late as the 1980s.

Application

The first application of this new triode to radio was probably as an audio amplifier, to increase the audio out of those tuned half-wave rectifier circuits called receivers. Audio frequencies were comfortable experimental areas. One of the first things noted was that all vacuum tube circuits ***invert*** their amplified voltages . Applications as RF amplifiers were few and far between at the beginning. Everyone had to know more about basic components, resonance, impedance, and so forth and there wasn't a lot of ***term standardization*** in the first quarter of the 1900s. De Forest thought that the principles of his new audion or triode were due to trace gasses in his vacuum. Another inventor, Irving Langmuir, would correctly say that the vacuum itself was responsible.

The first RF circuit applications took advantage of the high impedance of negatively-biased grids. Those grids drew so little power that they appeared as *high impedance loads* to any source. This was good for tuned-RF circuits since a high-impedance load did not degrade the Q of a tuned circuit.

An early experimenter, Edwin Howard Armstrong, would begin his revolutionary inventions with the ***regenerative detector*** that replaced the old crystal set receivers. Armstrong used ***positive feedback*** to increase the sensitivity of his detector at least a hundred times more than was possible with the half-wave rectifier detector of beginning radio days. Armstrong's detector was essentially unstable, best sensitivity a bit short of being an actual oscillator. So close to being oscillatory that

---

[3] Tetrodes and Pentodes count the filament and plate in number of electrodes. By the time of the Pentagrid, the name assigners probably ran out of descriptive names.

[4] Using transistor terms, vacuum tubes always worked similar to NPN transistors.

a feedback or *regeneration* control was necessary for the user to adjust. Despite its so-called shortcomings, it was a remarkable improvement in sensitivity and literally opened the world of radio to more distant communication ranges. The telephone infrastructure was also experimenting, but with ***negative feedback*** to both stabilize amplification and remove much non-linear distortion.

Overview

Standardizing How Vacuum Tubes Work

In the approximate two decades after 1906, thinking people started to change some of the names and terminology as well as improving the technology of making them. That was due to academicians and manufacturers. Academicians wanted *analytical models* so they could design circuits mathematically. Manufacturers wanted better performance or, of more importance, *repeatability in specifications* of what they manufactured. To suit the latter, the filament was encased in a coated sleeve with the coating able to coax more electrons being emitted. That was called the ***cathode*** and the filament of this ***indirectly-heated cathode*** tube type was called a ***heater.*** The plate was interchangeably called ***anode*** or just *plate.* Grids were called just that but, beyond the first or *control grid,* they had prefixes such as ***screen grid*** for the second one, ***suppressor grid*** for the third one. The first Fleming Valve fell into the ***directly-heated cathode*** category. That directly-heated tube variety would remain with the low-temperature, low-voltage filaments of so-called ***battery tubes*** used in portable radios until up into the 1960s.

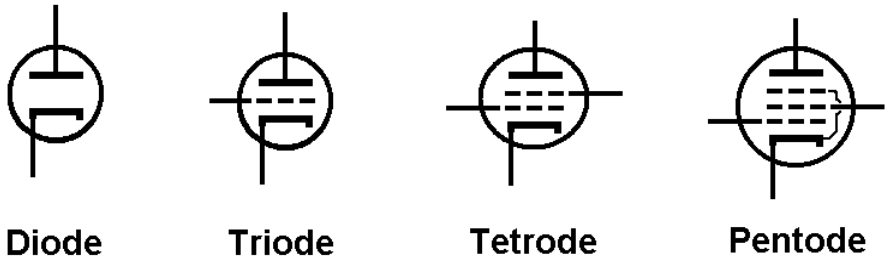


Figure 20-1 Four principal types of vacuum tubes in standardized symbolic form for indirectly-heated cathode types. Electron flow is from cathode (at bottom) to plate (at top) with *grids* between the two on Triode through Pentode. The *control grid* is physically closest to cathode, has greatest influence in cathode-to-plate current control.

To set up analytical models, three new terms were devised called *dynamic characteristics.* These could be used by circuit designers to approximate actual operation with tubes and hardware.

Transconductance, abbreviated ***gm***, is the small-scale change in plate current due to a small-scale change in grid voltage. The term ***small-scale*** is just that, a tiny change in one electrode induced to effect a tiny change in another electrode's characteristics. *Transconductance* and the plate output loading can be the determinator of tube stage voltage gain in analytical models.

Plate Resistance, abbreviated r_p, is the small-scale change in plate voltage divided by the

small-scale change in plate current, and expressed in terms of Ohms. In analytical models of linear amplifier models, plate resistance appears as the equivalent of a high-value resistor in parallel with the plate load impedance. Most modern pentodes will have plate resistance values of 200 KOhms or larger.

Amplification Factor, abbreviated ***μ,*** is the small-scale change in plate voltage from a small-scale change in grid voltage, all other electrodes and currents being kept constant. Note: The term ***small-scale*** is just that, peak-to-peak values something between very small to right down to noise level, values depending on the measurement AC voltmeters and AC ammeters. Dynamic characteristics are measured at AC, ***plate tube curves*** seen on many tube datasheets, done at DC.

Vacuum tube plate characteristics at DC are somewhat non-linear. Some tubes are deliberately designed to have non-linear *transconductance* for the purpose of controlling stage gain by means of negative DC bias on the grid. In using *small-scale dynamic characteristics*, it is possible to closely approximate actual hardware performance with different DC bias conditions.

Directly-Heated Cathodes versus Indirectly-Heated Cathodes

The first triodes were directly-heated. That is, their heater-filament was the cathode. Not too long after that, filaments became indirectly-heated by having a tubular sleeve over the heater-filament, the sleeve being the cathode itself. In addition, both types of heater-filaments were coated with a material to yield greater electron flow from these cathodes.

By separating the heater-filament from the cathode sleeve, the indirectly-heated heater-filament became the tube of choice for ready-made circuitry. The heater could be isolated from the cathode in the circuit and this prompted new tubes with the same characteristics as the old ones but with different filament voltages. The *All-American Five* transformerless table-top broadcast receiver was born, all filaments taking the same current of 150 mA and all in series (with a pilot lamp) so that their filament voltages added up to about 129 VAC RMS. Plate and screen high voltage was supplied by a half-wave rectifier directly from the AC line voltage.[5]

With the indirectly-heated cathode there was the advantage of keeping the filament supply away from the vacuum tube itself. If nothing else, the filaments could be capacitively bypassed to ground to remove them from the circuits. In general, the indirectly-heated cathodes resulted in slightly more uniform characteristics.

Directly-heated filament tubes remained for the dry-cell-powered battery receivers and other instruments using a B+ of at least 67 VDC (usually 90 VDC) but with filaments of only 1.5 VDC. With directly-heated cathodes the designers had to contend with characteristics that were modified for biasing by the 1.5 VDC gradient along each cathode surface. Those gradients did not exist in indirectly-heated cathodes. Fortunately, this was a minor problem and could be overcome in most cases.

---

[5] The usual octal-based tube line-up was a 12SA7 pentagrid as both LO and mixer, a 12SK7 as an IF amplifier, and (usually) a 12.6 VAC filament dual-diode triode as detector, AGC rectifier, and audio voltage amplifier plus a 50 VAC filament audio speaker driver and a 35 VAC filament half-wave rectifier. For 7-pin miniature glass envelopes, the 12BE6, 12BA6, 12AT6, 50B5, and 35Z5 were circuit-complimentary for the same functions. Literally millions of these transformerless AM broadcast receivers were made starting just before WWII and continuing on for two decades after WWII was over. There was very little change between models insofar as the point-to-point wiring was concerned..

General Operation for Biasing

The general naming of USA-designed tube types comes from the total number of electrodes of a tube. The *control grid* (or just plain ***grid***) has the most pronounced control over cathode-plate current. In the Pentode, the last grid, called the *suppressor grid*, is usually connected internally to the cathode although a few types have the suppressor brought out as a separate pin.[6] The second grid of Tetrodes and Pentodes is called the ***screen grid*** (or just ***screen***) and has an effect on the cathode-plate current but generally kept at a specific DC supply voltage as indicated in vacuum tube data sheets. In most Pentodes, the third grid is called the ***suppressor*** and is generally there to reduce the effect of secondary emission of electrons striking the plate and bouncing off. Most suppressor grids are connected internally to the cathode.

Most Diodes in-use are *dual-plate, directly-heated-cathode types* (example a type 5Y3) for full-wave DC power rectification from a center-tapped power supply transformer secondary. Old FM receiver demodulators are generally dual-diodes (such as the 6AL5 or 6H6) having isolated diodes but with common filament-heaters.

Self-Bias Through a Single Resistor

Figure 20-2 is the general form of self-bias of a single triode voltage amplifier running with low signal input and output RMS voltages. The cathode-plate current passes through R2 and creates a voltage drop (E2). That voltage drop is the bias of control grid to cathode..

The grid input is normally sitting at common potential or zero Volts via R1. As far as the triode is concerned, E2 provides a slight ***negative*** potential to the grid on a relative basis. Note the electron current flow direction represented by I2; ***relative*** to the cathode, the grid has a slight negative voltage.

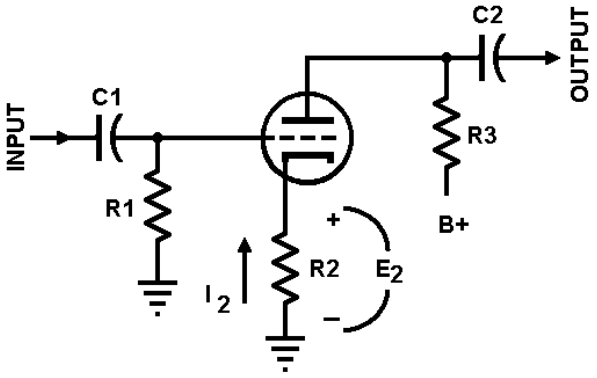


Figure 20-2 A triode connected for self-bias via voltage drop across R2

Series resistor R2 may or may not be bypassed by a capacitor to ground. That is dependent on the circuit design. That capacitor must have a value such that its reactive impedance is low at the lowest signal voltage frequency in order to bypass R2. If R2 is not bypassed, then it becomes a signal amplification ***loss***. That loss can be taken as an improvement on reducing distortion from the unbypassed cathode resistors. This is shown by:

---

[6] The purpose of the suppressor grid was to reduce the ***secondary emission of electrons*** from the plate and this distorting the plate current characteristics. In certain tube types the suppressor grid has been used as an amplitude modulator for transmitters in older designs, mainly to reduce the number of tubes used. Cathode-to-plate current at higher voltages without a suppressor grid would tend to make the higher electron velocity cause a *bouncing off* of electrons called secondary emission.

$$A_V = \frac{A}{(1 + b\,A)} \qquad \text{where:} \qquad (20\text{-}2)$$

A = Voltage gain without feedback

A_V = Voltage gain with feedback

b = Percentage of feedback (can be = to b / R3) where R3 equals the plate load.

For the difference between *feedback* and none at all, take the percent of the cathode series resistance divided by the total plate load in Ohms. If there is no cathode resistance, then the numerator becomes 1 and the A_V becomes the same as A. If there is a cathode resistance, then divide it by the total plate load resistance to obtain the value of b.

Assume a tube has an A of 100, total R_L of 10 KOhms and would have a cathode resistance of 220 Ohms. Value of b would be 22 x 10^{-3} and (b x A) would be 2.2. Total value of A_V would be 31.25 times voltage gain instead of the no-cathode-resistance case of 100 times.

Distortion reduction will work with any vacuum tube having a grid and with an unbypassed cathode resistor. It will not work with tube diodes since those have no grid to accept an input.[7] Actual distortion reduction ***must be measured*** since there are too many things to consider as to the cause of the bypassed self-bias resistor distortion.

Voltage Gain Amount

Voltage amplification with pentodes is quite easy. As a general rule, one can simply multiply the transconductance (g_M) by the load impedance. End result is the voltage gain for that stage. To be slightly more accurate, the load impedance would be in parallel with the plate resistance (r_P).

$$\text{Voltage Gain} = V_{OUT} / V_{IN} = \text{transconductance x } (Z_{LOAD} \,||\, r_P) \qquad (20\text{-}3)$$

If the load impedance (such as mid-band filter impedance) is known, then that simple formula is within ±5% of the correct value.

For triodes and some tetrodes this might be too coarse. In that case, a ***load line*** can be drawn on characteristic plate voltage and current curves (on datasheets) relative to the control grid bias voltage. See Figure 20-3 for the steps involved. It is quite simple. The ***slope*** of the straight load line begins at top left for zero plate voltage and ends at bottom right at maximum plate voltage but with zero plate current. Angle of the slope is equal to plate voltage divided by the load impedance to achieve the plate current with no voltage.

If the load impedance is 5000 Ohms, divide a convenient plate voltage (such as 200 VDC) by that to get 40 mA. See the dotted line in Figure 20-3 for that load line slope. Slide the load line

[7] Some diode circuits have the ***appearance*** of something similar to unbypassed self-bias resistors but those are generally for another reason.

(without changing its slope) left or right until the load line intersects a desired control grid voltage (such as -9 VDC for a 6AK6 pentode). That intersection is the bias for the grid and all signal inputs vary away from that bias level.

To find the resulting V_{OUT} for a particular V_{IN}, use the load line intersection to see the V_{OUT} from the x-axis scale at the bottom. This was already done for the 6AK6 audio speaker driver in Chapter 63 (using a slightly different plate characteristic curve set) for a -9 VDC bias point.

At the intersection of the Load Line and plate current curves, the control-grid to cathode voltages are:

$E_{C1\text{-}CATHODE}$	Plate Voltage
0 V	40
-3 V	68
-6 V	104
-9 V	135
-12 V	170
-15 V	192
-18 V	205

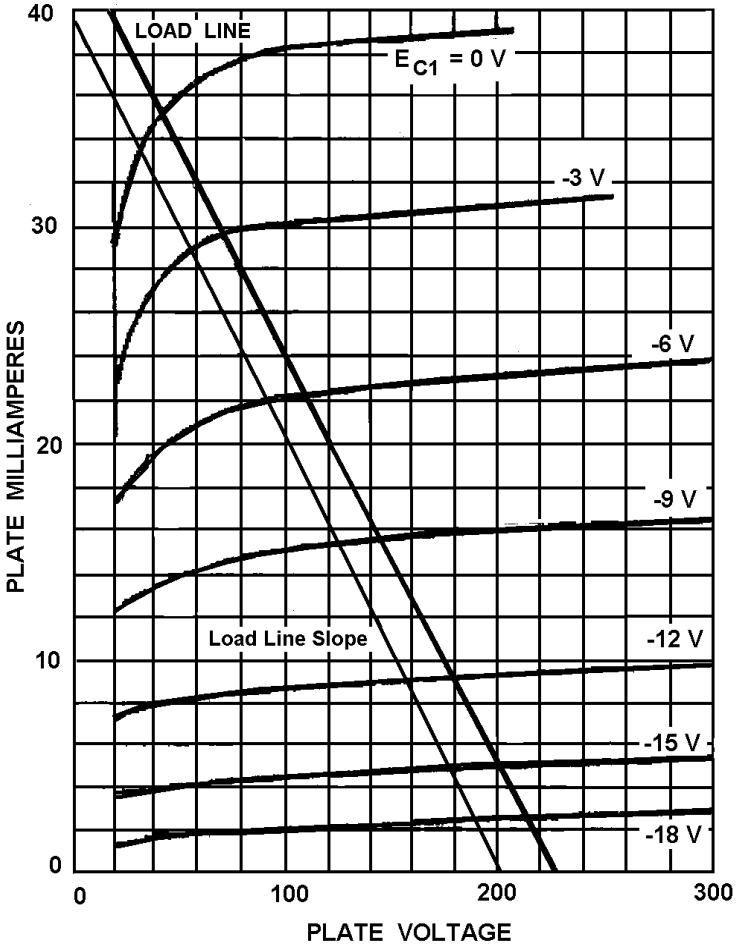


Figure 20-3 Load Line construction done for a 6AK6 small pentode on *plate curves* for a fixed screen and filament voltage. Control grid voltage curves are marked as E_{C1}, marked negative relative to cathode.

However, with an unbypassed series cathode resistor of 560 Ohms and 185 VDC cathode-plate, the control grid input voltage versus plate voltage can be taken as:

Eg-k	Ip,mA	Ep	Ip+Isc, mA	Ek-grd	Eg-grd
0	35	40	37.5	21.0	+21.0
-3	29.5	68	31.5	17.6	+14.6
-6	22	104	24.5	13.7	+7.7
-9	15	135	17.5	9.8	-0.8
-12	9	170	11.5	6.4	-5.6
-15	4.5	192	7.0	3.9	-11.1
-18	2.2	205	4.7	2.6	-15.4

Taking the -3 V and -15 V grid-cathode rows, the plate-screen current is 31.7 and 7.0 mA and plate current is 192 and 68 V. Grid to ground voltages are +14.6 and -11.1 which would be the peak to peak value of audio input; RMS audio input is approximately 9.09 VAC. Plate voltage change is 124 V peak-to-peak and plate current change is 25 mA peak-to-peak. Those translate to 17.3 VAC

RMS and 8.7 mA RMS for an approximate plate output of 151 mW. Gain of the audio output driver circuit is then 124 / 17.3 = 7.17 times or 17.1 db in voltage.

If the cathode resistor (for self-bias) is included, the voltage gain would be 124 / 12 = 10.3 times although the distortion would be more pronounced.

Resistance-Coupled Amplifier Charts

About the start of WWII appeared ***Resistance-Coupled Amplifier Charts***, primarily to reduce the need for bulky and heavy (and expensive) transformers for audio coupling. Such charts, primarily for low-power audio stages, could be taken verbatim for all values. Those needed only resistors and capacitors. Such charts were available for most low-power audio stages and used in various designs to eliminate lower-power audio coupling transformers.

Transmitting Applications

General

Receivers could be built from negative-grid-bias circuits almost entirely (except for certain musical-instrument audio stages). Transmitters required an intermix of negative grid bias with high power amplifiers needing zero bias voltage or a form of near-zero-bias where the tube control grids drew current. Most of such higher-power RF amplifiers would have the grid swinging into the positive grid-to-cathode potential over part of a stages RF cycle.

The First of the Amplifier Classes

These were three types, **A, B,** and **C.** Type A was considered for most low-power circuits, the control grid always in negative potential relative to the cathode. Type B had a zero-bias grid and, with a push-pull arrangement, was generally considered a *linear amplifier*. For half of a signal input cycle, the grid would draw current as it went positive, then nearly cut off on the other half of a signal input. Together in a push-pull arrangement they would be considered *linear.*

Type C had varying grid bias voltages, usually with the grid being at a negative potential relative to the cathode but would be swinging into the positive grid-cathode region during part of the RF cycle of power input. Class C amplifiers were the *brutes* capable of delivering Watts and KiloWatts (sometimes even Megawatts) of power into an antenna for transmitting.

A problem with Classes B and C was that they both delivered ***harmonics*** of the lower-powered RF cycle input. This was because of the distortion introduced when the control grid swung positive relative to the cathode; the tubes themselves lost their linearity due to loading the driving stage.[8] This was both good and bad. It was good for use as a ***multiplier*** stage to output a deliberate harmonic. It was bad for a power amplifier that should have output the fundamental. Fortunately,

[8] It should be emphasized that the control grid is normally a very high impedance in Class A due to its negative potential relative to its cathode. Once it starts to draw current the high impedance is lost.

L-C tuned circuits (relatively narrow-band) could select the proper harmonics for either case.

Later there were two in-between classes called ***AB$_1$*** and ***AB$_2$,*** where the bias point was very close to Class A but not quite on-the-money. Those were developed for the *high-fidelity* hobbyist who craved more audio speaker power with minimal distortion.

With the oncoming transistor stages there were more *classes* added, such as ***D, E,*** and ***F,*** all three using forms of pulse-width relative to power amplifier output along with some averaging filtering of generated harmonic powers. There were some types of tube circuits tried with pulse-width modulation but those were generally dropped quickly when power transistors were available.

Class B and C Input Loading

Once rather popular with amateur radio afficionados, those have been dropped from new designs. Design details are rather sketchy in those regards and much of it has become old and, now, unused for new designs. It is popular in newer designs to use low-power Class A biased circuits for driving the final amplifiers in HF equipment. If anything is done for non-linear biasing, it is with passive circuitry such as diode mixers to change frequency by heterodyning.[9]

Those who wish to use tube circuits for multiplication are free to consult old ARRL texts for more information on multiplying to achieve harmonic band-spreads.

Video Amplification

General

This is more concerned with the output or plate circuit of tubes. All vacuum tubes have some finite plate output capacity. That normally limits the upper frequency region due to excessive capacitive reactance loading of a load impedance. The point here is to lower the resistive part of a load impedance until the capacitive reactance is (somewhat) dwarfed-out by the lower resistive impedance. Unfortunately, that also lowers a video amplifier's stage gain, requiring more of them to be added to a design and thus increasing the production cost of video receivers.

Operating With Only R and C (and some L) in the Plate Circuit

A lot of old papers and application notes have been written and published on video amplification, to extend the upper frequency bandwidth using vacuum tubes. Nearly all of those are now obsolete due to lower load impedances possible with solid-state circuits.

As an example, suppose you have a load resistance of 100 KOhms but the combined tube plate, socket, wiring capacity is 15 pFd. The frequency where the total capacity is equal to 100 KOhms is about 106 KHz. Voltage gain at 106 KHz is about at half of the DC value. Dropping the

---

[9] It was once popular to assign amateur radio bands in multiples of harmonics to take advantage of the tube circuit multiplication. Original HF bands for amateur radio were 3.5 to 7.0 to 14 to 28 MHz, each band jump being a second harmonic of the next lower one. With successive World Radio Conferences newer HF bands were allocated so that there were no longer the emphasis on harmonics.

load resistance to 10 KOhms will drop the voltage gain to about one-tenth of what it was at 100 K resistance but the high-frequency end is now about 10 times higher, roughly 1 MHz.

A solution is to use a tube with higher g_M, higher plate current, and lower load impedance. This will not solve the low-frequency response but it does provide a wider bandwidth. Low-frequency response is still following the $g_M \times R_L$ so you might need two video amplifiers in series for a wide bandwidth. An alternative is to use a series inductance with the load resistance to make up for the total load resistance drop due to total output capacity. There are several papers on such load impedance adjustment involving R, C, and L. I won't go into such things since those haven't been used in new designs for over three decades but they are available and can be found.

The old Tektronix 540 series oscilloscopes were built with the final CRT vertical deflection plate coming from a multi-stage, push-pull (differential) ***distributed delay line*** load impedance. Those would have the total output capacitance as part of the delay line distributed capacity and could carry the higher-frequency up to better than 25 MHz. This seemed to be the ultimate in repeatable, production units including military hardware in the era of the 1950s.

Vacuum tube voltage amplifiers are limited in upper frequency response by another phenomenon: ***Electron transit time effects.*** That is, it takes a longer time for changes in tube electron flow to get from cathode to plate than in free space. The choice there is to make tubes physically with as little space between cathode-grid-plate as possible. Unfortunately, the limitation is roughly about 2 GHz at the maximum bandwidth.

Tubes for Microwave Transmission

Magnetrons

These essential RF power oscillators for microwave ovens are basically diodes sitting in an intense magnetic field. They are capable of very high RF peak power happening at around a microSecond (in radar transmitters) or at lower power (for microwave ovens). The combination of the diodes, magnetic field, and a load impedance of a cavity resonator will determine the operating frequency. Magnetrons can operate up to about 25 GHz but the microwave ovens are at about 2 GHz due to the frequency allocated to the ISM or Industrial-Scientific-Medical band.

Klystrons

Smaller cousins of Magnetrons, Klystrons have little external magnetic field and were originally conceived as the Local Oscillator for radar receivers, converting incoming microwave energy down to about 60 MHz band-center. Their internal structure is only slightly more complex than a Magnetron. A typical Klystron of WWII was about the size of an octal-based metal-envelope vacuum tube. After WWII higher-power Klystrons were used for UHF TV transmitter final amplifiers. Those may be several feet in length, dependent on RF power output.

Traveling-Wave Amplifier Tubes

Abbreviated ***TWT***, these are an old design but have about an octave of bandwidth. Their

internal structure resembles a vacuum tube electron flow going through a helix or spiral delay line extending nearly the length of a TWT envelope. TWTs are capable of very wide bandwidth and have low internal noise (relative to other microwave tubes). As of 2010 they are still the transponder group amplifier for communications satellites in orbit.

Display and Glow-Discharge Tubes

Cathode-Ray Tubes

One of the first of the non-amplifying type vacuum tubes built, a CRT is basically a small beam of electrons from an *electron gun* structure that can be focused on a phosphor coating at the faceplate. Deflection of the electron beam can be done electrostatically (as in many older oscillographs) or magnetically (as in most TV receivers). The first of the CRT deflections was electrostatic, done directly from a source since there was a lack of a wideband voltage amplifier. With the advent of lower-cost TV receivers of 10-inch and greater deflection, they were magnetically deflected. Electrostatic deflection was limited to about 7-inch picture sizes.

Original deflection angle was roughly ±3 degrees in the beginning, increasing to about ±45 degrees with color TV sets. Prior to being outsold by flat-panel TV displays, automatic production machinery made it possible to produce relatively low-cost CRTs and with triple dot-patterns of phosphors for color TV. Deflection at lower angles is proportional to the DC level of electrostatic deflection, hence they could be used as oscillographs at very fast rates.

With most CRTs in-use, the electron beam is constantly increasing in potential along its length. For electrostatic deflection the deflecting plates (a form of anode) may have up to 300 V of DC potential for small CRTs while the magnetic-deflection color TV CRTs may have up to 25 KVDC of its *ultor* or ultimate internal electrode. The ultor is formed from a conductive carbon coating on the inside of the bell or widening of the CRT.

Neon Pilot Lamps

Since before the end of WWII, pilot lamps had been standardized in bases and general sizes. The common pilot lamp (to indicate something was on) was a #47 incandescent, taking 150 mA at 6.3 VAC or VDC. By contrast, the NE-2 neon pilot lamp could do the same job with only a tenth of the electrical power of an incandescent bulb. While those were used in new equipment, they never caught on with industrial designers, despite their lower power demands.[10] Those did lead to a newer, faster, numeric indicator for a new instrument, a frequency counter.

Nixies

Originally a trademark of the Burroughs Corporation, a ***NIXIE*** was a clear glass enclosure

[10] A minor tale: When working at WREX-TV in Rockford, Illinois, the author replaced a number of neon pilot lamps in General Electric studio equipment in 1956. Engineers on day and night shifts did notice that since the older bulbs were rather used. The author was kidded about that for a long while.

containing at least 10 electrodes formed into numeric characters. Energizing just one of the electrodes would light up that one numeric and it would emit a visible glow. Some extra driver circuits could be added between electronic flip-flop counters and a Nixie such that the Nixie would show the state of the count.[11] Those would all be replaced with LED displays or other solid-state types later but a few hobbyists still like the Nixie for things like clock displays.

Magic Eyes

The first *magic eye* tube appeared in the 1930s as a tuning aid for multi-band receivers. Their purpose was to show when a station was tuned in exactly on-frequency, or at least within the passband of IF stages. A few were used in WWII military transmitters are very rough current indicators. Those worked like very small CRTs and were bypassed after WWII. Hobbyists saw a resurgence in Magic Eye tubes at the end of the old millennium but that has appeared to lapse.

Night-Observation Devices

The first *electronic eye,* or cadmium-sulfide junction, is older than vacuum tubes. Those increased in sensitivity by adding internal electron-multiplier stages in a vacuum, were used in electro-optical instruments. By the time that solid-state electronics had developed greater sensitivity, a few tube makers tried for ultimate sensitivity in the ***NOD*** or Night Observation Device, sometimes called a ***Starlight Scope***.

Digital Devices

The First Computer Circuits

Vacuum tubes were not very suited to digital devices in computing. At best, an analog computer was available as a kit from Heath Company.[12] Analog computation rather regressed in favor of the more reliable solid-state computation at the appearance of this kit. The first digital computer circuits used a combination of devices: Diodes for AND and OR gates followed by level shifting triodes to increase their outputs or to interface with the outside world. At best, tube digital circuits could reach 10 MHz counting rates (Hewlett-Packard model 524 series), taking a lot of electrical power to do so.

Non-Linear Devices as a Group

---

[11] Nixies did not come on the market immediately, were preceded by edge-lit Lucite panels to indicate numerics in the 1950s. A groove in a Lucite panel would cause an incandescent bulb (off of the edge of a panel) to illuminate the numeric carved into the Lucite. Any character could be displayed in that manner. After Burroughs came out with the Nixie, a number of different makers had their own versions, including a Great Britain version that was bi-quinary and easier to drive than single-electrode standard Nixie types. Previous to the Nixie, frequency counters used neon bulbs in a *vertical row* called a *thermometer display* (Hewlett-Packard 524A and 524B).

[12] It only lasted a couple of years in Heathkit catalogs.

The 1960s saw the greatest number of non-linear vacuum tube devices, mostly as test equipment and in video processing. Potentials in tube circuits were rather great and stray reactances limited speed. Some test equipment used special tubes (developed for the purpose) for (at the time) fast pulse forming and handling. The 1960s were also the date of introduction of practical transistor junction circuits which would take over the electronics world from then on. But, there were some oddities insofar as tubes were concerned.

Radiosondes (Weather Balloons)

A type of tube was especially designed to be made cheap and (essentially) thrown away on use. This was the ***pencil triode*** intended for weather balloons. The pencil triode was crimped into a sheet metal housing which was also its resonator and would become a UHF oscillator. Together with a single standard tube (later transistor) as a variable PRF oscillator connected to a variety of equally-cheap sensors, it would form the heart of weather balloons in the 1950s and 1960s. On the order of a half-million per year were dispatched in that time. Those formed the first of the true weather pattern mapping efforts of the world. Once into the 1970s their use would diminish as special geostationary satellites would do the mapping.

Using a one-shot battery, the radiosondes would be tracked by the receiver part of an converted radar set for positioning. As they rose under their largish balloon filled with helium, an aneroid barometer would act as a switch to both indicate altitude and to connect various sensors. They were designed for minimum damage should they be struck by another aircraft.

Small-Boat Limited-Range Radar

Those were built from some radiosondes, basically the pencil triodes used for oscillators. Their *radomes* would see later use as housings for recreational vehicle TV and FM antennas.[13] Their range was limited due to lower peak power but was enough to navigate through inland harbors and other waterways.

Other Special Tube Uses

Nuvistor

These were a miniature metal-glass envelope type of vacuum tube, most of them being triodes. They can also be thought of as a *last gasp* of the vacuum tube industry during the latter part of the 1960s against the semiconductor. Their major attribute was being small yet still had pins for replacement. It is very difficult to get a socket for a Nuvistor by the year 2010. Without a socket there isn't much need to include a Nuvistor in a new design, especially in a hobby project.

12 VDC Plate Supply Tubes

[13] Circular flat plastic housings having low height.

Another *last gasp* by the tube makers, these followed the Nuvistors. While their envelopes were mainly those of 7- and 9-pin all-glass types, their main attribute was a rethinking of the internal electron flow optimized for 12 VDC plate supplies. Many different types were made for the automobile receiver market but that bottomed out for tubes also. These do work, though.

More than One Type of Tube in the Same Envelope

These must have been the very last of the *last gasp* attempts to make inroads into the consumer electronics industry via tubes. Mainly they were a combining of two- and three separate functions, all isolated except for filaments. These were also short-lived on the industry market. Again, their separate tube sections did work well but, by the 1970 time frame, they were too late in reaching the market.

What the designers could not do is to make them replaceable easily without a socket. With so many functions and more pins than standard, they could not be cost-competitive with transistors using PCB mounting. Sockets for these tube types are mostly made of unobtainium by 2010.

5000. 6000 Series Tubes for Demanding Applications

These were improved versions of standard, registered types intended for military, mobile radio, or other harsh environments requiring greater filament life and/or more stringent *typical* curves of internal characteristics. In most cases the socket pin-outs are equal to their original types so replacement is not a problem. In general, such tube architectures were replaced (eventually) by their smaller solid-state counterparts.

A General View of Envelopes and Support Structures

Envelopes

In the USA at least, the first vacuum tubes were a hit-and-miss affair, being largely all-glass and having connections by wires. Those solidified to having bases which would plug into sockets. The first such bases were 4-pin affairs. With the internal structures including extra grids, either a plate or grid cap was placed on the opposite end of the glass tubing from the base. Eventually those were replaced with a more-standard 8-pin base called an ***octal base.*** Industrial standardization had begun. Interior structures tended to be largely spot-welded with sheet mica or ceramic insulators, vacuum retained with external connections going through metal-in-glass seals to retain a vacuum.

In the 1930s the octal-based tubes got metal envelopes, partly for shielding, partly for slightly-better heat dissipation. Glass envelopes were still favored among tube makers. By the end of the 1930s a newer envelope was developed entirely from glass. Base pins tended to remain being metal within glass, usually for 7-pin connections going to small sockets. Those were called 7-pin miniature tubes. Also developed were some non-standard all-glass envelopes fitted for more direct connections, called *acorn* tubes for their resemblance to the seed. The 7-pin bases were expanded to 9 pins to handle dual triodes, isolated except for the common filament connections. After WWII the all-glass envelope was expanded to include multiple tube structures within one envelope.

Manufacturers in the 1930s also developed ceramic-metal seals, coming to fruition after the 1930s with the *lighthouse* tubes for low microwaves and UHF. Lighthouse shaped tubes had very little spacing between internal tube elements to reduce the electron transit time effects. Lighthouse shaped tubes were of two types: Direct with the plate being at the narrow end of the lighthouse shape followed by the plate having the larger end with cooling fins added for heat dissipation.

RF power output tubes tended to be more conservative, most older designs having robust all-glass envelopes with newer types having ceramic-metal seals for the post-WWII market. The latter would survive for the longer-lived RF power amplifiers, from triodes to Klystrons for TV transmitters to TWTs for microwave work. Magnetrons tended to be large due to their (usual) integral magnet although they were essentially multiple diodes internally. Pencil triodes had the general diameter of a common lead pencil with the grid brought out to a wider disc. Pencil triodes would be crimped with a sheet-metal double cavity resonator for fixed-frequency radiosondes.

Oddities

The Nuvistor was one of the last of the post-WWII designs, intended as a replacement of solid-state transistors. Nuvistors were used in new designs for about a decade but replacement was difficult due to very small base pin-outs. The military got into the act with 8-pin subminiature tubes right after WWII for portable radios, subminiatures being half the diameter of the miniature all-glass envelope type.. The AN/PRC-6 handheld transceiver used a dozen 8-pin subminiature tubes plus one 7-pin miniature for the transmitter output tube. The subminiature type tube tended to disappear by the time of the Vietnam War and the all-solid-state (except for transmitter power output) AN/PRC-25.[14] One of the busts of military development was the one-soldier, one-radio type of *squad radio* at the start of the Vietnam War which had to be re-engineered to become practical. Even so, its early development time proved it a failure.

Photomultiplier tubes used in optical instruments were revised for Night Observation Devices and eventually took over most of the extremely low-light-level of illumination for night warfare. Those came out first as all-glass, then changed to glass-metal envelopes, eventually to all-metal and ceramic envelopes. Display CRTs were generally all-glass envelopes although a combination glass and metal envelope was tried, both for consumer and military-industrial electronics.[15]

In general, when confronted with new developments, established manufacturing tends to be somewhat conservative in trying out older constructions. The time of the vacuum tube was over by the entrance into the new millennium. The only life left for them was for replacement.

The End Result of Vacuum Tubes in the New Millennium

Except for specialty areas of NODs, communications satellite transponder relays, plus the

---

[14] The AN/PRC-77 had the same case and controls as the -25 except the transmitter power output stage used a solid-state transistor in place of the tube. Both the -25 and -77 used frequency synthesis with channelized tuning from the front panel.

[15] As CRTs got larger there was an apparent lack of attention to handling among some manufacturers, especially if the large CRT was already evacuated, hence the shift to a nearly-all-metal envelope structure. There is no other reason for an envelope to be made of metal.

radar or microwave oven output stages, there is no call for vacuum tubes in consumer electronics designs. There is a nebulous market in the niche for guitar amplifiers. That was bolstered by the ability certain types of audio distortion produced by tube amplifiers as opposed to semi-conductor amplifiers, both when over-driven. Vacuum tubes are relatively simple devices and they did usher in the wealth of electronics that dominates our senses now. But, they have a number of faults.

Basically, vacuum tubes are troubled by filament-heater burn-out. Although efforts were made in the 1970s to improve filament life, that failure is still the driving force for replacement. There are limitations to the size of electronics systems from that single failure rate. A secondary failure was losing the vacuum within due to leaks in various places of a single tube.

Size and potentials within tubes are a second limiting factor. From the laws of physics, tubes require a relatively high voltage for its electrodes. Add to that the factor that electrons will move only from cathode to plate through plate-current-current controlling grids. Amateurs tend to be conservative in naming their HF radios by old tube type numbers, largely from nostalgia and familiarity with such old type IDs. Thousands of different tube types were designed between 1910 and about 1980, all with different characteristics and uses. All were useful but the electronics market has dropped out in favor of just a few by 2010, evidenced by distributor catalogs.

The only thing left for vacuum tubes in most hobbyist projects is for replacement or re-construction. Even so, many tube types have tended to disappear.

Chapter 21

Semiconductor Basic Applications

A guide to semiconductor active device applications, this Chapter has an overall look at then as the devices of choice used for nearly all electronics in this new millennium.

A Bit of History

The first research into semiconductor technology might be taken as roughly a quarter-century before the first demonstration of radio as a means of communication. In 1874, Leipzig, Germany, Karl Ferdinand Braun published a paper on what would eventually become the first devices to enable rectification of an RF signal.

By 1879, Edwin Hall of Johns Hopkins University published a paper noting that the direction of current (and resulting Ohm's Law of Resistance) changed according to an external magnetic field. While nothing much was done about this discovery then, it became noteworthy in the next century as *Hall Effect Sensors* as special versions of semiconductors today.

In 1938 William Shockley, at Bell Telephone Laboratories, was trying out his *new idea,* to build a three-terminal device which might be able to amplify electrical signals using only semiconductor material. Not much was done during World War II days, but after, with partners Walter Brattain and John Bardeen, they made it work. Just before Christmas of 1947 the trio presented the first point-contact transistor to Bell Labs personnel and a patent for their invention started in process. That patent was granted by the middle of 1948.

At about the same time, but in 1943 in Paris, France, for a Westinghouse division there, Herbert Matare' and Heinrich Walker came out with a point-contact transistor they called a ***transitron.*** In 1948 they presented a more-polished version, presumably good for sales.[1] Not too long after, James M. Early, working at Bell Labs, came up with the *Early Effect* to explain the common transistor voltage-current curve origins.

The whole point of using dopants was to create a *free* electron or the *lack* of a free electron in the semiconductor material. An excess of electrons caused N-type junctions; the lack of a free electron caused P-type junctions. Combined with the otherwise-uninteresting semiconductor material of germanium or silicon, it now resulted in the astonishing action of controlling electron flow. Note that, by itself, germanium and silicon are only so-so insulators and so-so conductors of electron flow, hence the name of **semi***conductors.* Yes, crystalline carbon can be made into transistor junctions and has been done in the laboratory but the cost of fabrication was too far out of commercial reach to make it possible. Crystalline carbon is, of course, diamond.

What hindered these beginners in semiconductors was metallurgical instruments. Silicon has about 5×10^{22} atoms per cubic centimeter. A dopant of phosphorus (N-type) or boron (P-type)

[1] EDN magazine website blog at the beginning of February 2011 entitled *Who Invented Something Depends on Your Definition of "Something."*

can be done with just 5 x 10^{15} atoms per cubic centimeter. The difference is 1 x 10^{7} atoms, slightly less than the normal impurity from all causes in both germanium or silicon.[2] Before some new things were able to be investigated, metallurgical impurity measurements had to be much improved!

It should be noted here that, once fixed into a dopant area, the dopant will ***remain fixed***. It will not move until something catastrophic happens to a semiconductor.

Hole Flow in P-Type

The concept of having a *nothing* (lack of an electron) moving in a fixed crystalline structure bothers the author. If you can accept that, fine. We either have ***something*** (an electron) moving about in doped material or we have ***other*** electrons moving about to ***leave a hole*** there.

Temperature Effects

Movement of extra electrons (or lack of them) occurs mainly in the temperature range of about -60° C to about 180° C**.** Below -60° C there is almost no reorganization movement but by about +200° C both silicon and germanium start to become reasonable conductors.

DIODES

Take a rod of silicon and dope one end to P-type and the other end to N-type. What you get is an electron flow from N-type to P-type, but almost nothing in the opposite direction. This is the only similarity between tubes and semiconductors; one-direction conduction. There are some particular specifications that ***must*** be obeyed for any diode, from power rectifiers to small-signal types:

1. **Average Forward Current:** This determines the in-circuit heating effect with a forward current drop of about 0.7 VDC (silicon) to about 0.3 VDC (germanium).
2. **Breakdown Voltage:** Reversed polarity withstanding voltage. This is absolute.[3]
3. **Maximum Forward Current:** Applies largely to rectifier types at turn-on, initial surge to charge up filter capacitors. Usually given with a time scale for maximum current.
4. **Duty Cycle,** primarily for high-current diodes. Relates to power dissipation in forward conduction.

Chapters 22 and 31 cover power supplies. Chapter NN has specific supply information for 100 VDC regulators. Ordinary diode use is scattered throughout after this Chapter. What follows here is a brief description of some diodes for special applications from DC through VHF which are also used.

Voltage-Variable Capacitance Diodes

These are diodes designed to be used in ***reverse*** polarity such that junction capacitance is

---

[2] It is a bit like saying *a pinch of salt in a boxcar full of sugar...*or *better than 10 parts per million.*

[3] ***Zener diodes*** are designed to work ***at their breakdown voltage.*** This is the main exception from other diode types. A forward-biased zener diode is simply an ordinary diode.

inversely proportional to reverse voltage. They can be used the same as old mechanical variable capacitors for RF circuit tuning.

Zener Diodes

Again, these are used in **reverse polarity** for a specific voltage breakdown fabricated into the diode. Their use is mainly as shunt-voltage regulators for a voltage reference. They are quite stable within a certain specified zener current range. For the least noise across the reversed junction, a more complex ***band-gap reference*** is used, consisting of two or more transistors. Depending on the type, a forward-blocking current diode may be added to prevent conduction in the normal direction, limiting zener effect to the breakdown region.

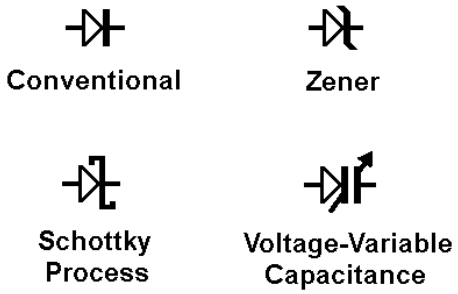


Figure 21-1 Diode Symbology

Schottky Diodes

These are diodes with junctions made in the ***Schottky process.*** That results in a lower forward-bias voltage that can extend almost to zero current. They are excellent for low supply voltage rails and specialized applications such as switching regulators. Note that the *Schottky process* can be done with digital circuitry to cut idling current almost to extinction.

TRANSISTORS

General

These are, as of 2013, three principle types: Bipolar-Junction, Junction Field Effect, and CMOS Field-Effect. In general they are all equal to vacuum tube triodes. The difference from tubes is that PNP and NPN junctions can be reversed; whereas vacuum tubes always require electron flow from cathode to anode. ***Great*** designer convenience.

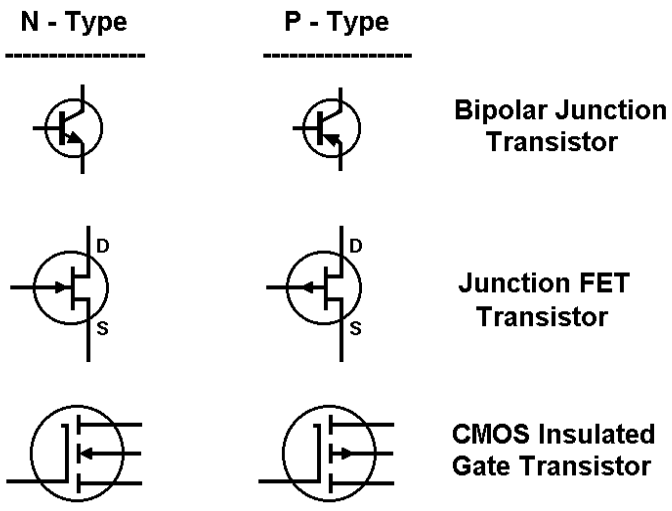


Figure 21-2 Symbology of common transistors. Note the arrows to mark type/polarity.

Bipolar Transistors

The name refers to the direction of electron flow within. An **NPN** transistor has the **emitter** connected to the lesser potential and electrons will flow from **emitter** to the **collector.** The **base** electrode controls this flow relative to the **emitter** potential and, for an NPN transistor will have its **base** acting as a diode

connected to the emitter. The collector-to-emitter equivalent circuit will be like a reversed-biased diode, the cathode at the emitter and anode at the collector.

A **PNP** transistor is the same except the electron flow is ***reversed.*** Electron current flows from collector to emitter and from base to emitter in a PNP, opposite to that of an NPN. The base-to-emitter junction will always behave as a forward-biased diode for either PNP or NPN. Only the direction of current will signify the real differences.

Note the ***arrow*** in Figure 21-2. For BJTs it follows the old *Franklin current flow*, opposite that of real electron flow. For FETs it is more like ***electron flow.*** JFET and CMOS symbolism came later.[4]

CIRCUIT APPLICATIONS, Basics

General

This begins with low signal level amplification, shifts to not-quite-so-linear high-level amplification, and finally to On-Off, non-linear digital circuits. It should be noted strongly that gain and impedance formulas given in here are ***only approximate*** for *first-guess* gains and impedances. Given the number of free circuit simulation programs now available, final gain and impedance numbers can be found more accurately through computer simulation.[5]

Three Ways to Connect Transistors

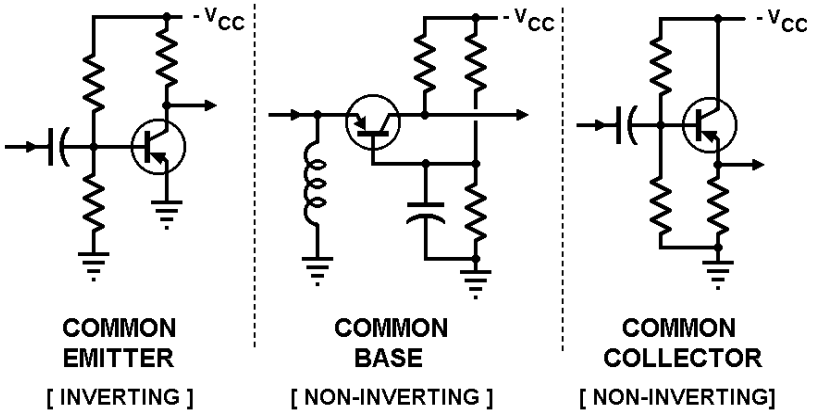


Figure 21-3 Three ways to connect a transistor.

Figure 21-3 shows the basic configurations for PNP PNP BJTs. Output has a generic resistive load resistance, from collector to supply for Common-Emitter and Common-Base while it is from emitter to ground for Common-Collector. Remaining resistors are to set the ***bias*** for determining collector current in a quiescent condition.

This applies to N-type FETs, substituting Gate for Base, Source for Emitter, and Drain for Collector. P-type FETs and PNPs require a ***negative*** voltage polarity. Note that Common-Base has an inductor to provide a DC return for the emitter.

---

[4] That may have resulted from higher-rank engineering types having to master old electronics texts where the old (reversed) *Franklin current flow* was used. [Speculation by the author]

[5] Primarily ***LTSpice*** from Linear Technology and ***TINA*** from Texas Instruments as of 2013. Free downloads and upgrades. There are variations in semiconductor specifications from manufacturers and those can be simulated fairly easily on a PC.

That return can be an inductor in the output of an L-C filter.

DC Biasing

Determining the Operating Point

This uses the ***Load Line*** method, very similar to use with vacuum tubes, to determine the DC current demand. On a set of ***collector curves,*** place a ***Load Line Slope*** such that it intersects one of the Base current lines. The Load Line ***Slope*** has an angle of the *load resistance.* Given a resistance of 1000 Ohms with 16 VDC for Vcc, maximum current is 16 mA. The slope will go from maximum current at zero collector current to maximum supply voltage with no collector current. An example set of *curves* with a *load line (dashed line)* is shown in Figure 21-4.

An ***Operating Point*** is chosen ***on*** the load line for a no-signal axis. In this example it is about 8 mA. At that *Operating Point*, collector-to-emitter voltage would be about 8 VDC. At the same time the base current can be found, very roughly at about 155 μA.[6] For this example, call it 160 μA.

In order to set the Operating Point, a transistor must be ***biased*** externally so as to make the base current close to 160 μA. This can be done simply with two resistors which ***waste*** anywhere from about 4 to 10 times the base current. The reason for such *waste* is to do some stabilization of base biasing for both β variations and for temperature. Note that resistors are generally less sensitive to thermal effects.

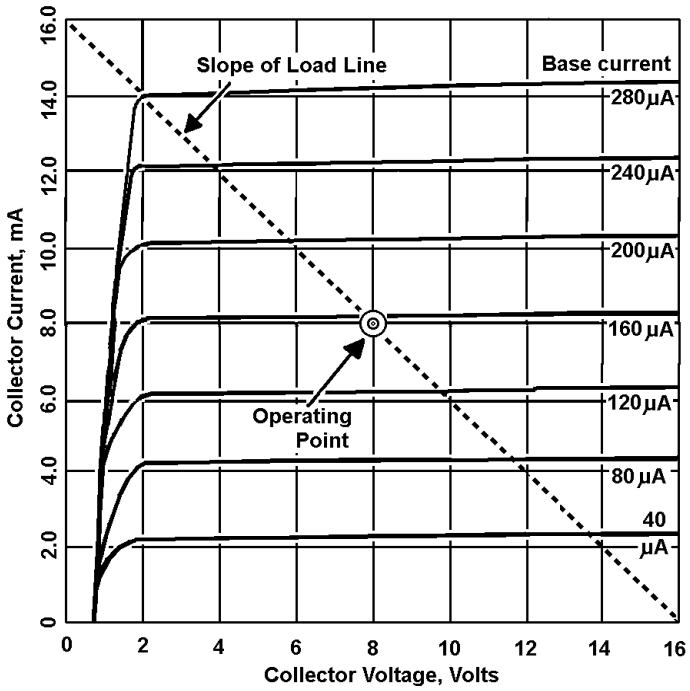


Figure 21-4 A fictitious set of *collector current curves* as an example. The *load line* may move anywhere horizontally but not vertically.

Simple Base Biasing

DC circuit given in Figure 21-5, formulas are simple:

[6] It is difficult to exactly measure base current from a collector current curve set. Curves are all ***typical*** but each transistor has a variation of Beta at DC and at an operating frequency. A datasheet for a transistor will show some other manufacturing percentage variations to confound the calculation problem as well as transistor operating temperature that also changes the forward-biased base-emitter diode portion of the transistor junction.

$$I2 = \frac{0.7}{k \cdot I_B} \qquad R2 = \frac{0.7}{k \cdot I_B} \qquad R1 = \frac{R2 \cdot (Vcc - 0.7)}{I_B \cdot R2 + 0.7}$$

k is approximated to anywhere from 4 to 10

Making R2 = 1 KOhm and then R1 = 18 KOhm would make I_1 = 850 µA and I_2 = 700 µA. I_B would then be the difference or 150 µA. That is fairly close to a correct Operating Point.

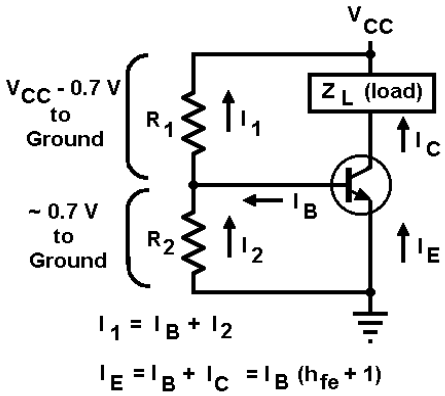


Figure 21-5 Simple bias circuit.

Self-Bias With a Series Emitter Resistor

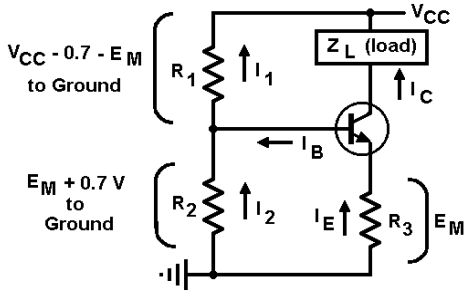


Figure 21-6 Self-bias circuit.

Figure 21-6 shows a third resistor in series with the emitter. At DC this provides a bit of negative feedback by trying to stabilize the base current. A problem is that there is the voltage across R3 which must be factored into the Simple bias circuit of Figure 21-5.

This begins by picking Em first, then I2 for a value of about 4 to 10 times Ib, finally R1, all within a manufacturing-tolerance resistor value limits. This is not a super-formula method but it is simpler and may be done fairly easy on a handheld calculator. The approximate formulas:[7]

$$E_M = R3 \cdot I_B \cdot (\beta + 1) \qquad I_2 = \frac{(E_M + 0.7)}{R2} \qquad E1 = Vcc - E_M - 0.7 \qquad R1 = \frac{E1}{(I_B + I_2)}$$

Voltage across the emitter resistor, R3, can be picked to (initially) be about 10 to 15% of the load resistance, RL. Choosing 150 Ohms would make Em about 1.224 VDC with 8 mA Ic and a β of 50. Adding the base-emitter voltage drop of 0.7 VDC makes the total of 1.924 VDC. Choosing R2 to be about 3.9 KOhms, I2 would be 493 µA. Summation of I_2 and I_B equals 653 µA. Finally, with a 22 KOhm R1 (rather than a calculated 21.54 K) the circuit settles in approximately correct.

Variations for FETs

While Figure 21-3 shows BJTs, the Bias circuits will apply to FETs. Only the electrode connections change names. Depending on the schematic form, resistors external to the transistor will have the same connections. As to configuration names, the following is worthy of note:

BJTs		FETs
Common-Emitter	=	Common-Source
Common-Base	=	Common-Gate
Common-Collector	=	Common-Drain

[7] These simple formulas are ***not exact.*** For an exceedingly exact group, see Agilent Application Note 1293 *A Comparison of Various Bipolar Transistor Biasing Circuits,* document 5988-6173, 22 January 2003. Available free over the Internet at www.agilent.com. While those are very nice, the number of (sometimes) hidden variables can befuddle their exactness for hobbyists. The best way to check your bias circuit is to load it into LTSpice and do a test for currents and voltages. That comes as close to reality as tens of thousands of dollars of fine equipment.

Note: FET Gates have very low current levels, hence FET Bias resistors can be selected as simple voltage dividers, ignoring Gate current.

Large- and Medium-Signal Gains and Impedances

This can be approximated directly from Load Lines but is quicker and easier to do with a SPICE program. A feature of both LTSpice and TINA (both free programs for PCs) is the ability to discern the level of harmonic generation from non-linearity. The ***magnitude*** of input impedance can use a single series resistor, adjusted in value until the input voltage is half that of the voltage source. Input resistance is then equal in magnitude to the series resistor value. Approximately the same simple technique can be used for output impedance, varying the final load resistance until the output voltage is about half of the no-load-resistance voltage.

Note that the ***magnitude*** is found, not the explicit (so-called perfect) value. There will be some transistor and other wiring capacitance in parallel which will reduce the *magnitude.* That can also be modeled but the amount of effort increases by roughly the cube over using a simple resistor in series method.

Small-Signal Analysis

General

Small-Signal levels are those measured in milliVolts or microVolts peak-to-peak. What is lovely about a PC analysis and design program is that it doesn't care about the absolute levels. To such a program it is merely a number entered for amplitude of a source generator. This makes it ideal for input stages of a receiver as an example. Harmonic content can still be calculated so it is possible to predict things like mixer spurious products or the interference from same. Such features are virtual spectrum analyzers on the screen.

In order to get a circuit model requires a basic circuit. What is given following are the simple ***approximation*** formulas for Gains and Impedances. Individual component values can be changed to fit the overall circuit. Having a circuit model allows fairly simple component changes to see their effect on overall performance.[8] You will **NEED** a model for accurate system answers.

In the Bias of BJTs the simpler R_{PI} value comes up.[9] This approximate value is based on the Base-Emitter current derived from the Operating Point. One takes the Base voltage relative to the Emitter and divides that by the Base current. In the BJT example given previously, the Base voltage to ground was 1.924 VDC. Dividing that by 160 μA Base current yields 12.025 KOhms.

The double vertical bars denote ***parallel connections.*** For resistors the total parallel value is the reciprocal of the sum of all individual reciprocals. On a scientific calculator with a **1/x** function key, that is very quick and easy to do.

---

[8] That cannot be stated enough. Computer-aided circuit analysis came into wide use in the 1970s and has grown ever since. It is ***ideal*** for the hobbyist...an unlimited number of parts which can be changed easily in order to optimize a circuit. The only problem is the learning curve needed to use such a program.

[9] That value is pronounced ***R-pi***. That name is unusual in that the symbol of ***pi*** was not used, rather to spell it as ***pi*** to avoid conflict with other variables given here and later.

Small-Signal Frequency

For resistive loads, the initial frequency is taken to be ***mid-band.*** This will be about 1 KHz for audio circuits. For ***tuned*** loads, such as parallel-tuned L-C circuits or filters, the initial frequency would be in the ***center of the passband***.

There are many different frequency-shaping effects with most circuits.[10] For initial calculations one can assume that coupling capacitors, emitter bypass capacitors, and supply rail filtering become short-circuits at this initial gain and impedance calculation. Once this initial calculation is done, it becomes the baseline for *other frequency* gain reduction and impedance variation.

Small-Signal Levels and Linearity

At input levels of 100 μV to 10 mV peak-to-peak the amount of distortion is about the same as if input levels 100 times higher. There are ways of reducing non-linearity; op-amps are the classic case for this. ***Negative feedback*** is the general way to reduce distortion, but at the expense of lesser voltage gain as compared to a circuit without such negative feedback. Using a series Emitter (or Source) resistor without any parallel bypass capacitor is the classic case.

Once a circuit is modeled with a PC program, the PC analysis program will have a form of *Fourier Coefficient* calculator built-in. This will show both the harmonic content and allow a quick calculation of ***Total Harmonic Distortion.***

APPROXIMATE Small-Scale GAINS, IMPEDANCES

General

These all assume the ***Load*** is ***resistive.*** For a reactive Load, use the ***magnitude*** of impedance as if it were a resistance at that mid-band-frequency. R_{PI} is the DC bias into the Base as mentioned.

Common-Emitter, no Emitter Resistor

Voltage Gain $= \dfrac{\beta \cdot R_L}{R_{PI}}$

Current Gain ~ β

Input Impedance ≈ $R_{PI} \parallel R1 \parallel R2$

Output Impedance ≈ R_L

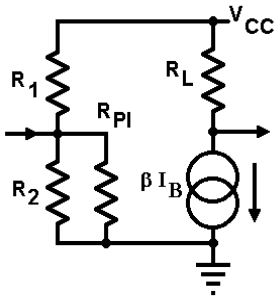


Figure 21-7 Model of Common-Emitter, no series resistor.

[10] There are enough variations to fill another book. For hobbyist work, it is quicker to approximate on the mid-band frequency, then model it on a PC program and refine the upper and lower frequencies that way. This doesn't change the Operating Point (set at DC) and the mid-band frequency characteristics are good enough to yield an overall ***Gain*** **Budget** for a hobby system.

Common-Emitter With Unbypassed Series Emitter Resistance

The analytical schematic (Figure 21-8) is the same as Figure 21-7 except it has a series resistance in the Emitter lead, R3. This requires a change in the Operating Point as well as the values of R1 and R2. As a result of R3, the voltage gain drops and the input resistance increases. Collector load is assumed to be the ***parallel*** of R_L and the external (driven stage) input resistance.

$$\text{Voltage Gain} = \frac{\beta \cdot R_L'}{R_X} \approx \frac{R_L'}{R_3}$$

$R_L' = R_L \,||\, R_{LOAD}$ [R_{LOAD} is Driven Stage]

Current Gain

$R_{12} = R_1 \,||\, R_2$

$$\text{Input Resistance} = \beta\left(\frac{R_L}{R_L + R_{LOAD}}\right)\left(\frac{R_{12}}{R_{12} + R_X}\right)$$

$= R_{12} \,||\, R_X$ $\qquad R_X = R_{PI} + (\beta+1)R_3$

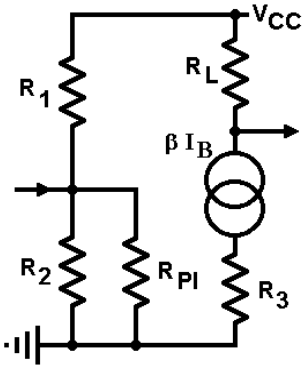


Figure 21-8 Analytical model of a Common-Emitter *With* a series Emitter resistance, R3.

R3 will reduce the voltage gain. As the input signal goes more positive, the Emitter voltage will also go (slightly less) positive, thus, for the same output voltage, the Base must go more positive than in the case of no series Emitter resistance. There is less loading on the input driver since the input resistance has risen above that of the stage without an R3.

Common-Collector or *Emitter Follower*

$$\text{Voltage Gain} = \frac{(\beta+1)\cdot R_{LL}}{R_{PI} + (\beta+1)\cdot R_{LL}}$$

Input Resistance = $R_1 \,||\, R_2 \,||\, R_{PP}$

Where: $R_{LL} = R_3 \,||\, R_L$

and $\quad R_{PP} = R_{PI} + (\beta+1)\, R_L$

$$\text{Output Resistance} \approx \frac{R_S + R_{LL}}{\beta+1}$$

Where: $R_S = R_1 \,||\, R_2 \,||\, R_{SOURCE}$

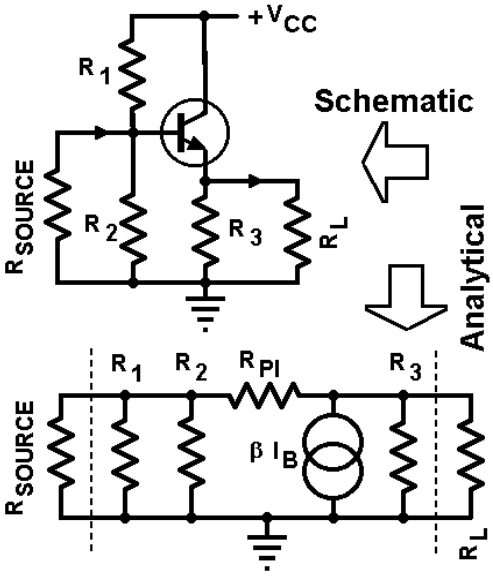


Figure 21-9 Schematic, Analytical diagrams of a Common-Collector or *Emitter-Follower.*

Voltage gain of a Common-Collector is in the normal range of 0.7 to 1.0. Input resistance takes a big jump over a Common-Emitter while output resistance falls. This is commonly used as a non-inverting buffer

such that output voltage tends to ***follow the input***, hence the colloquial name of *Emitter-Follower.*

In the circuit of Figure 21-9 note the additions of $\boldsymbol{R_{SOURCE}}$ and $\boldsymbol{R_L}$. This is because of their effect on output voltage. While there is little effect except at low values, it was mathematically impossible to divorce them entirely from these approximations.

Common-Base

This configuration has very low input resistance, moderate to high output resistance, power gain but the current gain is less than β. It is good for matching a low-Z source to high-Z. Miller Effect is very slight so wideband circuits can go to transition frequency.

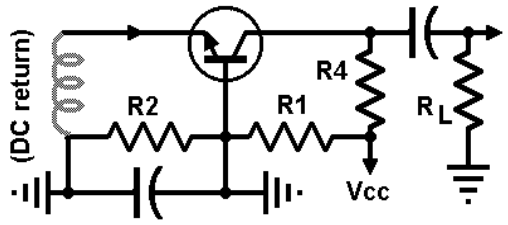


Figure 21-10 Common-Base. Shaded inductor represents a DC return for the Emitter to ground. R1, R2 bias for Base are same as for Common-Emitter.

Output Voltage Gain = gm (R_4 || R_L)

Input Resistance = R_{PI} / (β+1)

Output Resistance = R_4

Current Gain ~ β / (β+1) = α

Miller Effect is not present on Grounded-Base.

Variations For FETs

Since FET Gates operate more from Gate voltage to ground, the common figure of merit is ***transconductance.*** *Transconductance* is the small-signal ***rate of change of Drain current due to Gate-Source Voltage.*** It is abbreviated to ***gm*** and measured in ***mhos*** (Ohms inverted) or ***Siemens*** (the same thing but given a different name). For some clarity, ***gm*** is used here since it was present in vacuum tube analytical models first.[11]

There is very little Gate current relative to BJTs, enough that Bias resistors can ignore Gate current. There is ***no*** $\boldsymbol{R_{PI}}$ in the FET circuit models. Some FET characteristics show some thermal effects with Gate voltage; depending on the specific circuit, those may have to be dealt with.

Gate Bias can be calculated directly as voltages at the Gate. Bias resistor strings can be up to 250 KOhms, thus resembling vacuum tube amplifier stage inputs for Common-Sources.

Common-Source FET, no Series Source Resistor

Similar to Figure 21-5, the analytic circuit is shown in Figure 21-11. There is no current connection between input and output.[12] DC Bias voltage, Gate-to-ground, depends on the Operating Point and/or datasheet values. R_D is equal to reciprocal of gm, mhos of gm turning back to Ohms.

---

[11] See Chapter 20 on vacuum tube basic data for transconductance.

[12] There is, but it is confined to Gate-Drain inter-electrode capacitance and affects high-frequency gain.

Output Voltage Gain = $gm\ R_L'$

$R_L' = R_4 \,||\, R_D \,||\, R_{LOAD}$

Input Resistance = $R_1 \,||\, R_2$

Output Resistance = $R_4 \,||\, R_D$

This is a fairly straightforward configuration which is easy to calculate.

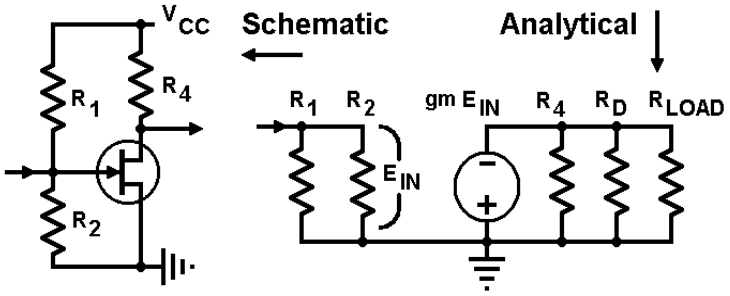


Figure 21-11 Schematic and analytical circuits for Common-Source FET amplifier.

Common-Source *WITH* a Series Source Resistor

This adds R3 between ground and Source electrode. Voltage gain for ratios of R4:R3 of about 4:1 becomes, approximately:

Voltage Gain $\approx \dfrac{gm \cdot R_L'}{1 + R_3 \cdot gm}$ Input resistance about same as no Source resistor.

Common-Drain or *Source-Follower*

This is a slight variation on the typical voltage-follower in that the Gate is Biased slightly Positive relative to the Source. For a P-Type FET the Gate would be Biased slightly negative relative to the Source.

While this has a slightly lower voltage gain, the added negative feedback provides less distortion over the conventional voltage-follower with R2 returned to ground. Typical voltage gains would be 0.5 to 0.8 over the signal source. Needed reciprocals shown in the Figure.

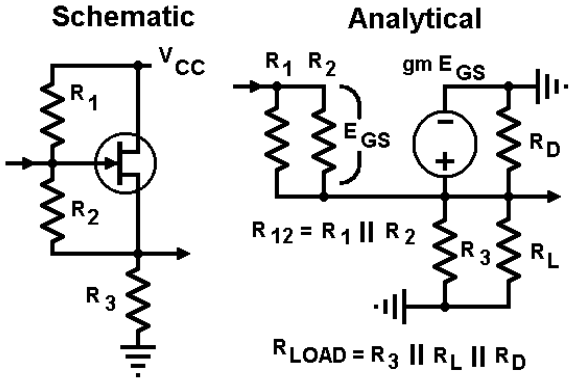


Figure 21-12 Schematic and analytical forms of the *Source-Follower*.

$$\text{Voltage gain} = \frac{\left[gm + \left(\frac{1}{R_{12}}\right)\right] \cdot R_{LOAD}}{1 + \left[gm + \left(\frac{1}{R_{12}}\right)\right] \cdot R_{LOAD}}$$

Input Resistance = $R_{12} + [(1 + gm\ R_{12})R_{LOAD}]$

Input resistance can become **very** large due to the second term-group multiplication.

Using Two or More Transistors

Darlington

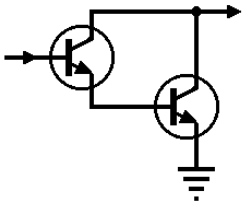


Figure 21-13 A Darlington pair.

Shown in Figure 21-13, the ***Darlington*** pair was an early method to increase both β and input impedance. Biasing requires twice the Base-Emitter DC voltage and total β is the multiple of all βs. Its name derives from its inventor, Sidney Darlington (1906 - 1997) who received a patent on it at Bell Telephone Labs in 1953.

Sziklai Pair

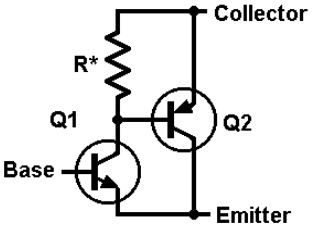


Figure 21-14 Sziklai pair

George C. Sziklai invented a replacement for the Darlington pair in 1957. Seen in Figure 21-14, it works the same as a Darlington. The resistor (marked R*) is there to handle some of the Collector current from Q1. While not much used and of no appreciable advantage, it shows the interchangeability possible between NPNs and PNPs in the same circuit. Connections of Base, Collector, and Emitter refer to the Darlington circuit.

BJT Cascode

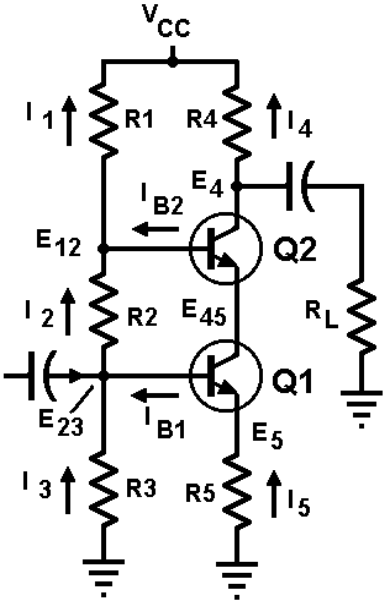


Figure 21-15 NPN Cascode.

Another tube circuit relic revitalized with transistors, the BJT transistor version of Figure 21-15 has a Common-Emitter stage, Q1, directly connected to a Grounded-Base stage, Q2.[13] Frequency response is relatively immune to *Miller Effect*, thus bandpass is the widest.

Emitter-Collector current flows through ***both*** Q1 and Q2. As a variation, the Base of Q2 may be connected to another DC source as a current/gain control.[14] To begin, one has to assume an Operating Point somewhere to determine Emitter-Collector current. Initially, $\boldsymbol{I_4 \approx I_5}$ and the following are true:

$$E_{23} - E_5 = 0.7\ V \qquad E_{12} - E_{45} = 0.7$$

$$V_{CC} > E_4 \qquad E_4 > E_{12} \qquad E_{45} > E_5$$

Current through the resistor string of R1, R2, R3 should be about 4 to 10 times the Base current. Voltage to ground at E4 is around 2/3 of V_{CC} to insure an adequate AC peak-to-peak output voltage swing. Emitter-to-

[13] Etymology of ***cascode*** is uncertain. Used with tubes in the 1950s, it may be a corruption of cascade.

[14] An article in QST, December 2007, entitled *The Hybrid Cascode - A General Purpose AGC IF Amplifier,* by Hayward and Damm, for a 100 db gain control in a superheterodyne receiver last IF. It used FETs for Q1 and BJTs for Q2 in each stage..

Collector voltage of both Q1 and Q2 should be greater than about 0.3 V to avoid clipping at AC. Note that R5 may be deleted. If so, E23 is 0.7 VDC. The capacitor at input, going to E23, is there as a reminder that, at DC, capacitors are open-circuits at DC. Choosing an Emitter current and setting I3 at about 10 times Base current is a beginning point; let Vcc = 18 VDC, Emitter current equal 2 mA, β equals 120 and Base current ~ 16 2/3 μA:

Let R5 = 1.2 KOhms so that E5 ~ 2.4 VDC (all or part of R5 can be bypassed for AC)
E_{23} ~ 3.1 VDC and, if I_3 ~ 167 μA, then R3 ~ 18 KOhms
E4 would be roughly 2/3 of Vcc so R4 can be about 2.7 KOhms for E4 = 6.6 VDC
E_{45} will then be somewhere between about 7 VDC and 4 VDC
Choosing E_{12} at about 5.3 VDC would make R2 ~ 12 KOhms (I_2 is ~ 184 μA)
As a result, E_{45} is about 4.6 VDC (holding to 5.3 V - 0.7 V)
(V_{CC} - E_{12}) = 6.7 V and I_1 is about 200 μA; R1 will be about 33 KOhms.

For AC mid-band, the junction of R1, R2, and Q2 Base is bypassed to ground by a capacitor. The Emitter of Q1 may go directly to ground (no R5) or it may be split into R5A (smaller) and R5B (larger), both in series. Mid-passband responses:

Voltage Gain = $R_L / (r_E + R_{5A})$ $\quad R_L' = R4 \parallel R_L \quad r_E = (26 \times 10^{-3}) / I_5$

Input Resistance = $R2 \parallel R3 \parallel R_N$ $\quad R_N = \beta (r_E + R5_A)$

Output Resistance = *R4*

Given the previous values with R5 split so that R5A = 56 and R5B = 1.2 KOhms, R_L = 10 K, the Voltage gain will be 30.8 times, Input Resistance = 3850 Ohms,
Output Resistance = 2700 Ohms.

FET Cascode

The FET version of the cascode is given in Figure 21-16. Most statements of the cascode of Figure 21-14 apply here but with some exceptions. Q2 Gate is bypassed to ground but can be used as a voltage-driven gain control.

For AC mid-passband small-signal operation, the following applies:

Output Voltage Gain = $gm (R4 \parallel R_L)$

Output Resistance = *R4*

Input Resistance = *(R2 || R3)*

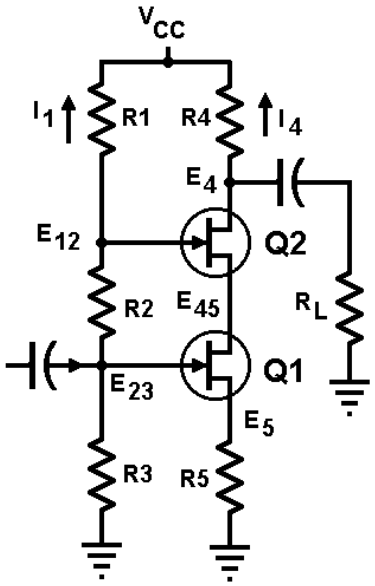


Figure 21-16 An FET Cascode, similar to Figure 21-14.

Differential Amplifiers

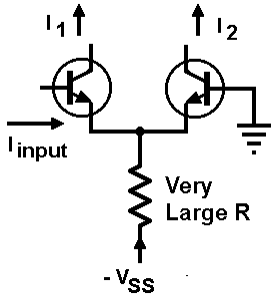


Figure 21-17 Ideal Differential circuit.

Figure 21-17 shows an ideal or theoretical differential amplifier which can create two outputs, each output having the opposite phase of the other. With a high-potential negative supply and a Very Large resistor common to both Emitters, a near ***constant-current*** supply goes to both Emitters. If both NPNs are the same type and the Base of the right-hand transistor is to ground, an input current to the Base of the left-hand NPN is amplified as I_1 at its Collector but with *current phase inverted.* With a constant-current Emitter supply, that amplified input current appears at I_2 with the ***same current phase.***

With equal Collector loads, the theoretical circuit can be a good driver for a push-pull audio amplifier or a building-block of many other ICs, achieving amplification and phase inversion in a simple circuit. A problem is that Very Large resistors are hard to make on ICs. The answer is a ***constant-current circuit.***

Constant-Current Circuits

Figure 21-18 (A) shows the simplest constant-current supply. With a high β value transistor, the Collector current curve has a wide voltage range with little variation in Collector current.

Figure 21-18 (B) is more practical in that the resistors can be of lower value. The transistor Base in (B) can be held more tightly to the Zener voltage and the Emitter series resistance can be laser-trimmed to set the constant-current output source more exactly.

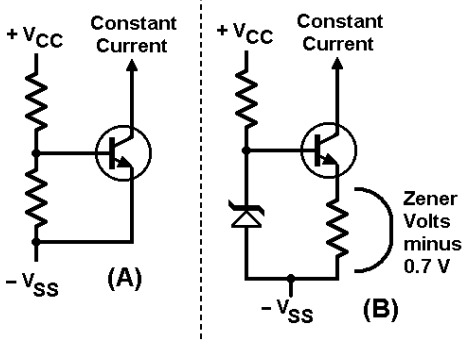


Figure 21-18 Constant-current circuits to eliminate large resistor values and large negative supplies

Current Mirrors

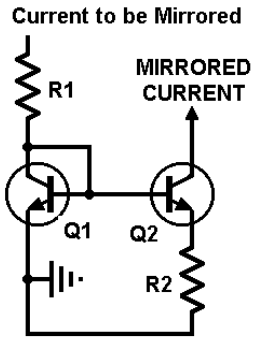


Figure 21-19 Rudimentary Widlar Current Mirror.

Current mirrors can provide a current that is proportional to an external resistor setting exceeding a 1,000:1 maximum-to-minimum range. They are uniquely a product of the semiconductor age and are built into a surprising number of ICs. Because they are built into ICs, it is difficult to make them out of discrete parts. They are worth mentioning because of their descendants that include (seemingly) simple temperature measurements and the ultra-stable ***bandgap*** voltage references.

The *Widlar Current Mirror* shown in Figure 21-19 is the first widely used circuit, born about 1965 by Robert J. Widlar, then at National Semiconductor.[15] According to legend, it grew out of a need to *not* make resistors on a chip greater than about 50 KOhms.

The rudimentary form at left shows Q1 to be connected as a diode, as if this diode anode were toward V_{CC}. In reality, base current through Q2 is smaller than through Q1 emitter current. It should be noted that such low

[15] IEEE Transactions on Circuit Theory, CT-12 (1965), pp 586-590

currents involved put this into the left side of normal transistor currents, the so-called *Early Effect.* Note: This is ***not*** in the normal amplifier load-line area of collector curves; it is in the left-hand edge where the collector current is beginning to rise to reach the stable lines.

After much calculation, the Widlar Current Mirror can have its component values picked and the circuit will be rather stable with temperature. There are variations on the Widlar circuit, notably the Wilson Current Mirror. All such came after the Widlar circuit.

Differential Amplifiers with 2 Signal Inputs

Given that a differential amplifier has a constant-current common Emitter supply, and that the constant-current value can be adjusted, it makes sense to combine everything in one IC as a ***Mixer*** or as a ***voltage multiplier.*** A Mixer is shown in Figure 21-20, the RF signal input as either push-pull or single-ended, the Local Oscillator (or AM source) as single-ended. This combines Figures 21-15 and 21-16(A).

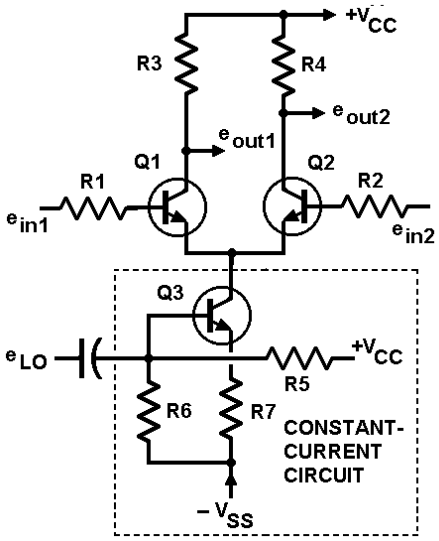


Figure 21-20 A Mixer.

The differential portion is Q1, Q2, and R1 through R4. The constant-current source is Q3, R5, R6, R7. If made in an IC the LO capacitor would probably be external. Using capacitive-coupling for RF inputs, this can be done using a single supply for Vcc with -Vss being ground.

Multiple inputs can be made, the only limit being the supply voltage range. With two differential RF inputs, this form is often called a ***Gilbert Cell.*** *Gilbert cell* mixers feature a high attenuation of each RF input so that output filtering can concentrate on the unwanted mixing product.

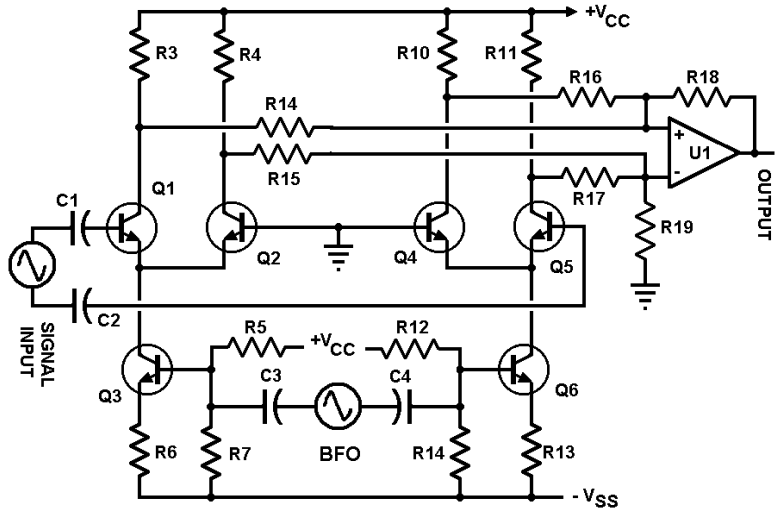


Figure 21-21 A two-input Gilbert Cell organized as a SSB receiver demodulator with signal input pre-filtered at RF, sum-mixing product attenuated by op-amp U1.

One specialty version is given in Figure 21-21 with full differential dual inputs. Some of the sum-mixing product is attenuated in U1, a general-purpose op-amp. Further sum-product attenuation can be done by following lowpass filtering.

A Gilbert Cell structure good from low audio to 60 MHz is the MC1350, originally from Motorola, now from Lansdale Semiconductor. A later version is the SA602 and SA612, each with a LF to VHF transistor structure for the Local Oscillator from NXP.

Some Different Semiconductor Structures

Silicon-Controlled Rectifiers or *SCRs*

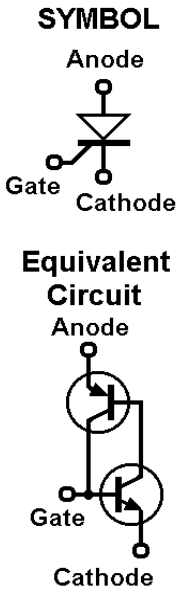


Figure 21-22

Silicon-Controlled Rectifiers, as a term, began at General Electric in 1957 as a ***four-layer diode*** off-shoot of a category of unusual semiconductor formations called ***thyristors.*** Originally invented by William Shockley in 1950, this four-layer-diode was later promoted by GE. They are niche-applications principally for AC power control. Their action is much like a flip-flop circuit in that they are inert up to a certain voltage, then when this voltage is exceeded, they conduct in a saturated manor and remain conducting until voltage is removed. Figure 21-22 shows the symbol and an equivalent circuit. An equivalent would require high-breakdown-voltage ratings above the peak-to-peak voltage of an applied AC signal. ***Triacs*** are dual SCRs in one 3-lead package specifically to handle AC. As part of this family, ***Diacs*** are a form of dual diodes to couple *Triac* Gates to an external phase-control circuit. All are going out of style for lower-power units, replaced by three-terminal regulators and switching supplies.

Unijunction Transistors

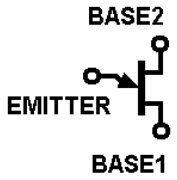


Figure 21-23 Unijunction.

Not exactly a combinatorial structure, this three-terminal device gets its ***uni*** from the control of main current via its Gate going having a positive, then negative, then (finally) positive in magnitude. Its symbolism is shown in Figure 20-21. Originally used for (inexpensive at the time) monostable and astable multivibrators, it too has gone out of style in newer designs. At best, a ***unijunction*** transistor is good for simple, beginning-instruction bench learning; used in a circuit, it can be replaced by a 555 Timer package with more-consistent, tighter-tolerance results and for less cost..

Example TYPE-COMPLEMENTARY Structures

Semiconductor *TYPE* Mixing

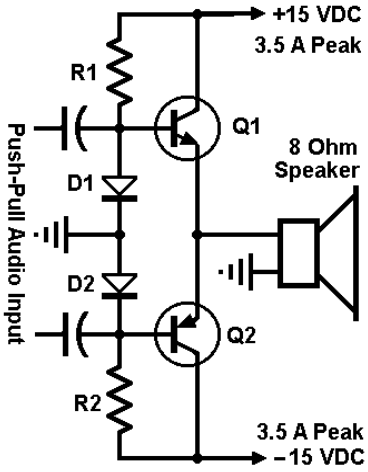


Figure 21-24 A symmetric 20 Watt speaker driver.

Using both NPN and PNP transistors (or N-type and P-type FETs) in the ***same*** circuit has been done for quite a while. It can be done in audio amplifiers driving low-impedance speakers ***without*** a transformer. Figure 21-24 illustrates such.

This has an NPN power transistor for Q1, a complementary PNP for Q2, with an 8 Ohm speaker having its return lead grounded. D1 and D2 match the Base-Emitter drop of Q1, Q2. R1 and R2 provide the Base bias current for Q1, Q2 and diodes D1, D2. R1, R2 should be fairly close in value to preserve symmetry and reduce distortion.

DC current normally flows from the -15 VDC supply up through the circuit to the +15 VDC supply Base bias can be set to

approximately Class AB (although closer to Class B). This is essentially a large voltage-follower to drive the speaker. Audio input can be another voltage-follower for less power, that coming from a lower-power differential amplifier.

At 20 Watts RMS and 8 Ohms impedance, the speaker will *see* 8 V RMS at 2.5 A RMS. Peak-to-peak voltage is ± 11.3 V with peak current at ± 3.5 A. The speaker will get positive swings from Q1 and negative swings from Q2. The advantage here is to have the speaker very close to ground potential for its return. Having that output ground return is an advantage in using negative feedback to improve linearity since the feedback is very close to ground level.

This can also be done a much lower power levels (such as to headphones) without an audio transformer since there are more complementary NPN-PNP combinations available in lower-current BJTs.[16]

Combinations in Both Built Circuits and ICs

Since complementary transistor and FET types may be purchased as symmetric equivalents, they can be used at will. For IC dies, it may be just a matter of changing photomasks and production steps to have both N- and P-types on the same piece of silicon. The main thing to remember is the initial DC current flow. That determines Base or Gate bias. The next thing is the AC characteristics to fit the intended use. Such complementary type use is just not possible in vacuum tubes.

Operational Amplifiers

The term *Operational Amplifier* was coined back in World War II days, describing simple vacuum tube circuits that could be used by itself (as in an analog computer) or as part of a gun-laying or autopilot system. Once transformed into silicon, its uses became legend in nearly ***all*** disciplines of electronics. Chapter 30 covers both Voltage-Feedback and Current-Feedback ***op-amps.***

Digital Logic Circuits

Digital circuits are mainly ***on-off switches*** in one sense but can be made with or without complementary symmetry sub-circuits. Chapters 23 to 29 cover most ramifications of digital circuitry in this book collection.

Hall Effect Sensors

There are many and varied sensors, usually simple in their principles, no assigned symbolism. These are used for sensing rotary motion (as in electric motors) to accelerometers for one to three degrees of freedom as small gyroscope substitutes.

---

[16] Complementary NPN-PNP lower-power BJTs are the 2N3904-2N3906 and 2N4124-2N4126. The author has used this for a 10 W speaker driver and as a 100 mW headphone driver. For a 16 Ohm headphone driver at 100 mW, the RMS speaker values are 1.27 V and 79 mA so the circuit would work with ± 5 V supply rails.

WHAT NEXT?

What has been obtained from this Chapter are just ***approximations*** of circuit values. Now it is time to go to a PC and SPICE-model circuits for more accurate estimations. Such circuit models allow all sorts of component variations without having to order anything. Once that is done a better parts list can be drawn up and your project physically begun.

The *in-between* task for the hobbyist is to ***analyze the circuit*** in detail on a Personal Computer. Both LTSpice (Linear Technology) and TINA (Texas Instruments) are available for free downloads over the Internet.[17] Most modern PCs can handle ***large circuits*** with either program and do frequency-domain and time-domain analysis. They can also do Fourier Analysis of time-domain signals to find out the magnitude of harmonics.

[17] Exact web site URL is unknown for the future. That is why both web sites have search functions. That will get you to the correct program download link.

Appendix 21-1

A Glossary of Microwave-Specific Semiconductors

MMIC

An acronym for ***Monolithic Microwave Integrated Circuit***, these may be any block of transistors or diodes designed by a manufacturer. Generally they refer to MMICs that operate at or above 1 GHz. Some may operate much farther down. Check their product descriptions.

Esaki -or- Tunnel Diode

Invented by Leo Esaki in 1958, this diode is specifically processed to include a ***negative slope*** within their forward-conduction voltage-current curve. This was the apparent start to several types of diodes with negative slopes. Esaki became a co-holder of the Nobel Prize in 1973 as a result of his work.

Gunn Diodes

Discovered by J. B. Gunn in 1962, it is another negative-slope diode, useful, as is the Tunnel Diode in making simple, low-power microwave sources. Anyone entering an establishment with an automatic door has probably been RF-sourced by an Esaki or Gunn diode-based oscillator..

PIN Diodes

Another specifically-processed diode, these are used for microwave switches, attenuators, and limiters. Voltage-current curves have ***no*** negative slope.

IMPATT Diodes

The name is an acronym for **IMP**act ionization **A**valanche **T**ransit **T**ime. A higher-power diode, it has been used as 3 KW source up to 100 GHz. Good as a frequency multiplier, higher-power oscillator.

Step-Recovery Diode -or- SRD

This one can work at lower frequencies, designed specifically for emphasizing harmonics such as in (frequency) ***comb-generators.***

Lambda Diode

Built up by a P-type and N-type JFET, Sources and Drains connected, it can operate at slightly higher frequency than a Tunnel or Gunn diode. Not much used, it has a portion of its voltage-current curve sloping negative.

Resources for Chapter 21

As of 2013 there were dozens of resources for modeling and analysis of semiconductors on the Internet. Too many to list here. The following give some of those resources with comments from this author on their efficacy.

[58] ***Introduction to Electronics,*** by Bob Zulinsky, [free download] 254 pp. A well-thought-out work that gets a bit deep in theory but details actual design equations. ***www.ece.mtu.edu/faculty/LJBohman/onlinetext/elnt200.html***

[59] ***Electronic Applications,*** by Bob Zulinsky, [free download] 218 pp. Another one which covers higher-speed design of transistors and DC-DC converters. ***www.ece.mtu.edu/faculty/LJBohman/onlinetext/elapp200.html***

Zulinsky is (or was) an Associate Professor at Michigan Technical University. He presents his work in a large, simplified form, showing how final formulas were derived. While both are free downloads, he asks for donations, something that is well worth it for those who are serious about electronics.

[60] Almost any book from 2000 onwards authored by Sedra and Smith, both at the Toronto, Canada, university. These are also well thought-out books with many examples. Prices can vary depending on the seller, ranging from about $35 to $515, with or without a CD-ROM disk. Newer ones carry the latest devices. Working title is ***Microelectronic Circuits: Theory and Applications.*** Principal authors are Kenneth C. Smith and Adel S. Sedra. Some editions have a third author, names varying per edition year.

There are, literally, many dozens of varying papers, essays, and lectures on the Internet for free perusal. Many were used to compare the succinct design equations given in this Chapter. Some are ultra-simple and some are almost mystical in their pursuit of mathematical perfection. There are a few in the middle category which should fulfill your theory needs.

Chapter 22

Basic Power Supplies

Power for your project, from simple to complex depending on use, stability, and accuracy.

General

Power supplies come in two major forms: AC-line-powered and battery. Prior to about 1950 most electronics were either powered by the AC line or from carbon-zinc dry batteries, the latter usually provided in packages having more than one voltage, overall designated as ***portable***. These batteries were not generally rechargeable so they were replaced at intervals depending on use of this portable electronics device.

With the coming of the semiconductor active devices, along with low current demand, portable power called for batteries, usually only one voltage. Perhaps the first portable power supply was a collection of several single 1.5 VDC carbon-zinc *flashlight batteries* in series. The next portable power source was the 9 VDC snap-contact battery pack, it being a collection internally of 6 carbon-zinc batteries spot-welded in series. Again, these were not rechargeable until electrochemistry came out with ***rechargeable*** battery packs, the first being made with nickel-cadmium or ***Nicad*** structures. Rechargeable batteries could now be reused and many electronics equipments came with simple chargers.

AC line-powered power supplies usually had a form of ***isolating*** step-up and step-down secondary windings to fit the required DC output voltages. An exception was the simple AM BC receiver designed for ***AC-DC*** power input. This transformerless tube receiver had no power transformer and took 115 VAC RMS directly for series-connected filaments and for about +120 VDC plate and screen supplies. After the semiconductor era was established, most AC line-powered supplies were all step-down secondary windings for transformers.[1]

All of the AC line-powered power supplies (and built-in chargers) described following are for total power demands less than about 300 Watts. That makes it possible for the hobbyist to rewind transformer cores to fit. Those who wish to copy power sub-station transformer designs can find design information in older texts. Those who wish to rewind their own power transformers are directed to Chapter 16 for specific transformer winding information.

[1] This explains the change-over from so many multiple-secondary ***plate-filament*** power transformers prior to about 1980 to the present-day step-down for all secondaries replacement units of today. The result for designers of the recent past was a need to usually design-in ***custom*** transformers in order to achieve the right filament and plate voltages. This was a temporary loss for electronics distributors who could now shift their stocking tasks to the new semiconductors. Power transformer makers were still in business doing their thing but their sales departments had fewer catalog pages to distribute.

AC-Line Powered Supplies

General Blocks of Circuits

At the AC input there is a transformer, primarily to achieve the correct RMS voltage, secondarily to act as an isolation element. Next is a diode rectifier to convert the AC to a form of DC having superimposed ***ripple.*** Following that is a filter capacitor to smooth out the ripple voltage depending on load resistance. What can follow that is a ***voltage regulator***, nearly all-electronics, to smooth out remaining ripple and establish a constant-voltage output.

Diode Rectifiers

Basic diode rectifier circuits are shown in Figure 22-1, all depicted with line transformer primaries unconnected.[2] In (A) only one side of the AC input is passed by the diode and the load capacitor charges only to the positive peak of the AC input sinewave, ignoring the negative portion of the AC period.

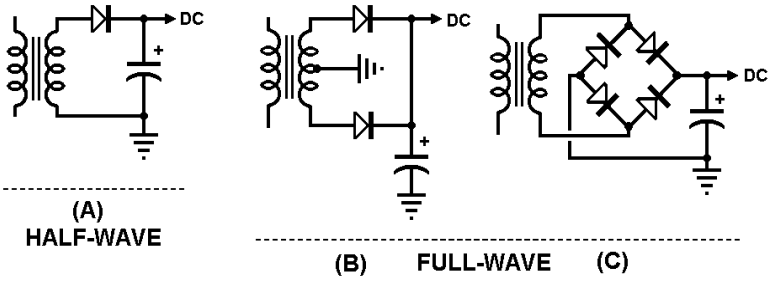


Figure 22-1 Basic AC-to-DC diode rectifiers.

In the Full-Wave circuits of (B) and (C) there is a choice depending on the transformer used. In (B) the transformer secondary is twice the voltage but it is center-tapped. Secondary voltage will be twice but the current is only an average of half. In (C) the transformer secondary remains the same, diodes steering the passed AC to the load.

Waveforms of Diode AC Rectifiers

This is shown in Figure 22-2 in two parts: In (A) for the initial turn-on and, in (B) for the steady-state running. Assuming a resistive load, the output voltage will have a slight ***ripple*** on its maximum value. That ripple comes from the AC peaks re-charging the capacitors after part of their voltage has bled off into the load. With enough capacity the ***ripple*** voltage will be low, too little and *ripple* becomes large. On the other hand, on initial start-up the capacitor is discharged (at its worst case) and must be charged from the first and subsequent AC cycles until it reaches the nominal output DC voltage. This initial start-up ***surge current*** can be quite high and must be taken into account on choosing the rectifier diode.

There is a reduction for initial *surge current peak* by means of a voltage-variable resistor (*Varistor*). A Varistor ***cold condition*** has a high resistance, reducing when carrying more DC output current. That will also drop the nominal, or average DC voltage output.

---

[2] It is recommended that a line transformer is always used. If not for stepping down the voltage, then to provide an isolation from the primary line voltage for personal safety.

The first capacitor current surge is due to the capacitor having no electrostatic charge. That must be built up. After that the capacitor discharges through the load resistance, charging up only

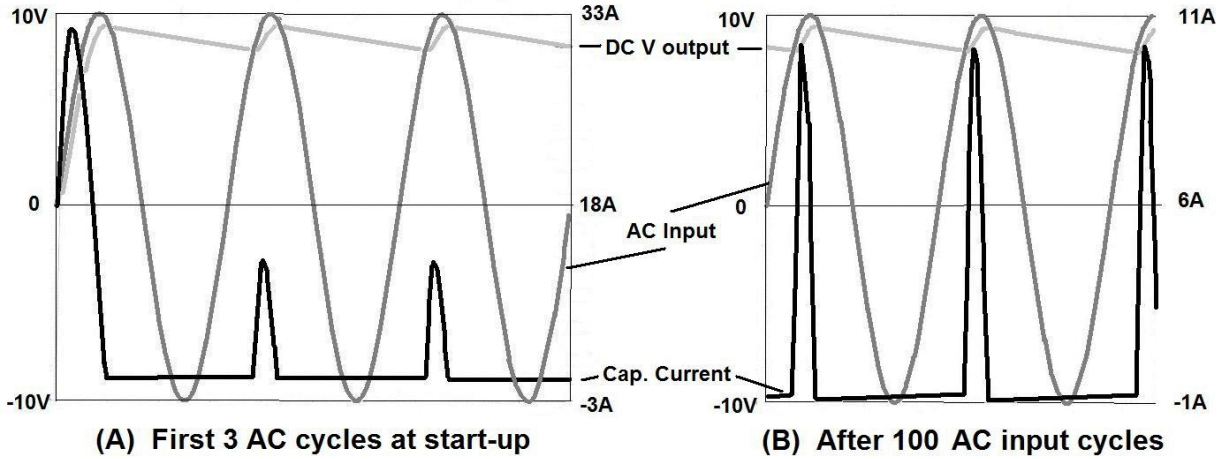


Figure 22-2 Voltage and current in a half-wave rectifier of Figure 22-1(A), using 1N5819 diode, 10,000 µFd capacitor and 10 Ohm resistive load. AC line input is 10 V peak shown for reference by medium-grey line. Capacitor current has darkest line, an initial surge in (A) of 31.7 A, settling to 10 A peaks in (B). There are two vertical current scales, 36 A total in (A), 12 A total in (B). DC output has the lightest trace, 8.27 V to 8.81 V.

when the AC input crosses the baseline. Combined action produces the sawtooth ripple voltage. See Figure 22-2 (B). Figure 22-2 is from an LTSpice program simulation using a ***Schottky Barrier*** diode having sufficient current capability and with less forward voltage drop than a conventional silicon diode. A *Schottky* diode has also been called a ***hot-carrier*** diode.

For a full-wave rectifier circuit the ripple voltage frequency is doubled and initial charge-up surge is reduced. In either circuit, the ***PRV*** or ***Peak Reverse Voltage*** depends on the opposite polarity of AC inpu*t,* the additive of opposite AC input voltage plus the maximum DC output voltage.

Total Voltage Losses

For a realistic scenario, a number of small voltage drops add up as follows:

1. Power transformer resistance, both primary and secondary.
2. Estimated series resistance (*ESR)* of the filter capacitor.
3. Total forward voltage drop of conducting rectifier diodes.
4. Series resistance elements between AC input and DC output.

Item (4) can include wire resistance for a low-voltage, high-current supply plus anything else such as a peak start-up current reducing Varistor. All of those resistances will add to make a simple R-C circuit described in Chapter 6.

Regulating Output Voltage

The simplest way to regulate output voltage is with a ***three-terminal series regulator.*** A number of those have been available since 1970 and have a number of built-in features. Both positive and negative versions are available. Common part numbers are the 7800 family (positive) or 7900 family (negative).[3] ***Switching regulators*** can be used but their switch action generates **noise** that can interfere with low-level circuits within a project box.

All series regulators absorb differences between the ***raw DC*** and regulated output voltage. Thatt requires a new term, ***voltage head-room. Head-room*** is the minimum DC output of an unregulated supply minus the minimum regulated DC voltage. Note that the term ***dropout*** is equal to *head-room, dropout* being more familiar to series resistances prior to regulation.

Full-Wave Rectifier Circuit Using 1N5819 Diodes
10 VAC Peak AC Line Input

Output Current, mA	Capacitor, µFd	DC Output Volts		First Input Surge, Amp	Steady-State Curr.Peaks, A
		Minimum	Maximum		
1000	10,000	8.27	8.81	26.2 A	5.36 A
500	5,000	8.48	9.05	15.4 A	3.34 A
250	2,500	8.62	9.21	8.56 A	1.88 A
125	1,250	8.70	9.27	4.50 A	1.07 A
62.5	625	8.77	9.37	2.28 A	0.60 A

Half-Wave Rectifier Circuit Using 1N5819 Diode
10 VAC Peak AC Line Input

Output Current, mA	Capacitor, µFd	DC Output Volts		First Input Surge, Amp	Steady-State Curr.Peaks, A
		Minimum	Maximum		
1000	10,000	8.12	9.38	31.7	10.0
500	5,000	8.24	9.54	17.4	6.53
250	2,500	8.30	9.63	9.08	3.68
125	1,250	8.35	9.67	4.61	2.02
62.5	625	8.37	9.70	2.35	1.06

In both tabulations the output was loaded with a single resistance to equal the load current at 10 VDC output. The capacitor selection was arbitrary, enough to reduce the ripple voltage on the output to an average of roughly 10% peak-to-peak relative to average DC output Voltage.

A more detailed value of filter capacitor can be obtained from desired ripple voltage, as follows:

[3] Part numbers usually have some added number to indicate output voltage. At the time of writing, so many different part numbers have become common that a whole collection of datasheets have been accumulated to find a correct series regulator IC.

$$C = \frac{60000}{R_{LOAD}} \text{ for 10\% ripple} \qquad C - \frac{120000}{R_{LOAD}} \text{ for } \sim 7.5\% \text{ ripple}$$

$$C = \frac{210000}{R_{LOAD}} \text{ for } \sim 4\% \text{ ripple}$$

Where:

C = Filter capacitance in uFd.

R_{LOAD} = Supply Load resistance in Ohms

These are all ***approximations*** for a full-wave rectifier operated with a 12.0 VAC RMS transformer input. Note that LTSpice AC inputs use the ***peak*** voltage input or 1.414 times the RMS value. Ripple voltage is taken as the difference of maximum to minimum output voltage divided by the average output voltage in DC.

Worst-case peak current demand on initial turn-on and the steady-state peak current when operating from nominal AC line voltage (secondary winding) can be found approximately by :

~ 10% p-p Ripple				~ 7% p-p Ripple				~ 4% p-p Ripple			
Filt. Cap. µFd	Start Curr., Amp	Run Curr., Amp	DCV Output	Filt. Cap. µFd	Start Curr., Amp	Run Curr., Amp	DCV Output	Filt. Cap. µFd	Start Curr., Amp	Run Curr., Amp	DCV Output
16000	62	18	14.1	33000	89	18	14.1	62000	120	18	14.1
8200	39	11	14.7	17000	71	12	14.8	30000	90	12	14.9
4100	22	7	15.0	8200	40	7.2	15.2	15000	64	7.5	15.3
2050	12	4.2	15.2	4000	24	4.5	15.5	7500	40	4.6	15.6
1000	6.2	2.3	15.4	2000	13	2.6	15.7	3500	20	2.7	15.8

As an example of a series-regulated supply for a fixed voltage, see Figure 22-3. Note that the rectifier, by itself, has a load of both the regulator and the output load resistance. At the input to the 7812 the DC voltage (at nominal AC input) is 15.48 to 15.92 VDC. Nominal headroom is then 3.48 to 3.92 Volts. Rectifier load will be about 63 Ohms.

To allow for AC line voltage fluctuations, take the AC input down to 11.4 VAC or 16.12 VAC peak. DC input voltage to the 7812 will be 14.67 to 15.09 Volts with a headroom of 2.67 Volts minimum. That is enough headroom based on a 2 VDC minimum. Normal current for regulator interior housekeeping is 5 mA typical, 8 mA

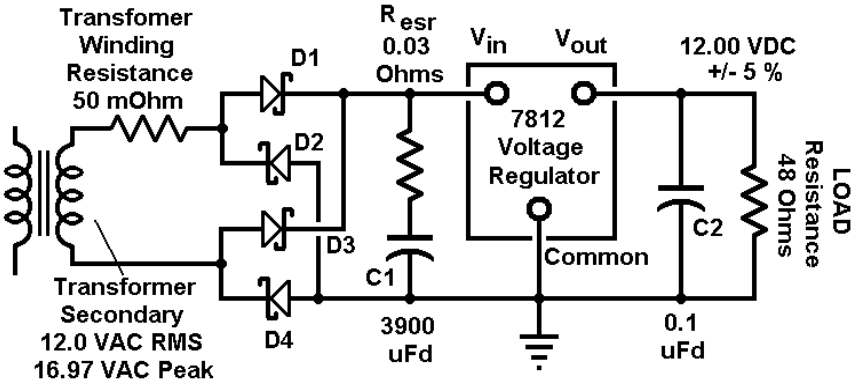


Figure 22-3 Small fixed DC (regulated) supply for 12 VDC.

maximum.

For a maximum power dissipation of the series regulator, take the +10% AC line voltage to find that sawtooth ripple voltage input is 16.29 to 16.75 Volts for an average of 16.52 VDC. At a 0.25 A load current, the series regulator would dissipate 1.13 Watts at maximum AC line voltage. That is low enough to mount it directly to a chassis metal using a TO-220 version of the 7812, a Fairchild KA7812 as used by the author.

Raw DC ripple is reduced at least 60 db by the series regulator so the 0.44 VDC peak-to-peak would result in about 440 µV peak-to-peak without C2 to dampen it further. C2 is included to set the internal regulator feedback phase for higher frequencies.

There are no contemporary specifications on temperature although a KA7805 is rated as operating from 0° to +125° C. For internal heat dissipation, for 250 mA output and an AC line voltage at maximum, the regulator would drop a maximum of 4.75 V at 0.25 A or 1.19 Watts. Without a heat sink, junction temperature would rise 65° C per Watt or 77.3° C at worst case. With a heat sink, junction temperature would be about 4° C above ambient.[4]

Adjustable Output Voltage Series Regulators

The National Semiconductor originated ***LM317*** (positive output) and ***LM319*** (negative output) have been around since just after the 7800 and 7900 family. The common pin is re-labeled as ***ADJ*** and the series regulator alone replaces that of Figure 22-3. Diodes D1 and D2 are used for protection, D1 primarily to discharge any load capacitance and shut-down.

The ***Adj*** pin has an internal reference voltage of 1.25 VDC ± 0.05 VDC. It is used with R1 and R2 to set the output voltage.

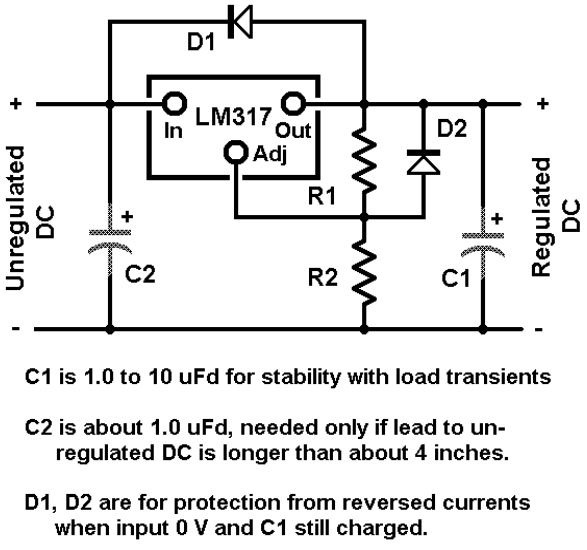


Figure 22-4 Basic LM317 voltage regulator circuit. See text for R1, R2 values.

If V_A = Reference Voltage (1.25 V)

I_D = Divider Current, Amperes

Then:

$$I_D = V_A / R_1$$

If R1 equals 220 Ohms, divider string current is 5.68 mA. Following that up:

$$R_2 = \frac{R_1(V_O - V_A)}{V_A} = \frac{R_1(V_O - 1.25)}{1.25}$$ Where V_O is desired output voltage

[4] Ambient temperature depends on the application. The author has not manually operated any electronics above an ambient of 118° F or 48° C, in the high desert just north of Los Angeles, CA.

If R1 is 220 Ohms and output voltage desired is 10 VDC, then R2 would have to be 1.54 KOhms. R2 could be partially a trimmer potentiometer or it could be fixed, made up of a parallel 6.8 KOhm and 2.0 KOhm resistors. For one-of-a-kind voltage regulators, using resistor combinations to set output voltage should suffice and be slightly less expensive in new parts purchases.

The minimum output current is specified as 10 mA maximum, 5 mA typical. A resistive voltage divider drawing 5 to 10 mA will satisfy the minimum load current with no load connected.

For an overall tutorial on solid-state series regulators, see National Semiconductor application note AN-1148 of May, 2002.

Shunt Regulation

This is hardly ever done with solid-state circuitry except perhaps by a zener diode used as a voltage reference. In tube days a moderate regulation was done by a gas-filled tube (OA2 through OD2) which had a reasonably-close voltage range over a limited gas-tube current. A general example is shown in Figure 22-5.

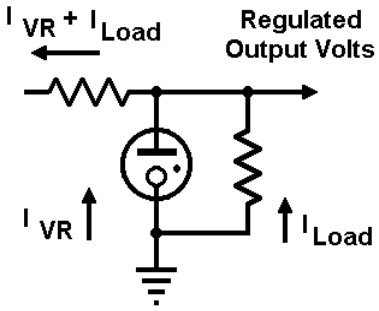


Figure 22-5 Example of a shunt-regulator circuit.

To actually regulate, one needed to know the load current represented by a simple resistor in Figure 22-5. The ultimate supply voltage has to be higher than the shunt-regulated circuit. A series resistance has to be inserted between the ultimate source and the regulated output, carrying the total of load and regulator tube currents.

As an example, if using an OA2 tube, it will hold to 1 Volt at 150 VDC, with a current of 5 to 30 mA. If the load resistance is 12 mA at 150 VDC then it is represented by 12.5 KOhms. Assuming a supply source voltage of 250 VDC nominal with a tube current of about 17.5 mA, the series resistance would be about 3.3 KOhms. Based on an AC line voltage variation of ± 10 % the OA2 tube current would vary between 10.7 and 25.9 mA.

Shunt regulators are rather inefficient for regulating any voltage but they ***are simple***. In the post-WWII era they were easy to use and, overall, ***cheap.*** In this brief example, an average of about 5.6 W is wasted to get 150 VDC output at 12 mA, a load of only 1.8 W.

Other Types of Regulators

Only one new type has surfaced, called a ***power-factor correction*** circuit, an adjunct to the switching type of regulated supply. Those are all intended to correct for the relatively large current demand occurring at, or just after, the AC input crossover time. Its main claim to fame is a reduction of harmonics and spikes of voltage that can interfere with other electronics. Most ***power-factor correction*** circuits are either specialized ICs or built into switching regulator ICs. Since such ***PFC*** circuits are relatively new and complex, they have been left out of this Designbook.

Around 1950 there were ***Ferro-resonant transformers*** used to dampen variations in the AC line input to rectifiers. Due to complexity and cost considerations, those were dropped from new power supply designs within a decade. Very few have been used in the half-century since 1960.

Battery Power

General

A ***battery*** of ***cells*** converts chemical energy to electrical energy. Cell voltages range from just over 1 Volt to about 3 Volts, depending on its ***electrochemistry*** and materials. Each cell has only three main components, two electrodes and an electrolyte. There are two main types, ***Primary*** and ***Secondary.*** Primary or use-once, then replace and dispose. Secondary is rechargeable. For voltages other than a single cell, two or more cells are connected in series,[5] dubbed a battery ***pack.***

When the author began with electronics as a teenager in 1947 there were only two kinds of *batteries* available for consumers: Lead-acid, rechargeable, as used for starting vehicles and carbon-lead-acid small *flashlight* cells. Small flashlight cells were often called ***dry batteries.***[6]

In the beginning of the post-WWII years, ***electrochemical technology*** for batteries and battery materials became cheaper and the Nickel-Cadmium began appearing for consumers.[7] The Nickel-Metal-Hydride type debuted next followed by an assortment of types: Alkaline electrolyte; Lithium-ion, itself subdivided into others depending on the electrode material; Zinc-Air (often used for hearing aids and the like); Silver-Zinc, Silver-Oxide, Zinc-Air. From the consumer market, battery packages extended from individual cylindrical cells to Button cells, Pouches (for certain Lithium cells), Prismatic (flatter rectangular cells) and specialized varieties for other purposes.

With the onset of PCs and other consumer products, plus the advantages of portability in so many electronics, ***Secondary*** or ***rechargeable*** batteries have taken over much of the text in application notes. All the various electrochemistry mixes have made for specialized charge rates and characteristics, even life of the various batteries. There is ***no specific recharge*** condition that applies to all rechargeable batteries. At best, one can get charging electronics that ***approximate*** most types.

Sizes and Connections

A long time ago when the vacuum tube was king, batteries were alphabetically labeled as **A** for filament, **B** plate-screens, and **C** for grid bias. That began to disappear around WWII and, in the early postwar period, alphabetic letters denoted **Sizes.** Most consumers now know the **AA**, **AAA, C, D,** and ***9 Volt***, plus various number-letter sizes for ***coin*** cells (used in small electronics) and ***prismatic*** sizes (used in PCs and peripherals). Letter-only types and the ***9 Volt*** package have become fixed in size internationally after about 40 years of manufacturing.

Letter-only sizes have plus and minus electrode connections flat at the ends of cylindrical

---

[5] Parallel connection is not done due to differences of manufactured source resistance, hence the series connection which makes the total voltage an integer multiple of a single cell.

[6] Dry batteries were mechanically sealed, as opposed to the open lead-acid vehicle battery. Vehicle batteries were later sealed shut.

[7] The carbon-zinc-acid or ***Leclanche cell*** and the nickel-cadmium types were both invented about the same time but the material required changing ***dry battery*** construction, along with materials.

physical cells and the ***9 Volt*** connections are opposed circular clips on the same side.[8] All the rest are a grab-bag of essentially-not-yet-standardized connections, usually flat metal connections.[9]

Cell Life in Current

There is one guideline, milliAmpere-Hours abbreviated as ***mAh.*** This applies to all cells and the common types of battery cells are tabulated following.

Table 22-1 Common Battery Cell Life expressed as Milliampere-Hours
Data from Eveready (Energizer) Engineering Handbook (Internet) dated 2001

Cell Type	Carbon-Zinc	Alkaline	NiMH	NiCd	Lithium (LiFeS2)
AAA	540	1200	800-1000		
AA	1100	2700	1700-2900	600-1000	3000
C	3800	8000	4500-6000		
D	8000	12000	2200-12000		
9 V	400	565	175-300	—	1200

The *Lithium* column of Table 22-1 refers to the Lithium Iron Disulfide type which has a 1.5 Volt nominal cell voltage, differing from other Lithium products. As a general rule, maximum current demand can be taken as one-fifth of the mAh rating and one can expect cells to hold nominal voltage within a 4 hour life at that current.

Cell Life in Voltage

Those depend on the electrochemistry involved. For the carbon-zinc, Alkaline and Silver-Oxide cells are all 1.5 VDC. Nickel-Cadmium and Nickel-Metal-Hydride, both rechargeable, are nominally 1.2 Volts. A Mercury battery cell (when one can be found) is nearly constant at 1.35 Volts over its life. Lithium cells, rechargeable, vary from 3.0 to 3.4 Volts depending on the Lithium formulation other than Lithium Iron Disulfide..

A ***standard cell*** or ***Weston*** cell, invented by Edward Weston in 1899, is a rare type, used as an International Standard between 1911 and 1990 for voltmeter calibration, is exactly 1.018 64 VDC under no load. It has a glass housing in the shape of a capital H with wire leads.

Battery ***packs*** for higher voltages are made from individual cells connected internally in series. Two examples are the *9 Volt* used in smoke detectors (6 cells) and a ***Lantern Battery,*** 6 Volts, (4 large cells in series). Old-style combination battery packs for vacuum tube receivers are no longer

[8] The **9 Volt** pack had the same circular snap-in connector type as predecessor small 45 and 67.5 Volt packs made earlier for small tube-type electronics. Those were never widely popular in their time.

[9] Vacuum tube portable equipment had a jumble of cheap connectors and contact arrangements that didn't begin to be standardized until the solid-state era had established itself. By then it was too late for the multiple packs of batteries of low and high voltages that didn't need but one voltage to go mobile. In the solid-state era there are fewer demands on battery connectors due to fewer connections.

mass-produced.[10]

Table 22-2 Nominal Cell Voltages at Charge

TYPE	Cell Voltage	Relative Energy for Same Weight	
Zinc-Carbon	1.5	0.13	[Primary]
Alkaline	1.5	0.40	
Lithium (LiFeS2)	1.5	-----	
Silver-Oxide	1.55	0.47	
Nickel-Cadmium	1.2	0.14	[Secondary]
Nickel-Metal Hydride	1.2	0.36	
Lead-Acid (vehicle batteries)	2.1	0.14	
Lithium Ion	3.0 - 3.4	0.46	[several types *]

* As of time of writing (2010) there were several types of Lithium Ion cells being made. Since this was a recent development, keep checking makers' engineering data on voltages.

In Table 22-2, the first four cell types are non-rechargeable primary types while the following four are rechargeable secondary types. ***Do not attempt to recharge primary cells.*** Silver-oxide types are all coin battery size. Lead-acid vehicle batteries refer to the usual self-starter/combination-ignition types, not those used in hybrid or all-electric vehicles for propulsion. Normally, those lead-acid batteries are sold fully charged, have high cranking currents, can last a long time at moderate loads and are sealed.

The four primary types are listed in descending order of price. Zinc-carbon is the least costly while the Silver-Oxide coin batteries are most expensive. Some foreign countries have banned NiCd cell sales for reasons of safety about cadmium in waste dumps.

Temperature Effects

Temperature affects electrochemistry. In general, the Lithium battery is the only one which can still operate at -40° C and then only with the higher voltage of 3 V per cell. All others in popular sizes begin to drop in capacity below 0° C. During the Korean War of 1950-1953, the U.S. Army made a *battery vest* containing all the zinc-carbon cells for the AN/PRC-6 handheld transceiver. The object was to wear this between shirt and outer jacket, keeping the cells warm by body heat. While it worked for the intended purpose, it was an added weight and bothered a wearer's movements. Use of this *battery vest* was discontinued shortly after the Korean War.

While recent additives in battery cell manufacture have decreased the temperature effects somewhat, it is a good idea to keep primary battery cells in a refrigerator (but not a freezer) for storage prior to use.

[10] A few are hand-made by very small companies for use with rebuilt old portable receivers, using internal modern cells. Those can be found by diligent search of the Internet.

Table 22-3 Relative Life of Cells Versus Temperature

	Output Relative to 100% @ +20° C		Storage Capacity to 80% Power	
Type	0° C	-20° C	+20° C	0° C
Alkaline	75	25	7 years	>12 years
Carbon-Zinc	70	25	~ 3 years	-----
Lithium (coin)	95	80	>12 years	>12 years
Lithium LiFeS2	95	80	>16 years	>16 years
NiMH	80	20	3 - 5 years	-----
Silver-Oxide	80	50	10 years	>12 years
Zinc-Air	75	30	5 months	>3 years [sealed]

Source is Energizer Engineering data, 2001 for all but Carbon-Zinc types. For storage capacity marked greater-than (>), this was interpolated from supplied graphs. Storage capacity is assumed to be 80% of power when cells were first put into storage. Carbon-Zinc output is obtained from old Signal Corps, U. S. Army, data. Note: For storage at 0° C, keep temperature close to freezing but never ***at*** freezing.

In Table 22-3 it should be evident that Lithium electrochemistries are superior to all other types. Silver-Oxide and Zinc-Air are both coin types with Zinc-Air intended solely for hearing aids. NiMH is next best, rated primarily for storage while Carbon-Zinc is slightly better for operating temperature.

It is suggested that cold storage of NiCds be done in the ***discharged*** state to avoid any ***memory effects.*** There is no *memory effect* as such, but there is a change in NiCd electrochemistry and that can cause shorting ***dendrite*** formation between electrodes. If dendrites form and remain they will drain the battery, usually irreversibly. It should also be noted that NiCds lose about 10% of their initial charge in the first 24 hours and can self-discharge at a rate of 10% a month thereafter.

Variations in Electrochemistry for the Same Type Cell

Nearly all USA-based manufacturers have two or three grades of common cells, along with rather grandiose names attached. For example, Energizer makes three grades of Alkaline manganese-dioxide batteries: ***Economy, Standard,*** and ***Premium.*** Those grades are rated on increasing current load demand. *Premium* would be for powering things like digital cameras with attached photo-flash, the combination resulting in a high peak current demand. *Economy* grade would be for seldom-used appliances or lower current demands such as toys and flashlights. Energizer controls the mixtures of things inside cells. As users of a commodity, all we can do is to ***read the labels***.

As another example, Lithium Ferric Dioxide cells are nominally 1.5 Volts. But another maker's AA-size cell might be Lithium Manganese Dioxide with a nominal 3.0 Volts potential.[11] One can't mix the two types without a distinct possibility of load damage from overvoltage. We ***have to read the labels.***

[11] Up to this writing, the only true variant in Lithium was the Ferric Dioxide type. All other Lithium types were the 3 V type.

Series or Parallel Connections?

For higher voltages, a series connection will have the sum of cell voltages and current capability of the least energy cell in the combination. A parallel connection is very risky. Firstly, the exact cell voltages cannot be obtained so the lowest cell voltage will drag all others down to it. Secondly, the lowest cell voltage will put its internal resistance as a *load* to all others until all others have exhausted their energy enough to make all cell voltages equal. The answer is ***Series, YES,*** but ***Parallel, NO.***

Internal Cell Resistance

This hasn't been discussed but it can be likened to a ***source impedance.*** It can be modeled as a series resistance to a constant-voltage source. At lower temperatures the cell resistance increases while at higher temperatures the cell resistance decreases. That is part of the electro-chemical-thermal effect of battery operation.

Secondary Battery Charging

General

There are two types of chargers, ***trickle*** and ***fast (or float).*** Trickle charging is suggested for NiCd cells at about ***0.1 C*** or one-tenth of a rated NiCd cell power output in milliAmpere-hours.[12] Fast charging is done at about 1.0 to 1.2 C for an hour, suggested for Lithium and NiMH cells. ***Never try to recharge primary (non-rechargeable) batteries.*** It does no good for most primary cells and may hurt the charger.

For prolonged periods of storage, NiCd cells are suggested to be ***fully discharged*** if no trickle charging is available. This applies only to NiCd cells, not to any of the other types.[13] The major reason is to prevent internal cell formation of dendrites from electrodes which may, in time, short them and render them unuseable. Note: That has been mistakenly called ***memory*** for NiCds.

Number of Charging Cycles per Battery Type

NiMH batteries can do about 1000 cycles of recharging for those rated up to about 2500 mAh, roughly 500 cycles for larger ratings. NiCds should be trickle-charged when not in use but some newer NiCds can be fast-charged. NiMH and Lithium (>2 Volts) cells, if so marked, can be fast-charged.

Vehicle Lead-Acid batteries can do more but their main drain is in seconds of peak current demand (starter motor) with the normal drain coming from other, lower-drain systems. It should

---

[12] Some older NiCds may require a lower trickle-charge current of about 0.025 C; check cell markings.

[13] The author cannot confirm that due to limited experience with non-trickle-charged NiCds but it is repeated in application notes and on the Internet.

also be noted that Lead-Acid vehicles are usually trickle-charged from the engine's generator when not connected to a starter motor. Typical vehicle Lead-Acid batteries can last 5 to 10 years before replacement, depending on use and operating environment.[14]

Internal Gas Build-Up During Charging

Over-charged NiCds have been known to break open and vent, mostly oxygen, if over-charged. NiMH cells will out-gas hydrogen but seldom change physical characteristics during charging. There are variations among all Lithium Ion cells (besides the LiFeO2 type) since these are under-going refinements as of this writing. Older Lithium Ion cells have been known to (almost) spontaneously catch fire whether or not they are being charged, used, or just sitting idle. Most of the modern Lithium cells are now made safer.

Normally, charged cells will get warmer when charged. NiCds will get slightly cooler than the environment up to the point of holding a maximum charge, then going into being warmer than the environment. That curiosity remains solely with NiCds and does not occur with other types.

Packs Versus Individual Cells in Charging

For battery packs of series-connected cells, cells must be of equal characteristics to be charged together. In general, packs of several cells in series have been made from the same manufacturing lot and more likely to be equal. For individual cells it is recommended that each is charged separately even though some consumer-grade chargers *appear* to be in series, usually an even number of cells.[15]

A Handy *Fast* Charger You can Make

An adjustable constant-current charger is shown in Figure 22-6, suitable for up to 3 Ampere charge (fast charging) to 75 mA (trickle charging) of 6.2 VDC battery packs. This is good for NiCd, NiMH, or Lithium Ion cells of near-equal characteristics. It is adapted from National Semiconductor Application Note 946, updated in 2002, written in 1994 by Chester Simpson.

This charger uses a switching regulator IC second-sourced by ON Semiconductor and Micrel. The IC is a ***buck regulator*** meaning it will step-down any input voltage.[16] An LM2576 comes in two packages, both variants on the familiar TO-220 case. It has a built-in 1.23 Volt reference which is used to compare the output voltage and change the switching frequency duty cycle of about 53 KHz in order to regulate output voltages. As a general rule of thumb here, switching regulators are not normally in this book since most circuitry in this book is for receivers. Higher-power switching circuits will generate RFI to interfere with low-level devices. For a battery charger usually left on over night or for periods when a receiver is not operating, this was considered okay.

---

[14] Usual problems are flaking-off of electrode material dropping to the bottom and shorting electrodes.

[15] There are too many charger types on the market by 2011 to list all of them. The author's claim is from a sampling of some of the available chargers on the market.

[16] ***Boost regulators*** will step-up an input voltage. Some switching regulators do both in the same package.

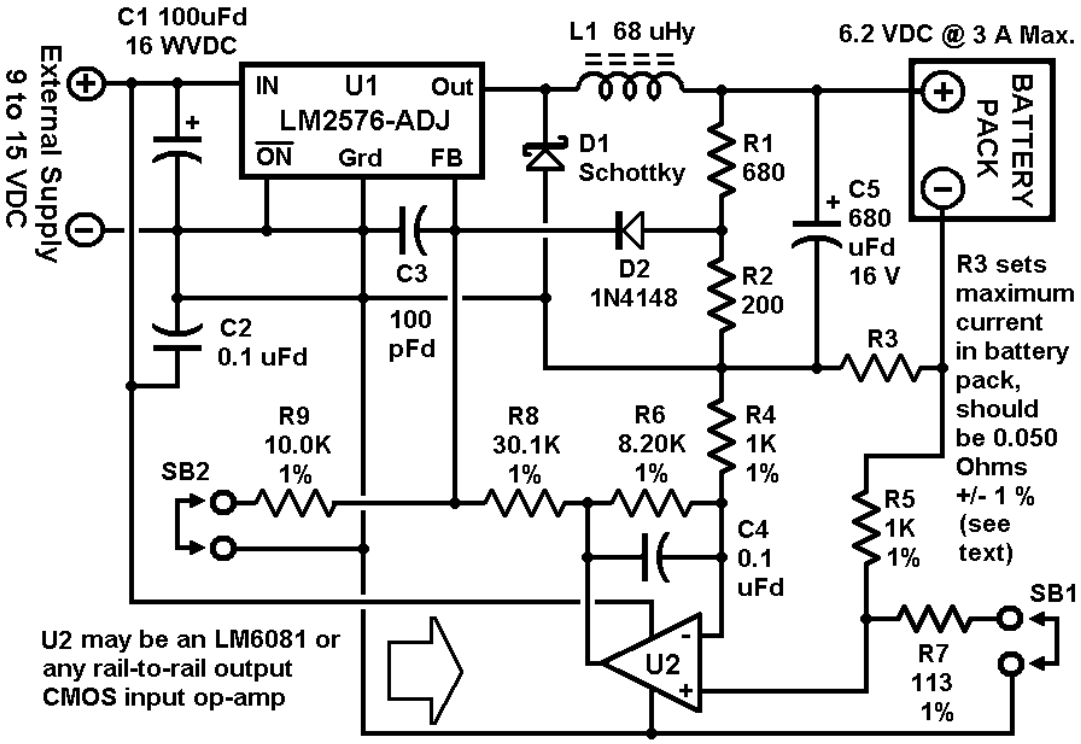


Figure 22-6 An all-purpose constant-current battery charger built with a switching regulator for a 6.2 VDC, 3 Ampere maximum current load battery pack. Shorting bars SB1 and SB2 set the pack's charging current in mA for a total of four different currents. Any input supply with more than 3 VDC above the desired charging current may be used.

Some details of Figure 22-6: Unless specified, resistors are 5 to 10% tolerance 1/4 Watt, capacitors are 50 WVDC. D1 is any low forward-voltage drop Schottky type, such as a 1N5819. R3 is used to sense current flowing into the battery pack and may be made from 23 3/8 inch length of #24 AWG coil wire if it is difficult to obtain from a distributor. Note that the milliOhm resistance of R3 is crucial to the exact charging current. The *shorting bars* for R7 and R9 can also be switches. These set the charging current according to the following table:

Table 22-4 Setting of Charging Current in Figure 22-6

Charging Current	Shorting Bars		Switches in place	
	SB1	SB2	Of SB1	Of SB2
3.0 Amperes	OUT	OUT	Open	Open
0.75 Amperes	OUT	In	Open	Closed
0.30 Amperes	In	OUT	Closed	Open
0.075 Amperes	In	In	Closed	Closed

How This Simple Charger Works

A basic constant-voltage switching regulator is shown in Figure 22-7 for a fixed 5 VDC output. Input voltage may be anything from 8 to 40 VDC with the input capacitor supplying extra input during switch cycles. The Schottky diode, series inductor, and output capacitor keep the output at a minimum ripple voltage. *Feedback* voltage is fed back to the LM2576-5.0 where it is divided down to 1.23 VDC and compared to a 1.23 VDC internal reference supply. If the output voltage rises, then the internal regulator seeks to lower the output voltage and vice-versa.

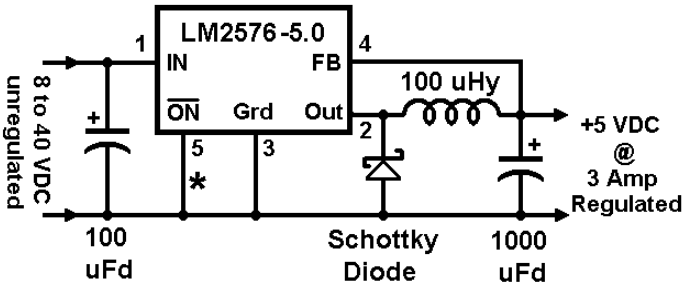


Figure 22-7 A simple fixed voltage regulator.

For the circuit of Figure 22-6 the charging *current* is dropped across R3 and develops 0.150 VDC at a 3 A current. Op-amp U2 is an amplifying integrator with a voltage gain of 8.2, and amplifies the 0.150 V R3 drop to about 1.23 VDC into the *FB* pin of U1. Polarity of connections to U2 are such that the current across R3 is maintained at a constant 3 A rate.

R1, R2, and D2 are there to keep an approximate 6 VDC output when *there is no battery pack connected to the charger.* Once a battery is connected, its slightly-lower voltage will take over and operate the charging current, D2 essentially becoming an open circuit. Feedback will be the amplified R3 voltage out of U2. Note: C3 is there to reduce some of the ripple present on the *FB* pin of U1.

When R7 is connected to common, it forms a voltage divider with R5 to drop the R3 charge current by a tenth. Maximum charge current would then be 0.3 A rather than 3.0 A. When R9 is connected to common, it forms another voltage divider for one-fourth of the U2 output into the *FB* pin of U1. The two dividers are isolated so the binary settings in Table 22-4 will hold.

The ***ON-not*** pin of the LM2876-5.0 in Figure 22-7 may be connected to a digital logic control input. When pin 5 is grounded, regulator output is always available. When pin 5 is held high, the output is cut off. That can be useful in adding to the charger of Figure 22-6, an additional voltage-sensing circuit added to make sure that *no* voltage will appear on the battery or just to turn off the charger. Looking at the 5-lead TO-220 case, pin 1 (input) is at the embossed circle; all others will be pin 2, Output; pin 3, ground; pin 4, Feedback; pin 5, On-not.

As already mentioned, a *simple switcher* buck regulator can be used in place of an analog 3-terminal regulator. National Semiconductor came out with several models in the 2000 decade, among the most popular are LM2586, LM2588, LM2592, LM2596, and LM2975. Datasheets for these types have explicit instructions for selection of output capacitor, inductor at various loads. Whether or not to use them is a matter of choice. Remember that these can produce RFI for low signal levels. Remember also that the saving in size is marginal compared to an analog 3-terminal regulator and, for small quantities, no great saving in cost per part.

Some Cautions Learned the Hard Way

For *Fast Charging* there isn't much need to track battery voltage. Fast charging is generally done in about an hour, that part of the hobby shop effort. That can be done in-the-shop while doing other hobby purposes.

Trickle charging can be done overnight, generally in a 14 to 18 hour period and ***should have a cell voltage monitor to shut off charging when a preset cell voltage is exceeded.*** This insures that NiCds will be protected from over-charging from a constant current source at the end of a charge cycle. Such will keep NiCds at peak performance.

The remainder of secondary cell electrochemistry is robust enough to tolerate cell voltage overloads.

Chapter 23

Digital Hardware Basics

Analog circuits have near-infinite level increments of input and output but digital circuits have only two levels referred to as 1 and 0. The gate structure is the heart of nearly all active digital packages, forming many different sub-functions which may be interconnected in many ways. This chapter gives the basics of available digital hardware, establishing the design guidelines for their use and application.

Gate Logic - The Brick and Mortar of Digital Devices

The basic building block of digital devices is the gate. A gate acts on two or more logic-level inputs to produce an output that is a logical function of the input state conditions. Logic level inputs have become standardized to active-high meaning a logic 1 is positive (between about half supply voltage to full supply voltage) and a logic 0 is near-common potential (less than 0.5 VDC). There are six basic gates:

AND - Output is a logic 1 when each input and all other inputs are logic 1.

OR - Output is a logic 1 when any input or any other input is a logic 1.

Exclusive-OR - Output is a logic 1 when one input is a 1 and the other input is a 0.

NAND - (Not AND) Same as an AND except the output is inverted.

NOR - (Not OR) Same as an OR except the output is inverted.

Exclusive-NOR - (Not Exclusive-OR) Inverted output Exclusive-OR.

The three ***NOT*** versions can come in handy when designing arrays of gates in that they can reduce the number of digital devices necessary.

Inverters are generally included as part of the gate group although their function is simply to invert the state of a single logic line. Buffers or Line Drivers are as simple but do not invert a logic state; their purpose is mainly to drive more following gate inputs, as in buffer amplifier for analog circuitry.

Electromechanical Analogues of Gates

For those unsure of the gate concept, Figure 23-1 shows the electromechanical equivalents to gates where the two switch inputs represent the input, the relay contacts representing the output. If K1's contacts close, the output state is equal to a logic 1.

In Figure 23-1 (A) both S1 **and** S2 must be up in order to energize relay K1 and close its normally-open contacts. The up position is analogous to an input logic 1. The number of switches in series could be increased to any number but the and function would require all of them to be up.

In Figure 23-1 (B) either S1 **or** S2 in the up position would energize K1. Any number of switches could be wired in parallel and still satisfy the OR function. Any one switch or the other up would energize K1.

Exclusive-OR function is represented by Figure 23-1 (C). Either S1 would be up and S2 down or S1 down and S2 up to energize K1. Both switches in the same position would not do anything; their position or states would be exclusive of the other.

The Figure 23-1 (C) circuit is essentially the same as three-way switches at the ends of a residential hallway; the switch at each end of the hall can turn the hallway light on or off.

Relay logic was used successfully for years in the telephone industry and, during World War II, at least three electromechanical computers were designed and built by Bell Telephone Laboratories for wartime production calculations. While slow, huge, and consuming considerable electric power, they were made possible by using logical functions similar to those in Figure 23-1.

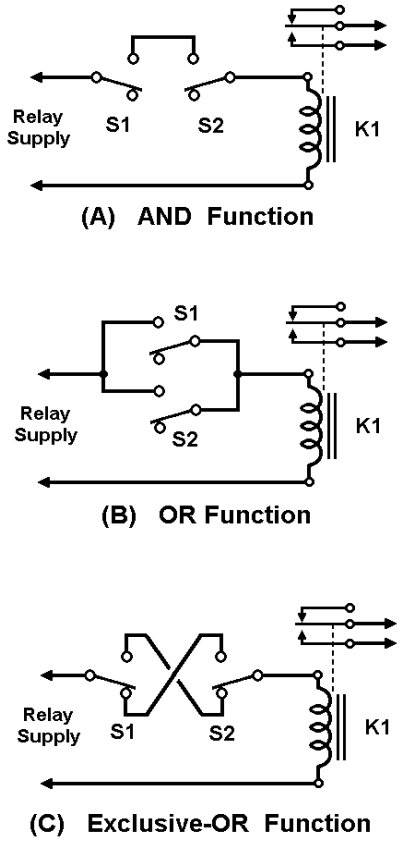


Figure 23-1 Relay equivalents of simple logic functions.

The First Electronic Versions of Gate Structures

Figure 23-2 shows the first electronic gates. In (A) any input, A or B, can be high or logic 1 to make the output high. In (B) either input A or B at a low voltage will cause the associated diode to conduct and its input to be a low level (above ground by only the forward-conduction voltage drop of the diode). Both inputs in (B) have to be high in order to cut off the diode, thus satisfying the AND function.

The resistors are chosen for the diode current during conduction, usually minimum to insure a reasonable forward-conduction voltage drop. Along with a following buffer amplifier (to feed many gates), these logic gates were the first discrete-component circuits in early computers and digital circuits. Integrating the diodes and bipolar transistor buffers on a single chip, they formed the first family of digital ICs, DTL or Diode-Transistor Logic. Now long obsolete, DTL was replaced by a more robust digital family, TTL or Transistor-Transistor Logic, born at Sylvania but popularized and expanded as a family by Texas Instruments.

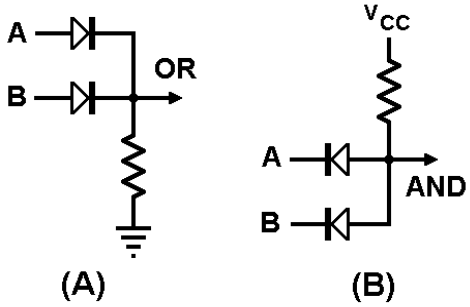


Figure 23-2 First electronic versions of logic gate functions.

Fairchild Semiconductor came up with a low-cost alternative to still-expensive TTL in the early 1960s. Dubbed RTL for Resistor-Transistor Logic, they were as fast as the standard TTL of its time and the TO-5 size round epoxy cases took slightly less PCB space than the DIP then becoming standard. They were reasonably cheap compared to the original space-rated applications for the Apollo space program.

Their on-chip circuitry must have been simple as indicated by equivalent discrete circuits depicted in Figure 23-3. The +3.6 VDC supply rail was at odds with TTL but probably picked as a compromise between speed and internal power dissipation. Their speed depended on low base-emitter junction capacitance; base resistor shunted by a capacitance would produce a simple lowpass filter.

RTL had simple functions per device such as a few gates, a J-K flip-flop, and even a monostable multivibrator device. While they stayed on the market for a few years, they lost out to customer demand for a more standardized digital logic family represented by TTL. TTL was growing with many different and multiple functions per package.

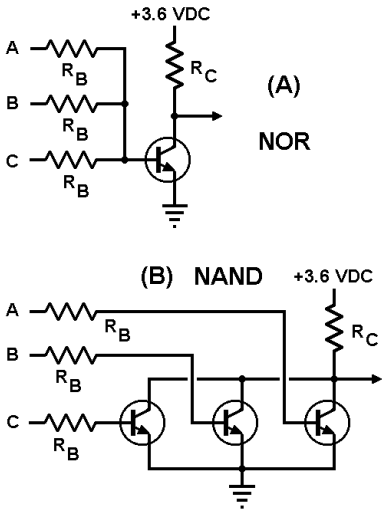


Figure 23-3 First Resistor-Transistor Logic circuits.

Generic TTL Gate Structure and Levels

TTL or Transistor-Transistor Logic is the surviving wide-use family among those that began in the early 1960s. The basic gate structure as shown in Figure 23-4 is the internal building block at the heart of all older TTL logic packages. Figure 23-4 is not an exact duplicate or model of what actually exists in silicon inside a digital device but it is good enough to permit some clarification of the necessary voltage-current of logic 1 and logic 0 levels. The following description concerns the original +5 VDC supply TTL gate.

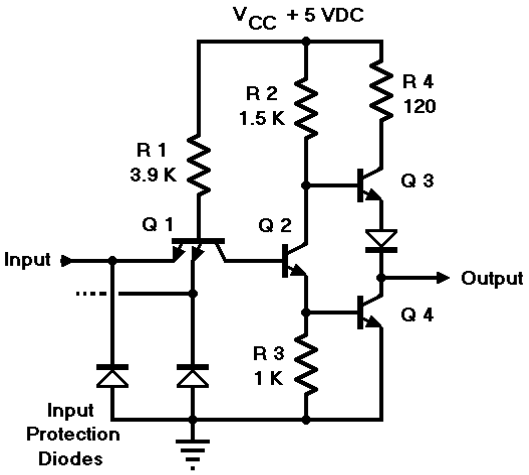


Figure 23-4 A basic TTL gate circuit.

Q1 is built with multiple emitter junctions and each one constitutes a separate input. If all inputs are at a voltage level of about +2.4 VDC or higher, Q1 will have its emitter-base junctions reverse-biased. R1 and the Q1 base-collector junction provide a maximum base current for Q2. In turn that provides base current for Q4 to conduct to saturation and the output is pulled down to ground. Q4 can be said to sink any external resistance's current to ground. The gate output will be about +0.2 VDC or the saturation voltage of Q4's collector-emitter junction.

If any one gate input is pulled down to +0.4 VDC or less, that input's emitter-base junction is forward-biased, Q1 conducts just enough to partly cut off Q2. Q2's collector voltage rises and the Q3 base-emitter junction is forward-biased. Q3 pulls up the output since Q4 has, simultaneously, gone to a low conduction state. R4 is a current limiter for the output at a logic 1 state and the diode between Q3 emitter and output helps to maintain the proper logic 1 output voltage. Operation of internal transistors is saturation and cutoff, primarily to insure constant logic 1 and 0 voltages.

The generic TTL gate structure is basically a voltage inverter. Any input logic 0 voltage will produce an output voltage at logic 1. If all inputs are at logic 1 the output voltage will be at a logic 0 level. This is also the basic definition of the NAND gate. To explain that acronym, there must be an explanation of gate types.

Basic Logic Gates in TTL

All digital structures perform certain tasks according to a logical combination of input states that produces a certain output state or states. The basic element is called the gate and the six basic gates are shown in Figure 23-5 in their standard symbols and truth tables of input-output for two-input gate structures.

The AND gate of Figure 23-5 (A) has an output (Q) state that is a 1 only if input A and input B are a logic 1. The output is logic 0 for all other input state combinations. The OR gate in (C) has a logic 1 output if either input A or B is a 1. The OR output is logic 0 if all inputs are 0.

The NAND gate output at Figure 23-5 (B) has the logical NOT or inverse of the output state of the AND gate. The NAND takes its name from Not-AND meaning it is an AND gate with its output inverted.

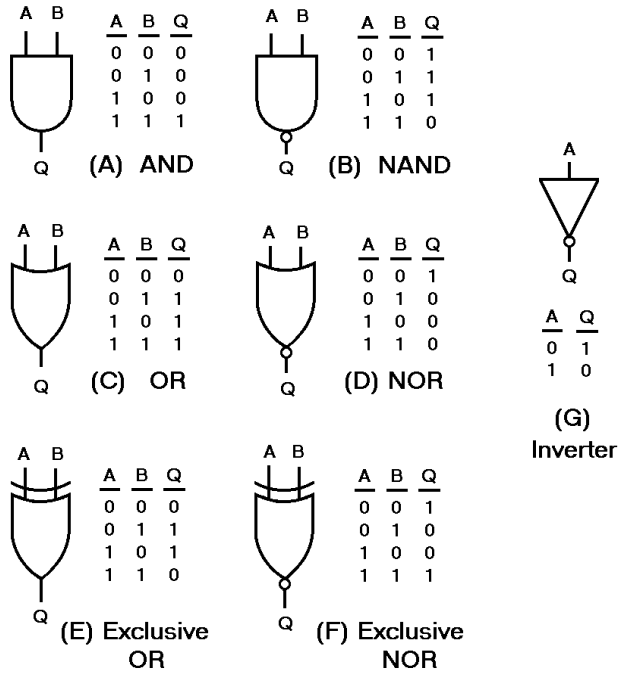


Figure 23-5 Logic gates and symbolism.

Note the little "bubble" at the Q output of (B). This is the inversion symbol indicator and can apply to either inputs or outputs. The NAND gate has the same overall shape as an AND but it has the little bubble at the output. In a similar manner, the NOR of (D) has the output states inverted from those of the OR of (C). The NOR gate takes its name from Not-OR.

The Exclusive-OR of Figure 23-5 (E) has an output state at logic 1 only if the input states are different states. Its output is 0 if both inputs are logic 1. The Exclusive-NOR of (F) does the same except its output is inverted. A common short form name of (E) is Ex-OR or XOR.

The inverter of Figure 23-5 (G) is included for reference of symbolism and truth table use. It might be described as an amplifier with a unity gain since it only inverts its input logic states. An inverter can do a form of amplification as will be seen shortly.

All seven symbols and their state combination truth tables apply to all digital logic families, past, present, and very likely the foreseeable future. The voltages and currents that apply to logic 1 and logic 0 depend on the device family.

TTL Logic Levels and Fan-Out

In the four decades between the 1960s and the new millennium, the popularity and robustness of the basic TTL structure evolved an accepted standard of voltage levels for logic states in the common 5 Volt logic devices. A logic 0 level is +0.4 V or less relative to common ground. A logic 1 level is +2.4 V or more up to but not exceeding the supply voltage.

Any voltage in the region of +0.4 to +2.4 V is to be avoided at all costs. If such a voltage exists at an input, the logical function of the digital device cannot be guaranteed by anyone. Note: Each TTL family may have slight variations on those limit voltage levels; double-check specification sheets.

Fan-out refers to the number of gate or device inputs that can be driven by a single gate output. Within the same device family this is generally a 10:1 ratio, ten inputs driven by one output with output voltages remaining within specification. Current demands below are for common gates from each TTL family:

TTL Family	Logic 1 Input	Logic 0 Input	Logic 1 Output	Logic 0 Output
54/74	40 µA	1.6 mA	400 µA	16 mA
54H/74H	50 µA	2 mA	500 µA	20 mA
54L/74L	10 µA	0.18 mA	200 µA	3.6 mA
54LS/74LS	20 µA	0.40 mA	0.4 mA	8 mA
54S/74S	50 µA	1.0 mA	1.0 mA	20 mA
54/74F	20 µA	0.6 mA	1.0 mA	20 mA

The term family has, by default, become synonymous with the Texas Instruments part numbering system which was 54nn and 74nn. The 54 denotes a military operating temperature range of -55 C to +125 C, 74 denoting commercial temperature range of 0 C to +70 C, and the nn being a numeral identifying the specific device function. That is the so-called standard family part number.

With demands for faster operation or lower power increasing, two new families were developed and identified with a letter between environment prefix and part number. Those were H for high speed and L for low power. With more development the internal basic gate structure utilized Schottky technology for the S family and its lower-power sibling LS or lower-power Schottky.

Many other semiconductor IC manufacturers developed advanced digital devices, all the way from newer basic gate structures to hybrid CMOS-transistor technology to LSI or Large Scale Integrated subsystems on a chip in a single package. After a quarter century of manufacture, sale, and incorporation into all sorts of different applications, the most-used device functions were cross-licensed and the designers tended to use the generic Texas Instrument part numbers as device identifications. For that reason, these common digital logic devices are heretofore referred to as 74xNN where the x can be any family designator. Device functions exist among nearly all families with the NN identification number denoting the package function.

It is up to the designer to choose which family, obeying the loading rules for that family, choosing the temperature range and the type of package (DIP or one of several SMD or Surface Mount Device styles). The digital function will be the same; but some functional pins may change position..

The pull-up resistor column in the tabulation is the nominal value for unused gate inputs. While unused gate inputs may float towards the VCC potential if unconnected, that also represents a relative high impedance to stray signals and spikes from adjacent circuitry to capacitively couple strays. Using a pull-up resistor mitigates such capacitive coupling by introducing a lower impedance to such coupled strays. In some circuits a deliberate ground-level connection (from a control switch) may force a logic 0 with the pull-up resistor assuring a logic 1 when the connection is opened. The small amount of VCC current drain from a pull-up is easily calculated.

If a particular logic family has a fan-out of 10, the amount of external load current to run other loads is found by simply multiplying the two input level currents by 10. In general, most logic

families can source and sink more current. The exact maximum loading current is specified on device datasheets as IOH for Output current in High (voltage) state or IOL for Output current in Low (voltage) state. This holds for DC on up to the maximum datasheet-specified frequency.

Open-Collector and Three-State Outputs

Some TTL-based devices are designed as open-collector. Output transistor junctions have the collector (of Q4) literally connected only to the output pin as shown in Figure 23-6. This has several advantages: The output can be connected to other loads such as an LED indicator; the output transistor may be designed with higher breakdown voltages for coupling to other circuits with higher supply voltages; the collectors may be wired-OR. For the latter, several outputs may be literally wired together with an external pull-up resistor such that any output with a logic 0 (Q4 pulls down to ground in saturation) will cause a logic 0 in the output. That fulfills an OR function for active-low outputs.

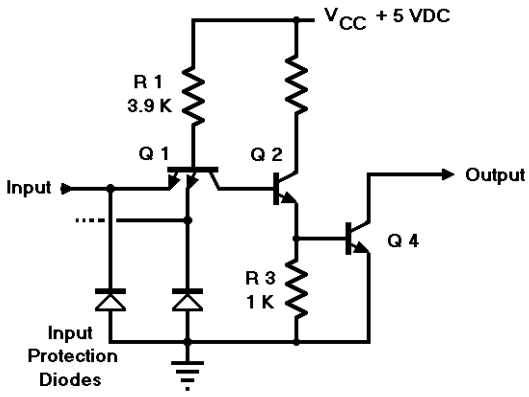


Figure 23-6 TTL Open-Collector

Some devices are designed with three-state outputs. A separate control pin and internal IC circuitry can switch all device outputs to a high impedance state, the third state of its name. When in a high impedance state the device has no effect on any loads or fan-out. Its use is principally for several sub-functions such as digital memory to be physically connected but only one being electronically "switched in" by a control circuit. Three-state acts much the same as open-collector but allows outputs to follow both logic 0 and 1 specifications when the control pin enables the device. Once in the Hi-Z (high impedance) state, the device acts as if its output isn't connected.

Active-High and Active-Low

By convention gate type descriptions and symbology follow active-high rules meaning that logic 1 voltage levels are above ground, logic 0 levels near ground, and the presence or absence of logic 1 input combinations is part of the functional description. Any gate or functional device input may be arranged as active-low.

Active-low refers to a function performed by a level near ground or logic 0. As one illustration, Figure 23-7 repeats the NAND and NOR symbols at the left and the functional equivalent of the same gates at the right. Note the little inversion bubbles at the inputs on the symbols to the right. Those inversion bubbles indicate inputs are active-low.

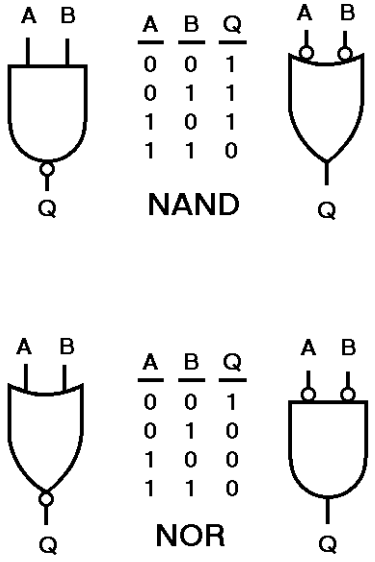


Figure 23-7 Equivalents.

Refer to the NAND at the left and observe the truth table in the middle. With active-high inputs both A and B must be high (logic 1) to change output Q. However, with active-low inputs (logic 0) as indicated by the input inversion bubbles to the right, either A or B low will produce an inverted output state change. The basic NAND gate component is an inverting-output AND with active-high inputs and

an inverting-output OR with active-low inputs. Similarly, the basic NOR gate is an inverting-output OR with active-high inputs but becomes an inverting-output AND gate with active-low inputs.

Once early digital logic designers got accustomed to that confusing functional equivalency they found a savings in package types. That is shown in Figure 23-8 for the same overall logic function. The logical function is the same although the one at the left is built with half of a 74x08 quad 2-input AND and one fourth of a 74x32 quad 2-input OR package. The one on the right uses three-fourths of a 74x00 quad 2-input NAND; one package type versus two types needed on the left.

The inversion bubbles are the indicator of active-low states. Since a NAND is an AND gate with inverted output, it can also be an OR with inverted input states. Basic functionality is the same (indicated by symbol shapes), but inversion requires some mental gear-shifting insofar as thinking and physical state viewing with an oscilloscope.

Many MSI (Medium Scale Integrated) and LSI (Large Scale Integrated) logic packages have some function pins requiring active-low inputs. Those inputs will have inversion bubbles on their package symbols or indicate active-low by an over bar of the function descriptor.

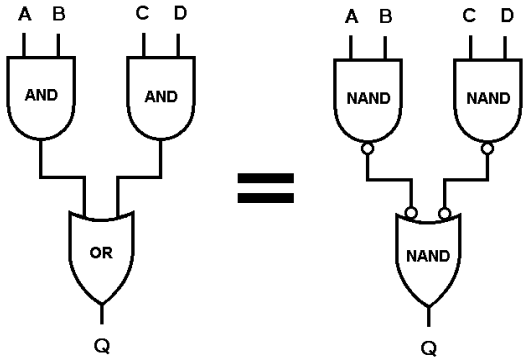


Figure 23-8 Combinatorial equivalency.

Propagation Delay, t_{PD}, and Power Dissipation

All gate structures, TTL or CMOS, have a finite time delay between input and output. There are no exceptions to this. The amount of delay is specified as t_{PD} on device specifications, almost always as a maximum. That is sometimes referred to as t_{PLH} (rise time, Low to High) or t_{PHL} (fall time, High to Low). Propagation delay will determine the speed of a logic device family, the maximum useable repetitive signal frequency.

Most of the propagation delay is due to parasitic capacitance within the IC physical structure but application of a device also requires a maximum capacitance of the digital circuit's output load that is in addition to other devices and their inputs. The external capacitance is specified on device datasheets for a stated propagation delay.

With the first of the TTL digital structures speed was proportional to power dissipation. CMOS and then Schottky junctions changed that to include signal repetition rate as the major determinant of supply power demand. Quiescent, no-AC-signal conditions with CMOS digital logic is so little it approaches insulation leakage current. Parasitic capacitance is the cause. Digital signals have fast rise and fall times and maintenance of such fast times requires internal power surges to charge and discharge parasitic capacities. Since operation of most gate structures involves transistor junctions in cut-off or saturation those surges or spikes of current must come from the external power source.

Local bypass capacitors are needed in high-speed digital logic circuit board layouts to handle those surges. The power supply output capacitance is too far away with too much wiring inductance between circuit and supply to be effective in supplying these transient current surges. As a general rule, bypass capacitors of at least 0.01 Fd with gate packages and at least 0.047 Fd with large-scale integrated devices should be located next to the digital package. Such bypasses should have the shortest-possible lead lengths to minimize self-resonance. Note: For hobby purposes with circuits at high repetition rates, it is better to use several smaller-value bypass capacitors than one large one;

there is no need to cut the parts count to the barest minimum as in commercial production.

Typical propagation delay, power dissipation, and signal rate per TTL family is:

Family	Delay per Gate	Power per Gate	Circuit Speed
74	10 nS	10 mW	DC to 35 MHz
74L	33 nS	1 mW	DC to 3 MHz
74LS	9.5 nS	2 mW	DC to 45 MHz
74H	6 nS	22 mW	DC to 50 MHz
74S	3 nS	19 mW	DC to 125 MHz
74F	3.7 nS	5.5 mW	DC to 125 MHz

This tabulation is a general guide only to show the relative delay-power-speed relationship between the five TTL families. Power demand does increase slightly with signal repetition frequency but nowhere near as great compared to CMOS devices.

The 74F is a later TTL family, originally from National Semiconductor, now from others, on par with 74S but having lesser quiescent current demand plus additional over- and undervoltage protection at inputs.

CMOS Digital Logic and Its Variants

RCA Corporation introduced the first CMOS digital device family under their part numbering scheme of CD4nnn, the package functional identifier as a 4-digit number. This family was characterized by extremely low DC power drain, high input impedances, signal speed about that of the 74L family with a 5 VDC supply. What was intriguing was a manufacturer's specifications at three supply voltages: 5, 10, and 15 VDC. maximum speed increasing with supply voltage. Unlike TTL the supply voltage was not restricted to being within ±10% of 5.0 V so it was ideal for varying-supply portable applications, aided by a very low quiescent supply current.

Many CD4nnn package functions were designed before RCA Corporation was purchased by General Electric in the late 1970s. The RCA semiconductor division was broken up and the CD4nnn family licensed out to other manufacturers, most packages reappearing as the 74HC family from National Semiconductor, the 74C. 74AC, 74ACT and CD4K from Fairchild Semiconductor. At the time of writing, Texas Instruments was still manufacturing the CD4000 plus the 74ABT, 74AC, 74ACT, 74AHCT families. The generic CMOS nomenclature is also known as HCMOS. Add to this the introduction of the new package styles (SOP, SSOP, TSSOP, etc.) and 3.3 Volt Supply device introductions and there is a large group of different device part numbers for any given logic function. The A pertains to "Advanced" CMOS which has internal structures improved for faster switching speed compared to the original family line. Shortly before the RCA Corporation sale and reorganization, their semiconductor division had come up with the B series of improved CMOS digital devices.

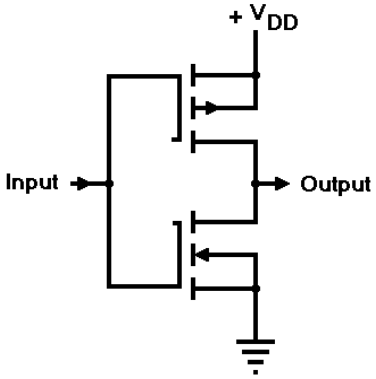


Figure 23-9 FET complimentary symmetry input-output.

CMOS digital logic is chiefly characterized by very low quiescent current, especially at inputs. The current demand of inputs at either 1 or 0 state is a few microAmperes. This is illustrated in Figure 23-9 showing a simplified gate structure of CMOS. Note that the FETs are IGFETs and that the transistor pair

is complementary. Insulated-gate FETs have very high input impedance, nearly all capacitive.

Modern CMOS gate structures incorporate diodes at inputs to act as electrostatic discharge (ESD) dampers as well as for negative polarity inputs and over-voltage positive inputs. The protection diode circuits do not affect normal logic level functions. The multi-supply-voltage CMOS devices require consultation of datasheets to make sure of correct logic state voltage levels.

Schmitt Trigger Inputs

The analog world has to interface with digital circuitry to get an AC or RF signal into being processed digitally. The godsend to this is the Schmitt trigger input, in effect a voltage switch with hysteresis so that its logic output state change occurs at an input level higher than the input level causing the opposite state change. Figure 23-10 shows that hysteresis effect.

For any typical 7414 hex Schmitt inverter, V1 is about +1.7 V and V2 is about +0.9 V. The transition times are 15 nSec, high to low or low to high. The V1 - V2 voltage delta (hysteresis) is about 0.8 Volts so any stray signals superimposed on the voltage input of a peak-to-peak value less than that delta would effect only the start and stop of transition times. The inverter output would stay within the proper voltage values for logic 1 and logic 0 states.

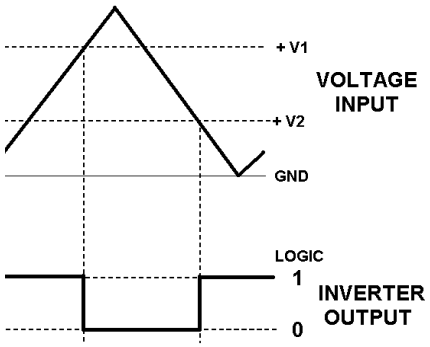


Figure 23-10 Hysteresis effect as with Schmitt trigger inputs.

Schmitt trigger inputs clean up input signals having superimposed stray garbage but their chief function is to make slow-rate, long rise and fall waveforms into fast-transition waveforms suitable for digital processing.

Comparators

Comparators are a special form of high-gain, wideband operational amplifier with internal hysteresis and with outputs suitable for driving digital logic devices. Their advantage is that they can accept negative input voltages as well as positive, making them useful as interfaces with many more sources.

Figure 23-11 is a general configuration. The Trigger Level Set can be a potentiometer or a fixed voltage divider made from resistors or just grounded as for an AC input. Figure 23-11 has the output logic polarity the same as the signal input. Reversing the Signal Input and Trigger Level Set to the comparator's inputs will make the output inverted.

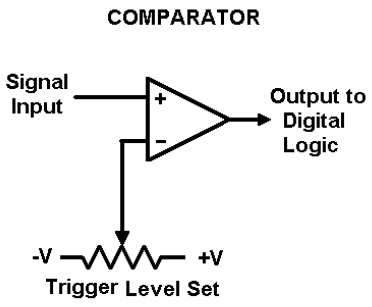


Figure 23-11 Comparator

Almost any operational amplifier can be used as a general comparator provided the outputs are clamped to stay within digital logic levels. Open-loop voltage gains of most op-amps are better than 60 db, therefore they are essentially limiters without feedback. Due to the relatively slow slew rate or rate of change of output (relatively low bandwidth), a conventional op-amp used as a comparator ought to have a Schmitt trigger between its output and the digital logic input.

Comparator outputs usually have a faster slew rate with a specified overdrive, the amount of input signal exceeding the specified input-output changeover voltage limits

Driving Digital Inputs with Discrete Solid-State Circuitry

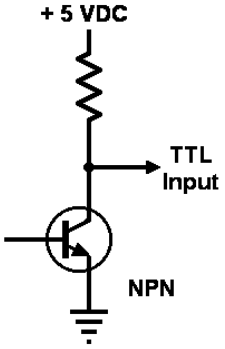


Figure 23-12 An NPN driving a logic input.

TTL device inputs are best driven by an NPN transistor with no emitter resistor as shown in Figure 23-12. The resistor value would be the same as a pull-up resistor (for the TTL family). When the NPN is not conducting the TTL input would be logic 1. The NPN needs only to conduct enough current to drop the collector-to-emitter junction just into saturation. The collector-to-ground voltage (about 0.3 V) would be logic 0.

Having the collector-emitter junction just at saturation, but no more, will insure the highest bandwidth of the direct-coupled NPN. A PNP with its emitter at + 5 V and collector connected to a resistor to ground would work but the resistor has to be a much lower value to satisfy the higher logic 0 input current. A PNP would have to conduct much more current to pull the top of the resistor up to minimum TTL logic 1 input voltage.

CMOS digital inputs can be driven by either NPN or PNP direct-coupled transistors. All they require is a minimum-current pull-up or pull-down (resistor to ground) resistor to insure the TTL to reach the proper logic level voltage when the transistor is in cut-off.

What to do With Unused Digital Inputs?

Unused TTL inputs ought to be tied-high with pull-up resistors. In older TTL devices an unconnected input would, eventually, float upward to a logic 1 level. In the case of gates such as needing only a 2-input NAND but a 4-input NAND gate is available and unused, the unused inputs can simply be paired with used inputs on the same gate. If not, just tie the unused inputs high or low directly. With CMOS devices the unused inputs should always be tied either high or low. Those should never be left open due to their high input impedance; unconnected inputs can float towards either ground or supply voltage with no control.

Driving External Discrete Circuitry from Digital Devices

That is not a problem if the device's specification maximums and minimums are obeyed. TTL outputs can drive more current at logic 0 state than at their logic 1 state. For the original TTL family devices that is 16 mA at logic 0 versus 0.8 mA at logic 1. CMOS devices are fairly equal at each logic state and average about 1 mA for TTL-equivalent voltage levels. Datasheet specifications should be checked before doing such circuit designs.

Early in TTL device use, outputs were classified as sinking current at logic 0 and sourcing current at logic 1. At logic 0 the output transistor junction to ground is at saturation. A maximum current (specified) will flow out to a load resistance connected to the supply rail; it will be a sink for the load. At logic 1 current will flow into an output if that load is a resistor to ground; the output is said to source current at logic 1.

With open-collector or high-impedance outputs, the device breakdown voltage specifications must be obeyed. Open-collector outputs are sometimes used to drive small relay coils. With any inductive load subject to absolute on or off current, there will be a large negative voltage spike developed when the current ceases and magnetic field held around the inductor by current collapses. Reversed polarity diodes across the relay coils is mandatory to avoid exceeding negative voltage breakdown limits. That is also true for discrete transistor circuits.

A Change-Over in Digital Device Families

On entering the new millennium it was safe to say that the original 54/74 TTL family was obsolete. Included was the Low power branch (74L), High-speed bipolar (74H) types. Some of the Schottky branch (74S) devices were missing from the market. What is left in bipolar TTL is the 74LS (Low-power Schottky) and the 74F (introduced by National Semiconductor). Part of the semiconductor industry's reorganization included National Semiconductor dropping its logic and memory device line, most of which was acquired by Fairchild in 1997, and Motorola Semiconductor spinning off into ON Semiconductor and Freescale.

The CMOS logic devices advanced with the demand for lower power supply drains to satisfy portable equipment. The 74AC and 74HC families were popular in the international semiconductor market, high speed in the 74AC and flexible supply voltage range in the 74HC families were in demand from the 1980s onward. But, adding to the confusing part numbers assigned by manufacturers, the original RCA Semiconductor CD4000s part numbers were kept as part of some second-sourced, popular CMOS IC devices. As an example, the Philips and Texas Instrument 74HC4046 Phase-Locked Loop IC is pin-for-pin and functionally equivalent to the RCA CD4046 introduced nearly three decades prior to the 1997 Philips and TI datasheet dates. The clue to similarity (or equality) is in the four-digit portion of the part designation; the older TI digital device part numbers rarely went over three digits.

Semiconductor manufacturers have always tended towards identifying themselves in part numbers by a prefix letter or letter pair. Following is an approximate tabulation of some part number prefixes and manufacturers. This may help:

SN	Texas Instruments (in some parts a CD prefix also used)
MC	Motorola (ON Semiconductor and Freescale, includes analog devices)
TC	Toshiba
DM, LM	National Semiconductor (logic line taken over by Fairchild)
AD	Analog Devices (primarily for complex mixed-mode systems on a chip)
MAX	Maxim Semiconductor
DS	Dallas Semiconductor
LT	Linear Technology

Add to the above so-called generic replacements of devices by NTE with NTE's own kind of part number scheme as well as the Radio Shack store chain's own coding and we hobbyists have no choice but to accept catalog data as correct when we buy parts.

The suffix on logic device part numbers usually identifies the type of package but that is up to the manufacturer to specify and hobbyist to find in detail. In general, a P or N suffix denotes a plastic dual-inline package (PDIP) which is most popular to hobbyists (and of least cost).

Some of the legacy devices of the 1970s never survived to the new millennium. Most of that is due to lack of customer demand or the growing competition of systems on a chip that replaced whole circuit boards with a single package and a handful of passive components. There are many references to logic devices in older textbooks which have simply become extinct on the market.

Table 1 following is a function versus generic part number of basic 5 VDC logic devices which have survived into the new millennium, given by the two- and three-digit 7400 bipolar/CMOS and four-digit 4000 CMOS numbers based on mid-2006 catalog data of Allied, Mouser, Digi-Key, Newark, and Jameco mail-order retailers.

Table 1 - Digital Device Functions and Family Branch Numbers

Device Function	CMOS Family Cdnnnn	TTL Family * 74xnn, 74xnnn
Hex Buffer	4050	--
Hex Inverter	4069	04
Hex Inverter Buffer	4049	--
Hex Inverter open-collector	----	05
Hex Inverter with Schmitt Trigger	4584	14
Quad 2-Input		
NAND	4011	00, 37
NAND open-collector	----	01, 38
NAND open-collector Buffer	----	26
NAND with Schmitt Trigger	4093	132
NOR	4001	02, 28
NOR open-collector	----	33
OR	4071	32
AND	4081	08
AND open-collector	----	09
Exclusive-OR	4070, 4507	86, 136, 386#
Exclusive-NOR	4077	--
Exclusive-NOR open collector	----	266
Dual 3-Input		
NOR plus Inverter	4000	--
Triple 3-Input		
NAND	4023	10
NAND open-collector	----	12
NOR	4025	27
OR	4075	--
AND	4073	11
AND open-collector	----	15
Dual 4-Input		
NAND	4012	20, 40
NAND open-collector	----	22
NAND with Schmitt Trigger	----	13
NOR	4002	--
OR	4002	--
AND	4082	21
Dual 5-Input		
NOR	----	260#
Single 8-Input		
NAND	4068#	30
Single 13-Input		
NAND	----	133#

Table 1 - continued

Device Function	CMOS Family Cdnnnn	TTL Family * 74xnn, 74xnnn
Monostable Multivibrator		
Single 1-Shot	4047	121#, 122
Dual 1-Shot	4098, 4528, 4538	123, 221#
Bistable Multivibrator		
Dual D Flip-Flop	4013	74
Dual J-K Flip-Flop	4027	73#, 76#, 112
Counters/Dividers		
Binary Up Counter (preset)		161, 163
Dual Decade Up Counter	4518	390#, 490#
Dual Binary Up Counter	4020	393
BCD Up/Down Counter	4510, 4029**	190#, 192#
Binary Up-Down Counter	4516, 4029**	191, 193
7-stage binary ripple-through	4024	--
12-stage binary ripple-through	4040	--
14-stage binary ripple-through	4020	--
4-Bit Shift Registers		
parallel-in, parallel-out	4035	194, 195
Dual Serial-in, parallel-out	4015	--
8-Bit Shift Registers		
parallel/serial-in, serial-out	4014, 4021	--
serial-in, parallel-out	4034	164
parallel-in, serial-out	----	165
parallel-in, parallel-out	----	299, 323 (bidirectional)
Latches/Storage Registers		
4-Bit	4076	75, 77, 175, 256
8-Bit	4099	259, 373, 374
8-Bit Data Bus Driver, 3-State Outputs		
Non-Inverting	----	241, 244
Inverting	----	240
8-Bit Data Bus Transceiver, 3-State		
Non-Inverting	----	245
Inverting	----	640
Multiplexers - Digital		
Quad 2-to-1, Non-Inverting	----	157
Quad 2-to-1, Inverting	----	158
Dual 4-to-1, Non-Inverting	4555	153, 253
Dual 4-to-1, Inverting	4556	352, 353
8-to-1, Inverting	4051	151, 251 (3-state)
Multiplexers - Analog or Digital		
Quad Bilateral Switch	4016, 4066	--

Table 1 - continued

Device Function	CMOS Family Cdnnnn	TTL Family * 74xnn, 74xnnn
Decoders/Demultiplexers		
3-line-to-8-line, Inverting	----	138
Dual 2-line-to-4-line, Inverting	----	139,155, 156 (3-state)
BCD-to-7-Segment Decoder-Driver	4555, 4558	47, 48
BCD-to-7-Segment Latch-Decoder-Driver	4511	--
BCD-to-7-Segment LCD, Latch-Driver	4543	--
BCD-to-Decimal Decoder	4028	42
BCD-to-Decimal Decoder, open-collector	----	145
Arithmetic		
4-Bit Full Binary Adder	4008	83, 283
4-Bit Magnitude Comparator	----	85

# Indicates production may have stopped as of or before 2013 based on absence or scarcity in a family branch in that year's catalogs.

* x for 74 x nnn refers to family branch type: LS, F, AC, HC, S, ACT, HCT, or none (original).

** indicates Selectable decade or binary

Appendix 23-1

Evolution of the *74* Families

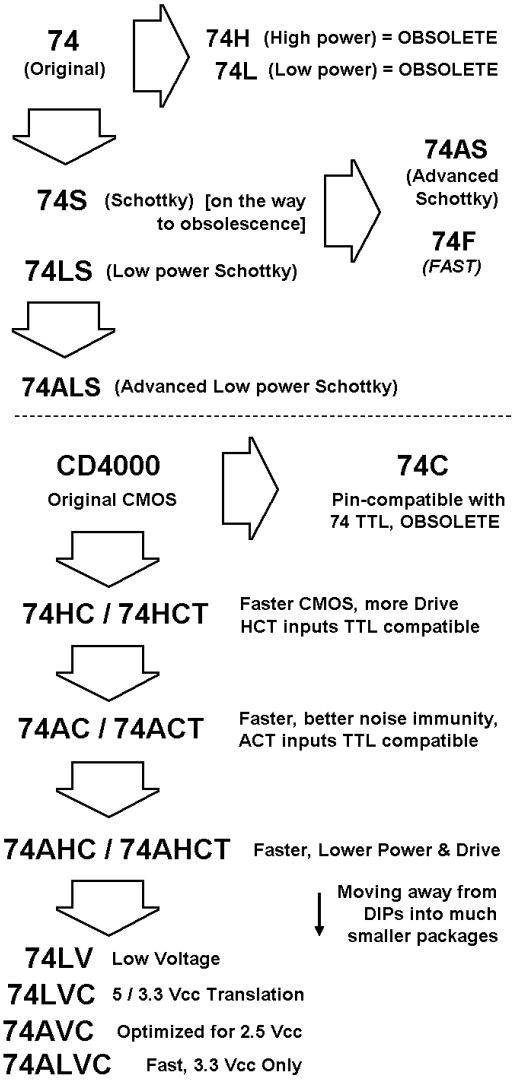


Figure 23-13 Evolution of the two major digital logic families, TTL at top and CMOS at the bottom.

In the 1960s there were several *number families* of bipolar logic devices.[1] As market demanded, the *7400*s became the family of choice. The original *74 devices* are still available on the market as of the first decade's middle in the new millennium but the *first generation*, the 74H and 74L, are no longer in production although a few retailers have old stock on hand.

Of the third generation, 74S and 74LS, only the 74LS is still in production, the 74S succeeded by the fourth generation of 74AS and 74F. There are indications that the 74AS is bound for an early obsolescence due to the slightly faster 74F series. As one generation succeeded another, the power demands grew less for the same operating speed.

The *CD4000s* were only slightly behind the *74*s but slowly became desired for their phenomenally low quiescent current demand relative to bipolar logic. Designs were going away from the high-speed demands of mainframe and minicomputer-oriented electronics into consumer electronics and the ubiquitous personal computer. CMOS had many advantages but operating speed was not one of them at first. With improved internal technology, the pin-compatible, function-compatible (with 74s) series *74C* tried, but did not quite take with enough logic designers; the 74C became obsolete early and the newer 74HC and 74AC of the CMOS 3rd generation replaced it.

By the time of the fourth generation of CMOS, the internal structures are incorporating bipolar junctions for achieving the best of both. The ***T*** in the CMOS line indicates input voltage compatibility with the old TTL standard voltages. The fifth generation is a merging of bipolar and CMOS with supply voltages dropping.

The *why* of the de facto standardization on the *7400* function-package numbering scheme is relatively unimportant. Nearly every major semiconductor manufacturer had their own line of

[1] From the author's observation, having worked at design since then, and also having to put up with an ever-changing logic device family tree where only a few branches remained by the new millennium.

saturated bipolar logic function families. Texas Instruments' nomenclature system became the *most comfortable* for designers in the 1960s-1970s.[2] It became the functionality *standard* with TTL as the most-used type of logic component. Sylvania seems to have originated TTL in their *SUHL* (Sylvania Universal High-Level) family in 1963 followed by TI in 1964; the TI numbers predominated.

ECL, Emitter-Coupled Logic, was innovated early for high-speed operation, possible because the transistor junctions within were ***not*** operating between saturation and cutoff. Internal IC circuits had much less capacitance to charge or discharge during state transitions, therefore operation could be more rapid than TTL. ECL was plagued with the unfortunate (and uncomfortable) choice of a ***negative*** *supply voltage* and *reversed logic levels* that was not compatible with all other families. ECL did find a niche application in the high-speed *supercomputer* mainframes. PECL, or Positive ECL, was an attempt to merge high-speed ECL with the rest of the digital world but was not popular except in specialized applications such as *prescaler dividers* for UHF and microwave region frequency control. Motorola seems to have originated ECL as their proprietary *MECL* (Motorola ECL) family in the 1960s.

For hobbyists, the godsend was the 74LS family once the Schottky transistor junctions were incorporated. Operating speed was high enough but not too high to be seen on less-expensive oscilloscopes. Power demands were reasonably low enough to keep supply costs down; greatly helped by the series regulator ICs which have become the standard power supply regulator (a subject in itself). Best of all, the prices per package were more affordable; 74LS was adopted for the much larger but less glamorous non-computer logic systems, thus helping to bring the prices down.

A major breakthrough for ultra-low quiescent power consumption logic was CMOS.[3] That nearly had an early demise due to susceptibility damage from ESD or ElectroStatic Discharge from handling without grounding. This was essentially solved in the ***B*** *suffix* series of the *CD4000s* through internal diode protection. Unless specifically designed for TTL compatibility, the general CMOS families can operate over an extremely wide range of DC supply voltage, the higher the faster. With improved MOS transistor junction technology, CMOS operating rates increased until they became compatible with 74LS.

PMOS and NMOS, *positive* and *negative* MOS junctions have been used in subsystems-on-a-chip, including some microprocessors. There were no logic families innovated for PMOS and NMOS devices other than those subsystem designs. Such devices were made to (generally) be electrically compatible with TTL families.

The popularity of the original *CD4000* line has grown enough that some later-generation *74*

---

[2] Subjective observation. When faced with complex tasks to perform under very tight time constraints, designers would opt for a well-rounded collection of functionally-capable components, availability, reliability, cost, constructability, and enough characteristics information to allow them a chance to make their systems work. This is not to say that other manufacturers had inefficient or ill-conceived or unreliable logic families...there were, perhaps *too many different families and choices* in the 1960s-1970s. The market was competitive but there were no *rules* that made one product line supposedly better than others. CMOS was perhaps an exception but it took nearly a decade after introduction to become a de facto low-power standard family. As to which logic technology or family is *better* boils down to subjectivism and how much an individual is impressed by marketing influence techniques (i.e., *brainwashing* potential customers).

[3] Most digital logic systems will do little but *sit there*, consuming DC power while waiting for some stimulus. While operating speed was been the driver for choosing a particular family, most applications (other than the GHz-clock-rate PCs' processors and memory) actually operate at low signal rates. DC power consumption during quiescent periods adds to problems of heat dissipation. In portable, battery-powered applications a low DC power consumption is essential to retain a long operating time.

families incorporate those number nomenclatures after the *74xx*.[4] The 74C family started that but it was too early in the transistor junction technology evolution to capitalize on it. With the demise of the RCA Corporation and licensing out of their original *COS/MOS* line, the nomenclatures get a bit confusing.

The fifth generation of logic families is noted for incorporating ***BiCMOS*** structures or a CMOS input followed by a bipolar transistor output. Such junctions are in the ABT, ALB, ALVT, ACT, BCT, and LVT newcomer families. While capable of operating faster than ECL and much less power consumption, BiCMOS has been somewhat restricted to bus buffers and bus transceivers in digital logic.[5] While BiCMOS has gotten much publicity for operating speed, that technology has been rather restricted to application-specific very large scale integrated circuits such as for wireless radio, cellular telephony, digital television and their *set-top boxes*.

The trend in new devices/families of logic at the beginning of the new millennium is dropping the supply voltage (and input/output signal levels), first to 3.3 VDC, then to 2.5 VDC, and finally to 1.8 VDC rail potentials.[6] Unfortunately for the hobbyists, physical size shrinkage down from DIPs to SOIC (Small Outline Integrated Circuit), SOT (Small Outline Transistor), TSSOP (Thin Shrink Small Outline Package), all at 0.050" lead spacings, down to innovative package name descriptors having 0.025" lead spacings. Surface mount technology may result in very compact electronics but it does make life difficult for hobbyists not having 10-power stereo microscopes nor micro-manipulators in their workshops.

Fortunately for hobbyists, the 74LS and CD4000 families have survived the four-decade period since their introduction, are sourced by many semiconductor manufacturers.

---

[4] Texas Instruments' *Logic Cross-Reference, 1st Edition, 2003*. 311 pages (available on-line) with cross-references to part numbers of Fairchild, IDT, Motorola, ON Semiconductor, Pericom Semiconductor Corporation, Philips Semiconductors, STMicroelectronics, and Toshiba. While TI logic part numbers use their familiar *SN*74 prefix, some parts use a ***CD74*** prefix.

[5] Many of the original 74ABT family devices have become obsolete (rather, out of production) less than a decade after introduction. That may be due more to a saturation in so many new devices created for the personal computer market, once again making things uncomfortable for designer choices by having too many to consider.

[6] Some consumer item devices operate down to 1.3 VDC levels, such as the relatively slow-speed clock oscillator-dividers in wristwatches and wall clocks using a single battery cell. Those are specialty ICs rather than logic families.

Appendix 23-2

Determination of Power Supply Drain

Once you have developed a digital logic subsystem, you must provide it with a DC power supply source of a certain current capability. With TTL determining the current requirements per device is straightforward: Look it up in the big datasheets. The ***quiescent*** (no signal) condition is generally good enough; TTL is bipolar and consumes DC power at a fairly steady rate, signal or no-signal.

CMOS devices are decidedly different. Their quiescent current drain demands are minuscule, rarely exceeding 0.5 mA per package at no-signal conditions. Their power demands are ***frequency dependent,*** increasing proportionally with signal frequency increase. The reason for that is charging all the stray capacitances within and without the device. To ease the designer's burden, CMOS device manufacturers specify their ***power dissipation capacitance*** or $\mathbf{C_{PD}}$ on datasheets in pFd. To determine the current demand, use the formula below to calculate it:

$$I = C_{PD} \bullet V_{CC} \bullet f + I_Q \qquad (23\text{-}1)$$

Where:

I = DC current in μA

C_{PD} in pFd as given on datasheet

V_{CC} = Logic supply voltage, Volts

f = Maximum operating frequency, MHz

I_Q = Quiescent (no signal) current demand, μA

As an example, let's use the 74HC192 counter from the datasheet information of STMicroelectronics (then SGS-Thompson) of October, 1992, giving the power dissipation capacitance as 68 pFd at 5 VDC supply and quiescent current of 40 µA. When operating at 10 MHz, the calculation is:

$$I = 68 \bullet 5 \bullet 10 + 40 = 3.44 \bullet 10^3\ \mu A = 3.44\ mA$$

It's easy to see that current demand goes up in direct proportion to frequency. Note: If the frequency range is at all broad, use the highest frequency; that way the power supply won't be strained. A 74LS192 was rated at 17 mA *typical*, 31 mA maximum current demand.

For multiple gates per package, the calculation is a ***summation*** of the worst frequency per gate for that package. For the 74HC00 2-input NAND datasheet of 30 June 2003 from Philips, their current-demand-at-frequency is a bit more complex as follows:

$$I = C_{PD} \bullet V_{CC} \bullet f \bullet N + \sum (C_{LOAD} \bullet V_{CC} \bullet f) \qquad (23\text{-}2)$$

Where:

I = Total current demand in μA

C_{PD} = Power dissipation capacitance in pFd (from datasheet)

V_{CC} = Power supply voltage

f = Frequency, MHz

N = Number of gates handling f

C_{LOAD} = Load capacitance, pFd, output of each gate

Σ = Sum of individual gate calculations in parentheses

The load capacitance has to be estimated. It is the sum of input capacitance of each device being driven plus the wiring capacitance to ground. A 74HC00 input capacitance is 3.5 pFd and, if this gate is driving four other gates of the same type, that alone would be 14 pFd. There might be 16 pFd of wiring capacitance for a total of 30 pFd.[7] If all four NAND gates are used at rate **f**, 10 MHz in this example, then N would be 4. Each gate within the summation parentheses term would have:

$$30 \bullet 5 \bullet 10 = 1.5 \bullet 10^3 = 1.5 \text{ mA}$$

$$4 \bullet 1.5 = 6.0 \text{ mA total within the summation parentheses}$$

The full calculation of formula 35-2 would then be:[8]

$$I = 22 \bullet 5 \bullet 10 \bullet 4 + 6.0 \bullet 10^3 = 4.4 \bullet 10^3 + 6.0 \bullet 10^3 =$$
$$10.4 \bullet 10^3 \ \mu\text{A} = 10.4 \text{ mA}$$

By contrast, a 74LS00 will draw about 4.5 mA, but does that during quiescent periods as well as at its maximum switching speeds. A 74HC00 will draw, at worst, about 20 µA quiescent current.

Power dissipation capacitance is derived by AC testing. The Philips *User Guide [for HCMOS]* explains this and many more.[9] Most other manufacturers have similar product family documents available although their C_{PD} values may differ slightly. If in doubt, take the highest value. If only one C_{PD} value is given, just multiply it (in pFd) times fastest operating rate in MHz times supply voltage for the device current demand in µA.

---

[7] There is no good way to estimate wiring capacitance since it will vary considerably depending on trace widths (on PCBs) and whether or not a ground plane exists on the opposite board side, length and crossings in wire-wrap plus wire sizes and insulation, the amount of bundling in prototype board wiring, and so forth. At best, one can assemble something and actually measure that finished capacitance by a meter capable of resolving pFd capacities. Usually, but not always, a high load capacitance also means long lead lengths, either of which can slow down fast rise and fall times.

[8] Philips datasheets' power demand calculates microWatts of power; that requires squaring the supply voltage. Formula 35-2 simply divides the power by voltage (eliminating the squaring) with the result being current demand.

[9] 25 November 1997, available on the Philips Semiconductor website.

If the total current demand calculated is as much as twice that of the finished circuit, no real mistake has been done for hobby purposes. Since a hobby application is not a production design task of making everything the most cost-effective, there is some security in being able to expand the project later, adding more circuits and thus drawing more power.

One thing that cannot be skimped is ***bypass capacitors***, particularly in doubling and tripling minimal lead length ceramic capacitors ***at the logic package***. Never depend on the power supply's output capacitor to do all the bypassing except at audio frequencies and below.

Chapter 24

Selection, Arithmetic, Switching

Basic gates can be arranged to select specific binary state combinations, do binary arithmetic, and switch lines or whole binary words. Those can be found by inspection; formal methods of seeking bit patterns aren't required of hobbyists. There is an introduction to specific logic devices which can help reduce parts count plus some of the building blocks for later digital chapters.

General

Selection of certain bit patterns in binary words are necessary to produce some output or initiate an internal circuit action. Certain ***bit states*** or ***state combinations*** need to be used and the selection process can be done by simple observation for most hobby purposes. Formal methods are not necessary in most cases.[1] As a definition, a binary ***word*** is a group of bits together to form some information out of or into a circuit. Computers have *data words* in sizes of 16 to 64 bits long. A *byte* is commonly 8 bits wide, a term to that came into vogue with advent of the Personal Computer. Not necessarily a binary word per se, but byte fits into the word definition.[2]

AND, OR Gates for a State Selection Problem

In this first example, let's say we have a small binary word of only 3 bits. Three binary bits will have 8 possible combinations of 1s and 0s. The goal is to select 3, and only 3 state combinations for an output and to try for the simplest method. A semi-graphical tabulation of the bit states is:

D	C	B	D	C	B	
0	0	0	1	0	0	
0	0	1	***1***	***0***	***1***	← Desired bit states
0	1	0	***1***	***1***	***0***	← shown in dark
0	1	1	***1***	***1***	***1***	← italics

[1] The most-used method among many digital designers is the *Karnaugh Map* or familiarly, *K-Map.* A Karnaugh Map procedure and search is not intuitive to everyone nor is it necessary for most hobby projects.

[2] The first personal computers had instructions and data of only 8 bits in size. The text characters into and out of those computers was also 8 bits in size. Such small computer words were, perhaps, considered *bite-size* relative to the 60- and 64-bit data words of big mainframes of the late 1970s, hence the coining of *byte.*

Converted to hardware in Figure 24-1 are three ways to produce the same output. Note the Boolean logic algebra statements at each output.

Figure 24-1 (A) is the most formal of the logic state combinations, three AND conditions of 3 bits each ORed for the output. Note that bit D is always high for each of the ANDs.

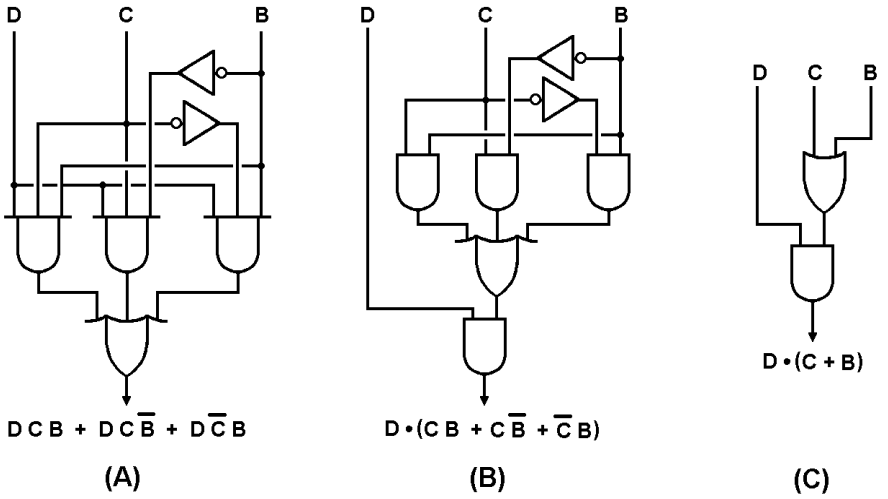


Figure 24-1 Gate arrangements and Boolean algebra statements for outputs in the first example. All produce *identical* outputs!

Note also the two inverters in Figure (A) and (B). Since there are only three inputs and AND gates need all inputs high to complete the AND, the inverters supply the ***B-Not*** and ***C-Not*** inputs representing the logic 0s to be ANDed.[3]

Since bit D is always high, the gate arrangement of Figure 24-1 (B) will work. The three combinations of bits B and C are ANDed, then ORed, that output ANDed to bit D. Two binary bits will have 4 state combinations so it is wasteful of hardware to deliberately select 3 of the 4 as in (B). Inspect the little tabulation on the previous page. In the three desired states of DCB, either bit C or bit B are always high; they are never both low. A single OR gate in Figure 24-1 (C) produces that, its output ANDed with bit D for the desired output. Only two gates in (C) and no inverters.

The Boolean logic algebra notation can be seen in Figure 24-1. In bit state logic, the + denotes an OR. The optional little dot • denotes an AND. No dot means the same AND as in the 3-bit AND groups in the output statement of Figure 24-1 (A). That statement is pronounced *D and C and B ORed with D and C and B-Not ORed with D and C-Not and B.* It isn't necessary to pronounce it but the written convention is convenient once it becomes familiar. The enclosing parentheses in (B) means that bit D is ANDed with the three 2-bit groups ORed within the parentheses. Figure 24-1 (C) logic equation of *D and C or D* is an equality to the other three statements.

The importance of detailed inspection of bit states lies in the simplification of the circuitry and thus the least hardware and wiring necessary to accomplish the task. There are some digital logic devices that might help such state combination selections and it is useful to examine them. First thing, though, the circuit of Figure 24-1 (C) could not have NAND gates substituting for the OR-followed-by-AND pair. Direct substitution of NANDs with active-high logic state inputs allows only the AND-followed-by-OR sequence. Two inverters would have to be added to (C) if they were to be done with NAND gates. Figure 24-1 (C) will be used a bit later in a *binary-to-BCD converter.*

---

[3] Inverters turn the logic 0s into logic 1s suitable for ANDing. The inverted states of bit B and C are normally pronounced as B-Not and C-Not. In written notation they can be written with an overbar as shown in Figure 24-1.

Using *Decoder* ICs for State Combination Selection

The 74x155 *dual 1-of-4 Decoder/Demultiplexer* is useful for this sort of work, especially with state combinations that do not have easily-simplified state combinations.[4] It was designed to be a device with rather universal applications so it has a perhaps-confusing *truth table* description. *Truth tables* abound in digital device datasheets and, despite their seemingly odd phraseology, just show all of the input versus output state combinations possible with the device. A 74x155 is really *two* 2-bit decoders on one chip, sharing two inputs and having two separate *enable* inputs to each decoder.[5] Internal gating separates the two input bit combinations into four *active-low* outputs for each decoder. Active-low outputs allows ORing with NAND gates.

Figure 24-2 shows only half of a 74x155 Decoder used for the conditions of the first example.[6] Using only the ***a*** half with E_A enable (pin 1 of the DIP version) as the third, bit D input, the ***a*** decoder half can select any one of the four input bit states when bit D is high.

Pins 1 and 2 are internally made to an active-low AND gate with the pin 1 input internally inverted (to make it active-high). Pin 2 is hardwired to ground for a permanent logic 0 input. Pins 3 and 13 are both active-high with internal inverters as needed. The 74x155 functions as if it had eight NAND gates inside, four for each half of the decoder. Flexibility of the *enable* inputs allows a single 74x155 to be connected as a 1-out-of-8 decoder for a 3-bit input. That leads to a second example.

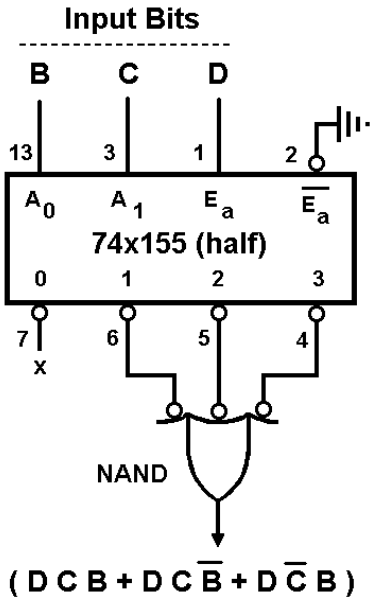


Figure 24-2 Half of a 74x155 and a 3-in NAND selecting the same state combinations as in the first example.

Taking Advantage of Existing Devices

Lets suppose we have the same three bits as in the first example but need to select four state combinations which are not such relatively easy discrete-gate decoding tasks. The tabulation following shows bit states and which

---

[4] As a reminder, the ***x*** in the 74x155 device nomenclature means it is available in all branches of a device family, **LS, F, AC,** or **HC**. The logical function and pinout is same, only the operating speed and power consumption differ among the family branches.

[5] ***Enable*** or ***Chip Select*** (usually abbreviated to ***CS***) inputs are useful in larger circuits allowing the same device to be used in many different ways. This can reduce the number of different digital device types used in a production application.

[6] Signal names are those of Motorola's in their 1992 *FAST and LS TTL Data* book, page 5-151, on the 74LS155. Texas Instruments' 1973 *The TTL Data Book for Design Engineers, First Edition,* page 313, has the same pinout and functions for a 74LS155 but calls the pin 1 input for *Data* and the pin 2 input for *Strobe.* The direct decoding inputs are termed *Select A* and *Select B* (pins 13 and 3, respectively). That is not unusual (there is no specific industry standard on all names) and can be confusing to a beginner in design. Study truth tables.

combinations are to be selected (in bold italics):

Pattern[7]	D	C	B	Output Pin
0	***0***	***0***	***0***	9
1	0	0	1	10
2	0	1	0	11
3	***0***	***1***	***1***	12
4	1	0	0	7
5	***1***	***0***	***1***	6
6	***1***	***1***	***0***	5
7	1	1	0	4

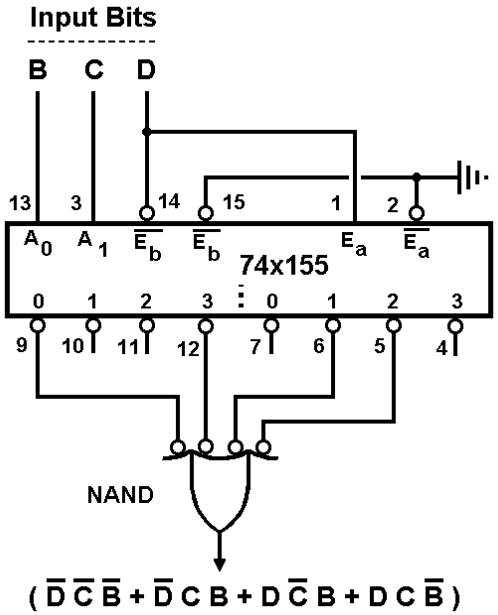


Figure 24-3 74x155 as 1-out-of-8 Decoder to select patterns of example two.

Figure 24-3 shows the 74x155 wired as a 1-of-8 Decoder. Pins 14 and 1 are tied together and used as the most significant input bit (D). When bit D is low, the left-half outputs are active. When bit D is high, the right-half outputs are active.

Pins 15 and 2 are both tied to ground (logic 0) to enable all outputs. Tied together, they could be used to enable the decoder from another source when logic 0. A 4-input NAND gate ORs the desired outputs' active-low to complete the four state combinations.

Logic gates could have been used to get the same output. The decoder-selector of Figure 24-3 results in a minimum number of digital devices (one and a half DIPs) and the least interwiring. A 74x156 performs the same function as a 74x155 but the outputs are *open-collector*. Open-collector outputs can be used to drive other circuitry such as lamps, relays, or LEDs.

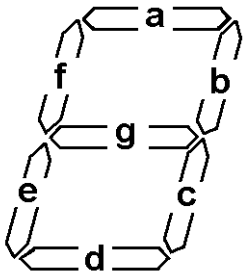


Figure 24-4 7-Bar Segment Display.

A 74x138 1-out-of-8 Decoder/Demultiplexer is commonly used for decoding three binary bits. It has three inputs all active-high. In addition, the device has three *enable* inputs, one active-high with the other two active-low. With the addition of a single inverter, four 74x138s could decode the 32 state combinations of 5 binary input bits. All outputs are active-low.

A 74x139 has two entirely-separate 1-out-of-4 Decoder/Demultiplexer. Each decoder has its own 2-bit inputs and only one active-low *enable.* All outputs are active-low. The 74x145 accepts 4-bit active-high BCD (Binary Coded Decimal) inputs and decodes those to 10 separate active-low open-collector outputs representing the ten decimal states of four BCD bits. The 6 state combinations of four bits not used in BCD are not decoded. There are no enable inputs on the 74x145 (inputs and outputs require 14 pins, leaving only 2 pins left for power and ground). A 74LS145 is guaranteed to operate at 24 mA when output is active and will withstand 15 VDC breakdown on inactive outputs.

---

[7] It is convention in digital logic to number-identify sequential binary state combinations beginning with 0 rather than 1. That way the number is equivalent to the binary word's *weight.*

A 74x154 was made in earlier times as a 1-out-of-16 Decoder with four-bit active-high inputs and two active-low enable inputs. It required a 24-pin DIP and has fallen out of favor with digital designers. It probably ceased production in the 1990s.

Decoding For Numeric Display Applications

The ***Seven-bar Segment*** display unit has become a standard for decimal numeric state indication to humans, usually in the form of LEDs to illuminate each segment. Generally-accept segment designations are shown right. ***Driver*** devices have internal gate arrays to select the correct segment for each decimal state. Such drivers take their inputs as 4-bit ***BCD*** states for each decimal digit. BCD or Binary-Coded Decimal is simply the first ten states out of 16 for a 4-bit binary data word. Table 1 following shows the relationship of decimal - binary state - segment illumination:

Table 1 - Seven-Bar Segment Display

Decimal	BCD Bits				Segments ON when logic 1						
Digit	D	C	B	A	a	b	c	d	e	f	g
0	0	0	0	0	1	1	1	1	1	1	0
1	0	0	0	1	0	1	1	0	0	0	0
2	0	0	1	0	1	1	0	1	1	0	1
3	0	0	1	1	1	1	1	1	0	0	1
4	0	1	0	0	0	1	1	0	0	1	1
5	0	1	0	1	1	0	1	1	0	1	1
6	0	1	1	0	1	0	1	1	1	1	1
7	0	1	1	1	1	1	1	0	0	0	0
8	1	0	0	0	1	1	1	1	1	1	1
9	1	0	0	1	1	1	1	0	0	1	1

It might be a fascinating personal puzzle to make a gate array that would decode each segment for each BCD digit. However, that is not productive since a number of ICs have been made for decades which do that in one package per digit. The 74x47 and 74x48 (open-collector outputs) do that in TTL; CD4511 does that in CMOS-input, bipolar-output devices. So do the CD4028 and 74x42 and have internal 4-bit latches suitable for retaining the input state. All are rated for 15 VDC breakdown on outputs and can drive common-anode LED seven-segment units.

A distinct advantage for numeric readability is *leading-zero blanking* capability, optional by adding a single wire between adjacent drivers. A *test* control pin allows all seven segments to be lit at once although this is a moot point with LED display illumination.[8]

LCD, Liquid Crystal Displays, can generally use the same 7-segment driver ICs by low-frequency pulses applied to *intensity control* pins on the driver ICs. LCDs require an AC waveform for excitation. Type and kind of waveform depends on the LCD display.

BINARY ARITHMETIC

[8] Prior to LED availability, small incandescent lamps provided illumination. Incandescent lamps are not reliable at their maximum illumination output ratings; LEDs are. The *lamp test* feature is an outgrowth of that. *Push-to-test* holders for aircraft lamps came into being for that reason during World War II on military aircraft.

Binary Addition

While that *seems* of little use outside of software in microprocessors and microcontrollers, it is of use in certain sub-systems of digital logic circuits. Binary arithmetic is almost absurdly simple as can be seen following for the four possible input conditions of two binary bits involved in addition:

Bit A ➛	0	1	0	1
Bit B ➛	0	0	1	1
Sum ➛	0	1	1	0
Carry out ➛	0	0	0	1

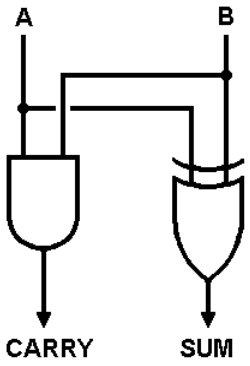


Figure 24-5 - Half Adder

A circuit for that function is equally simple as seen in Figure 24-5 forming a ***half adder***. An Exclusive-OR generates the Sum output since the output is high only when the inputs are *unequal*. An AND gate is needed for the carry-out since than will be high only when both inputs are high. It is called a *half adder*[9] because it does not accept a *carry-in*.

A ***full adder*** is shown in Figure 24-6, that one made up of two half adders plus an OR gate for the carry-out. The arithmetic table above would be valid for a full adder when the carry-in is logic 0. A carry-in at logic 1 would extend that table by:

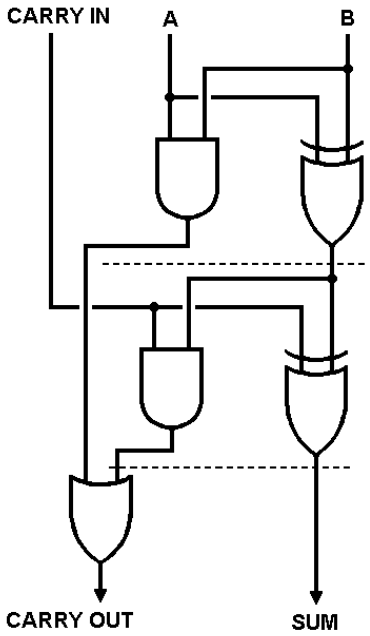


Figure 24-6 Full Adder, two half-adders plus an OR gate.

Bit A	0	1	0	1
Bit B	0	0	1	1
Carry-In	1	1	1	1
Sum	1	0	0	1
Carry-Out	0	1	1	1

Note that a carry-out is always generated when both input bits are high, regardless of the carry-in state. A carry-out occurs with a carry-in high and both input bits unequal. The arrangement of two half adders plus OR in Figure 24-6 satisfies the arithmetic rules. There will be only one bit as sum and only one bit as the carry-out.

The 74x83, 74x283, and CD4008 are all 4-bit wide binary adders handling four bit pairs in parallel, with a carry-in to the least-significant bit and a carry-out from the most significant adder. The 74x283 is a slightly faster version of the 74x83 with better *carry look-ahead* internal gating. When adding any

[9] That term was coined about 1950 when digital circuits were still designed using discrete components. It is possible it was done to distinguish it from *full adder* which could accept a carry-in.

number of binary bits in parallel, the carries will propagate adder-by-adder towards the most significant bit. While that is important only to computer central processor unit (CPU) designers, ***all*** digital devices exhibit some delay from input to output and that is a factor of time-sequential digital circuits.

Binary Addition and Subtraction

Assume two 4-bit binary words representing decimal 7 and decimal 13 are added. A pencil-and-paper exercise will appear this way:

	Binary 7	0	1	1	1		
+	Binary 13	1	1	0	1		
	Sum	1	0	1	0	0	= decimal 20, (16 + 4)

The binary addition can be widened to any number of bits but the basic rules of addition will apply the same to each bit of the words. There is no specific digital device to handle binary subtraction but the circuits could be built. The significant difference between the two is subtraction requiring a ***borrow***, not a carry. Single-bit binary subtraction arithmetic, A minus B, table is as follows:

Bit A ➛	0	1	0	1	0	1	0	1
Bit B ➛	0	0	1	1	0	0	1	1
Borrow-in ➛	0	0	0	0	1	1	1	1
Difference ➛	0	1	1	0	1	0	0	1
Borrow-out ➛	0	0	1	0	1	0	1	1

Note that the difference in binary subtraction will have the same states as the sum in binary addition. Borrow-out gating would be the dissimilarity. Rather than building specific circuits for binary subtraction, we can use a convenient binary equivalent to subtraction by ***adding*** the ***two's complement*** of the subtrahend,[10] ignoring any carry-out of the most significant bit. The two's complement of a binary number is ***all bits inverted plus 1 in the LSB.***[11]

Using the above example, find 20 minus 13 in binary. The difference should be binary 7. Binary 20 is the minuend and binary 13 is the subtrahend. Take the 13 and find the complement:

Binary 13:		0	1	1	0	1
Bits Inverted:		1	0	0	1	0
Add 1 in LSB:	+					1
Sum is the Two's Complement:		1	0	0	1	1

Note that adding the 1 preserves the odd or even status of the original. That status preservation is a requirement of the two's complement addition in place of subtraction. It does not normally need to be done physically by adders. Let's set up the problem of subtracting 20 - 13 by doing 20 + the

[10] In case you've forgotten things learned in grade school, the *subtrahend* is subtracted from the *minuend.*

[11] All bits inverted is called the *one's complement*; exchanging all 1s with 0s, all 0s with 1s.

two's complement of 13 using binary addition rules:

Binary 20:	1	0	1	0	0
+ Two's complement of 13:	1	0	0	1	1
Result:	0	0	1	1	1

(ignre the carry out of MSB)

The result is correct at binary 7. This method works directly ***only*** with positive quantities. If the problem was 7 - 13 (subtrahend magnitude larger than minuend magnitude), the result would require taking the two's complement of it as well as storing a negative sign bit somewhere.[12] Binary subtraction by two's complement addition in hardware is best relegated to problems of code/format conversion or division control of phase locked loops' dividers that must have an IF offset.

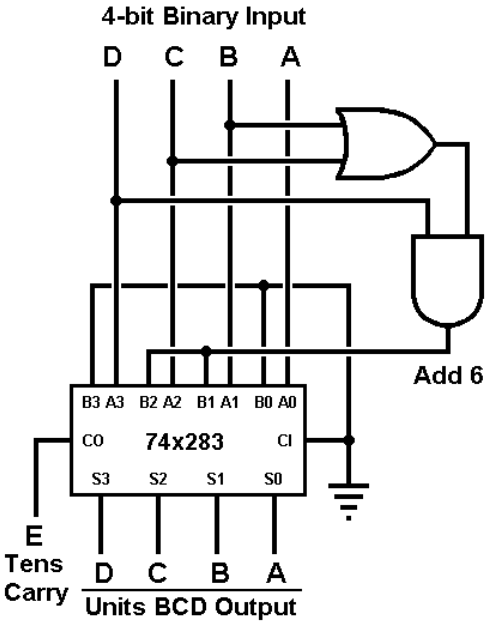


Figure 24-7 Converter for Binary to BCD using a 4-Bit Adder.

Simple Code or Format Conversion

BCD is binary but in only the first 10 states possible with a 4-bit word or data quantity. There are 6 binary state combinations which don't fit the decimal scheme for that digit but are significant for the next-higher decimal digit. Conversion from BCD to binary is straightforward: Use the BCD as binary. Conversion from binary to BCD requires a binary addition to convert all binary state combinations above decimal 9. To explain, the following truth table shows the relationship:

Binary Value	Binary Bits D C B A	BCD Bits E	BCD Bits D C B A	
0	0 0 0 0	0	0 0 0 0	
1	0 0 0 1	0	0 0 0 1	
2	0 0 1 0	0	0 0 1 0	
3	0 0 1 1	0	0 0 1 1	
4	0 1 0 0	0	0 1 0 0	
5	0 1 0 1	0	0 1 0 1	
6	0 1 1 0	0	0 1 1 0	
7	0 1 1 1	0	0 1 1 1	
8	1 0 0 0	0	1 0 0 0	
9	1 0 0 1	0	1 0 0 1	
10	1 0 1 0	1	0 0 0 0	(add 0110 to get BCD)
11	1 0 1 1	1	0 0 0 1	(add 0110 to get BCD)
12	1 1 0 0	1	0 0 1 0	(add 0110 to get BCD)

[12] That is easier handling it in software for a microprocessor. Binary 7 minus (addition by two's complement of) binary 13 would result in 11010 (00111 + 10011) and the two's complement of that is 110, the correct result for a minus decimal 6.

13	1 1 0 1	1	0 0 1 1 (add 0110 to get BCD)
14	1 1 1 0	1	0 1 0 0 (add 0110 to get BCD)
15	1 1 1 1	1	0 1 0 1 (add 0110 to get BCD)
		_	BCD bit E is in next higher Decade

Done in digital hardware, the Figure 24-7 circuit selects all binary state combinations from 1010 to 1111 and adds binary 0110 to create the 1 ½ digit BCD output. Any input state combinations from 0000 to 1001 get all logic 0s into the adder. The OR-AND gate arrangement is the same as in Figure 24-1 (C) of the first example.

One problem is that there are ***five bits*** on the output, a consequence of the BCD coding. If BCD adders were available, bit E, called the *tens carry out*, could become the carry-in for the adjacent more significant BCD adder.[13]

BCD addition can be done for one digit by an adder at the input, checking the sum of that to see if it is greater than 1001 and converting that first sum with the circuit of Figure 24-7. But, some augend-addend combinations can produce a carry-out from the first 4-bit adder. An example is decimal 9 (1001) added to another decimal 9 will produce a sum of binary 10010 (decimal 18). What to do about that new 5th bit from the adder's carry-out?
Some example additions are useful:

			Carry-in	0 0 0 1	
Decimal 9	1 0 0 1		Decimal 9	1 0 0 1	
Decimal 5	0 1 0 1		Decimal 9	1 0 0 1	
Binary Sum	0 1 1 1 0		Binary Sum	1 0 0 1 1	
Add 6	+ 0 1 1 0		Add 6	+ 0 1 1 0	
Result	1 0 1 0 0	[BCD 14]	Result	1 1 0 0 1	[BCD 19]

If the addend and augend in BCD did not add up any higher than decimal 9, the sum would go directly out with no *add 6* and no carry-out to the next higher digit. But, in order to properly sense all conditions where the add-6 conversion is necessary, the two gates of Figure 24-7 would require an additional OR gate after the AND. That OR gate would OR the *greater-than-9* condition and the carry-out of the first adder. Gates have expanded to three total, those having to come from other places on the board (two-input gates typically come in DIPs of four each).

Notice three facts: The least-significant bit does not need conversion; it can go straight through (or around) an add-6 conversion adder. The maximum sum of a BCD adder will never reach decimal 20, even with a carry-in high. The add-6 conversion is needed only if the binary sum *is greater than nine.* The *greater than* is a clue for using a digital device still in production, the 74x85 *Magnitude Comparator.* An alternate version of a BCD Adder using that is shown in Figure 24-8.

[13] BCD Adders were designed and made in TTL but the customer demand became low enough to cease production.

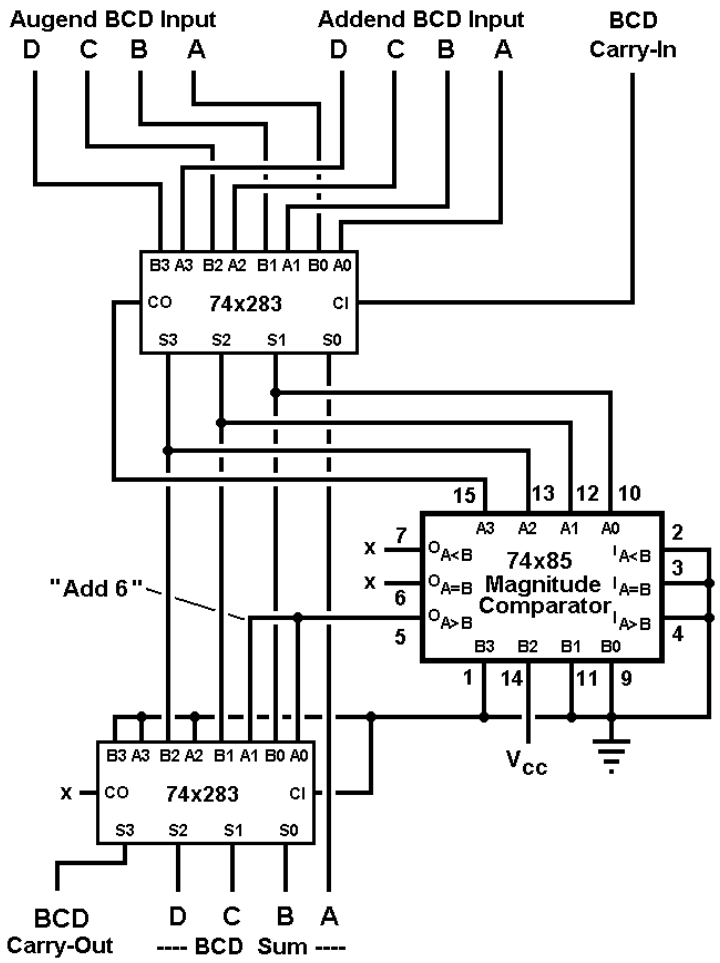


Figure 24-8 BCD Adder using a 74x85 Magnitude Comparator instead of gates to detect *greater than nine* conditions out of the top binary adder.

The 74x85 was intended for cascading with other Magnitude Comparators for any width (length or size) binary word. In Figure 24-8 it is not cascaded so those inputs (pins 2, 3, 4) are at logic 0.[14] The ***A>B*** output (pin 5) provides a logic 1 for the *add 6* conversion in the bottom 4-bit adder. The ***B*** inputs are hardwired as if for a binary 4. That is really a binary 8 since the comparison has *shifted right* once.

The least significant bit out of the topmost 4-bit adder does not need any conversion, therefore there is no need to convert all 5 bits of the first adders sum. Think of the topmost adder output as being a 4-bit output with its carry-out being the most significant bit.

It appears as if the bottommost adder is wired for an *add 3* but the conversion addition has been shifted right by one bit. The BCD carry-out comes from the most significant bit output (S3) of the bottommost adder; that 4-bit adder's carry-out pin (Co) can remain unconnected.

The circuit of Figure 24-8 is not optimum but it can be a stand-alone circuit for one BCD digit, repeated as needs be for more decimal digits. No *borrowed* gates necessary. The only way to make a BCD adder in a single IC is to burn one into a PLA (Programmable Logic Array) or PLD (Programmable Logic Device), beyond the scope of this book.

There is much more on format conversion circuitry in Chapter 99 later, including how to use an EPROM (Erasable Programmable Read Only Memory) on replacing a dozen binary adders plus assorted gates...and getting some extra functions, too.

Time-Sequenced Gating

Gating is often used to select a particular segment of time such as forming a pulse or

[14] All pin numbers are for the Dual Inline Package.

waveform from fixed counter state sequences. Such an example would be forming the vertical and horizontal blanking and synchronization pulses for an analog TV system.[15] Such gating can be used to set or preset feedback in variable-length dividers and counters. Another example is the forming of repetitive characters as in transmitter identification via morse code or in this first example, a ***CQ Generator.***[16]

Repetitive Two-Character Morse Code CQ Word Generator

Morse code characters are simple to use on transmitting and receiving but they have one drawback for digital generation and decoding: Variable character length.[17] The 8-level coding for electronic text transmission of ASCII (American Standard Code for Information Interchange) is fixed-length, any character. So is the much older 5-level code used in electromechanical teleprinters, sometimes called *Baudot coding*.[18] Morse code characters are composed of groups of *dots* and *dashes* with variable blank spacings in between for letter separation and word separation. The basic unit of measure is *dot length. Dashes* are three dot lengths long. Spacings between dots and dashes in a single character group is one dot length. Spacing between characters is at least three dot lengths. Spacing between *words* (contiguous characters) should be at least five dot lengths.

The word abbreviation *CQ* is used several times in *calling* (sending a request for contact) other stations or anyone that can hear the signal. In Morse code that word abbreviation is dash-dot-dash-dot, dash-dash-dot-dash (usually pronounced *dah-dit-dah-dit dah-dah-dit-dah)*. It is rhythmic and fairly easy to pick out, even in noisy reception conditions. While such a word is almost second-nature to an experienced Morse code operator, it is used here as an example of sequential design using gates since the code is familiar and of reasonably short length. The same principles can be applied to larger character groups if desired.

What is important for the circuit designer is setting up the problem for examination, then examining state combinations towards organizing the gating, finally looking even more closely for circuit simplification or alternatives.

---

[15] As this is written the USA is in the midst of changing over from the NTSC (National Television Systems Committee) analog television system to DTV (Digital Television). Several large-scale ICs were already available for Sync and Blanking signal generation, used in test equipment and closed-circuit TV systems. DTV is a bit more complex and is not covered in here; ICs for those require very large-scale IC design.

[16] This following example appeared in somewhat different form in *Ham Radio Magazine* of October, 1979, as part of a *Digital Techniques* tutorial series written by the author.

[17] Character part spacing and coding representation of letters, numbers, and punctuation herein follow the *International Morse Code* standard as published by the ITU, the International Telecommunications Union, a United Nations body.

[18] The French inventor did not have much to do with the old 5-level code but the surname was attached to it by so many for so long that it became synonymous with any 5-level teleprinter coding.

Setting Up the Gate Determination Problem

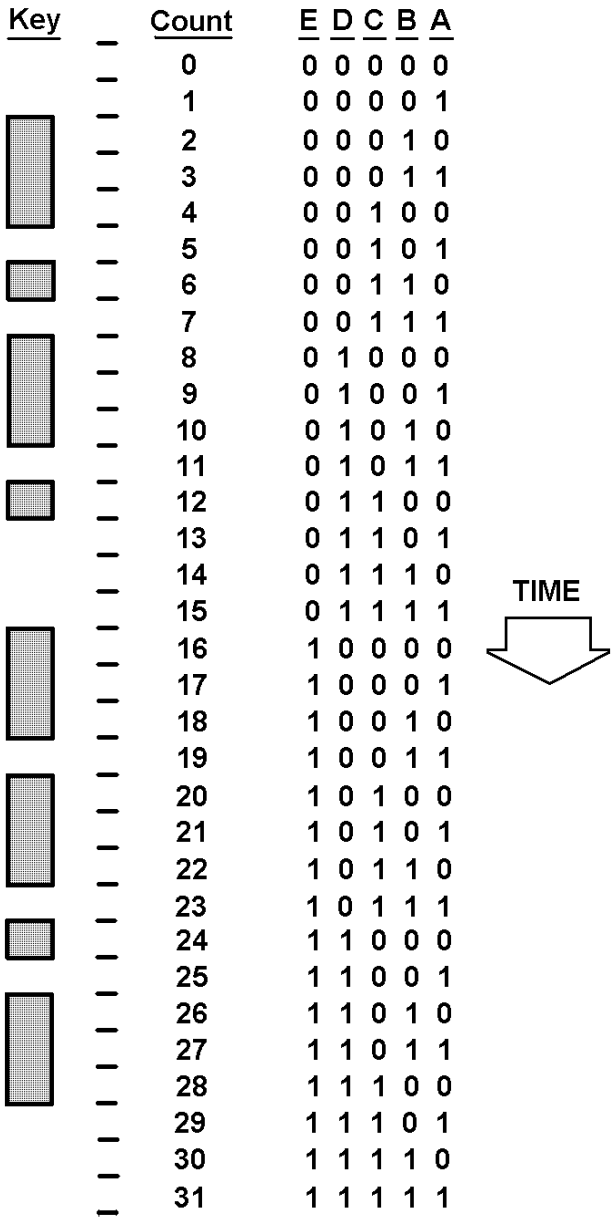


Figure 24-9 CQ Character and counter sequence.

With 14 dot times of Morse code characters and reasonable spacing, a 5-bit binary counter can provide the timing and state combinations from which the final key-down output is generated. Figure 24-9 shows the key-down[19] (grey) sequence and counter states, bit E (most significant bit and slowest stage) through bit A (least significant bit and fastest stage).

There are 3 dot times between the code characters. If the two letters are allowed to repeat, there will be 5 dot times between repetitions.

Note the bit E relationship to the letters. Bit E is logic 0 for all of letter C; it is logic 1 for all of letter Q. The characters were arranged that way to make timing and gating a bit easier to organize in hardware. That is permissible since the binary counter is there only to time the dots, dashes, and spaces. It could start anyplace. In this example it starts two dot times before letter C is begun.

Knowing that Bit E can be used as a separator between letters allows some control over the individual character gating. If Figure 24-9 is split in two for a 16-state count set, there is some visibility into what gating is needed for what. That is done in Figure 10. Note the overlap of key-down conditions for both letters. Gating could be arranged for both letters, letter C, and letter Q. Note also that, for both letters, Bit A is always 0. Except for one dot time of letter Q, Bit A is always 1 for individual letters. This is a good start in seeing what is required in a gating circuit.

---

[19] For the benefit of non-morsepersons, *key-down* is the state of transmitting. Key-up would be no transmissions and presumed listening for replies.

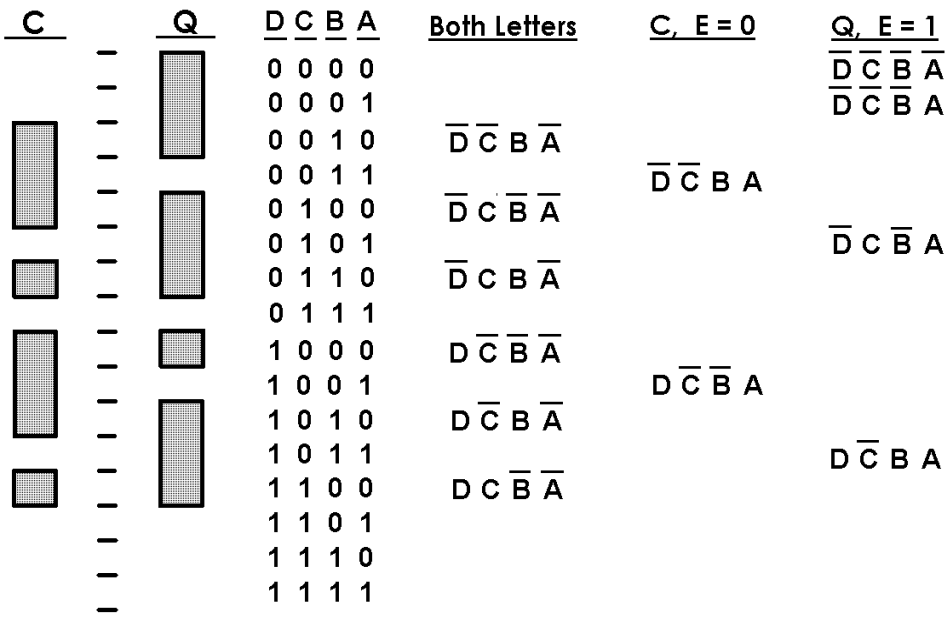


Figure 24-10 Reassembly of Figure 24-9 to separate letters C and Q.

A *first-cut* at a gate array is shown in Figure 24-11. This follows the *both* and individual letter logic state combinations, both letters to the left, letter C in the middle, and letter Q at right. Note that one NAND gate output is used for both letters and for letter C.

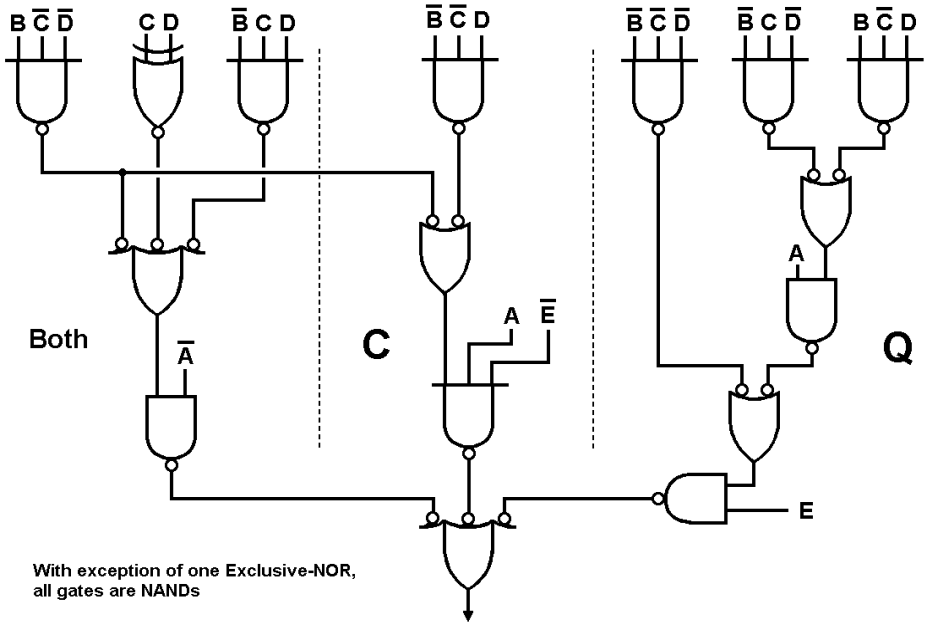


Figure 24-11 Gate array following tabulations in Figure 24-10.

All 5 of the counter bits ***must*** have inverted states available, either from the counter directly or from added inverters. The 6th of a Hex Inverter device could be used with an Exclusive-OR in place of the Exclusive-NOR. That Exclusive-NOR takes care of four of six state combinations of Bits DCB as 010, 011, 100, 101; B can be a *don't-care* in these four so the combinations of Bits DC would be 01 and 10, the active-output combination for an Exclusive-OR.

An Exclusive-NOR is used here so that a 3-input NAND gate can be used as an OR. If an Exclusive-OR were available from elsewhere, it could be followed by an inverter to allow the active-low ORing of a NAND. If 5 inverters are used to derive the *not* states of the counter, the 6th one could be used that way; inverter packages come in groups of six. The circuit is not optimum.

Another Version Using the *Inhibit* Feature of NAND Gating

Figure 24-11 will require 6 devices to make as shown: 6 2-input, 9 3-input NANDs plus an Exclusive-NOR. Two 2-input NAND gates and three Exclusive-NORs are left over for other things. It is often useful to try some partial state-combinations to see if the gate array can be simplified. One version of that is shown in Figure 24-12 (gates are numbered for explanation purposes).

What is rather different is using gates 1 and 2 as *inhibits* for gate 5's input. If any NAND input is low, the output is forced into a logic 1 state. Gates 1 and 2 effectively remove the first two and the last two DCBA states of Figure 24-10. Inhibit inputs to gate 5 in Figure 24-12 are marked with an asterisk Any state of the remaining 12 with Bit A low will have the both-letter states to pass on to gate 15. Gate 2 output is inverted to yield the all-low AND (via gate 11) needed for letter Q.

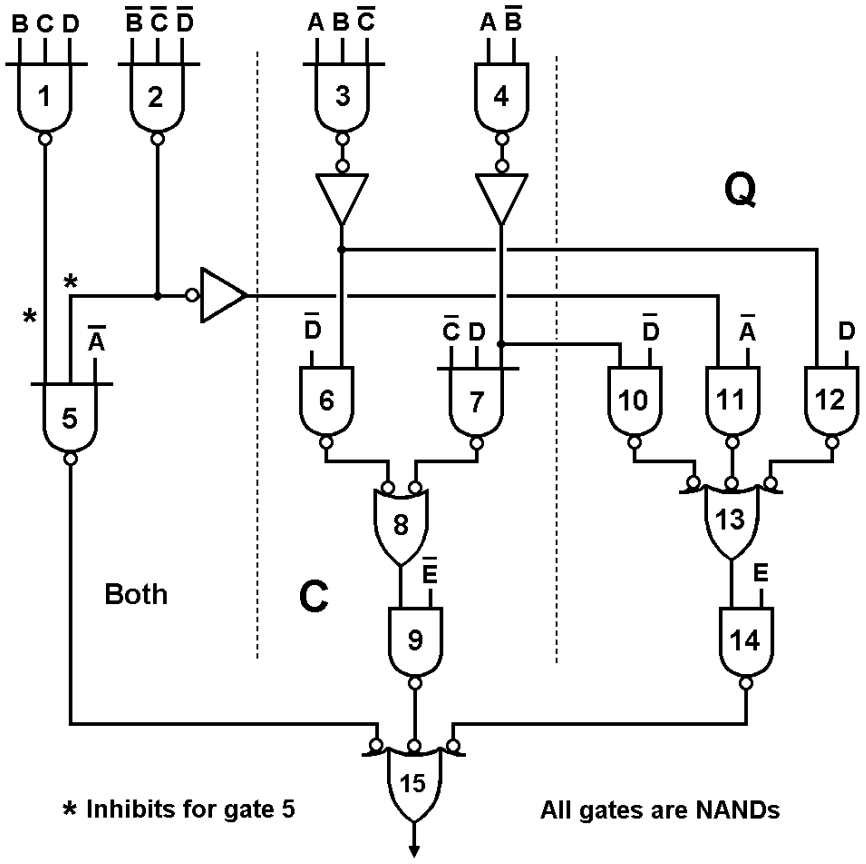


Figure 24-12 A second version using partial state combinations.

Gates 3 and 4 outputs are used in both letters and use inverters to restore the outputs to active-high for ANDing. That technique will reduce the need for four and three-input NANDs in gates 6, 7, 10, and 12. Whether or not that is an advantage overall depends on availability of other gates or using spare gates from Figure 24-12 for other purposes. As it is, the circuit of Figure 24-12 requires 6 digital devices: 8 2-input NANDs, 7 3-input NANDs, and 3

inverters. That's not a saving over the circuit of Figure 24-11 but it shows there are alternate ways to do the same functional task. There are more ways to do that as explained following.

A Simpler Character Generator Using Decoders

A 74x154 provides an active-low output for each of the 16 states of a 4-bit input. That fits the 16-bit timing of each letter character. All that is required is to OR the active-low outputs to get all the dot-time pieces needed. This does ***not*** require any inverters to get the inverse of counter states.

Gate 6 ORs the common dot-times from active-low outputs 2, 4, 6, 8, 10, and 12. Gate 1 ORs 3 and 9, then ANDs that active-high output with E-not in gate 3. Gate 2 is really an inverter (the only one required) and wired so since gates 1, 2, 3, and 5 can be one 74x00 quadruple 2-input NAND package.

The interconnections are all done almost as if the truth tables in Figure 24-10 were a wiring diagram. Total package count is 4 devices with half of a dual 4-input NAND to spare.

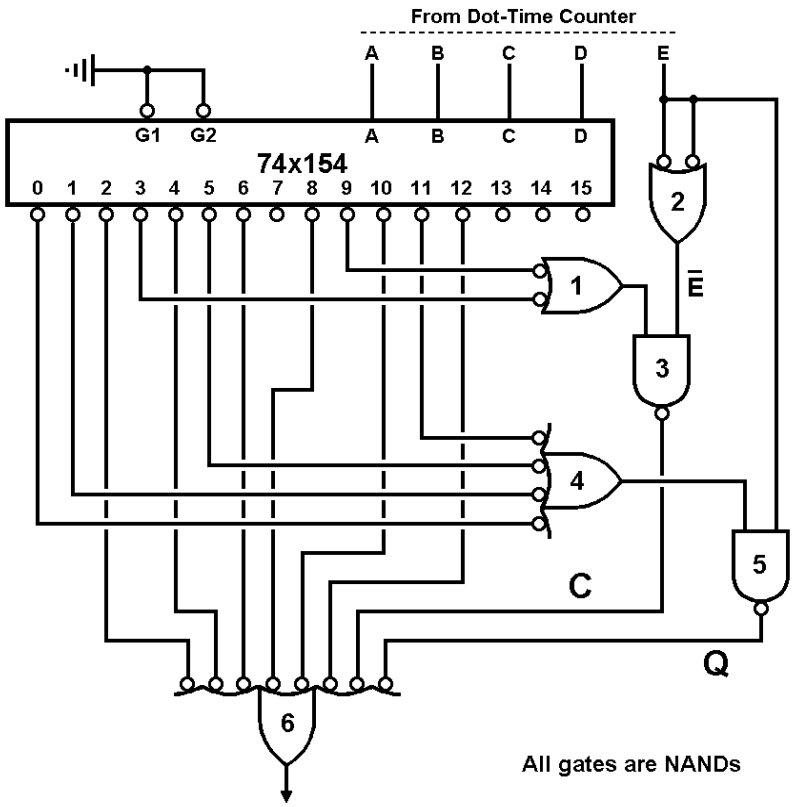


Figure 24-13 A Decoder/Demultiplexer simplifies gating..

Multiplexers Can Provide A Different Way To Assemble Pieces

A ***multiplexer*** can be thought of as an electronic version of a rotary switch, the rotor of the switch governed by the state of the controlling or switch bits and the stator connections being the signal input connections. Had the old 74x150 1-of-16 ***Mux*** been available, the simplest circuit for a two-letter character generator would have been possible.[20] Instead, the simplest circuit of Figure 14 uses two 74x251 1-of-8 multiplexer devices. The 74x151 is the same device but without the 3-state outputs. Both devices have the single output available active-high and active-low

[20] The author's tutorial of October 1979 in *Ham Radio Magazine* used a 74150. Since then that device has been made of unobtainium. Part of the cause might be that a 24-pin DIP is required for the needed 22 pins of the package. DIPs larger than 16 pins seem to be frowned upon by circuit board layout designers.

simultaneously. A 3-state output is convenient here in that ***no*** ORing is required of either output.

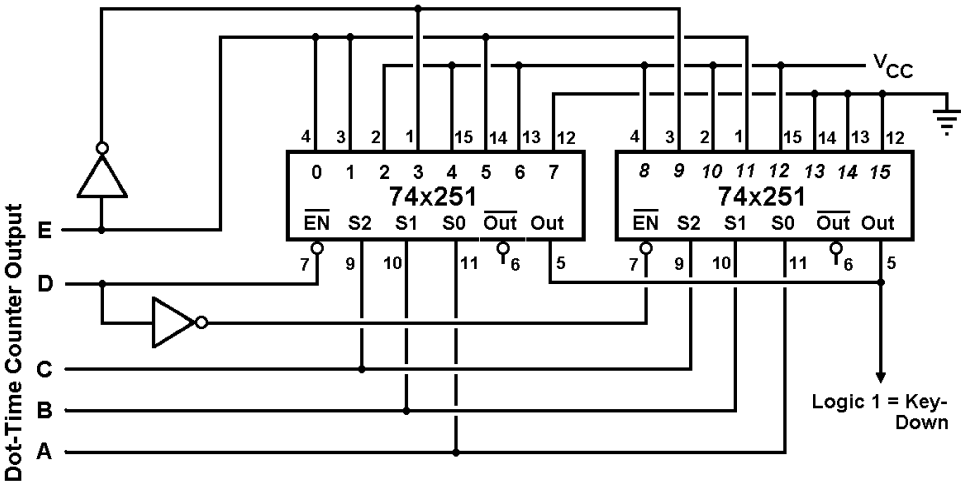


Figure 24-14 - Simplest character generator using Multiplex switches.

Counter output Bits A, B, and C provide the direct addresses for the two muxes. Bit D is the *enable* for each one. When Bit D is low (1st 8 count states), the left-hand mux is active. On the next 8 count states Bit D is high, the inverter makes that a low state to enable the right-hand Mux. The outputs of each 74x251 are in the *high-impedance* state when their *EN*-not input is logic 1. In a high-impedance condition the output pin has, effectively, opened and thus has no affect on the rest of the circuit.

The right-hand Mux has its inputs denoted by italic numerals within the box symbol. This is to fit the 16-bit time pattern of each character.[21] Inputs 7, 13, 14, and 15 connected to ground will be at logic 0 for any 16-bit time period. Inputs 2, 4, 6, 8, 10, and 12 correspond to the dot-times common to both letter characters and will be logic 1 at all times connected to the supply voltage. Inputs 0, 1, 5, and 11 will be logic 1 only when counter Bit E is high (letter Q). Inputs 3 and 9 are high only when counter Bit E is low (letter C).

For an output Key-Down at logic 0, connect the output to both 74x251's pin 6. Figure 14 requires only 2 ½ packages, perhaps only 2 if there are spare inverters in an adjacent circuit. Again, no inverters are needed for Bits A, B, and C out of the dot-time counter stages.

Multiplex devices are gate arrays themselves, consisting of ANDs of the signal input pin and the decoded address supplied by the *Sn* pins. Think of them in the same way as selector switches of signals, only in the much-faster electronic sense.

Summing Up the Versions of the Character Generator

All such beginning-to-be-complex arrays should be examined in some detail on what will fit the function to be performed. *Truth tables* of the various state combinations involved can be examined and used without formal methods of logic state determination. Several different IC devices have been used in the examples and all are useful in various time-dependent gating requirements. What hasn't been touched on in detail is something that will affect high-rate operation, ***propagation delay time*** of devices and arrays of devices**.**

---

[21] That is not standard practice but rather for explanation on this version's example. The DIP pin numbers are correct and the same for both 74x251s.

Time and Time Again

All digital devices have propagation delay. If you are designing for high-rate signal applications, engrave that permanently on your synapses. In high-rate conditions, too much delay in a device or propagation delay build-up can cause a failure of function. In the two-letter character generator just shown in various examples, the Morse code rate is relatively slow, even at 20 word-per-minute rates; several microseconds delay could exist and not be noticed . At 40 MHz clock rates of a programmable divider that is part of a phase-locked loop, the shortest possible propagation delay time is paramount. The period of a 40 MHz signal is only 25 nanoseconds.

The full datasheet on a digital device always contains values of propagation delay or various forms of t_P, usually with the transition being from *low-to-high* (t_{PLH}) *state* or *high-to-low* (t_{PHL}) *state.* Those propagation delay specifications may seem confusing but they require study and perusal to apply to any circuit operating at high rates. They are very important.

In general, the available digital devices are described in their part numbers, in order from fastest to slowest, **F, AC, S, LS, HC,** and the original, no-letter.[22] At one time, ***ECL*** (Emitter-Coupled Logic) was the fastest but that genre of TTL has gone obsolete.

While propagation delay is desired to be as short as possible, that delay exists is ***necessary for some circuits***, particularly bistable multivibrators constructed from gates. The next chapter covers that in more detail but the following solves a real problem with the simplest of *flip-flops*, the *RS* or Reset-Set *FF*.[23]

Switch Contact Bounce and What To Do About It?

All electromagnetic contacts exhibit some kind of ***contact bounce***.[24] That is a literal bouncing of the contacts as they close or open, varying from a few to up to 50 milliseconds in duration. This is particularly noticeable in *dry circuits* (low currents involved of a few milliamperes or less). The result to a circuit connected to a mechanical or electromagnetic switch contact is a brief period of many on-off pulses. Generally, closing a contact results in more contact bounce than opening the contact.

The RS Flip-Flop Applied as a Switch *Debouncer*

---

[22] Left out are the old H (for high-speed) and L (for low-power), both of which have become obsolete. The ACT and HCT variations of CMOS-based devices need to be checked out independently of AC and HC depending on the manufacturer. A-suffix and B-suffix *CD* number CMOS devices have delays dependent on their wide range of supply voltage, generally the higher, the faster. There are some non-family-identification, manufacturer-specific part numbers which require their datasheets to find the time-delay truths.

[23] Purists, especially some university instructors, may be horrified by using colloquial terms such as *flip-flop* or its abbreviation *FF*. On the other hand, those of us who are in the workshops, in the labs, and in the field use some familiar names regularly. Electrons, fields, and waves don't care about *proper* pronunciation and nomenclature of electronics hardware that allows us to work with them.

[24] The sole exception is the mercury-wetted contact type, available in reed capsules of *reed relays.* Mercury is injected into the glass capsules prior to final sealing and that coats the reed switch contacts.

Two NAND gates cross-connected as in Figure 24-15 make up the basic Reset-Set Flip-Flop. The figure shows the inputs connected to an SPDT switch with pull-up resistors. Assume the switch is in the rest position when its arm is down. Input B will be at logic 0 through the switch.

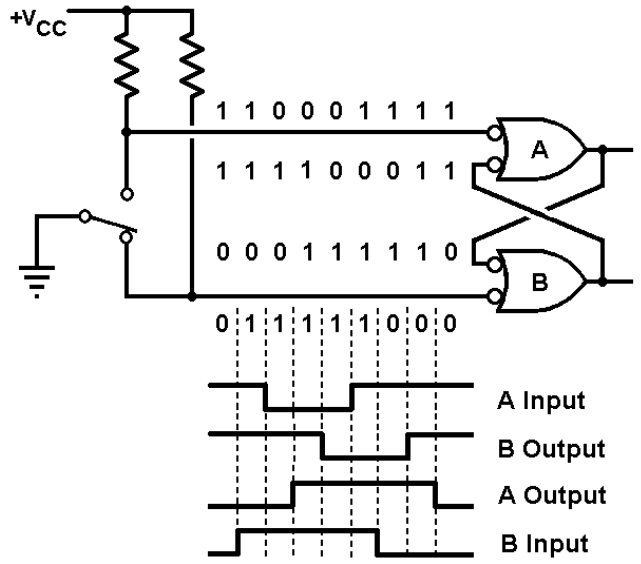


Figure 24-15 RS Flip-Flop used as switch contact debouncer. Time scales and delays are exaggerated to explain operation.

Remember that, with NAND gates, ***any input at logic 0 will cause the output to go high.***[25] Since B input is low, B output will be high. Since NAND gate A's input is high via the pull-up resistor, output A will be low (the inverted AND condition built into the gate). Both inputs to gate B will be low and B's output will be high. This is a stable state.

When the switch is thrown upward there will be a small amount of time when the rotor is not touching either stator contact. The pull-up resistors will maintain a high input to both gates. Gate B still has a low input from gate A so B's output remains high. Gate A gets both inputs high so its output remains low. Still a stable condition. When the switch first makes contact in the up position, the input of gate A is forced low. Output of gate A goes high. Both inputs of gate B are now high, it ANDs with output B going low. Since B's output is also to one input of gate A, gate A maintains its high output. That is the other stable state, justifying its classification as a ***bistable*** *multivibrator* circuit.

Switch bounce occurs for a few milliseconds after the first contact closure. Digital gates change state in a few nanoSeconds. The first contact closure to ground (logic 0) to gate A input forces its output high, gate B's output ANDs and goes low, a total of two propagation delay times. Gate A's output will not change if the switch input bounces high again after those two delay times because the input from gate B remains low.[26] When the switch is thrown back down, the action reverses to reach the resting stable state. Again, if there is some switch bounce after two delay times, it won't matter since gate A's output will be low to one input of gate B, insuring that gate B's output remains high. Using one of the RS Flip-Flop's outputs to control a circuit insures that the control line has no extraneous pulses on it after switch throws in either direction.

---

[25] That is the so-called *NAND Rule*. That rule seems to have fallen out of favor, perhaps because of NAND gates being so widely used. Any AND gate output will remain low until *all* inputs are high. Since a NAND gate is an AND with inverted output, any input at low will, naturally make the output high.

[26] The average of two propagation delays of the old obsolete 74L devices is about 66 nanoseconds, still far faster than the mechanical bouncing of switch contacts occurring in milliseconds. CMOS and 74LS gates bring the double delay time down to about 18 nanoseconds.

Equivalent of an Interlocked Mechanical Push-Button Switch Using RS FFs

The basic Reset-Set bistable does not have to be made only from 2-input NANDs. There's no limit on inputs to the Reset input provided those *extra* inputs are kept high using pull-up resistors. That expanding ability allows equivalents to the mechanically-interlocked push-button switch assemblies often seen on older electronic equipment.[27]

Figure 24-16 shows a 3-station assembly of push-buttons where pushing one button will reset all others. The *reset* inputs are on the upper gate of each RS flip-flop gate pair. The *set* inputs are on the bottom of each pair. As one button drives the desired set side input low it also drives the reset side input low of the other two flip-flops. Numbered outputs assume that the activated button makes that output a logic 1.

A 3-station assembly requires three 3-input and three 2-input NAND gates or two packages. With 4-input NANDs on the reset side there would be four stations. Expansion could go to eight stations using 8-input NANDs on the reset side. A practical limit is 8 stations although there can be more stations by ORing groups of the reset lines from switches.[28]

A problem is power-on initialization of one and only one station locking in. There's no telling which one of the RS flip-flops will be set or reset on initial power-up. That can be solved by using an R-C-R combination in place of one pull-up resistor. The capacitance should be large enough to hold that one button line low until the rest of the switch lines' pull-up resistors reach supply voltage.

An LED and its current-limiting resistor can be connected to the reset-side output to indicate which station is active. A 5 to 10 mA LED current is sufficient for most such indications, good with most gates.

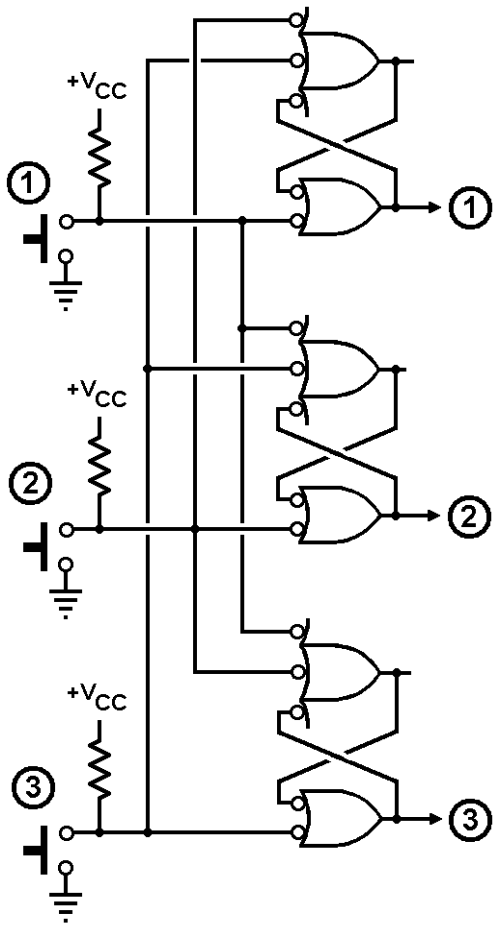


Figure 24-16 Three NAND gate pairs arranged as a three-push-button interlocked switch.

---

[27] Once upon a long-ago time one could order such mechanically-interlocked switch assemblies as *catalog items*. Those switches, along with rotary switch assemblies and parts, are now special-order from a few makers, too expensive for most hobbyists.

[28] Caution: That gets complicated quickly.

Chapter 25

Flip-Flops, Counters, Dividers, Registers

The functioning of the two basic bistable multivibrators, JK and D, made solely from NAND gates are explained via text and timing waveforms showing gate propagation delays. The various digital devices available on the market are described with typical applications as counters, dividers, and shift register generators.

General

The ***RS*** or Reset-Set flip-flop was introduced at the end of the previous chapter. This chapter concerns itself with the ***JK*** and ***D*** types of flip-flops which have become the standard building block in counters, dividers, and shift registers. Early in electronic computing there was a virtual alphabet soup of flip-flop types and names; those fell into disuse or obsolescence when flip-flops were built into integrated circuits. Both the JK and D flip-flop have a ***clock*** input to initiate changes of state. That clock input is usually the reference point for specifications on IC device propagation delay Some counter/divider ICs have ***reset*** control pins to reset all stages to logic 0. Some counters and dividers have ***preset data*** inputs which allow setting each internal flip-flop state to those data inputs under control of an ***enable*** or ***load*** control pin. ***Shift registers*** are strings of flip-flops which can literally shift the whole state combinations within right or left one bit for each clock input. Shift registers, sometimes referred to as *RS* ICs come in forms of serial-in serial-out, parallel-in serial-out, serial-in parallel-out, and parallel-in parallel-out. The latter form is generally called a ***latch***, a small temporary memory storage device.

JK Edge-Triggered Flip-Flop

JK flip-flops have three inputs: J, K, and clock. If both J and K pins are low, the flip-flop output does not change. If J and K pins are different states, flip-flop output assumes the state of the J input on a clock edge. If J and K are both high, the flip-flop ***toggles*** on a clock edge; i.e., it changes to the opposite of the state held just before the clock edge. The clock edge can be either positive-going (low to high level edge) or negative-going (high to low level edge) depending on the type of JK device. Several JK flip-flops can be used together as counters/dividers or as shift registers by the way the J and K pins are interconnected with other flip-flops' outputs or added gates.

A Look Inside a Typical JK Flip-Flop

Figure 25-1 is one equivalent-function NAND gate arrangement for a JK flip-flop having a

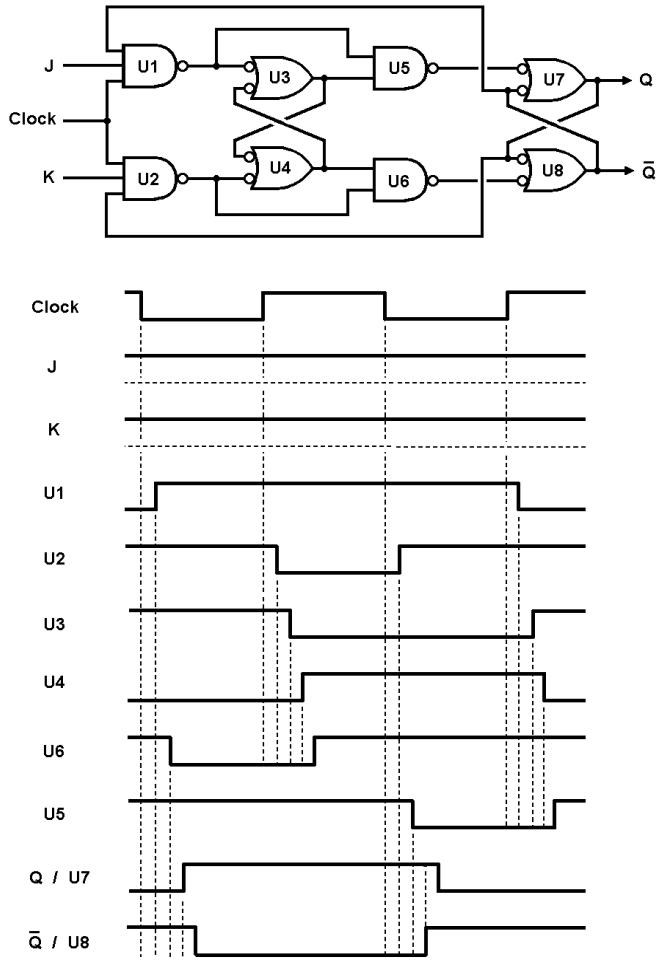


Figure 25-1 A JK flip-flop made entirely from NAND gates with waveforms of each gate's output. Propagation delays are exaggerated.

negative-going Clock edge input.[1] The waveforms are shown with exaggerated delays and time scale in the *toggle* mode; both J and K inputs are hardwired high and the outputs are at half the frequency of the Clock input. The arrangement is sometimes called the *master-slave* for the sequence of RS flip-flops (U3-U4 and U7-U8) although it's a moot point as to which is the *master* and which is the *slave* with the state feedback shown (U7 to U2 and U8 to U1).

In this toggle mode the negative clock edge starts a state-change sequence of Clock➛U1➛U6➛U7(Q)➛U8(Q-not) or Clock➛U2➛U5➛U8(Q-not)➛U7(Q) depending on the previous clock-edge Q state. Gate propagation delays are exaggerated in time scale to both illustrate the sequence of gate state changes and that there will be a finite time delay between the clock edge and an output state change.

If the J and K inputs were both tied low, gate outputs of U1 and U2 would be forced high and could not AND with either the Clock input or the state feedback from U7, U8. That is from the *NAND rule* of any low input will force the gate output high. That fulfills the type JK functional specification on equal input states of J and K.

If the J and K input states are unequal, Q output will follow the state of the J input. That can be worked out with pencil and paper following the input-output states of each NAND gate. Such an arrangement can be used for shift registers and in counters using state feedback for *skipping* certain states.[2]

Inside an Equivalent D Flip-Flop

[1] Figure 25-1 and following equivalent NAND arrangements were in similar form in the author's article in *Ham Radio* magazine, April 1979, entitled *Flip-Flop Internal Structure.*

[2] More on dividers having divisions not a power-of-2 number are covered later in this chapter.

Figure 25-2 has a D flip-flop equivalent made entirely of NAND gates. The waveforms are depicted in its original purpose as a latch or shift-register stage. A difference from that of the JK in Figure 25-1 is that the clock *positive-going* edge initiates the gate state changes. An equivalent circuit for a negative-going clock edge could also be done; showing different clock edges illustrates that flip-flops, counter/dividers, and shift register ICs come with either clock polarity and that designers must be aware of that when choosing which device.

A D flip-flop will have Q output taking the state of the D input on the clock edge. The D (or *Data*) input can change asynchronously (relative to Clock edge time) but must be stable a minimum period called *setup time* before the clocking edge. The same applies to JK flip-flops in regards to their J and K inputs. Setup time is device-specific and stated on flip-flop specification sheets.

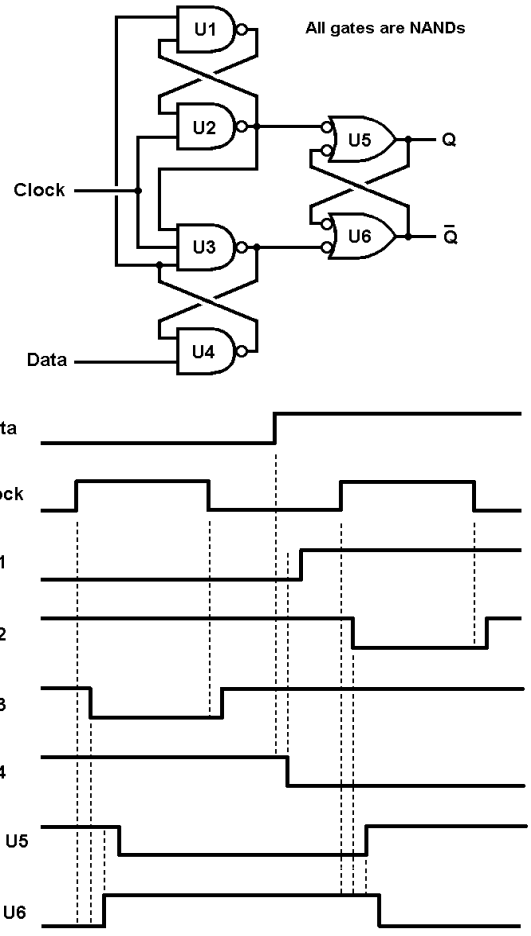


Figure 25-2 A D flip-flop made entirely from NAND gates. Delays exaggerated.

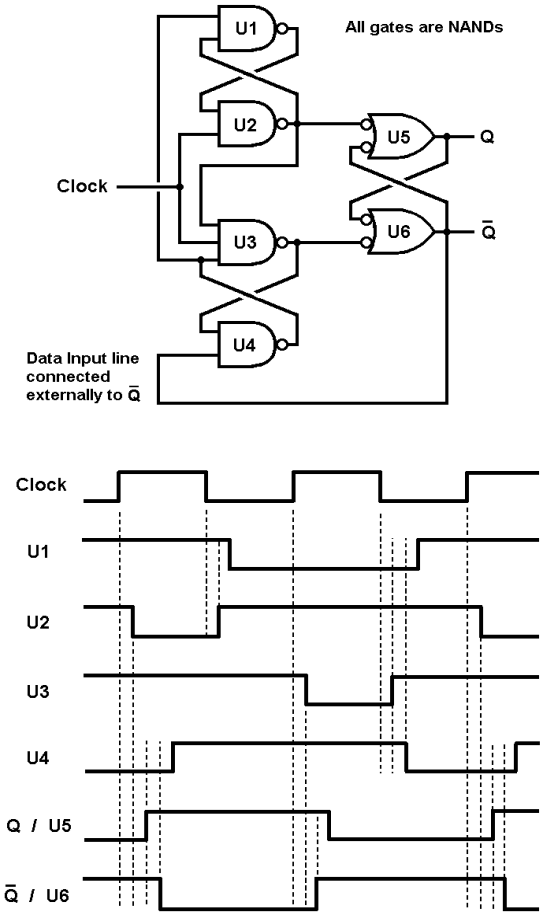


Figure 25-3 A D flip-flop equivalent connected as a *toggle* or divide-by-2 and its operating waveforms.

In Figure 25-2 the rest state at left of the waveforms has Q output high and D input low. The state change sequences are Clock➛U3➛U6(Q-not)➛U5(Q) and Clock➛U2➛U5(Q)➛U6(Q-not) depending on the Q state prior to clock edge. Once Q output has assumed the D input state, only U2 or U3 will change output state on a clock; the U5-U6 RS output flip-flop will not change state since the

cross-connection and U2-U3 output state combinations won't alter their output.

A Toggling D Flip-Flop

Figure 25-3 has the same circuit as Figure 25-2 but the D input is connected directly to Q-not output. This is what would be done on a circuit board in using one flip-flop from a 74x74 dual D flip-flop as a divide-by-2 stage.

Note that the clock positive-edge to Q or Q-not delay is still only 2 or 3 gate delays, same as for Figure 25-2. However, U4 output won't change state until 4 gate delays have elapsed. U4 ANDs U3 and U6 (Q-not) outputs to determine the output state of U3. That may or may not be the reason for the setup time specifications given to this type of flip-flop.

Which One To Use?

That depends on the designer. Both D and JK are useful as toggle or data-shifting flip-flops and there is only a slight advantage in certain dividers in favor of the JK. General use and a slight speed margin are in favor of the D flip-flop.

COUNTERS AND DIVIDERS

Counters and dividers are basically the same thing, a chain of flip-flops. The names given to them refer to the *application* rather than their function. A *counter* application is where *events per unit time* are accumulated, as in a frequency or time period counter test instrument.[3] A *divider* is a *frequency* divider whose output frequency is an integer division of the input frequency.[6]

Binary Dividers

The simplest form of divider is a sequence of flip-flops set to *toggle* on the preceding stage's state. The frequency at the end of the chain is the input divided by an integer power-of-2 number; i.e., 4, 8, 16, 32, 64, etc. Each toggle flip-flop will introduce some propagation delay. Note that the last stage output is three propagation delays behind the input signal. That delay build-up is approximately equal to the number of stages in the chain. That can be much reduced by the ***anticipatory carry*** gating technique shown following. *Anticipatory carry* uses AND gates to set up

---

[3] About 1960 Berkeley Instruments produced one of the first time-frequency counter instruments and trade-named it an *EPUT Meter*, the EPUT an acronym for Events Per Unit Time. A catchy name at the time, it didn't last. *Period* and *Frequency counters* became the widely-accepted names for such instruments later.

6. *Fractional-N* division is sometimes used in frequency synthesizing subsystems but that is really an integer divider set to select one of two very close value integers such as 5 or 6 under command of a controlling input. This chapter will concentrate solely on integer division in dividers.

a next-stage clock input based on all the preceding stages' combined logic 1 states just prior to the carry-out. The maximum delay time of any stage's output is just one gate delay relative to input.

A *carry* or *carry-out* in counters or dividers is the output edge which will toggle the next stage. In the binary dividers of Figures 4 and 5 the carry always occurs after all the preceding stages are high; that is the *anticipation* part. Note that in Figure 25-5 the outputs of U1, U2, and U3 are very nearly time-coincident, each one only one gate delay from the input; each flip-flop will then toggle at the same time. Counter/divider ICs of this type are called ***synchronous***; those without such internal gating are, naturally, ***asynchronous.***

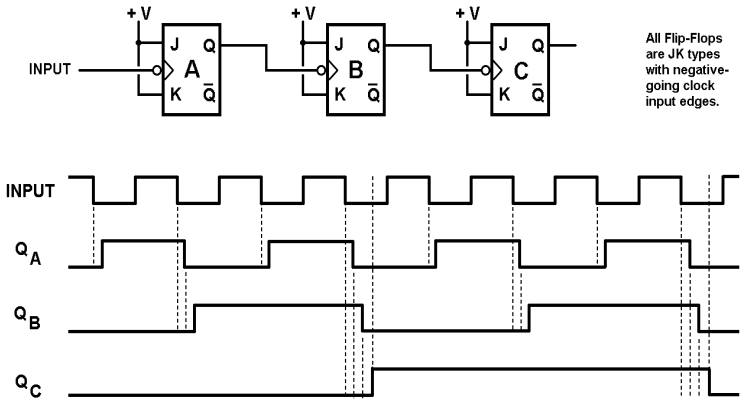


Figure 25-4 Three Flip-Flop chain for an output frequency division by eight. Negative clock input JK flip-flops are connected to toggle in sequence.

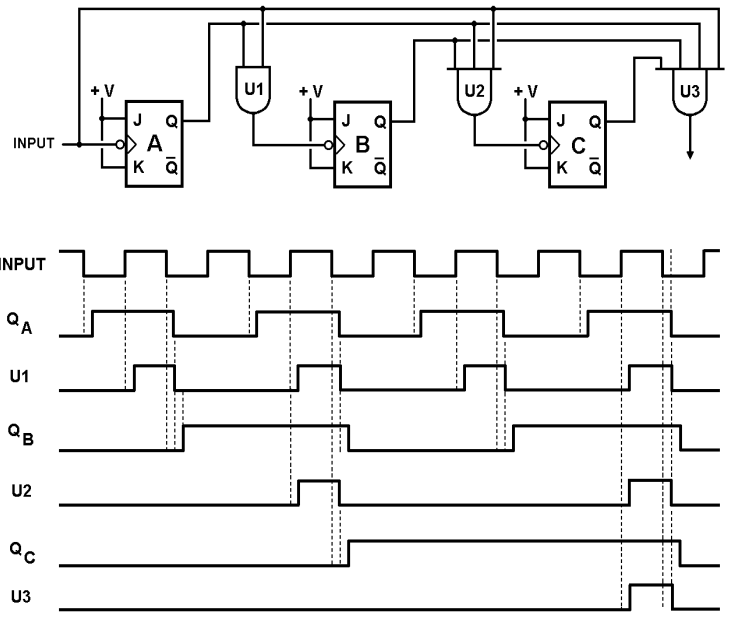


Figure 25-5 *Carry anticipation* AND gates are added. U3's output lags the input by only one gate delay.

Synchronous dividers/counters have the least propagation delay from input to output and are well-suited for high rate input (up to low VHF) multi-bit dividers such as in PLL systems.

Division by Integer Numbers Other than Powers-of-Two

This can be done with JK or D FFs although it takes slightly fewer parts with JK types. The number of flip-flops has to be as many as required for the *next-higher* power of two but is always more than the next-lower power of two. The divided output frequency will be asymmetrical.[4]

Figure 25-6 shows a divide-by-3 that does not require any added gates. The J of FF A is hardwired to logic 1. At start (state 0), both A and B Q outputs are low. Since Q_B is hardwired to

[4] Odd divisions ***can*** be done with symmetric outputs but the input frequency ***must be symmetrical.*** Those will generally take one more flip-flop than the minimum.

both FF's K input, those will also be low.

At the first clock edge, Q_A will go high. FF B cannot change (even though it too is wired to the input clock) because, at the time of the negative clock edge, Q_A is low and both J and K of FF B at a low state will inhibit any toggling of B.

At the second clock edge, Q_B will go high from toggling on the input clock; both J and K are high at that point. FF A will not toggle because J is high and K is low so Q_A retains the state of its J input. At the end of the 3rd state (state 2 on diagram[5]), both FFs J and K inputs are high so both of them toggle.

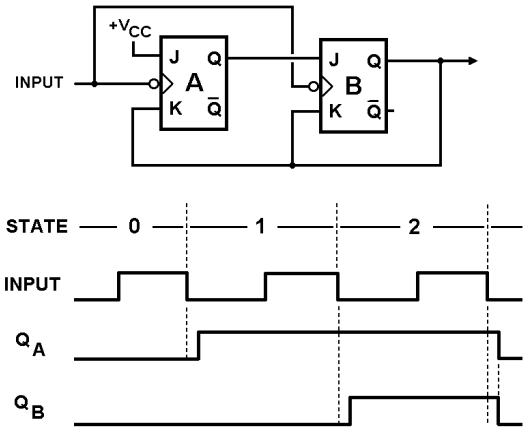


Figure 25-6 A divide-by-3 circuit using JK FFs with negative-going clock edge triggers.

It's possible that power-on conditions could cause the FFs to come up with Q_A low and Q_B high, a state combination not in the division cycle. If so, both J and K of FF A would be high and A would toggle on the first input to reach state 2. Such ***initial conditions*** must always be checked on dividers and counters which skips state combinations and not allowed to cause the divider to lock up. A divider or counter that brings itself out of initial conditions into one of the correct sequence state combinations is called ***self correcting.***

A Divide-by-5 Circuit

Figure 25-7 shows how three negative-going clock input JKs can divide by five. The state sequences are *not* a linear binary one, missing state combinations 011, 101, and 110. That is a consequence of not using any gates to set-up state changes, relying only on the J and K inputs to control the state sequencing.

At state 0 (all Qs low), FF A is ready to toggle after the input line goes high; both J (from FF C's Q-not) and K (connected with the input clock) are both high. FF B will not toggle because its J (from FF A Q output) is low just before the negative going edge of the input and K is low (from FF C's Q). FF C also has both J and K low at that time.

At the end of state 1 FF A toggles again due to both J and K inputs high just before that edge. FF B is ready to flip its Q to high to reflect the J input state (K is low from Q_C). FF C is inhibited by both its J and K inputs at low state.

At end of state 2 FF A is set to toggle again but FF B will be inhibited by both J and K being low. FF C is set up to toggle because its J is high from Q_B and K high from Q_A-not. FF C will not toggle on the negative edge at end of state 3 because its J and K states are dissimilar and FF B Q output will control the output then. FFs A and B will toggle at that state 3 negative edge since both their J and K inputs are high.

At the end of state 4 FF A will not toggle on its negative edge because is J, K inputs are both

[5] It is convention in binary states to begin with zero, not one. The last state in a cyclic pattern will always be one less than the number of counts.

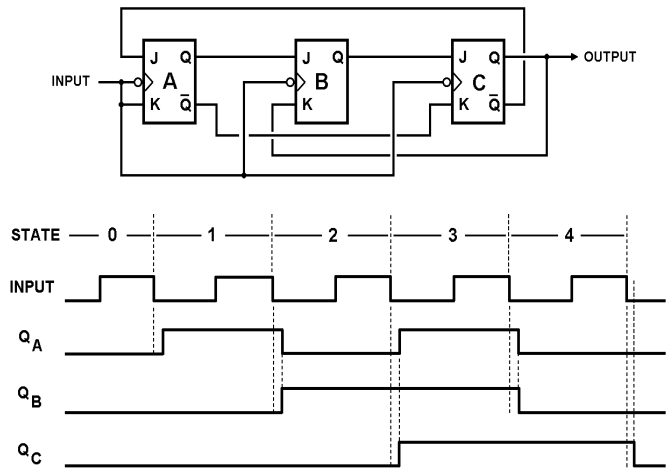


Figure 25-7 Divide-by-5 by three JKs, no gates.

low. FF B is set up for dissimilar J and K states and its J input is low, therefore no state flipping. FF C is set up for toggling via the common clock line and both its J, K inputs high.

Note the repetitious *set up* statements. Any change in output state of a flip-flop is governed by its controlling inputs and those set up states determine what it will do on a clock edge. In most subassemblies of flip-flops as well as dual FF ICs there is a specific ***set up time*** where the specified inputs are stable ***before*** a clocking edge occurs. That is one of the causes for an upper bound on input frequency on all of them.

Is the circuit of Figure 25-7 *self correcting*? Yes it is, one input clock edge from coming up on the illegal states of 011, 101, and 110 will cause the three flip-flops to flip/flop to one of the legal state combinations.

A Divide-by-5 in Linear Binary State Sequence Configuration

Adding one 2-input AND gate to the three FFs of Figure 25-7 and rearranging the J and K connections will permit a linear sequence of state combinations in a binary progression. Figure 25-8 shows the circuit and its waveforms. Preceded by a single flip-flop, the four could be used as a BCD counter having a linear-progression binary state combination sequence of 0000 through 1001. Note that the sequence for Figure 25-8 is 000 through 100.

In Figure 25-8's circuit, flip-flops A and B are wired as if a chain divider. FF C will not toggle since its J and K inputs will always be of dissimilar states from AND gate U1 and the K state feedback from Q_C. At the end of state 3, but before the negative-going clock edge, both Q_A and Q_B will be high so the output of AND gate U1 will be high. FF C's K input will be low.

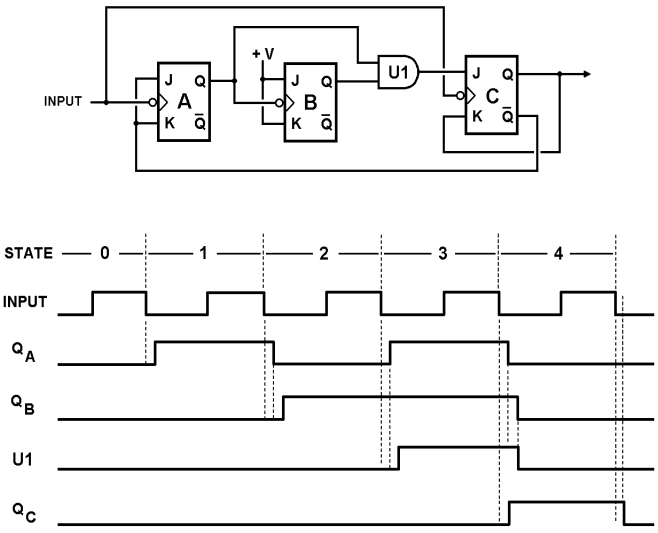


Figure 25-8 Adding one 2-input AND gate makes the 3-FF circuit into a linear binary state sequence progression for following a single flip-flop as a BCD counter.

At the end of state 4 the feedback from Q_C-not has gone low and will inhibit FF A from toggling (both J, K low stops all flipping).

The J input to FF C has gone low from no ANDing of the Q outputs from FFs A and B while the K input has gone high (from state feedback). With dissimilar states of J and K, J will control the Q output of C and the 5th clock edge will cause all Qs to be low. Illegal state combinations of 101, 110, 111 will all flip into legal states with the first input clock.

There are still a few decade counter ICs in both CMOS and TTL, all in one package. It wastes circuit board real estate to make them out of 2 ½ DIPs as in Figure 25-8; an off-the-shelf decade counter in one DIP would be the thing to use. A divide-by-7 circuit is possible using three flip-flops and one 2-input gate but there's a way to use one DIP of a few types to get any division from 3 to 16 using ***presettable*** divider ICs.

Presettable Dividers, Division as Desired

A 74x161 or 74x163 four flip-flop divider is a synchronous up-counter with ***preset*** inputs plus a ***terminal-count*** or carry-out gate. With the addition of a single inverter it can be connected to divide by any integer number from 3 to 16 whose ***complement***[6] binary state combination is at the preset input pins. A diagram and pinout is in Figure 25-9 for a fixed divide-by-7. These also contain extra gating to chain any number of 4-bit dividers for total division values from 3 to 256 (two ICs), to 4096 (three ICs), to 65,536 (four ICs), and so forth. The division number value can be odd or even, prime or factorable.

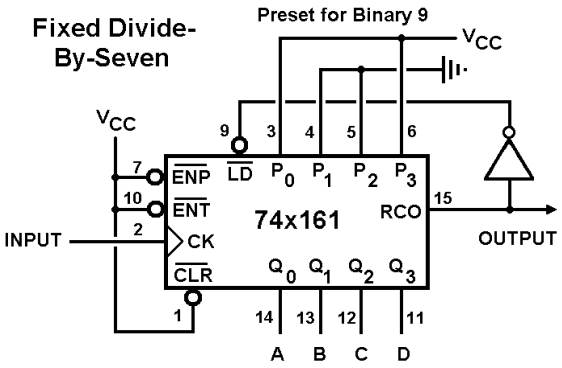


Figure 25-9 Fixed divide-by-7 using 74x161 and one inverter. Preset input equals 16 - Division value.

Counting is done on the positive-going edge of the Clock pulse. The ***ENP-not*** and ***ENT-not*** are active-low count-enable control inputs. For a single IC fixed divider both should be high. The ***CLR-not*** is an active-low reset input; for a fixed divider this also should be tied high. A 74x161 and 74x163 are virtually identical in operation, differing only in clear/reset and

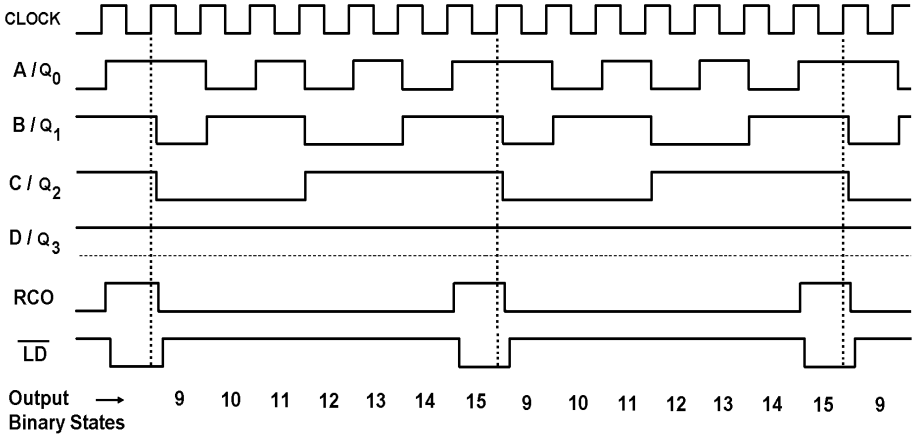


Figure 25-10 Waveforms of the fixed divide-by-7 in Figure 25-9.

[6] Complement of 16 in this case. The Preset binary input must be equal in value to (16-N) where N is the division desired.

enables being synchronous with clock (‘163) or asynchronous (‘161). ***RCO*** or Ripple Carry Output will be high only on the maximum state combination of all-ones. Counting is synchronous with the clock positive edge.

Both the ‘161 and ‘163 are *up-counters*. Counting always increases from the rest state of all-zero (reset) or the preset input state which can be anything from all-zero to binary 14. Actual division of the input frequency when the preset is loaded is equal to the maximum (16 for a single IC) minus the desired division number. In this case the preset is 16 - 7 = 9. Thinking in binary, the preset state can be the *two’s complement* of the division; Division number in binary bits is 0111, one’s complement would be 1000 and two’s complement is 0001 more or 1001.[7]

In the 74x161 and 74x163, the ***LD-not*** (Preset Load) will load all stages with the preset input state combination on the clock positive edge regardless of output state prior to the clock and, in effect, disables the clock from counting during that active-low load pulse. The following tabulation has the preset input state combinations for the division values:

Division	Preset Decimal	Preset State D C B A	Division	Preset Decimal	Preset State D C B A
5	11	1 0 1 1	11	5	0 1 0 1
6	10	1 0 1 0	12	4	0 1 0 0
7	9	1 0 0 1	13	3	0 0 1 1
8	8	1 0 0 0	14	2	0 0 1 0
9	7	0 1 1 1	15	1	0 0 0 1
10	6	0 1 1 0	16	- don’t bother -	

Except for the case of divide-by-7, there’s not much point in using a 4-stage counter for division less than 8. Similarly, with the *LD-not* tied high, either device will free-run in dividing by 16.

Are these ICs self-correcting? In a sense, yes. If they power-on in a state lower than the preset value, they simply count upward until maximum state is reached and *RCO* goes high to

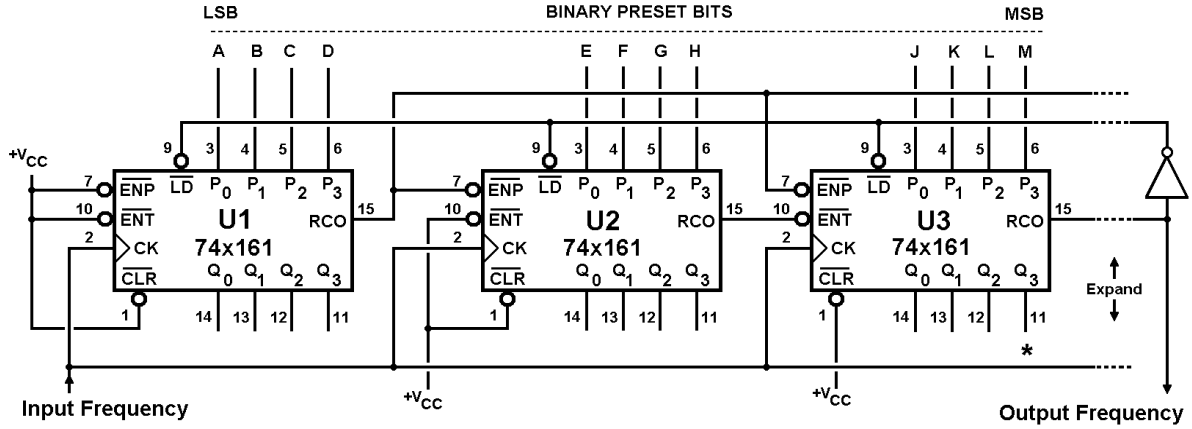


Figure 25-11 12-bit Divider suitable for any division up to 4095. Expansion as needed to the right, copying U3 connections at dotted lines for more 74x161s.

[7] See previous chapter for *two’s complement* in binary bits.

initiate a preset load. The usual frequency divider applications do not require resetting via the ***CLR-not***.

Ganging Up for Higher Divisions

The circuit in Figure 25-11 may be set for any division up to 4095. It takes advantage of the ***ENT-not*** control input to prevent any toggling of U2 until U1's ***RCO*** output went high. The same for U2 and U3. The output frequency signal is a positive pulse whose width is equal to the period of the input frequency. If that is too narrow, take the output frequency from Q_3 of U3 (pin 11, marked by an asterisk).[8]

Since this chain is 12 binary stages, the binary preset bits have to be set for a value that is 4096 minus the division number desired. If the desired division was 2345, the preset would be 4096 - 2345 = 1751. Note that the LSB is to the left, the input flip-flop of U1.

For expansion to higher divisors, copy the U3 connections for more 74x161s to the right at the dotted circuit lines. For four devices (16 bits), the maximum division is 65,536 and the preset bit settings would be (in binary) 65,536 - N where N is the desired division. For five devices and 20 bits the maximum division is 1,048,576. Expansion beyond 4 devices becomes unwieldy for several reasons: The loading of the common clock line (input frequency); loading of the RCO output from U1, mostly from added circuit trace capacitance; the difficulty in determining the correct preset bit states for a given large divisor.[9] The last can be relieved by using a ***down-counter***; more on those in a few pages.

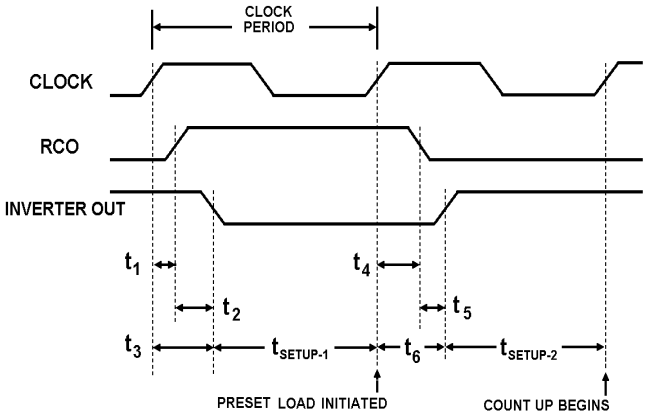


Figure 25-12 Key time intervals in determining maximum operating frequency.

Checking Maximum Division Frequency

Designers have to study datasheets with an eye to the task of determining which statements are constructed to entice sales of products versus what they can do when connected in recommended circuits. Lets use **74LS161** datasheet specifications as an example.[10]

Texas Instruments rates the minimum maximum- operating-frequency as 25 MHz, 32 MHz

---

[8] That width is dependent on the desired division and the time that U3-Q3 stays high.

[9] The 74x160 and 74x162 were designed as BCD divide-by-10-maximum counterparts to the 74x161 and 74x163. Due to lack of sales, most semiconductor logic manufacturers have ceased or are dropping the BCD versions. Any '160 or '162 devices on the market for hobbyist purchase are probably old stock. It is generally easier to organize and troubleshoot both fixed and variable-preset divider chains using BCD counter devices.

[10] Referenced datasheet in footnote 10. The **LS** branch has been popular with hobbyists.

average.[11] That is for the a 74LS161 ***not connected*** for divisions less 16. In using the preset load feature the times in Figure 25-13 have to be pulled out of datasheet tables.

To explain Figure 25-12 further, the left half of the waveforms are governed by the delays associated with RCO and the added inverter delay. A common **74LS04** hex inverter is also used in this example. The inverter forms the active-low preset load control signal. When ***LD-not*** is low it inhibits toggling from the internal clock and overrides internal flip-flop action by the loading of the preset bit states. In order to resume up-counting, there must be a minimum time delay from the end of the ***LD-not*** low state to the ***next*** positive clock edge. That small interval is called ***setup time.*** Think of *setup time* as all internal gates are settling down from their previous state changes, a getting-ready for the next external state-change signal input.

In Figure 25-12, $\mathbf{t_1}$ is the positive-going delay (*PLH)* of Clock to RCO, stated as 35 nSec maximum, and $\mathbf{t_2}$ is the negative-going delay (*PHL*) of the inverter, stated as 15 nSec maximum. The sum of those two, $\mathbf{t_3}$, is the delay from Clock to LD-not or 50 nSec. The interval that LD-not remains low prior to the next positive-going clock edge is $\mathbf{t_{SETUP\text{-}1}}$ and stated by TI as 20 nSec *minimum.* Motorola data for the 74LS161 states that as 25 nSec minimum.[12] This yields a total time of 70 or 75 nSec and the 1/period frequency (for the left half of waveforms) would be 14.3 MHz or 13.3 MHz depending on whose specifications you use. If in doubt, use the lowest frequency..

For the right-half waveforms, $\mathbf{t_4}$ negative-going (*PHL)* delay is also 35 nSec and $\mathbf{t_5}$ positive-going (*PLH*) delay is also 15 nSec. The second setup time (defined by TI on their datasheet as *inactive state*) is also given as 20 nSec minimum. That results in the same maximum clock input frequency minimum as 13.3 MHz. Can such a circuit handle a higher frequency? Most probably, yes, but the upper bound can't be guaranteed. To use a higher input frequency, go to a faster family branch such as a **74AC161.** In the TI datasheet, the configuration from which Figure 25-12 was taken has a single statement that maximum input frequency is $1 / ((\text{CLK to RCO } t_{PLH})+(\text{ENP } t_{SU}))$ = 1 / 55nSec = 18.2 MHz. That is without any inverter for using the RCO as a preset load..

For the 74AC161, using Figure 25-12 with a 74AC04 hex inverter, the left-half waveform clock period would be 27.5 nSec total and the right-half waveform would be shorter at 20.5 nSec. The lowest frequency would be 36.4 MHz.[13] A 74AC161 is guaranteed to 100 MHz (on the datasheet) but that is individual devices without any interconnections.

The synchronous counting built-in to all '161s allows expansion as desired for the same maximum input rate. The most difficult task is burrowing through all the datasheet tables to find the specification values. Remember that those values are based on specific loading capacitance stated. Increasing the circuit capacitance can increase some of the delay times, lowering the

---

[11] ***Average*** or ***typical*** values are apparently based on yields of thousands of devices; the percentage is seldom stated.. The only guarantee is that a device will meet minimums. The same is true of specified delays where the only guarantee is that the delays will not be longer than maximums stated.

[12] This is not unusual. The datasheet from Motorola (now ON Semiconductor) is from their 1992 *FACT and LS TTL Data* book, pp 5-160 to 5-165. As several manufacturers do, pin function names may be changed such as **PE-not** (Parallel Enable) instead of Texas Instruments' **LD-not** or **TC** (Terminal Count) in place of TI's **RCO**. If in doubt on pin names check the number on the device pinout.

[13] ON Semiconductor datasheet, publication MC74AC161/D of August, 2002 - Revision 6. Values used are for 5.0 VDC supply and ambient temperature of 25 °C.

maximum input frequency.[14]

Down-Counters Can Eliminate Preset State Complementing

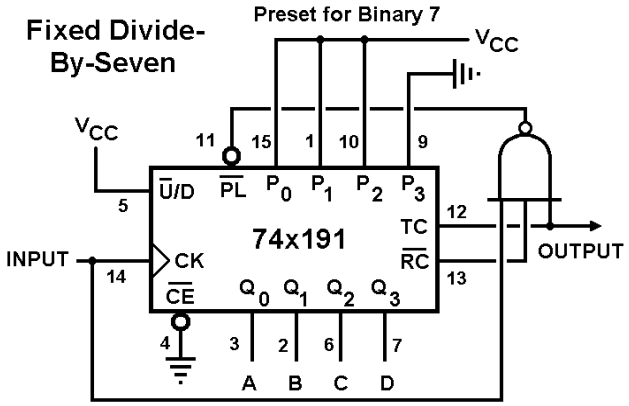


Figure 25-13 Divide-by-7 example using 74x191 and a NAND to load the preset states.

The **74x191** is very similar to the 74x161 in that it has a positive-going clock edge and synchronous clocking of all internal flip-flops, an active-high *total count* output, and an active-low preset load input. The major difference is that the '191 can be commanded to count either up or down. In counting up the total count output is during the maximum binary state combination of all-ones, as in the '161. In counting down, that total count output occurs at the minimum binary state combination of all-zeroes. Connected as a single device or a chain of several in down-count mode as a frequency divider, the states of the preset data input correspond exactly to the desired divisor value.

The circuit in Figure 25-13 and its waveforms in Figure 25-14 can be compared to the previous. The ***TC*** output in down-count mode is active-high when all stages have counted down to all-zeroes. ***RC-not*** is active-low coincident with ***TC*** high and the Clock low. A NAND gate is used to get a ***PL-not*** started as soon as possible after the positive-going edge of the clock input. While the ***RC-not*** could have done that without a gate or inverter, ***RC-not*** doesn't go low until the latter half of the clock cycle. Using the gate allows a slightly higher maximum input frequency even though it has its own propagation delay. An alternative would be an inverter

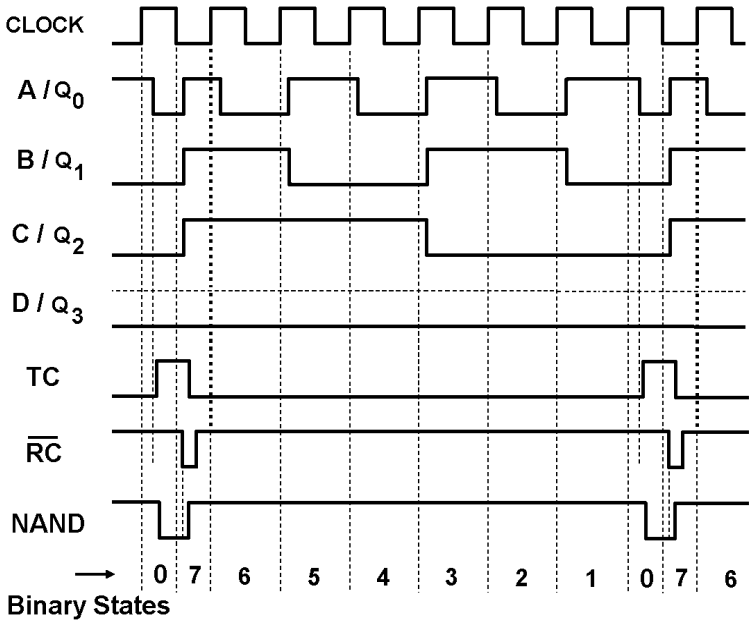


Figure 25-14 Waveforms of Figure 25-13 circuit.

[14] One of the main reasons that engineering design departments budget time for building up prototypes prior to the *production prototype.* Hopefully the circuit board layout is with a minimum of interconnection length and therefore minimal stray capacitance to ground. Sometimes it isn't and that needs examination to see if change is needed or not. The hobbyist seldom has the time for that so it is best to plan for the worst-case conditions.

from ***TC.***

When all flip-flops reach all-zeroes, ***TC*** goes high. The gate ANDs ***TC*** with the high state of the input clock and the still-high ***RC-not***. Preset loading on the 74x191 is asynchronous so the binary 7 would be set into all flip-flops with only the internal load-to-Q delay. ***TC*** would then drop low (the all-zeroes state is gone) and ***RC-not*** also go high (if it ever went low).

One clock cycle during the ***TC*** high interval will be both 0 and then 7, the preset value. The actual width of the ***LD-not*** out of the gate is less than about 50 nSec with 10 MHz into a chain of four 74x191s dividing down to 250 Hz. That makes it difficult to see with an analog oscilloscope at slow sweep rates.

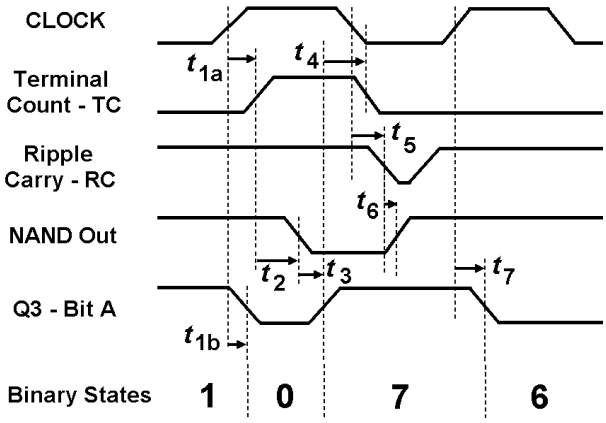


Figure 25-15 Time delays during 0/7 transition state in circuit of Figure 25-13.

Maximum operating input frequency can only be approximated based on accumulation of time delays. Figure 25-15 is an expansion of the critical time period during achievement of count 0 and the preset loading of 7 for the Figure 25-14 circuit. Time intervals are derived from device datasheets and examples of a 74LS191-74LS10 and a 74AC191-74AC10[15] combination. The time intervals in nSec are tabulated below using *typical*[16] and maximum values to the positive-going slope edges (*PLH)* and negative-going slope edges (*PHL*).

Time *t*	LS *typical*	LS Maximum	AC *typical*	AC Maximum	
1a	28	42	7.5	12	
1b	24	36	6	10.5	
2	10	15	4	6	
3	22	33	5.5	9.5	
4		-?-		-?-	Not specified
5	16	24	5	8.5	
6	9	15	4.5	7	
7	24	36	6	10.5	
2 to 6 (min. LD-not width)		35		1	Pulse width minimum
Low state of Clock input		25		3	minimums

For t_3 that depend on the preset for fixed dividers going from all-zeroes to a 1 state; if bit is preset to 0 there is no delay. Datasheets don't seem to specify the delay from a non-zero state combination

[15] LS data from Motorola 1989 *FAST and LS TTL Data* book; AC data from Fairchild datasheet DS009940 (November 1988, Revised November 1999) for '191, DS009915 (November 1988, Revised January 2000) obtained via www.fairchildsemi.com. AC family values for 5 VDC supply and 25 °C ambient temperature.

[16] Take *typical* values with scepticism. Only the maximum time delays are guaranteed.

of Qs to the high-to-low state of a Terminal Count for t_4; there obviously is some but the designer has to guess at it or breadboard the circuit to find their own values. The latter is probably easier than trying to estimate internal device delays from the datasheet device logic schematics (not recommended by manufacturers since the schematic is intended to show logical functioning). For t_{1A} and t_{1B} take the longer of the two as a safe value.

To estimate maximum divider input frequency, add times 1, 2, and 3. Assuming a symmetric Clock input, add times 5 and 6 to that. For a 74LS family using maximum delay values, that is (42 + 15 + 33) + (24 + 15) = (90) + (39) = 129 nSec. Estimating that the time delay to change a Q output combination out of the all-zeroes Terminal Carry (t_4) is about the same as for t_{1B} then 36 nSec is added to that for 165 nSec total. The 1/f result is 6.1 MHz as a maximum clock input frequency, estimated.[17]

The specified minimum low-state width of the Clock input is given as 25 nSec for the 74LS combination. Part of that would be included in times 5 and 6, but not guarantee on the remainder until the next positive-going slope of the Clock. To be extremely conservative, adding 165 and 25 would get 190 nSec as a clock period or 5.3 MHz clock frequency as a maximum input.

Going through the same procedure for a 74AC family combination results in (12 + 6 + 9.5) + (8.5 + 7) = 43 nSec. Estimating the TC ending delay at 10.5 nSec brings that to 53.5 nSec. Adding only 3 nSec (minimum clock low-state width) makes that 56.5 nSec or 17.7 MHz maximum clock input frequency. If the *typical* values for the AC family were taken, the total time would be 35.5 nSec or 28.2 MHz clock frequency. The 74AC is obviously a faster family but the stand-alone, no-preset-loading toggle rate of a 74AC191 is100 MHz.

If the faster family is used, note the datasheet specifications for capacitive loading of outputs, 50 pFd. The 74LS family capacitive loading is only 15 pFd. Having less circuit capacitance will speed up operation some, but not entirely.[18]

Chaining the 74x191 For Larger Divisions

A recommended interconnection for larger divisions is shown in Figure 25-16.[19] This uses the Count Enable control input (***CE-not***) to inhibit any action in more significant devices until a Terminal Carry has been output from a less significant device. The time delay buildup from any ***TC*** to the Parallel Load (***PL-not***) is no greater than in the circuit of Figure 25-14. The input frequency signal is the common Clock line to all devices in the chain.

---

[17] The author has used 10 MHz clock input successfully, but that is ***not*** any sort of guarantee that everyone can do that with 74LS devices at any time, using any purchased devices. The parts you buy to use may have been on the far side of an average production yield curve, meeting maximums but not meeting the *typical* (or average) values given on specification sheets.

[18] Delays *inside* a device exist but device users don't get that information. Device manufacturers have to use some kind of standardized loading with logic families' outputs and we don't know what real, guaranteed effect there is on reducing propagation delay values with less loading capacitance. It's always a good idea to minimize circuit capacitance and 50 pFd would be many inches of wire or PCB trace length. Keep lead lengths short when working with times of a few nanoseconds; frequency components of such signals *begin* in the UHF region.

[19] From Motorola 1992 *FAST and LS TTL Data* book, page 5-211. The same circuit arrangement is also found in datasheets from Fairchild and National Semiconductor.

Division is set by the Preset Bits with the binary state combinations the same as the binary value of those 16 bits. Bit A is the least significant bit and U1 will be operating through the entire division process; its Count Enable (***CE-not***) pin is hardwired low. Down-mode counting is also hardwired to the supply voltage; it would be grounded for up-counting. Inverter U5 and 2-input NAND U6 can be sections of other gate devices, 74AC04 and 74AC00 respectively. A dual 4-input gate (74AC20) could be used for U7A (two inputs tied together) and U7B or part of a 3-input NAND (74AC10) used for U7A and half of a 74AC20 used for U8A.

This variable divider is in use by the author as part of a PLL to control the Local Oscillator of a receiver. The input frequency is a symmetric 13.60 to 18.65 MHz output of a single toggling flip-flop. Division ranges from 27,200 to 37,300 so the step increments are 500 Hz apart (when PLL is locked). The Preset Load negative pulse from the 4-input gate is estimated at about 20 nSec width, difficult to see when triggering an oscilloscope synchronized to the 500 Hz PLL reference frequency. The one fault of a down-counter with asynchronous Preset Load is that poor visibility, both for debugging the circuit board and for any troubleshooting later. However, the advantage is that the division number is precisely related to the binary state combination numeric value, easing any adjustment of tuning versus frequency control.

In Figure 25-16 the Output is shown connected to the Q output of the last stage of U4. This is partially dependent on the divisor's minimum and maximum numbers and the resulting most-significant-bit time-symmetry during a complete count cycle.

A Slightly Different Up-Down Counter

The **74x192** and **74x193** appeared in the original TTL form together with the 74x190 and 74x191. The '192 is a decade counter while the '193 is hexadecimal with a count of 16. The big difference is that each have two inputs and two outputs, an input-output for up-count and an input-output for down-count. These are entirely separate as to direction. The share only the four flip-flops which are brought out as **Q** states.

Data preset inputs are provided along with an asynchronous preset load, active-low. A separate control input is an asynchronous active-high Clear resetting all four flip-flops to all zeroes. The Up-count and Down-count inputs must be kept high during no counting. Either input signal will toggle the

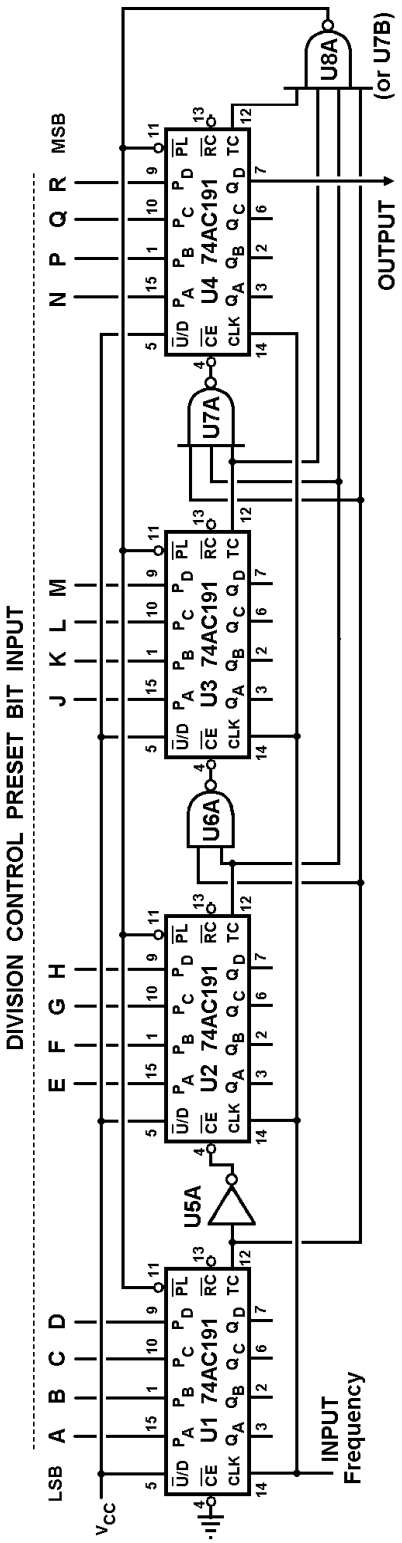


Figure 25-16 16-bit Down Count using 74AC191s.

count on the positive-going slope of the input after it has first gone low. There is no clock input signal.

Each output, *Carry* for the Up-count and *Borrow* for the Down-Count, goes low during a terminal-count state, synchronous with the low part of each input. The terminal-count state for Up-count is all-ones; for Down-count all zeroes. The positive-going slope of each output is synchronous with the clock and will toggle the next-most-significant device's inputs as appropriate for the counting direction. No extra gating is needed to chain several in series.[20] The propagation delays of 74LS193's Clock to positive-going output is 24 nSec maximum (16 nSec *typical*) but that is additive for each device in the chain.

The 74x193 makes a good Up/Down **totalizer** but not good as a divider continuously dividing according to the data preset inputs. The data preset inputs are good for *setting* particular state combinations under command of the active-low Load input.

A Stand-Alone Down Counter

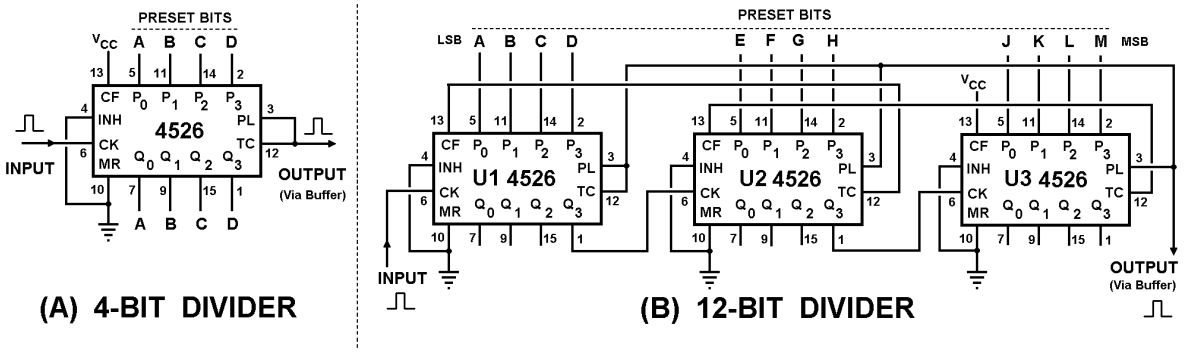


Figure 25-17 Recommended circuits using the *4526* device as a frequency divider.

An **MC14526B** (ON Semiconductor) or **HEF4526** (Philips) can do division without added gates, but at a price in lesser input frequency. Designed with P-channel and N-channel MOS transistor junctions, that series is part of the ***A*** and ***B*** series originally brought out by RCA.[21]

Figure 25-17 (A) has a single device dividing up to 16. The input is a positive pulse of at least 250 nSec width. ***TC*** or terminal count is a positive pulse in synchronism (but delayed) with the input and occurring when all ***Q*** outputs are zero. For single device dividers the ***TC*** is connected to

---

[20] The complete datasheets showing the timing relationships and (idealized) waveforms from Texas Instruments have been available from the 1970s on up to mid-2006. Unfortunately, most manufacturers, including second-source licensees, either have discontinued or are about to discontinue the decade versions of all the *160* and *190* models, namely 74x160, 74x162, 74x190, and 74x192. Only the 16-bit versions remain as of 2006 although there is *old stock* available at some retailers.

[21] RCA Solid State Division (Somerville, NJ) brought out their ***COS/MOS*** line of ICs in the first half of the 1970s, all featuring then unheard-of low, low quiescent current demand logic devices. Not to be outdone, Motorola Semiconductor (part of which became ON Semiconductor later) debuted their own line of the A and B devices, usually with an **MC1** in place of the **CD** prefix of the *4000* series number nomenclature. That whole group featuring operation from 5 to 15 VDC supply, moderate operating speed (about that of the 74L family in TTL). Philips Semiconductor later used a **HEF** prefix for the same numbers.

PL, the Preset Load input and the output may be taken from that; a buffer stage is recommended for any connections to TTL inputs. Preset bit input is the binary states of the division number. Using the same worst-case propagation delay criterion for the 74x160s and 74x190s as dividers, the maximum delay is calculated to be 1320 nSec at 5 VDC supply plus a necessary 125 nSec clock pulse positive width.[22] The input frequency would then be about 690 KHz maximum with guaranteed maximum delay times. Using *typical* value figures (approximately half that of maximum), the maximum input frequency would be 1.3 MHz. That would be roughly inversely proportional to supply voltage to make it about 6 MHz at 15 VDC supply.

In Figure 25-17 (B) three devices are interconnected for a maximum division of 4096 using the ***CF*** or Cascade Feedback input from the next most-significant device's Terminal Count output. To reduce the time delay buildup from ripple carrying to the next higher stage, U1 forms the Preset Load active-high signal with just the Clock➛TC delay, carried over to U2 and U3 PL inputs and the frequency output. The feedback from U2-TC to U1-CF keeps the U1-TC from going high until both U1 and U2 Q outputs are all-zeroes (Terminal Count condition). Similarly, U3-TC to U2-CF keeps U2-TF from going high until U3's Q outputs are also all-zero. Preset Load will only happen when U1, U2, and U3 Q outputs are all at Terminal Count.

The worst-case propagation delay from input frequency positive-edge to U1-TC low-to-high transition is about 900 nSec at 5 VDC supply. The delay from clock to U3-Q outputs is about the same value. Adding 125 nSec for the minimum clock high during would set the shorted clock period at about 1025 nSec at specified maximums or a frequency of 976 KHz. With *typical* values of delays that frequency would double. Division number is equal to the binary state value of the Preset inputs, Bit A being the least significant bit and Bit M the most significant bit.

Philips Semiconductor recommends Preset inputs as always grounded via a 10 KOhm resistor. Switch settings for a logic 1 would put the V_{CC} directly on the Preset pin (and draw 500 μA current when forcing a logic 1).

A Couple of Legacy Bidirectional Counters

The **CD4510** and **CD4516**, decade and 4-bit-binary counters respectively, are capable of both Up and Down counting mode based one the logic level at one control pin. Those two have been on the market since the early 1970s and are second-sourced by Texas Instruments (under the original RCA nomenclature), Philips (HEF4510, HEF4516), Toshiba (using a TC prefix), and Motorola/ON Semiconductor (MC14510, MC14516). Using the circuit of Figure 18, two devices plus a 2-input NOR gate can divide by any

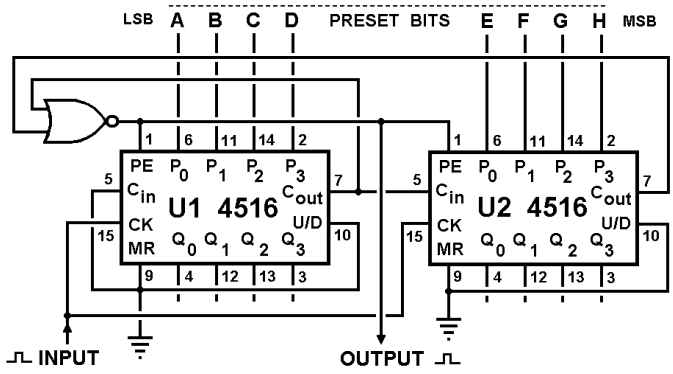


Figure 25-18 Chain of hardwired-down-mode 4516s as an adjustable 8-bit frequency divider.

[22] Based on datasheet information MC14526B/D, April 2006, Rev. 5 from ON Semiconductor and HEF4526B Product Specification File under Integrated Circuits, IC04, January 1995.

number from 2 to 255. Preset bit logic state is the same as the binary of the divisor number. Output pulse is positive-going, of the same width as the input frequency positive-going pulse.

Pin 10 is the count-direction control, low for down-count, high for up-count. Maximum input frequency is about the same as the **4526** described previously.[23] The two devices are part of the **COS/MOS** original MOS family having a wide supply voltage range and are not capable of higher speeds; the 74AC family and some of the 74HC are faster using newer CMOS transistor junctions.[24]

The Carry-out (pin 7) is active-low on the terminal count (15 or all-ones for 4516, 9 or binary 1001 for 4510) with a low-going output width equal to one clock cycle for up-count mode; all-zeroes for both devices in down-count mode.[25]. The 4510 is a decade counter with BCD outputs but was absent from the **ST** (SGS-Thompson) and ON Semiconductor lines in the year 2005.

An 8-Bit Binary Down-Count Divider in a 16-pin DIP

The **40103** is interesting in that it dispenses with the flip-flop stage outputs as pins, having all 8 preset inputs available. One DIP and no extra gates can do division from 2 to 255, easing circuit board real estate requirements. The caution in use is that the actual division is ***N+1*** where *N* is the binary number input to the preset pins. Philips Semiconductor produced this as the **74HC40103.**[26] The Philips datasheet is quite comprehensive though longer than most.

With one 2-input NAND gate and an active-low start pulse, it can provide a high-going time interval as long as ***N*** input clock periods. While not strictly a frequency divider circuit, that can provide a precisely-controlled pulse width generator for other functions.[27]

A *Universal* 4-Bit Presettable Counter

The **CD4029B** has controllable BCD or 4-bit binary count mode, Up- or Down-count mode, and preset ability for all four flip-flops, all in a 16-pin DIP. This device, still on the market as of 2006, might prove to be a building block for hobbyist circuits from its functioning universality. It was designed as part of the original ***COS/MOS*** group and has some limitations on operating speed, about 4 MHz in chained, synchronous totalizing, almost 1 MHz as a programmable divider using a 5 VDC supply.

---

[23] Figure 19 is taken from ON Semiconductor datasheet MC14516B/D dated March, 2006 - Rev. 5, page 10.

[24] Some of the 74HC and 74HCT families are pin- and functional-equivalents of the original *CD4000* line but not all. One has to check the datasheet propagation delay values to make certain of that.

[25] Texas Instruments, Toshiba, Philips and ON Semiconductor provide full count timing charts as part of their datasheets (TI's is apparently a direct copy of the original RCA datasheets). That same cannot be said of the 4526 device datasheet from the same manufacturers.

[26] Datasheet 9397 750 13812, Rev. 03 - 12 November 2004; part of a new look in Philips semiconductor datasheets.

[27] Circuit on the Philips datasheet.

SHIFT REGISTERS AND SHIFT COUNTERS

Basic Shift Registers

A 4-bit right-shift *serial-input, serial-output* shift register using D flip-flops is shown in Figure 20. Assume a start condition of all Q outputs at logic 0 and the serial input at the left has a logic 1. At the first clock pulse the left-hand flip-flop will set its Q output to 1. The right-hand three D flip-flops will still have logic 0; their D inputs were all at 0 (from left-hand adjacent Q outputs) just before that first clock pulse.

If the serial input now goes to logic 0 and a second clock pulse asserts itself, that single logic 1 will appear to *travel* right one stage. Reading from left-to-right the Q state combinations after succeeding clock pulses will be:

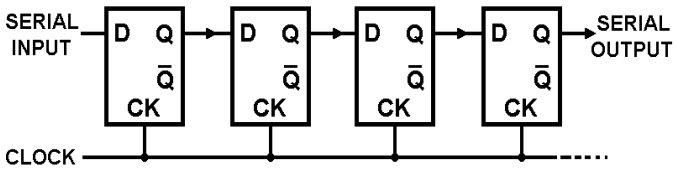


Figure 25-19 4-bit right-shift register.

Time Period	Serial In	Q outputs	
Initial state	**1**	0 0 0 0	
After 1st Clock	0	**1** 0 0 0	
After 2nd Clock	0	0 **1** 0 0	
After 3rd Clock	0	0 0 **1** 0	
After 4th Clock	0	0 0 0 **1**	
After 5th Clock	0	0 0 0 0	➛ (to *bit bucket*)

That initial serial input logic 1 has traveled to the right one stage at a time for every common clock pulse. After the 5th clock it will go into the *bit bucket*[28] and never be seen again.

If there were a way of *setting* states into the four flip-flops, such as the *set direct* input of a 74x74 a dual flip-flop, whatever state combination was set into them prior to the clock pulses, would appear to *travel* to the right. Set the initial state as 1100 then, after each clock pulse the state combinations would change from 1100 to 0110➛0011➛0001➛0000.

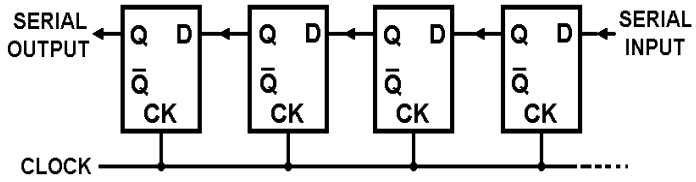


Figure 25-20 Left-shift counterpart.

By reconnecting inputs and outputs in Figure 25-19 left-for-right, Figure 25-20 creates a *left-shift* register. If an initial condition was set to 0011, subsequent clocks would make the state combinations 0110➛1100➛1000➛0000. Most unidirectional shift-registers are right-shift while bidirectional shift-registers can shift either way on command of a control pin.

In binary arithmetic, a left-shift by one bit position is a multiplication by 2 and a right-shift is a division by 2. Both unidirectional and bidirectional shift-register devices are available along

[28] Colloquial term equal to waste basket or dumpster.

with combinations of *parallel-load* (all stages set as in presets of dividers) and parallel-output (all stages brought out). One doesn't normally think of a shift-register as a *counter* but those can be made and have some advantages in certain applications.

Shift Counters

Taking the arrangement of Figure 25-19 and connecting the output stage back to the input, a ***shift*** or ***ring*** counter is created, one such in Figure 25-21. If the Q output pattern of 0110 is first set into the flip-flops, that pattern will continuously circulate and it becomes a divide-by-4 from the input (clocking) frequency. If the initial pattern of 0101 had been set in, it would have become a divide-by-2.

A problem with the arrangement of Figure 21 is that it is ***not self-correcting.*** There are only four valid state combinations of 1100➛0110➛0011➛1001. There are 12 other state combinations which could have been set during power-on and would be invalid for any Q output. A way out of that is to change the state feedback from U4-Q to U4-***Q-not***. This version is shown in Figure 25-22 and is called a ***Johnson counter*** (also *twisted-tail* for the U4-Q-not to U1-D). A thing to remember with ring counters is that the Q will set up to the D input state *on the next clock*. The D input does not do anything by itself except to prepare the flip-flop to switch or not switch states on the next clock edge.

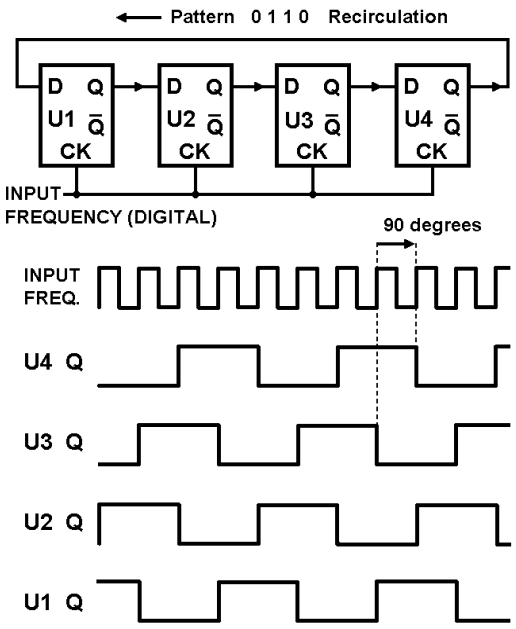


Figure 25-21 A divide-by-4 shift counter

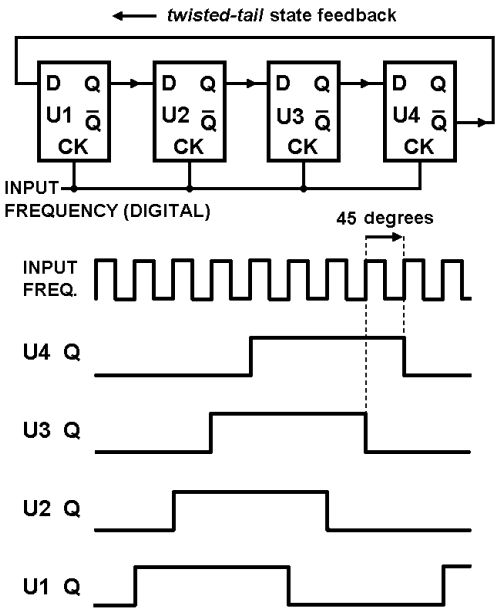


Figure 25-22 A Johnson (or *twisted-tail)* shift counter for better efficiency.

A *Quadrature* Output Generator for SSB Applications

SSB (Single Side Band) phasing modulators and demodulators need an accurate, stable two RF waveforms which are as close to *quadrature* or 90° phase relationship as possible. Figure 22 could do that but Figure 25-23 is simpler. The state combinations are: 00➛10➛11➛01. All four state combinations of two bits are in the sequence. That means the shift counter is ***self-correcting.***

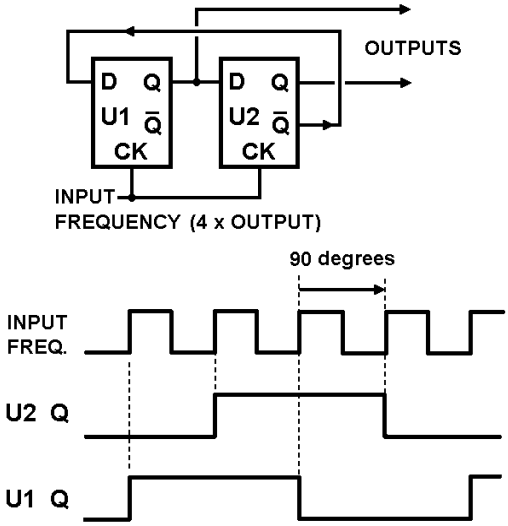


Figure 25-23 Quadrature phase from a 2-bit Johnson counter circuit.

All Johnson counters have a count cycle that is two times the number of flip-flops in terms of clock input. The 2-bit of Figure 24 has a cycle of 4 clocks, the 4-bit of Figure 23 has a cycle of 8 clocks. A decade counter in the Johnson manner requires 5 flip-flops. The number of state combinations which are invalid (for a count cycle) grows exponentially as the number of shift register stages increases beyond two.

As the number of register stages grows, the problem of self-correction becomes worse. Invalid states have to be detected and stage bits set to break out of invalid states during a clock cycle. The CD4017 decade counter (5 stages) and CD4022 octal counter (4 stages, cycle of 8 clocks) have their internal schematics presented in datasheets showing correction via an added AND gate and a NOR gate feeding the D input of the third-from-the-left flip-flop. Beyond five stages the self-correction gating becomes cumbersome and synchronous binary dividers take less hardware.

An advantage of the Johnson counter back in the late 1960s (when the *COS/MOS* family debuted) is the ease of decoding count state combinations. Both the CD4017 and CD4022 use 2-input AND gates for each of the individual count outputs. That is simpler than the gating schematics seen on datasheets for BCD-to-Decimal decoders. Forty years later with both CMOS and Schottky transistor junctions highly improved, plus better semiconductor wafer production technologies have made the Johnson counter arrangement rather old-fashioned except for niche applications such as the mentioned quadrature generator.

Chapter 26

Pseudo-Random Shift Generators

Some specific Pseudo-Random Shift Generators for creating known random noise and for security purposes to encode and decode digital data.

The *Maximal-Length* Pseudo Random Sequence Generator

The ***PRSG*** *(Pseudo Random Sequence Generator)* is a curious offshoot in the shift counter category of ***maximal length*** count cycles. The count cycle is 2^N-1 clock periods long where N is the number of flip-flops in the register. The state feedback is obtained by exclusive-ORing two or four of the shift register Q outputs. Figure 24 is an example of a simple 4-stage PRSG. Its count cycle length is 15 clock periods. The single Exclusive-OR provides a logic 1 to U1-D input only when U3-Q and U4-Q are at different states. Note the power-on conditions to skip the all-zeroes state condition, the only one not allowed in the cycle. The 15 state combinations in sequence after each clock edge:

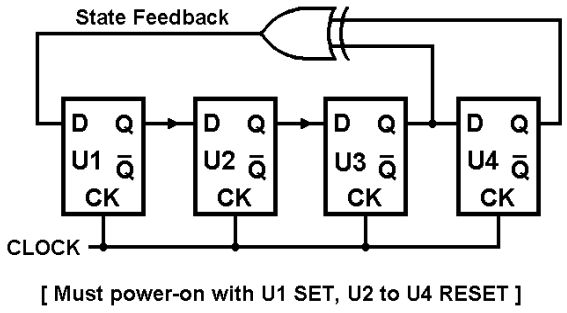


Figure 26-1 A 4-stage maximal length shift counter using Ex-OR state feedback.

1000➔0100➔0010➔1001➔1100➔0110➔1011➔0101➔1010➔1101➔1110➔1111➔0111➔0011➔0001

The pattern of just four stages is hardly *random* since the pattern length is only 15 clock periods. With 8 stages the pattern cycle is 255 clocks and begins to take on a random nature. At 16 stages the cycle is 65,535 clock periods long and has definite appearance of randomness. See following for a very long cycle length PRSG and more details of a relatively simple circuit to generate it.

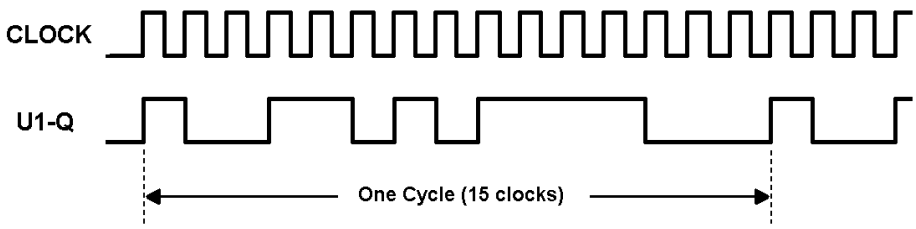


Figure 26-2 Waveform of pattern produced by Figure 26-1.

Of what use is it? Very long patterns taken from one register stage have the property of ***seeming to be random***, hence the name of *pseudo random*. As such they are good for scrambling or unscrambling data streams on-line if the clock is synchronous with the data rate. That could be adapted for cryptographic purposes.[1] If the sequence is paralleled to a digital-to-analog converter, the equivalent of a known-characteristic pseudo random analog signal can be used in development of noise-canceling circuitry; the advantage is that the pseudo-random noise is known and synchronizeable whereas naturally-random noise is not. If the pseudo-random noise can be synchronized with an oscilloscope display, it is easier to determine what is taking place in such a developmental circuit.

Blending Analog and Digital for a Simple Sinusoidal Signal Generation

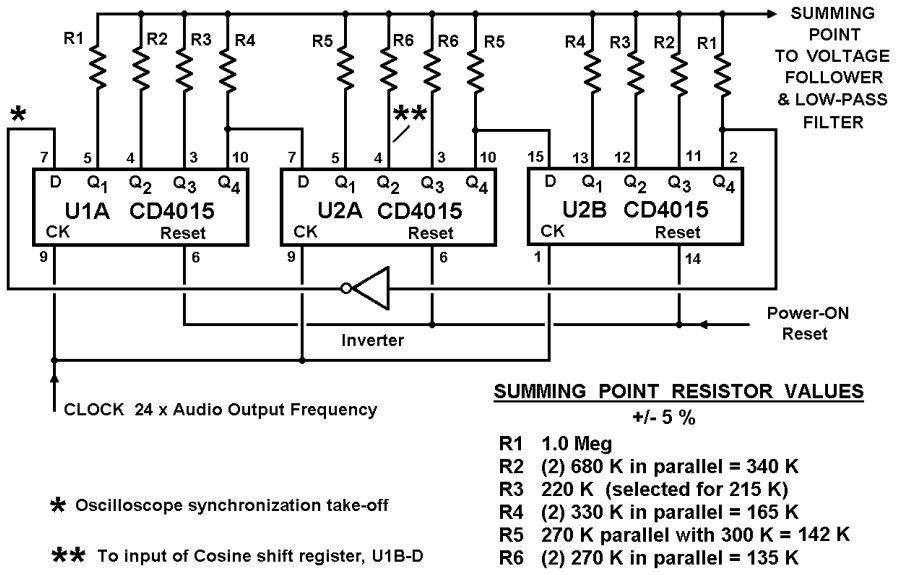


Figure 26-3 Master Johnson counter for Sine-Cosine audio frequency generator; clock source not shown.

The first of the CMOS digital devices had low current drive outputs but the logic state voltages under very low currents are nearly the ground and V_{CC} levels. A CD4015B dual 4-stage shift register (of the original *COS/MOS* family) has a logic 1 output voltage of 4.95 VDC, logic 0 voltage of 0.05 VDC at less than 1 mA load current and 5 VDC V_{CC}. That 50 mV difference from ground and supply rail persists even at the 15 VDC supply voltage ratings.

Figure 26-3 has the master Johnson counter of a synchronized sine and cosine sinusoidal waveform audio signal source.[2] The master shift counter is a 12-stage register with a cycle time of 24 clock periods made from 1 ½ CD4015 dual 4-bit shift register devices.

A sinewave is approximated by resistors R1 through R6 whose summing point is made to

---

[1] Judging from some of the papers seen from the 1950s and 1960s on this subject, the author concludes that some considerable work was done for that very purpose. With the advent of the high-speed microprocessor and GHz-clock-rate computers, longer, more-secure cryptographic keys were developed via mathematics and software rather than through hardware.

[2] Circuit by Gary Steinbaugh, of E A Transform, Loveland, Ohio, appearing in *EDN* magazine *DesignIdeas* section of 13 April, 2006, pages 106 and 108. The original article used ±1% resistors.

an op-amp voltage follower (for minimal loading) and then into a lowpass filter to remove the steps in the summed waveform. Five percent tolerance resistors can be used as shown. The summing point *sees* a voltage divider formed by R1 through R6 duplicated with the most-positive voltage being 4.95 VDC when a Q output is at logic 1, least-positive voltage 0.05 VDC when a Q output is at logic 0; +5 VDC V_{CC} supply assumed.

With all shift register stages at logic 0 there is no virtual voltage divider since all resistors are in parallel to +0.05 Volts. After the first 6 clocks, six of the register outputs are at logic 1 and the other six at logic 0; the virtual voltage divider is then 34.4 KOhms from summing point to +4.95 V and 34.4 KOhms from sum to +0.05 V for an even split at summing point of +2.50 V. After 12 clocks all register stages are logic 1 and all Q outputs put all 12 resistors to +4.95 V. In the next 12 clocks fewer and fewer resistors are to the more-positive Voltage and the summing point potential drops until it reaches +0.05 Volts. Resistor values are graduated for a sinusoidal output. If all resistors were equal, output would be triangular. The virtual voltage divider is tabulated following with *R-High* being the resistance from summing point to +4.95 VDC and *R-Low* being the resistance from summing point to +0.05 VDC, all assuming a +5.00 VDC supply and 50 mV drop across the CMOS source-drain voltage drop of register output circuit.

24-Clock States	R-High	R-Low	Fraction[3]	E-Sum
000000000000111111111111	----	17.2 K	0	0.050
100000000000011111111111	1.00 M	17.5 K	0.017	0.133
110000000000001111111111	248 K	18.5 K	0.069	0.388
111000000000000111111111	117 K	20.2 K	0.147	0.770
111100000000000011111111	68.3 K	23.0 K	0.252	1.285
111110000000000001111111	46.1 K	27.4 K	0.373	1.878
111111000000000000111111	34.4 K	34.4 K	0.500	2.500
111111100000000000011111	27.4 K	46.1 k	0.627	3.122
111111110000000000001111	23.0 K	68.3 K	0.748	3.715
111111111000000000000111	20.2 K	117 K	0.853	4.230
111111111100000000000011	18.5 K	248 K	0.931	4.612
111111111110000000000001	17.5 K	1.00 M	0.983	4.867
111111111111000000000000	17.2 K	----	1.000	4.950
011111111111100000000000	17.5 K	1.00 M	0.983	4.867
001111111111110000000000	18.5 K	248 K	0.931	4.612
000111111111111000000000	20.2 K	117 K	0.853	4.230
000011111111111100000000	23.0 K	68.3 K	0.748	3.715
000001111111111110000000	27.4 K	46.1 K	0.627	3.122
000000111111111111000000	34.4 K	34.4 K	0.500	2.500
000000011111111111100000	46.1 K	27.4 K	0.373	1.878
000000001111111111110000	68.3 K	23.0 K	0.252	1.285
000000000111111111111000	117 K	20.2 K	0.147	0.770
000000000011111111111100	248 K	18.5 K	0.069	0.388
000000000001111111111110	1.00 M	17.5 K	0.017	0.133
000000000000111111111111	----	17.2 K	0	0.050

A ***Cosine*** sinusoidal waveform can be created simultaneously by copying the U1A-U2 circuit, omitting the inverter, calling the second shift register U1B-U3, and taking the U1B D-input

[3] The *fraction* column pertains to the virtual voltage divider. A Sine will give you ± values which must be shifted and normalized (divide Sine+1 by 2) to get comparative values. The pattern tabulation actually starts at 270° which has the least-positive summing point voltage.

from U2A pin 4 (Q3 output) as indicated by the double asterisk on Figure 27. The U1B-U3 register would repeat the same pattern but be lagging by exactly 6 clock periods, 90 degrees with each clock period being 15° long.. With the same resistor values and an identical voltage follower and lowpass filter from its own summing point, that 90° lag would hold regardless of the clock frequency.

A ***Three-Phase*** (0°-120°-240°) generator can be built by copying the other two shift registers and summing resistors, plus a third, identical voltage follower and lowpass filter. The *120°* register could take its D input from U2B's D input and the *240°* register could take its D input from the *120°* register's 8th stage. The exact D input points can be anywhere as long as the phases are in their relative positions at 120° increments.

A power-on Reset (active-high) must be provided to set all register flip-flops to zero; there is no assurance that all will assume the same state at power-on. Once reset, the twisted-tail state feedback of the Johnson counter configuration assures a self-correcting state combination sequence.

The lowpass filters can be of any type desired although the -3 db response point should be at least 5 times the upper audio output frequency and at least 20 db per octave attenuation above cutoff frequency. Their function is to minimize the 24 *steps* present at the summing point. Those steps contain the majority of the higher harmonics. Spectral content without filtering will be as follows:[4]

Fundamental	0 db
2nd harmonic	-22.0 db
3rd harmonic	-32.5 db
4th harmonic	-37.4 db
5th harmonic	-43.5 db
6th harmonic	-51.5 db
7th harmonic	-58.3 db
8th harmonic	-53.4 db
9th harmonic	-66.1 db
10th harmonic	-77.7 db
11th harmonic	-72.0 db
12th harmonic	-60.8 db
24th harmonic	-34.3 db
36th harmonic	-64.8 db
48th harmonic	-71.0 db

Harmonic content calculated by piecewise approximation using 240 pieces per cycle with each step taking 10 pieces, one a slope to the first step level and the other nine at equal amplitude levels for the flat part of the step. hase relationship of harmonics to fundamental is not shown but is an essential value to hold waveform shape. Attenuation of harmonics after the 5th by a lowpass filter will minimize the steps and smooth the overall waveform into a sinusoid shape. Shift of the harmonic's relative phases also introduces a time delay of the smoothed sinusoid from the stepped waveform input.

Since all active or passive lowpass filters have some time delay, use of identical lowpass filters for the two-output quadrature generator or three-output, three-phase generator will insure that the two (or three) waveforms all maintain their relative phase differences even though a lag appears on viewed waveforms using the oscilloscope trigger point shown by the single asterisk on Figure 27-4.

---

[4] From the author's program that sums the Fourier coefficients of individual pieces having rectangular pulses with either flat or sloping top amplitudes. This is a time-to-frequency transform for a repetitive waveform. The sum of all Fourier coefficients represents the piecewise approximation and the reverse (frequency-to-time) transform is checked against the original data for accuracy. The calculation is not absolute since the division of each step into 10 pieces is itself an approximation. However, an op-amp voltage follower's frequency response (unity gain at around 1 MHz for low-cost legacy devices) plus an active filter made of op-amps will justify that approximation for audio frequencies in the 400 to 2000 Hz frequency range.

The other inverters of a hex inverter device can be used to pick off oscilloscope trigger square waves.

This Johnson counter form of audio generator does not have the spectral purity of a Wein Bridge oscillator nor the touted versatility of triangle-square-sine *function generators* using specialty IC devices. It can assure generation of extremely well-aligned 90° or 120° multiple-phase outputs with as tight control over repetition frequency as desired; the 24-times clocking frequency can be generated by a variety of means from a crystal-controlled source to an R-C astable multivibrator.

A Very Long PRSG

Pseudo Random Sequence Generators are shift register counters with Exclusive-OR stage feedback such that their full patterns are 2^N-1 clock periods long, N being the number of stages in the total shift register. With the correct Exclusive-OR input connections their full pattern appears as a *random* sequence of 1s and 0s. An example of two generators with their outputs combined for

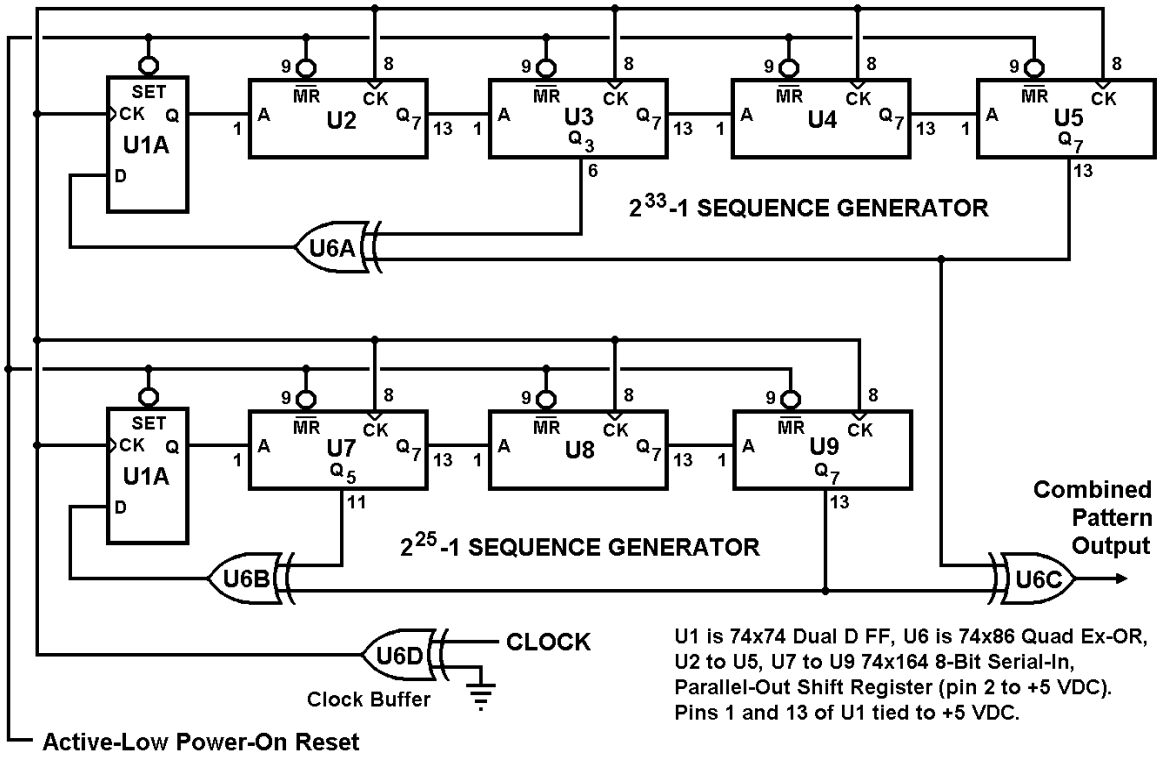


Figure 26- 4 Two long sequence generators combined for an even longer pattern length.

an even longer full pattern is shown in Figure 26-4[5].

The particular circuit was chosen for a customer's request for a maximum length coincident with a circuit board space of only 9 DIP spaces. The *taps* (state feedback inputs to register's first stage) permitted just one Exclusive-OR; with one reserved for combining the two patterns, the remaining Ex-OR of a quad package could be used as a clock buffer. 74LS devices were used. See the second table for useable taps.

A choice of 74LS74 D flip-flop as the first stage permitted the power-on setting of the first stages at logic 1 with the remaining stages (74LS164 8-bit serial-in, parallel-out shift registers) cleared to logic 0 at power-on. The PRSGs will not run if it begins at all-zeroes (the first stage is part of the register chain even if a separate device type).[6] For testing purposes a pull-up resistor of 1 KOhm was provided on the power-on reset line for a manual push-button reset to ground.

The choice of a 2^{25}-1 and 2^{33}-1 revealed they had no common factors in their pattern length. See Table 1 for lengths and length value factors for N from 2 to 36. The 2^{25}-1 generator had factors of 31, 601, and 1801 while the 2^{33}-1 generator had factors of 7, 23, 89, and 599,479. None of the factors were common so the combined pattern would not accidentally repeat. The combined pattern length would be 288,230,367,528,222,721 clock periods long before it repeated itself! If the clock is 10 MHz the combined pattern repetition would not occur for 8,006,399 hours or nearly 333,600 days or 913.347 years.[7] There would be no periodicity apparent in the *short term* (as in an 8-hour working day) and the bit stream from the combined generator's output could be considered as close to random as possible. The customer was satisfied and the author built a duplicate for himself for general workshop service.

Synchronization for Viewing or Communications

For oscilloscope observation purposes or as a DSSS (Direct-Sequence Spread Spectrum) spreading generator, both shorter pattern lengths and a synchronization pick-off in the bit pattern stream is desirable. All maximal-length PRSGs using Exclusive-OR state feedback will have one and only one state condition of all-ones in a complete pattern. This allows a fairly easy gate pick-off using 74x30 (8-input) or 74x133 (13-input) NAND gates from the parallel-output pins of 74x164 shift register devices. That all-ones gate output would be only one clock period long.

---

[5] From the author's submission to *Electronics* magazine *Designer's Notebook* published on 8 November 1978 pp 134-136 and later reprinted in *Designers Casebook 3*, pp 91-93 under the title *Uniting number generators for long bit patterns.*

[6] This is true only if Exclusive-ORs are used for state feedback. If an Exclusive-NOR (74LS286) was used, it would start from an all-zeroes condition (no separate flip-slops needed as first stages) but the all-ones state condition would be invalid and would have to be detected through gating with automatic clearing to zeroes of the register stages. The mixture of a single flip-flop followed by the shift register devices seemed to be the most viable at the time.

[7] Using 3,600 seconds per hour, 24 hours per day, and 365 1/4 days per year; the quarter day being for Leap Years at such a long period.

MAXIMAL LENGTH PSEUDO RANDOM SEQUENCE COUNTER LENGTH AND FACTORS

N	2^N-1	Factors in (2^N-1)
2	3	3
3	7	7
4	15	3 5
5	31	31
6	63	3 7 3
7	127	127
8	255	3 5 17
9	511	7 73
10	1023	3 31 11
11	2047	23 89
12	4095	3 7 5 3 13
13	8191	8191
14	16,383	3 127 43
15	32,767	7 31 151
16	65,535	3 5 17 257
17	131,071	131,071
18	262,143	3 7 3 73 19 3
19	524,287	524,287
20	1,048,575	3 5 31 11 5 41
21	2,097,151	7 127 7 337
22	4,194,303	3 23 89 683
23	8,388,607	47 178,481
24	16,777,215	3 7 5 3 17 13 241
25	33,554,431	31 601 1801
26	67,108,863	3 8191 2731
27	134,217,727	7 73 262,657
28	268,435,455	3 5 127 43 29 113
29	536,870,911	233 1103 2089
30	1,073,741,823	3 7 31 3 11 151 331
31	2,147,483,647	*-prime-*
32	4,294,967,295	3 5 17 257 65,537
33	8,589,934,591	7 23 89 599,479
34	17,179,869,183	3 131,071 43,691
35	34,359,738,367	31 127 7 122,921
36	68,719,478,735	3 7 5 3 73 13 19 3 37 109

Factors have been grouped to show their repetition as N increases; that repetition continues to higher N values.

EXCLUSIVE-OR REGISTER STAGE TAPS FOR STATE FEEDBACK OF *N* REGISTERS[8]

Always Exclusive-OR the *LAST* stage and the stages tabulated below.

N	Stage(s)	N	Stage(s)	N	Stage(s)
4	3	15	14	26	6 ⊕ 2 ⊕ 1
5	3	16	15 ⊕ 13 ⊕ 4	27	5 ⊕ 2 ⊕ 1
6	5	17	14	28	25
7	6	18	11	29	27
8	6 ⊕ 5 ⊕ 4	19	6 ⊕ 2 ⊕ 1	30	6 ⊕ 4 ⊕ 1
9	5	20	17	31	28
10	7	21	19	32	22 ⊕ 2 ⊕ 1
11	9	22	21	33	20
12	6 ⊕ 4 ⊕ 1	23	18	34	27 ⊕ 2 ⊕ 1
13	4 ⊕ 3 ⊕ 1	24	23 ⊕ 22 ⊕ 17	35	33
14	5 ⊕ 3 ⊕ 1	25	22	36	25

Ns of 8, 12, 13, 14, 16,19, 24, 26, 27, 30, 32, and 34 require three Exclusive-OR gates, two with inputs from stages indicated and third one Ex-ORing the output of the two for feedback.

The above tabulation is not the only stage tapping possibility. Several different stage taps are possible above N of 7 and are documented elsewhere. Different stage number taps will yield the same pattern length but the sequence of 1s and 0s will be altered from tap connections shown above.

Predetermining a Pattern Start State Condition

A shift register device with parallel-input as well as parallel-output plus serial shifting operation can use the parallel input to preset a state condition for the start of a PRSG bit sequence. This is very similar to the preset loading of a counter. Useful devices for that are the 74x194 and 74x195 4-bit and 74x299 and 74x323 8-bit shift register types. All but the 74x195 are bidirectional in serial shifting. Parallel inputs as presetting can be used in place of the power-on state condition setting.

One use for that would be in DSSS sequences that must repeat even though their bit patterns seem to be random. Another is cryptographic applications such as enciphering and then deciphering ASCII text bytes; a problem there leads to some way to synchronize the receiving end with enciphered text in order to decipher it into clear text.

Conversion of a Binary Bit Stream to Analog Noise

The genesis of this Appendix began with the author briefly working with an HP-3722A *Gaussian Noise Generator* in 1968. Then very new, this Hewlett-Packard (now Agilent) instrument

[8] From Xilinx application note XAPP 052, July 7, 1996 (Version 1.1), page 5. Lists taps of N from 3 to 168 (2^{168}-1 PRSG has a pattern length of $374 \bullet 10^{48}$ clock periods).

used an adjustable maximal length 2^N-1 shift counter with an innovative *digital waveshape filter* and a following 40 db per decade active lowpass filter to create analog noise of known AC characteristics. The *noise spectra* is constant at any clock frequency (by virtue of the *digital filter*) and the peak and RMS values of pseudo random noise also constant below the analog lowpass cutoff frequency. That pseudo randomness is also repeatable. It's a strange concept to realize that artificial *noise* can be quantified and even calibrated for in an instrument. It is true, explained in detail in the instrument's Operating and Service Manual.[9] While long since obsolete, the principles of operation are still valid and presented following.[10]

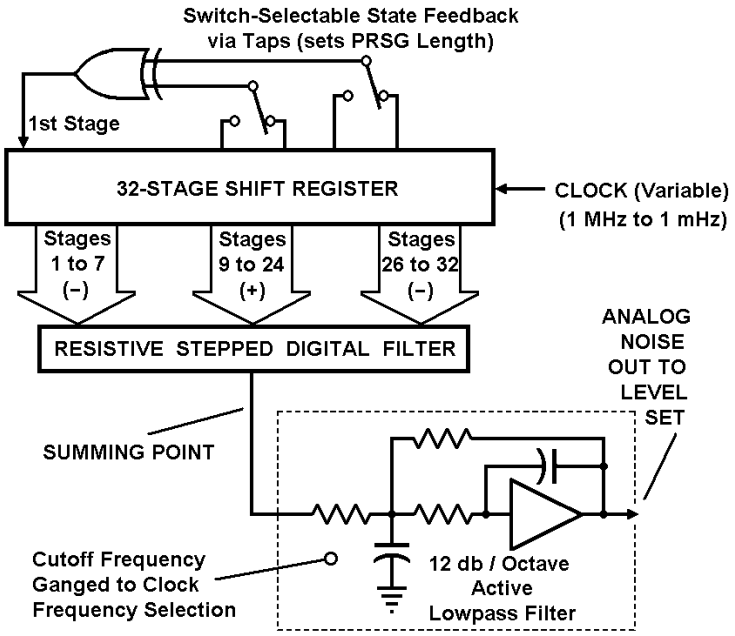


Figure 26-5 Simplified Block diagram of HP-3722A Noise Generator

Figure 26-5 shows the general block diagram of the 3722. The 32-stage shift register performs two functions: The first is as a variable-length LFSR, switch-selected for pattern lengths of 15 clocks (4 stages) to 1.048,575 clocks (20 stages); the second function is to input to a resistive *digital filter* which forms a *Sin(x)/x* step impulse waveshape approximation (very broadband) of a single logic 1 circulating through the shift register.

Of that second function Hewlett-Packard used a complementary pair of PNP-NPN s per shift register stage operating between +9 VDC and ground. The common point (collectors) was made to a fixed ±1% resistor of varying value per stage all connected to a common *summing point* having a low-value resistor to ground. The common collector point of each saturated PNP-NPN switch

---

[9] *Operating and Service Manual, Noise Generator 3722A, Hewlett-Packard LTD, South Queensferry, West Lothian, Scotland,* April 1968, document number 02182-1 03722-010, primarily Section 3 (principles of operation). Design of this instrument probably began in the early 1960s judging from the fact that there isn't one integrated circuit in its 5 1/4-inch high rack mount cabinet-chassis; all flip-flops and gates within were made from discrete parts, two-transistor flip-flops four to a printed circuit plug-in card.

[10] There's some *reverse engineering* involved here, but its one of admiration and respect for the designers of this innovative but now-obsolete instrument. The Test & Measurement Division of Hewlett-Packard (now Agilent) did good. If pseudo-random noise instruments are in use today, they are probably adaptations of programming to the more expensive *arbitrary waveform generators* on the market.

would then switch between about +8.7 VDC and +0.3 VDC depending on the logic state of the pair's input from a single shift register stage. This is very much like the CMOS two-phase sinusoidal waveform generator's method described on page 37-20; at low load currents a CMOS logic device output switches between near-ground and near-supply-voltage potentials.

The digital filter should produce a relative voltage output as in Figure 26-6. The steps correspond to the 1/8th Pi fractional multiples along a *Sin (X) / X* function curve.[11] The fractional multiples are based on finite time steps corresponding to the 16 clock periods between X = ±1, the

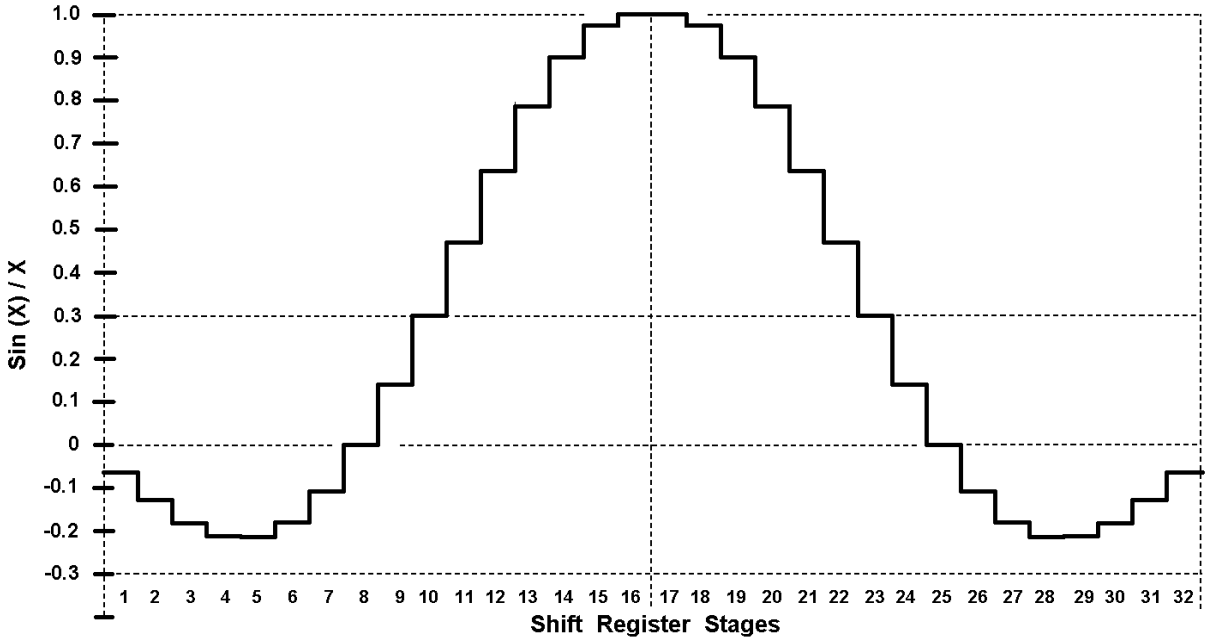


Figure 26-6 The *Sin (X) / X* stepped waveform of the Digital Filter's output as created by a single clock period circulating through the shift registers stages (horizontal axis).

zero amplitude steps. If a single shift register logic 1 state circulates through all 32 stages, it should produce the waveform shown in Figure 34.

While not a *filter* as most of us think of it, this stepped *impulse filter* is a form of lowpass with its equivalent -3 db bandwidth at **1/20th of the clock frequency.** All it needs is an analog lowpass filter following it to remove the higher harmonic content responsible for the step rise and fall times. Such an analog lowpass filter would also integrate the discrete amplitudes of the step flat

[11] In that function, the *X* is a multiple ***radian*** fraction of Pi, the same as the argument for the Sine and for the denominator. At shift register stages 10 and 23, the value of X would be (6/8) x Pi = 2.3562 and Sin (2.3562) [in radians] = 0.70711 so [Sin (X) / X] = 070711 / 2.3562 = 0.30011. ±1 Pi radians is the same as ±180° so (6/8) Pi radians is equal to 135°. It is surprising to the author that so many others fail to state that this [Sin (X) / X] function is done in radians, not degrees.

portions into a smoother curve. The digital filter's *cutoff frequency* ***always follows the shift clock frequency*** and it always contains the ***same number of harmonics*** of the pseudo random summed binary bit pattern regardless of the pattern length and clock frequency.

Normalized Values of *Sin(X)/X* for Resistor Network Summation Filter

Register Stage	Direct	Offset
16 17	1.0	1.0
15 18	0.97450	0.98899
14 19	0.90032	0.91788
13 20	0.78421	0.82273
12 21	0.63662	0.70065
11 22	0.47053	0.56382
10 23	0.30011	0.42343
9 24	0.13921	0.29088
8 25	0.0	0.17620
7 26	-0.10878	0.08658
6 27	-0.18006	0.02786
5 28	-0.21388	0.0
4 29	-0.21221	0.00138
3 30	-0.18097	0.02711
2 31	-0.12862	0.07024
1 32	-0.06497	0.12267

Note that extreme end values of *Sin(X) / X* are negative at the ends of the step waveform of Figure 34 and in the *Direct* column of the tabulation above. In the HP-3722A this is solved by taking the inverted logic state from registers 1 through 7 and 26 through 32 into the PNP-NPN switches and then adding a specific current from a -12.2 VDC source to the summing point. The same thing can be done with all-non-inverted shift register outputs (as in CMOS logic) and a DC offset.

Let R2:R1 = R4:R3 and the input is from the *Offset* column in tabulation above. Assume maximum level is +1.0 VDC and minimum at ground level. To achieve the same positive-to-negative swing as from the *Direct* column (1.21388 V peak-peak), the ratio of R2 to R1 (which determines the op-amp circuit gain) would be 1.21388:1. A ***positive*** offset voltage at R3 input would be 0.21388 / 1.21388 = 0.17620 VDC. If the input varied from 0 to +1.0 V, the output would vary from -213 mV to + 999 mV (almost, but not quite 1.0 V[12]). The output voltage would be the same as the *Direct* column above.

Capacitors could be added across R2 and R4 to get the first stage of a two-stage lowpass filter needed to achieve 12 db/octave (40 db/decade) attenuation above

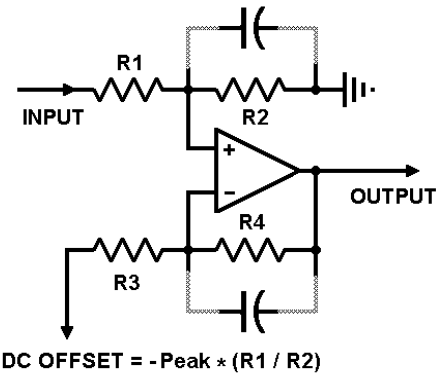
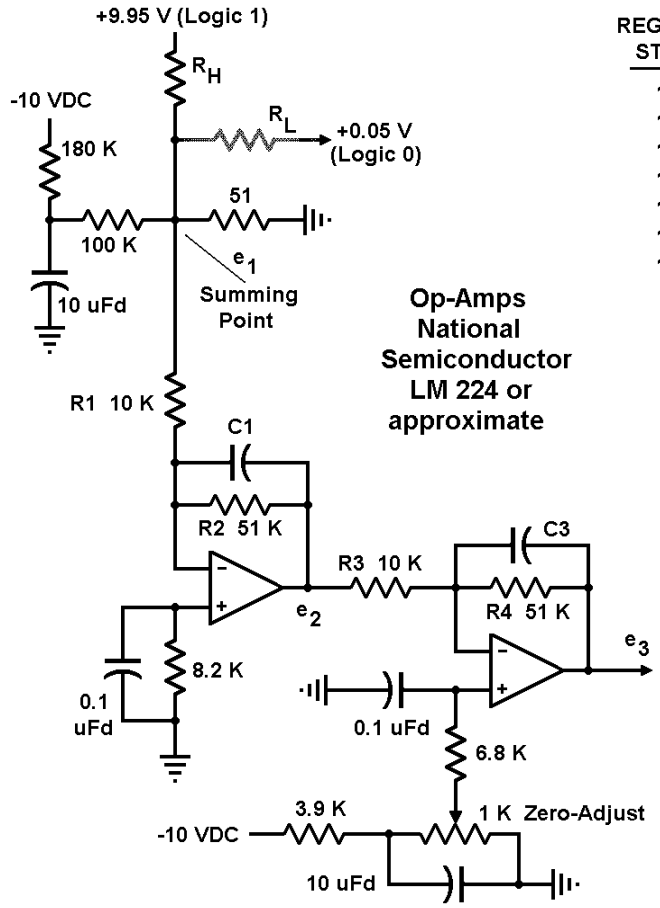


Figure 26-7 Op-Amp summing voltage multiplier and offset circuit for a positive to negative output.

[12] A numeric nicety for purists; 999 mV is within 0.01% of exactly 1.000 V. There must always be a slight output variation from exactly the R2/R1 *gain setting*. See the chapter on op-amps for details.



SHIFT REGISTER STAGES	R_H ± 5 %	e_1* mV
16, 17	10 K	49.20
15, 18	10.3 K (5.6 K + 4.7 K)	47.77
14, 19	11 K ((2) 22 K paralleled)	44.74
13, 20	12.4 K (6.8 K + 5.6 K)	39.71
12, 21	14.3 K (18 K parallel 68 K)	34.45
11, 22	18 K	27.39
10, 23	24 K (12 K + 12 K)	20.56
9, 24	34 K (12 K + 22 K)	14.53
8, 25	56 K	8.821
7, 26	115 K (100 K + 15 K)	4.188
6, 27	310 K (300 K + 10 K)	1.373
5, 28	- none -	0
4, 29	2 Meg (1 M + 1 M)	0.078
3, 30	315 K (300 K + 15 K)	1.27
2, 31	133 K (100 K + 33 K)	3.48
1, 32	82 K	6.026

* For a single Logic 1 recirculating through registers

Output Voltage 14.5 V Pk-Pk (full pattern)

Filter f_C	C1, C3	CLOCK
20 KHz	100 pFd	100 KHz
6 KHz	330 pFd	33 1/3 KHz
2 KHz	1000 pFd	10 KHz
600 Hz	3300 pFd	3 1/3 KHz
200 Hz	10 nFd	1 KHz

Figure 26-8 Summing and analog lowpass filtering suggestion for a noise generator.

cutoff frequency. This is included in one version of both a register stage summing and analog integrating amplifier shown in Figure 26-8.

There are 30 resistors from as many shift register stages (indicated on tabulation) to the 51 Ohm resistor at e_1. R_H refers to those resistors having a stage at Logic 1 or +9.95 V (74HC family shift registers used running from +10 VDC supply). R_L is the same resistor but at stages at Logic 0 or +0.05 V; the total parallel resistance to +0.05 V will vary depending on the pattern. No more than 20 stages are at Logic 1. The small negative current (from -10 VDC supply) counter-acts the finite Logic 0 voltage present at the *summing point.*

The two op-amps form a 12 db/octave lowpass filter whose cutoff frequency is ***one-fifth*** of the clock frequency.[13] The output noise bandwidth is effectively that of the lowpass filter response. The capacitors (tabulated in Figure 26-8) are approximate and selected in the author's circuit for the five noise bandwidths indicated. The op-amps are operating from ±10 VDC regulated supplies; ±12 VDC and even ±9 VDC supplies could be used. Output voltage, at full pattern with 1,048,575 clock

[13] From the operating principle description in the HP-3722A manual.

periods in length, will be about ± 7.75 V at peaks with the 1 K *zero-adjust* trimmer potentiometer set for equal maximum magnitude voltage swings. Its arm would have about a -1.2 VDC potential to balance the second op-amp.

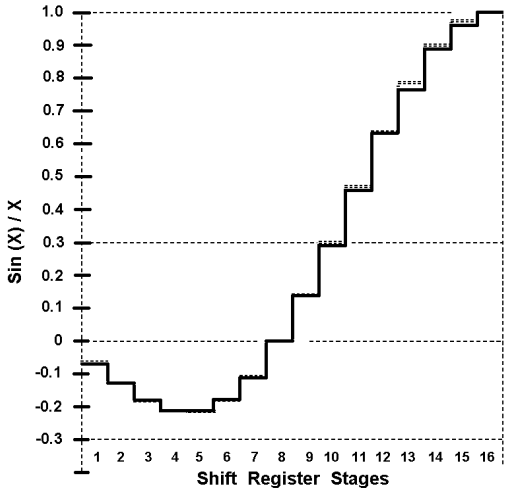


Figure 26-9 Plot of step voltage levels normalized to unity, compared to true *Sin(X)/X* values (dashed step lines).

Note: Unlike sinusoidal or square-wave generators, the output voltage will vary widely due to its pseudo random nature. According to alignment instructions in the HP-3722A manual, an output of about ± 7.75 V peak swing would read as 3.16 VAC ***RMS*** on an AC voltmeter having ***true RMS reading*** specifications.[14] Voltage gain (at low frequencies) of the two op-amps in Figure 26-8 is about 26, each stage gain determined by the ratios of R2:R1 and R4:R3.

Figure 26-9 shows only half of the impulse response waveform since the other half is identical except reversed in horizontal axis.

After passing through the analog lowpass filter most of the abrupt steps will be changed to rather smooth transitions from one step to the next. The time delay of the 12 db/octave lowpass is about one period of the clock frequency (lower right tabulation of Figure 26-5). At the fastest clock rate there is internal op-amp limitation due to *slew rate,* the specified volts/μSec response of the op-amp to a step input.

Some form of output voltage attenuation of the analog noise source should be provided for use as a test instrument for circuit development. That is left up to the designer's choice. A suggestion is to use one of the two op-amps out of a quad op-amp package (such as the National Semiconductor LM224) as a voltage follower into a complementary pair of PNP-NPN discrete emitter followers. Feedback from common emitters to the negative input of the op-amp would yield a very low output impedance yet having the same maximum ± voltage swings out past 20 KHz.

The HP-3722A had an 18-position front panel selector switch for clock frequencies from 1 μSec to 333 1/3 seconds (internal analog filter limited to 333 1/3 mSec lowest period setting) but that would be of unlikely interest to radio hobbyists.[15] Likewise the analog lowpass filtering down to 0.6 Hz is a bit low for hobby work. A suggestion to those wanting to build one, clock frequencies

---

[14] For Gaussian output an HP-3400A voltmeter is suggested in the manual, plus an HP-3440A DVM with an HP-562A recorder. The latter to measure the voltage at ***each step of a 1023 clock period length pattern.*** Calibrators are told to take the square of each of the 1023 step voltages, add them all, divide that by 1023, then take the square root of that value. The result should be the equivalent RMS voltage. A computer-controlled DVM could do that now but the manual was printed in 1968 and few calibration laboratories could afford the recently-made-available HP-9100 programmable desk calculator. It is difficult to find a *true RMS* voltmeter that a hobbyist can afford. *True RMS Voltmeters*, including their construction, is covered in the first chapter of Metrology..

[15] Original purpose of the instrument was varied but included very slow-rate geophysical phenomena and building structure analysis.

from 100 KHz to 1 KHz (noise bandwidths 20 KHz to 200 Hz) should be sufficient.

The binary bit pattern lengths will probably serve best at a minimum of 1023 clock periods. Below that they tend to emulate repetitive sounds rather than acting like pseudo random noise. A maximum of 20 stages' feedback for a pattern length of 1,048,575 clock periods. At a 100 KHz clock (20 KHz noise bandwidth) that length pattern would repeat every 10.486 Seconds. That long length could be hardwired-in once the instrument is finished debugging and test. Shorter patterns are preferred for debugging. Stage feedback and pattern lengths for 2-input Exclusive-OR:

PRSG Length	**Feedback from Stages**	**Pattern Length, Clock Periods**
10 Stage	10 and 7	1,023
20 Stage	20 and 17	1,048,575
31 Stage	31 and 28	2,147,483,647

Note: PRSG Lengths of 30 and 32 require three Exclusive-ORs, 30-6-4-1 and 32-22-2-1 stage inputs, respectively.

External oscilloscope synchronization is useful. Using Exclusive-OR state feedback with an inverter at the output, a 10-stage PRSG could sync on ANDing 9 stages, a 20-stage ANDing 19 stages, and a 31-stage ANDing 30 stages. That would provide one sync pulse per complete pattern at any clock rate. The inverter to make the state feedback into an Exclusive-NOR would allow a reset of all shift registers to all Logic 0s and prohibit all shift register stages reaching all-Logic-1s.

Input for an external clock source would also be useful. The user would have to remember that the analog filter cutoff frequency is selected by the internal clock rate switch even if the internal clock is not used.

Chapter 27

Other *Stables* (Monostable, Astable Multivibrators)

Monostable and astable multivibrator circuits can be used for a variety of applications in radio and television. The available *one-shot* (monostable) devices and their circuits are discussed in here plus variations on them to form *free-run* (astable) circuits. The ubiquitous *timer* IC subsystem-on-a-chip is covered as well as some *function generator* ICs.

General

Bistable multivibrators were described in the previous chapter. This chapter concentrates on the *monostable multivibrator* or ***one-shot*** (so dubbed because it generates one pulse for each trigger input) and the *astable multivibrator* or ***free-run*** digital oscillator (constantly producing a repetitive AC output).[1]

Vacuum tube one-shots have been extensively covered in detail in early college-level electronics textbooks.[2] While fascinating to some, that won't be repeated here. Available integrated circuits can do that task with the only design necessary being the selection of a few external passive components. The duals of those first tube one-shots and astables will be explained briefly to outline their control by a single R-C combination. Those same R-C combinations (with differing values) are used with the integrated circuit versions.

[1] The origin of the term *multivibrator* is obscure (and unimportant) but it dates back before the 1920s when vacuum tubes were relatively new and digital logic, not really named as such, existed mainly in telephone automatic switch centers. One would suppose the analogy then was made to the musician's *tuning fork* tone standard. That device *vibrated* when struck. Nearly all working electronics labs eschew the term *multivibrator*, not using it much in practice (except perhaps in reports written to impress their few readers). For that reason I refer to monostables as ***one-shots*** and the astables as ***free-run***. To be correct and precise, I suppose the astable ought to be called a *digital oscillator* but that gets confused in many minds with the crystal-controlled *gate oscillator* (also described in this chapter) and L-C oscillators which are part of some large-scale integrated systems-on-a-chip.

[2] One of the best (in this author's opinion) is Jacob Millman and Herbert Taub's *Pulse, Digital, and Switching Waveforms,* McGraw-Hill Book Company, 1965. Good, basic stuff that does what the subtitle says (*devices and circuits for their generation and processing*) and includes some basic discrete transistor duals of vacuum tube circuitry. The early full operation and maintenance manuals for Tektronix *530* and *540* oscilloscopes included detailed voltages and waveforms of non-linear vacuum tube circuitry.

One-Shots - Monostables

The Discrete Transistor Version

To do a chronological cross-over from tubes to ICs, Figure 27-1 shows the intermediate solid-state, discrete-part version that was essentially based on a dual-triode vacuum tube one-shot. In the quiescent period, Q2 is conducting and Q1 is generally cut off by the Q2 current drop across R5.

When a positive pulse of amplitude greater than the quiescent drop across R5 arrives, Q1 will conduct and its collector voltage will drop (waveform B). When the collector voltage drops, C1 is pulled low and must recharge to its new voltage through R3. With the right selection of resistor values and currents, the base of Q2 is also pulled low at that time, sufficient to cut off Q2. With Q2 cut off, its collector voltage will rise to V_{CC} (waveform D). For the meantime, assume that Q1 remains conducting after the trigger pulse (waveform A) has dropped to zero.

C1 will eventually charge to equilibrium potential through R3 such that the base voltage of Q2 (still cut off) will rise at an exponential rate as in waveform C. At some point in time, the base voltage will exceed the base-emitter cut-off limit (the little circle on waveform C and the asterisk marking voltage differences) and Q2 will again conduct. As Q2 begins conducting, the common emitter voltage (waveform E) will rise, causing Q1 to cut off.

C1 and R3 in Figure 27-1 are the components chiefly responsible for controlling the length of time of the positive output pulse. Note the exponential voltage curve, very much like the usual *time-constant* curve of an R-C circuit responding to a step-function. Had C1 been allowed to charge at a constant-current rate, waveform C would have been a straight-slope ramp.

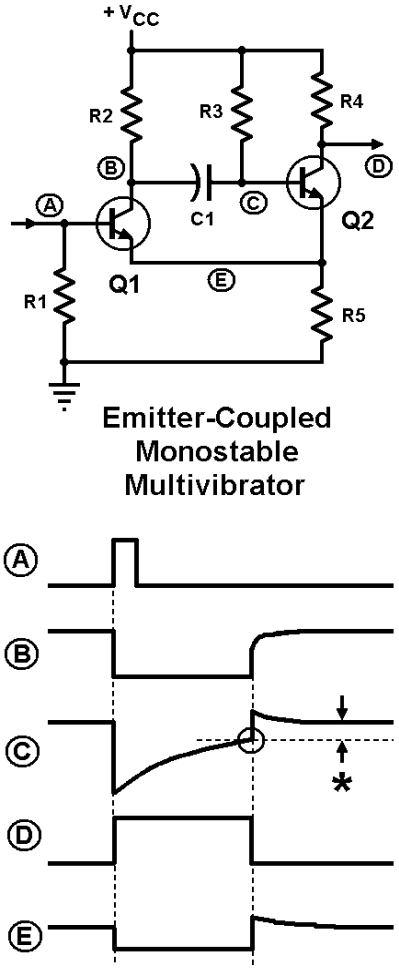


Figure 27-1 *conventional* emitter-coupled one-shot of discrete parts

Digital Logic ICs

All of the available TTL and CMOS one-shot devices accept triggers of positive-going or negative-going (selectable on-chip) edges and produce a pulse width governed by an external resistor (***Rx***) and capacitor (***Cx***) connected as in Figure 27-2. The pulse width itself is determined by PW = K • R • C with width in seconds, resistance in Ohms and C in Farads. Alternately, one can use µSec, Ohms, and µFd or nSec, KOhms, and µFd. The value of K is between 0.1 and 1.0 and is device

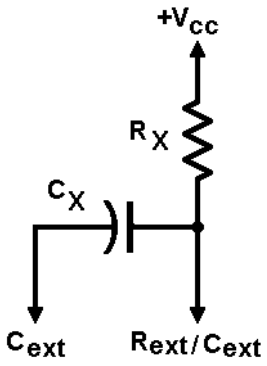


Figure 27-2 Pulse width determination

specific. Each device has specific minimums and maximums for ***Rx*** and some have specific maximums for ***Cx.*** There is no minimum for ***Cx*** as each one-shot device has some internal capacitance which sets the absolute minimum output pulse width. The 74LS122 has an internal ***Rx*** available.

Table 27-1 - Available One-Shot IC Devices, values for +5 VDC supplies[3]

Device	Per Pkg	K ①	Rx min.	Rx max.	Cx max.	Min. Trig. Width	Schmitt Trigger
CD4047	1	2.48	22 K	2.2 M	0.1 µFd	200 nSec	No ②
CD4098	2	0.5	5 K	10 M	0.1 µFd	190 nSec	No
CD4528/ HC4528	2	0.27	5 K	100 K	0.1 µFd	150 nSec	No
CD4538/ HC4538	2	1.0 ③	5 K	10 M ④	- none - ⑤	70 nSec	No
74HC121	1	0.8	2 K	1 M	- none -	120 nSec	Yes ⑥
74LS121	1	0.55	1.4 K	100 K	1000 µFd	500 nSec	Yes ⑥
74LS122	1	0.45	5 K ⑦	260 K	- none -	100 nSec	Yes ⑥
74LS123	2	0.38	5 K	260 K	- none -	50 nSec	Yes ⑥
74HC123	2	0.45	5 K	- none -	- none -	30 nSec	Yes
74LS221	2	0.7	1.4 K	100 K	1000 µFd	40 nSec	Yes ⑥
74HC221	2	0.8	2 K	1 M	- none -	120 nSec	Yes ⑥

① For **Cx** greater than 1000 pFd.
② CD4047 is pin-connection option as a free-run multivibrator.
③ Linear curve for K•Rx•Cx from 100 µSec to 1 Sec (most linear of all).
④ Practical limitation from Cx leakage and PCB insulation resistance.
⑤ Minimum of 2000 pFd for HEF4538B only, all other 4538s allow 0 pFd minimum..
⑥ Positive-going edge (***B*** input) only; HC123 and HCT123 have Schmitt on both triggers..
⑦ R_{INT} of 10 KOhm provided; Rx does not have to be used.

[3] Data originally from National Semiconductor application note AN-372, *Designer's Encyclopedia of Bipolar One-Shots,* May 1984, by Kern Wong; expanded to include CMOS and all data checked against device datasheets dated from 1998 to 2005 from Philips, Fairchild, Texas Instruments and Motorola/ON Semiconductor.

All but two (74x121, 74x221) allow ***retriggering***, a feature that allows a new trigger input to generate an output pulse *before* the first output pulse has finished.[4] With retriggering it is possible to generate a very long output pulse with a burst of trigger edges whose spacings are shorter than the set pulse width for a single trigger. Several devices also have ***clear*** (or reset) inputs to cancel any output pulse before it has been completed. While the output pulse width (all have active-high and active-low) width is essentially a straight line corresponding to the PW = K • Rx • Cx equation, device package, external wiring capacitance, and internal delays begin distorting that curve below 1000 pFd Cx such that the effective K decreases. This is shown on all datasheets. The minimum possible pulse width is between 30 and 50 nSec for the ***74***s (both bipolar and CMOS) but the ***CD4***s are limited by the slower output rise and fall times to roughly 0.2 µSec.[5] There are some slight differences between those of the same type number and their family variants; e.g., 74LS123 and 74HCT123 with the HCT having a different K and Schmitt trigger inputs for both positive- and negative-edge trigger inputs.

For protection against sudden cessation of power supply voltage, an additional silicon diode (1N914 or 1N4148) is suggested as in Figure 27-2. This is primarily for one-shot output widths which are long and in case of sudden supply voltage failure such as breadboard accidental cessation of power. Reasons are covered in detail as to the possible failure conditions in datasheets for CD4538, 74LS123, 74HC123, 74HC121 and 74HC221. This isn't an absolute requirement but it will protect the timing control portions of a device even if it has internal protection diodes.

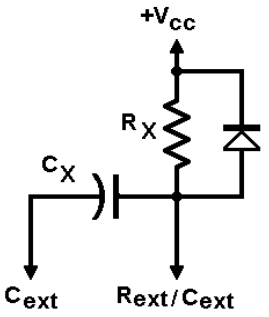


Figure 27-3 Added protection diode connection.

All of the devices in Table 1 have a choice of positive-going edge or negative-going edge triggering. The designer has to consult the datasheet to ascertain what to do about the *unused* trigger input, whether to tie it high or low. *Edge triggering* implies a form of flip-flop inside the IC and that is the case with every device tabulated; most have two flip-flops internal. The CD4047 was apparently intended as a sort of all-purpose device since it has more flip-flops inside it, most of them devoted to the optional free-running mode.

In general, the pulse width stability is better than ±5% over a temperature range of 0° to 70° C and, with regulated 5 VDC supplies, the supply voltage variation of pulse width can be disregarded. Expectance of good pulse width stability with values of ***Cx*** at 1 µFd or greater depends on the capacitor leakage; electrolytic or tantalum capacitors should ***not*** be used for ***Cx*** for that reason and for their polarization. Similarly, with ***Rx*** approaching 1 MOhm for long pulse widths, stability will be decreased from circuit board insulation resistance (and odd residue thereon) and possibility of stray noise pickup (particularly with CMOS families and their high-impedance inputs).

Quick-Design Non-Critical Pulse Widths and Delays

[4] This was not done in vacuum tube circuits, hence is seldom mentioned in earlier texts. With the ability of ICs to include more internal circuitry, the retriggering feature became standard. Those interested can consult earlier datasheets to see internal schematics and read applications information on how retriggering was done.

[5] For pulse widths in the small µSec range, the author suggests considering a trimmer capacitor to set the proper value of Cx.

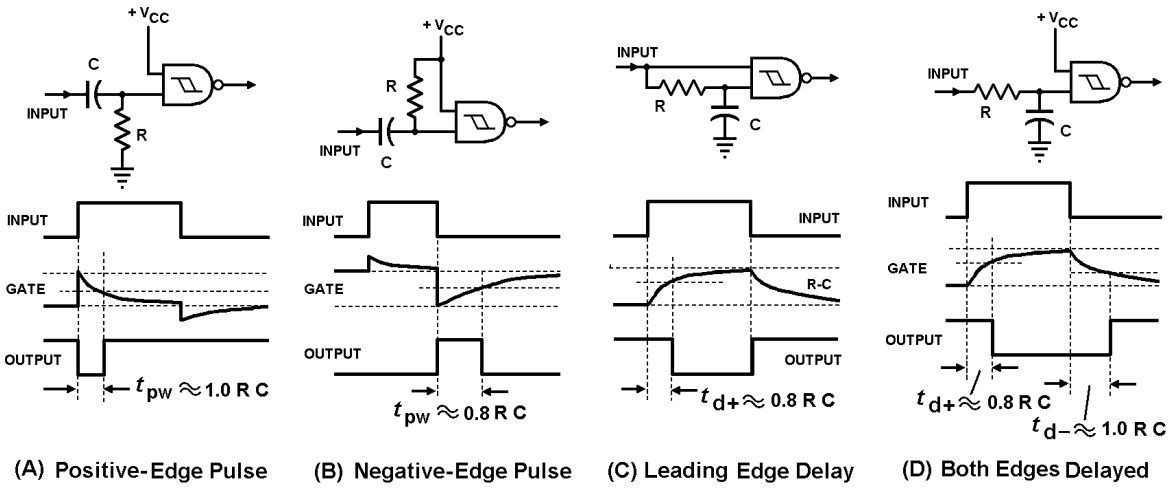


Figure 27-4 Various CMOS NAND Schmitt trigger circuits for +5 VDC supply operation

CMOS family Schmitt trigger gates and simple R-C networks can be used for both delay and pulse width formation if the timing requirements are not critical. As in Figure 27-3, the input pulse drives an R-C integrator to create an exponential voltage rise and fall fed to a high-impedance Schmitt trigger input of a NAND or Exclusive-OR gate. The result is a delayed input pulse appearing at the gate output with some slight change in pulse length; dissimilar Schmitt thresholds change the output length. Formulas given in application notes and datasheets are:

(A) POSITIVE - EDGE TRIGGER

$$t_{PW} = R \bullet C \bullet L_N\left(\frac{V_{CC}}{V_{CC} - V_P}\right)$$

(B) NEGATIVE - EDGE TRIGGER

$$t_{PW} = R \bullet C \bullet L_N\left(\frac{V_{CC}}{V_N}\right)$$

(C) LEADING EDGE DELAY

$$t_{D+} = R \bullet C \bullet L_N\left(\frac{V_{CC}}{V_{CC} - V_P}\right)$$

(D) DELAY BOTH EDGES

$$t_{D+} = (\text{as C}) \quad t_{D-} = R \bullet C \bullet L_N\left(\frac{V_{CC}}{V_N}\right)$$

All units in Seconds, Ohms, Farads, Volts

V_P = Positive - going Schmitt threshold V_N = Negative - going Schmitt threshold

Approximate formulas in Figure 27-4 are based on V_P of 2.8 Volts and V_N of 1.8 Volts. Specification limits of V_P are 2.2 to 3.3 Volts and V_N of 0.9 to 2.3 Volts at +5 VDC supply. Actual times, assuming ±5% R and C, would be about ±30% of desired. For other supply voltages calculate as needed and take the *Natural logarithm* for the multiplier. The ***B*** suffix of the CD4000s is preferred for higher supply voltages since the input protection diodes will handle the overshoot and undershoot spikes present in Figure 27-3 (A) and (B). Note: The input pulse must be longer than about 2•R•C times in order for the exponential gate input voltage to curve through the Schmitt

thresholds. The circuits in Figure 27-4[6] have approximate formulas for +5 V_{CC} only due to the variation of the Schmitt trigger level voltage between devices. These are reputed to be good for widths and delays from 5 µSec to 1.0 Sec.

The circuit of Figure 27-5 generates a 10 µSec positive pulse after about 1 Second of the logic supply reaching 5 VDC. It combines Figure 27-3's (D) followed by (B). Part values are not critical and the 100 KOhm drain resistor could be omitted if desired. It is useful for initializing shift counters and some microprocessor/microcontroller circuits.

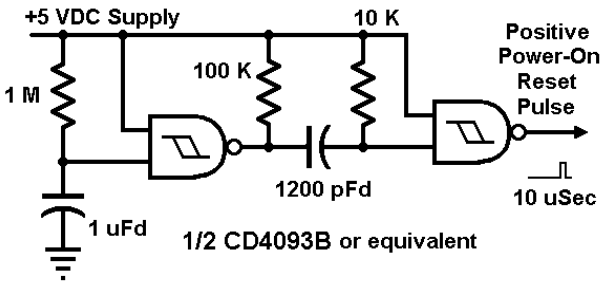


Figure 27-5 Power-on reset pulse generator.

Free-Running Schmitt NAND Gates

A single 2-input CMOS NAND gate with Schmitt trigger inputs can function as a free-running multivibrator as shown in Figure 27-6. The values for R and C (within limits shown) are calculated from:

$$t_{PERIOD} = R \bullet C \bullet \mathrm{Log}_E\left[\left(\frac{V_P}{V_N}\right)\left(\frac{V_{CC} - V_N}{V_{CC} - V_P}\right)\right]$$

Where:

Units of Seconds, Ohms, Farads, Volts

V_P = Positive-going Schmitt threshold

V_N = Negative-going Schmitt threshold

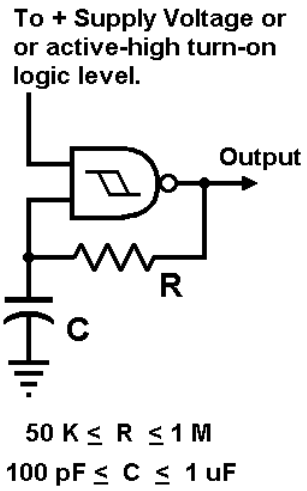


Figure 27-6 Free-run astable using a single 2-input NAND Schmitt.

For +5 VDC supplies and typical thresholds of 2.8 and 1.8 Volts (for a CD4093 2-input CMOS NAND with Schmitt trigger inputs), this approximation could be used:

$$t_{PERIOD} \approx 0.8 \bullet R \bullet C \qquad \text{[within about } \pm \text{ 30\%]}$$

[6] From RCA Corporation *COS/MOS Integrated Circuits* databook, 1977, application note ICAN-6346 (pp 630-634) and Fairchild Semiconductor datasheet for CD4093B, January 1999, and Texas Instruments CD4093B datasheet of September 2003. Most manufacturers refer to V_{DD} (*Drain* voltage) rather than V_{CC} (Collector voltage) but mixed TTL and TTL-compatible logic circuits share the same supply voltage.

Within the limits shown on Figure 27-6, the free-running frequency could range from 250 KHz down to 1.25 Hz. If the upper input to the NAND gate is tied to supply voltage, the circuit always free-runs; it is self-starting on power-up. An active-high control could be applied to the upper input for an on-off switching as desired. Symmetry of the output waveform depends on the Schmitt threshold voltages of the gate.

Two One-Shots As A Fixed-Width, Variable Period Pulse Generator

Given the ability of available one-shot devices with both positive and negative edge triggering, it should be intuitive that *two* in cascade can provide both a delay in the start of a pulse and the pulse width itself, both independently adjustable. Figure 27-7 takes that a step further in providing a free-run (astable) mode. This sort of circuit can be quickly generated from one dual one-shot device for testing purposes if desired.

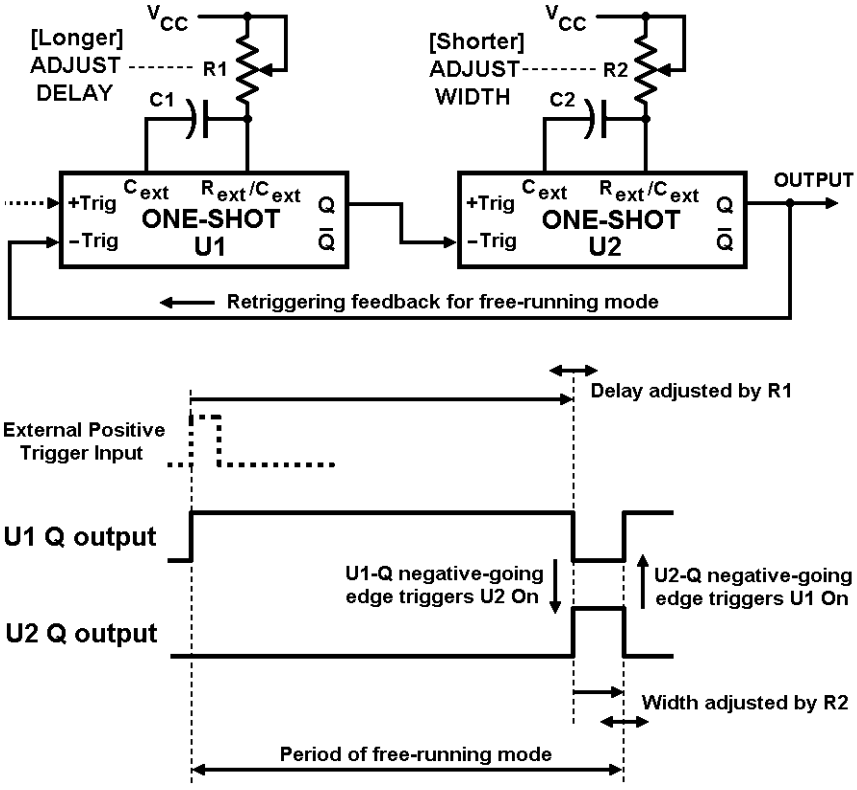


Figure 27-7 Free-running pulse generator with independent On and Off time adjustment.

The triggering of U1 from an external repetition period generator is shown in dotted lines. Assume that U1 has turned on. When U1 ends its set on-time, the negative edge from U1-Q triggers U2 on. When U2 finishes its on-time, the negative edge from U2-Q triggers U1 back on and the free-running cycle has begun. The repetition time is the total of U1 and U2 on-times; there is no independent control of repetition time.

There is the possibility of a power-on condition where both U1 and U2 came up off. There would be no triggering feedback in that case. One solution is to use the Figure 27-4 power-on circuit to temporarily trigger either one-shot. Another is to use different R-C combinations in the supply voltage decoupling; one one-shot would come on faster than the other. Note that this power-on condition is fairly common on all time-related sequencing circuits which have states that should be avoided; analog oscillators are biased so that they are self-starting and don't need this power-on set or reset.

Commercial pulse generator test instruments have three independent controls: Period (or frequency), delay, and output width, generally obtained from a free-running oscillator (period)

followed by two one-shots (delay and width). A delay feature is good for allowing an oscilloscope synchronization directly with the astable and the delay allowing for better observation of the pulse effects on a developmental circuit.

A Surviving CMOS All-Purpose Device, the CD4047

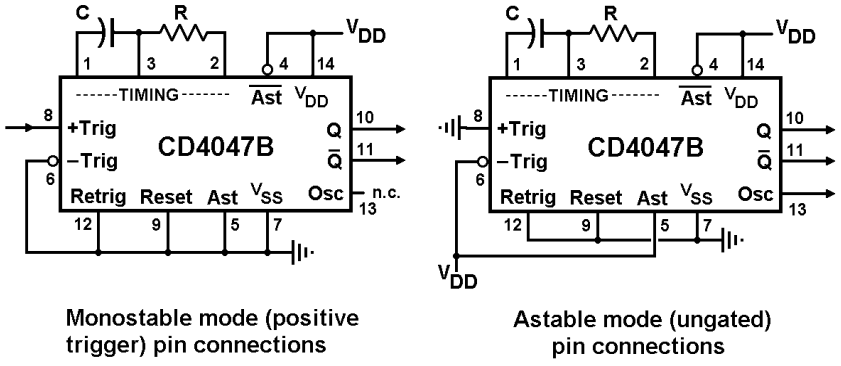


Figure 27-8 Pin connections to the CD4047 connected as a one-shot or as a free-running, ungated multivibrator.

The CD4047 was one of the original RCA *COS/MOS* devices and apparently intended to be an all-purpose pulse building block.[7] It may be connected as a one-shot with both positive- and negative-edge triggering or it can be a free-running multivibrator (ungated or gated) with a 50% duty cycle on Q and Q-not outputs. Timing is reasonably precise (within ±5%) but at relatively slow speeds (about 200 KHz top in astable mode, pulse widths limited to 2 μSec or longer). Note that timing resistor **R** in Figure 27-8 goes to pin 2, not to V_{DD} (or V_{CC}) as with all other devices in the monostable group of devices in Table 1.

For one-shot pulse width setting, use width (seconds) = 2.48•R•C (Ohms, Farads) and for free-running frequency period (seconds) = 4.40•R•C (Ohms, Farads) for any R from 10 KOhms to 1 MOhm, any C value from none to 1 μFd (at least 100 pFd minimum to stay within the time-constant formula). In the free-running mode, an internal flip-flop divider insures 50% duty cycle but pin 13 will provide twice the output frequency.

For a negative-edge trigger in one-shot mode, the trigger goes to pin 8 but pin 6 (positive-edge trigger) must be tied to the supply voltage. It has retriggering capability through pin 12 but not with the edge triggers. In the astable mode the output may be gated through either pin 5 (active-high gate) or pin 4 (active-low gate). The datasheets on the CD4047B have a table of connections and inputs, outputs for all 7 modes of operation. Figure 27-8 shows only the two most-commonly-used modes' pin connections. While this device is certainly versatile and mature in the market and production, there are some other versatile devices lumped under the *Function Generator* and *Timer* category which might be better as free-running R-C oscillators.

---

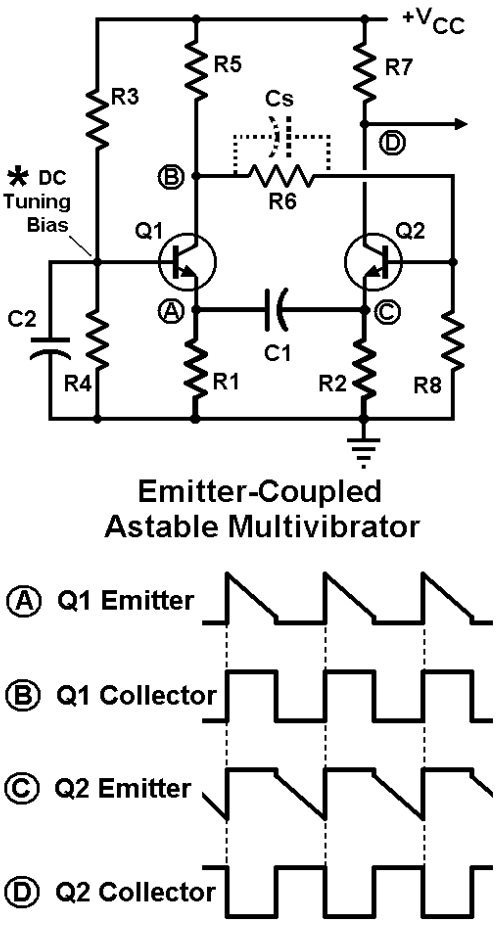
[7] Supposition. The first *COS/MOS* group (so-called ***A*** series) debuted when TTL was just beginning to become the logic device standard but the Schottky junction TTL (74S, 74LS) was not yet on the market. As a result, the planning at RCA Corporation Solid-State Division did not make pin-compatible CMOS devices as drop-in replacements for TTL. The 74HCT and 74ACT CMOS devices which were pin, function, and drive compatible with old TTL wouldn't arrive until over a decade later. Just the same, those early CMOS devices won acceptance in the market in that interim due to relatively microscopic current demand and many applications didn't require the speed of the early TTL families.

Discrete-Form Emitter-Coupled Astable

There are as many forms of an astable as there are monostables. Figure 27-9 shows an emitter-coupled version because the free-running frequency is determined primarily by C1, R1, and R2. Having only one capacitor allows easier control of the frequency tuning; resistors can be chosen to tight tolerances and good temperature-range stability but capacitors tend to be expensive and large, especially with large values.

C2 in Figure 27-9 is a very large value to act as a decoupling for bias current to the base of Q1. The R3-R4 and R6-R8 bias-setting resistors are selected (with R5 and R7 collector resistors) to have both Q1 and Q2 swing barely between saturation and cutoff.[8] As such, the slightest amount of noise or disturbance on power-on will begin the oscillation.

Figure 27-9 shows three cycles of the repetitive waveform. To the right of the left-hand dotted line, Q1 has gone into cutoff and Q2 has begun conducting. The voltage across C1 has reversed itself, Q1-emitter near ground and Q2-emitter a positive voltage from Q2's emitter current through R2. But, C1 can only change that charge through R1 so the voltage at point A will be falling at an exponential rate. As Q1-emitter voltage is falling, the base-emitter voltage (base to ground fixed in voltage) will rise to allow Q1 to conduct again.



Emitter-Coupled Astable Multivibrator

Figure 27-9 Two-transistor discrete-component emitter-coupled astable.

When Q1 begins conducting, its collector voltage will fall (point B waveform) enough so that there is insufficient base current (through R6-R8) to keep Q2 conducting. Q2 will cut off, C1 has to change its charge potential, this time charging through R2. Q2 emitter voltage is now falling at an exponential rate (point C waveform) but, eventually, its base-emitter junction voltage will rise enough again to allow Q2 to conduct. When Q2 conducts, the charge across C1 will again reverse and Q1 will conduct. The cycle has been completed. C_S is a *hurry-up* capacitor to compensate for stored charge in the Q2 base-emitter junction, practical value less than 56 pFd.

[8] That is explained in excellent detail in textbooks such as *Millman and Taub* already cited (familiar name, formal is *Pulse, Digital, and Switching Waveforms*). There are several other texts which go into such painstaking detail. It is a task to design a discrete circuit that meets strict environmental limits with a 50% h_{FE}-range transistor type. Figure 27-9 isn't quite the circuit description on page 448 of *Millman and Taub* but is a form the author has done that is only slightly easier to apply with practical transistors.

The asterisk and notation *DC Tuning Bias* points to the possibility of an external voltage giving some control of the repetition rate. As the DC base voltage of Q1 is raised, the point at which the base-emitter junction causes Q1 to go into conduction happens sooner. See waveform A and consider the baseline to be slightly higher. There is a limit to this *control voltage* that is set by all the resistors and their interaction relative to the charge-discharge time of C1 through R1 and R2. While that is seldom more than ±10% of center frequency (if that), it does point out that voltage control of a free-running oscillator is possible without resorting to variable-capacitance diodes.

An Astable Multivibrator of Wide Frequency Range, Controlled by Voltage

Figure 27-10 shows the block diagram of the wide-range variable-frequency generator in the AN/ASM-416 military test set.[9] While the frequency range is only 1 Hz to 1 KHz, it can be precisely controlled by a single, multi-turn potentiometer, holding a constant amplitude at any frequency.

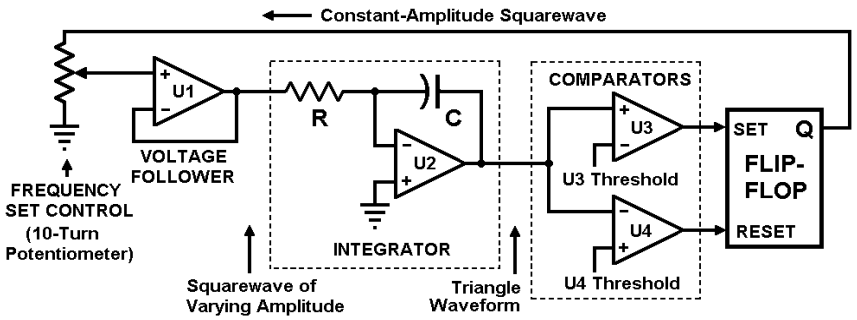


Figure 27-10 Linear frequency control wide-range oscillator, part of the military AN/ASM-416 test set.

The Set-Reset flip-flop is triggered by comparators U3 and U4 whose inputs come from a triangle waveform output by integrator U2. With no shunt resistance across ***C***, the integrator acts as a constant-current charging-discharging (through ***R***) source for the capacitor. The end result is a straight-slope triangular waveform. Frequency is varied by controlling the amplitude of the

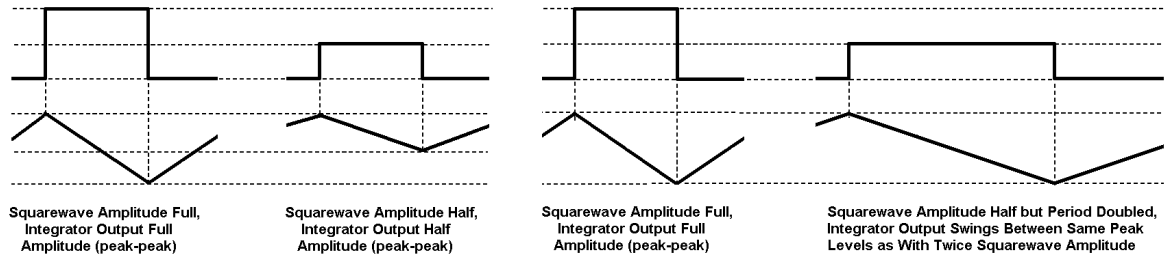


Figure 27-11 Method of control of repetition rate in the circuit of Figure 27-10.

[9] The author designed this in 1970 at RCA Corporation *Electromagnetic and Aviation Systems Division*, Van Nuys, CA. The test set would calibrate and maintain a Voice Warning Adapter to give Army Aviation pilots an audible warning of up to 20 fault conditions on 5 Army Aircraft. Needless to say, it had to work under severe military temperature environments and use only approved Mil-Spec components. Done with all discrete components on most of a PCB, it could have been replaced by a single Signetics NE555 Timer IC that was first available in 1971.

squarewave input to the integrator via the potentiometer, shown in Figure 27-11. In Figure 27-11-(A) and -(B) a constant-period squarewave with full and half amplitude will produce a triangular waveform at the integrator output whose peak-to-peak amplitude is maximum and half of maximum. The voltage out of any integrator is the product of the input amplitude and the R-C time constant at any given frequency. Think of the *angle* of the ramp slope as proportional to input voltage amplitude.

The flip-flop ***Q*** output is clamped to fixed voltage limits so its peak-to-peak voltage swing is constant at any frequency. A linear-taper potentiometer adjusts the flip-flop squarewave amplitude into a voltage follower (to isolate potentiometer from integrator input loading[10]). If the squarewave amplitude (now from the flip-flop) is halved, as in Figure 27-11-(B), the slope of the ramp voltage over time is also halved. But, the comparators (U3 and U4), producing the Set and Reset triggers of the flip-flop always output on the *same* ramp voltage input. The end result is that halving the squarewave amplitude will also halve the repetition frequency. Repetition frequency is directly proportional to the change in squarewave amplitude.

Repetition frequency could have been changed by varying ***R*** or ***C*** of the integrator (squarewave amplitude input kept constant). A variable capacitor for a 1 Hz to 1 KHz control range was out of the question. A variable resistor (potentiometer) for ***R*** could have done it but that made a difficult alignment adjustment due to necessity of long wire leads to the integrator in the test set physical structure. In a prototype breadboard circuit, the precision wirewound potentiometer tended to change the R-versus-time curvature at shorter periods. Varying the squarewave amplitude was decided as the safest approach to both linear-frequency control (no custom dial was needed) and calibration in light of the interior physical structure.[11]

Voltage control of repetition frequency was enabled accurately.

TIMERS

The Start of it All

The Signetics NE555 *Timer* debuted in 1971 and, by 1982, would take the title of the most-produced and sold integrated circuit in the world. It is basically a flip-flop and two comparators, a high-current (200 mA) output and an NPN *discharge* transistor. Quite stable over temperature (50 PPM/°C), it can operate over a 4.5 to 16 VDC supply voltage change yet drift only ±0.1% per

[10] A voltage follower has a very high input impedance and a very low output impedance. That effectively isolates the potentiometer's division ratio from being changed by shunt loading to ground of a directly-connected integrator input impedance.

[11] The physical structure of the system you are building has a large effect on *how* variability of functions is controlled. In older electronics this was solved by *mechanical coupling* of the variable control device which was placed close to the circuit being varied. In newer electronics there are many techniques which allow variability through DC control or by means of microcontrollers doing everything from adjusting frequency (via PLLs or DDSs) to controlling audio volume to doing complicated modulation mode switching.

supply Volt change. With external components, it can be configured as a one-shot, an astable free-running oscillator, a linear ramp generator, and many other circuits based on a basic building block of a small 8-pin DIP. It is second-sourced by nearly every semiconductor manufacturer as an active item, is available two per package (14-pin DIP, usually dubbed 556), in its original bipolar form or in CMOS form. Designed by the late Hans Camenzind in 1970, its development story is told in detail in an excellent text, *Designing Analog Chips.*[12]

Interior of the *555*

Figure 27-12 has a block diagram representing the inside of a typical 555 Timer. Note the similarity to the circuit of Figure 27-10.[13] Comparator *set-points* (threshold voltage for changing state) are fixed by the 5 KOhm resistor string to 2/3 V_{CC} for comparator A, 1/3 V_{CC} for B. For one-shot operation, the trigger input on pin 2 is negative and must be less than the set-point voltage. That trigger pulse width must be less than the one-shot's output pulse width. Internal flip-flop set and reset are toggled by the comparator outputs. There are two outputs, one through an inverting stage to pin 3 for up to 200 mA of current[14], the other through an NPN stage with collector to pin 7 intended for discharging a timing capacitor. The external reset control pin 4 is active low and will reset the flip-flop; that is intended for better retriggering in monostable mode. Tie pin 4 to V_{CC} when no external reset is used.

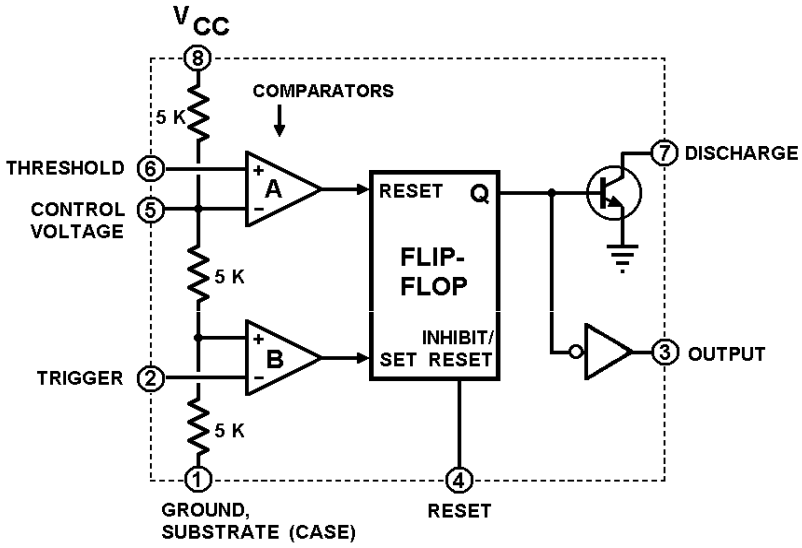


Figure 27-12 Representative block diagram of the 555. Pin numbers are for the 8-pin DIP

---

[12] *Designing Analog Chips* by the late Hans Camenzind, Virtualbookworm.com Publishing, 2005, ISBN 1-5893-9718-5, 242 pages. A treasure trove of details on the actual transistor and diode junctions plus some details on circuit design itself. This was also available in PDF form as a free download at www.designinganalogchips.com in 2006. Hans Camenzind passed away in late summer, 2012.

[13] Figure 27-10 circuit is not original, was somewhat obscure in 1969 but not unknown. It is almost intuitive that comparators can be used to toggle flip-flops, as they had done in many monostable ICs. Figure 27-12 block diagram is similar to what appeared in the 1972 Signetics *Linear Data Book, Volume 1*, Philips application note AN170, *NE555 and NE556 Applications,* December 1988, ON Semiconductor MC1455, MC1455B datasheet, February 2006.

[14] There are limitations on the maximum amount of output (pin 3) current, depending on supply voltage. See a datasheet for maximums and on-voltage, off-voltage levels with various loads.

555 As a One-Shot

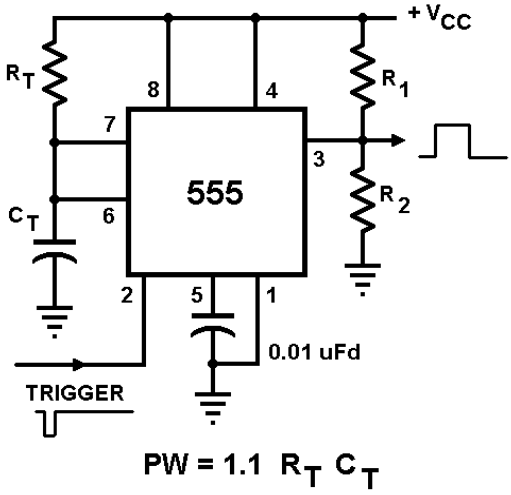


Figure 27-13 555 as a simple one-shot

In the circuit of Figure 27-13 a negative trigger of at least 50 nSec width at a minimum amplitude of 1/3 the supply voltage will produce a positive-going output pulse width = 1.1 • R_T • C_T using Seconds, Ohms, and Farads values. A 0.01 µFd and 100 KOhm yields a pulse width of about 1.1 mSec.

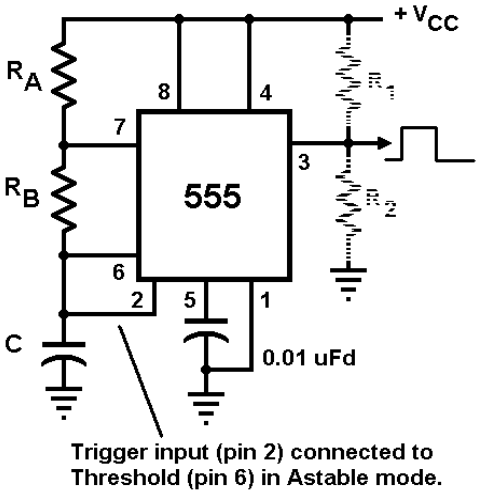
R1 and R2 are arbitrary and represent the load resistance. For a high output voltage, *sourcing* current (from ground through load resistance into pin 3), its value is fairly constant at (V_{CC} - 1.4) Volts from 1 mA to 20 mA. The low output voltage, *sinking* current (from pin 3 out through load resistance to V_{CC}) is less than 0.2 V at sink currents up to 10 mA with a +5 VDC supply; up to 25 mA with +10 VDC and +15 VDC. Low output voltage rises to about 1.0 V at sink currents of 20 mA and more with a +5 VDC supply. Output pulse rise and fall times are about 100 nSec. Propagation delay from trigger leading edge to output pulse width leading edge is 100 to 150 nSec; rise- and fall-times, propagation delay will be lessened by using R1 or R2 values lower than a following stage's higher input impedance. The *discharge* transistor (pin 6, open-collector NPN) is rated at 200 mA maximum.

AC-coupling a trigger input is generally recommended. The ON Semiconductor datasheet (February 2006) shows sequential 555 one-shots with trigger coupling of 0.001 µFd and 27 KOhms. Supply current at 25°C, discounting load current sourcing or sinking, is about 3.5 mA at +5 VDC, 6.5 mA at +10 VDC and 10 mA at +15 VDC. Pulse width versus supply voltage variation is so little that passive component tolerances will determine stability.

The 0.01 µFd capacitor to ground at pin 5 decouples the set-point voltage input to comparator A's negative-going input (Figure 27-12) to prevent stray spikes and general supply voltage transients from inadvertently changing the one-shot (or astable) on time. It also does decoupling to comparator B's positive-going input. Maximum value is not critical; a higher value could be used.

555 As a Free-Running Oscillator

That application has the 555 triggering itself as seen in Figure 27-14. Timing capacitor voltage will rise from 1/3 supply voltage to 2/3 supply voltage during its charging time (and back down). The duty cycle can be adjusted by varying the ratio of R_A to R_B.



$$PRF = \frac{1.44}{(R_A + 2R_B)\,C}$$

Figure 27-14 555 Connected as a free-running multivibrator.

$$\text{Duty Cycle} = \frac{R_B}{R_A + 2\,R_B}$$

The lower the relative value of R_A, the more the duty cycle approaches 50% (equal on and off times of the output). The limitation is the 200 mA maximum conducting current of the *discharge* transistor.

$$\text{Minimum } R_A \text{ Value, Ohms} \approx \frac{V_{CC}}{0.2}$$

Note that this also applies to R_T of the one-shot in Figure 27-13.

If R_A is 1.0 KOhm, R_B a 0 to 50 KOhm potentiometer and C is 0.047 µFd, the frequency range would be 300 Hz to about 30 KHz. Duty cycle would be 49.5% at the low frequency end and less than 0.1% at the high frequency end.

Recommended external resistance (total) for timing setting is 1 KOhm to 10 M Ohms and external timing capacitance 0.001 to 100 µFd. That applies to both monostable and astable modes. While the author hasn't confirmed this, the limit on accurate timing seems to be 100 KHz maximum in astable mode, 10 µSec minimum delay in one-shot mode.[15]

External DC Control of Timing

A 555's *control voltage* input (pin 5) can be used to vary the timing action in either one-shot or free-run modes, but with some cautions: The nominal voltage on pin 5 is centered on 2/3 of V_{CC} and it goes directly to the three-resistor string (5 KOhms each) shown on Figure 27-12. Applications data from manufacturers limit the maximum external voltage swing, peak-to-peak, on pin 5 at about 1/3 V_{CC}. This is the case of the one-shot-based pulse width modulator shown in Figure 27-15. Note: datasheet specifications are not given as to the linearity or accuracy of external voltages applied to pin 5; while it is called *control voltage*, that refers to the action of sensing the exponential charge-discharge waveforms by the two comparators.

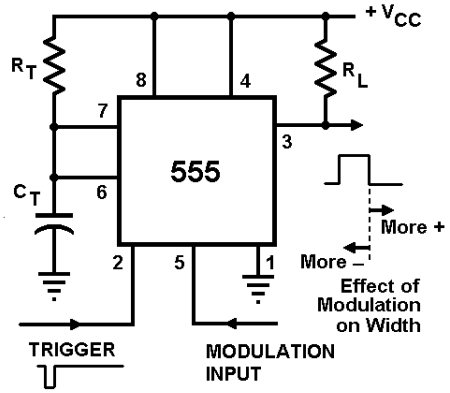


Figure 27-15 Pulse-width modulator

The circuit at right is from the ON Semiconductor datasheet for their MC1455. Their example has an external trigger of about 4 KHz repetition rate with R_T = 10 KOhm, C_T = 0.02 µFd, and V_{CC} = +15 VDC. At no modulation the pulse width would be about 220 µSec, making the output duty cycle about 88%. From the oscilloscope photos with the datasheet, the modulation appears to be ***negative-going*** and only a half cycle of a 100 Hz modulation sinewave shown.

---

[15] Based on datasheets and application notes. The bipolar-based 555 is not a speedy device (relative to 74AC and 74F logic families) but it isn't a slouch such as the early CMOS devices. Minimum one-shot pulse width of about 0.5 µSec is probably possible but the capacitor values will have to be trimmed on-the-bench. Little has been written about either the frequency response of the internal circuitry nor the stray capacity at the various pins that could yield clues on higher-frequency operation.

While that is an interesting application, it is not good for modulation linearity. The exponential ramp voltage across the timing capacitor will introduce distortion in the modulated *carrier*. If that ramp were linear rather than exponential, the distortion would be minimal and modulation voltage would be directly proportional to pulse width. That is possible in the circuit of Figure 27-16.

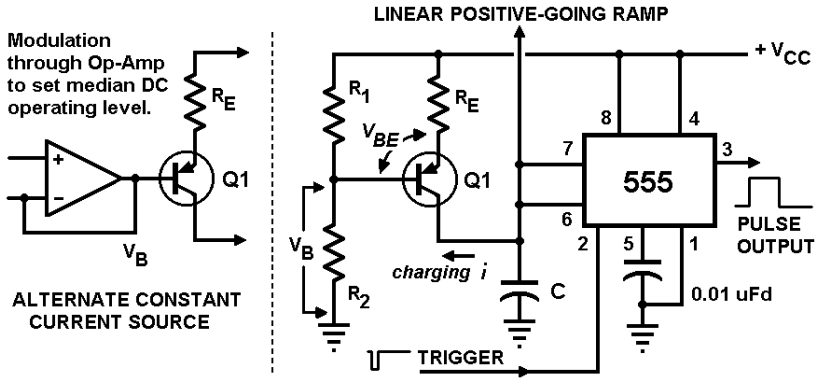


Figure 27-16 One-shot with timing resistor replaced by a constant-current source (Q1).

Every 555 connected as a monostable will have its on time coincident with a ramp voltage across the timing capacitor. That ramp begins at near-ground level and rises to 2/3 V_{CC} where it trips comparator A, turns on the discharge transistor (output on pin 7) and turns the output off; the timing capacitor is discharged quickly to ground level. In the simplest passive component version, the ramp is an exponential shape because capacitor charging current has to go through a series resistor. By replacing that timing resistor with a *constant-current* source as in Figure 27-16, the ramp becomes very linear in time with a constant rate-of-change of voltage with time. Q1 can be any high-h_{FE} PNP and its collector current is determined by:[16]

$$I = \frac{V_{CC} - V_B - V_{BE}}{R_E}$$

That form of constant-current generator works because the base-emitter voltage drop, V_{BE}, stays fairly constant at about 0.6 Volts. R1 and R2 provide a voltage-divider whose supply drain is at least 100 times the base current of Q1 (for stability). With a current gain of at least 100 and the fact that both collector and base current flow through the emitter, R_E can set the collector current if it has a good reference voltage on the Q1 base.

The alternate form on the left of Figure 27-16 is a bit better suited to apply modulation such as voice to control the one-shot width. The op-amp can be configured as a linear audio mixer and be able to set the quiescent, no-modulation output DC level and thus the constant current.

The amount of constant current can be found from equation (6-1) on page 6-1, repeated here:

$$\frac{V}{T} = \frac{I}{C} \qquad \text{or} \qquad I = \frac{V\,C}{T}$$

Where:

All units are Volts, Seconds, Amperes, and Farads

[16] From the ON Semiconductor MC1455 datasheet of February 2006, page 7.

In the case of the Figure 27-16 circuit, the *Volts* will be 2/3 of supply voltage. Assume a supply voltage of +15 VDC so the Volts will be 10. Given the 0.02 µFd timing capacitor and 220 µSec pulse width of the ON Semiconductor page 7 circuit, $C = 2 \bullet 10^{-8}$ and $T = 2.2 \bullet 10^{-4}$. The calculation is:

$$I = \frac{10 \bullet 2 \bullet 10^{-8}}{2.2 \bullet 10^{-4}} = \frac{2 \bullet 10^{-7}}{2.2 \bullet 10^{-4}} = 909.1\ \mu A$$

To determine the emitter resistor value, one has to keep in mind that the collector to ground voltage of Q1 must be at least 10 Volts (comparator A set-point for a 15 VDC supply) in order to keep any current flowing. Calculation can go in a different way by picking an approximate value of the emitter resistor as 2.2 KOhms. At 918.2 µA (includes base current of 1/100th collector current), the voltage drop across 2.2 K would be 2.02 Volts. The base voltage to ground would then have to be 2.62 Volts less than the supply voltage or +12.38 Volts. That is easily over the limit. R1 could be 2.2 KOhms and R2 13.5 KOhms (two 27 K resistors in parallel) with that voltage divider drawing about 1 mA from the supply.

In using the alternate version, the op-amp output voltage would be +12.38 VDC, directly into the base of Q1. That would result in 2.02 Volts across the emitter resistor and the 909.1 µA into the timing capacitor. As the modulating voltage swung more negative (at op-amp output), the constant current would decrease and the pulse width would lengthen. In swinging more positive the constant current would increase and width would shorten. Note that there are limits of the modulating voltage swing and those are determined by Q1 cutoff (maximum positive modulation swing) and Q1 saturation (maximum negative modulation). The latter would be in going below the +10 V comparator A set-point.

This form of timing control could be used with a front panel control calibrated linearly in time versus position. The voltage across the timing capacitor is a very linear ramp and could be used (via high-impedance coupling) for other things such as an oscilloscope horizontal sweep.

Pulse width modulation is used in DC-to-DC and AC-to-DC *switching* converters. DC (or rectified AC) is switched at rates in the range of 50 KHz to 2 MHz with the high frequencies filtered to take the average DC component as the output. Varying the pulse width allows control of the DC voltage and deviation from a reference voltage is taken as an error signal to change pulse width and thus keep the DC output voltage steady.

Voltage Controlled Oscillators [of the digital kind]

A Discrete-Component VCO

Figure 27-17, highly simplified, is based on the now-obsolete Motorola MC4024/MC4324 voltage-controlled astable multivibrator.[17] The externally-supplied control voltage is made to a pair

[17] From the Motorola datasheet for the MC4024 of 1973.

of constant current sources, i_1 and i_2, such that their current is proportional to the control voltage. Constant-current source i_3 does not change; current is such to supply either Q1 or Q2 conducting, but not both.

Assume Q1 is cut off. Q2 conducts and its collector voltage, coupled through emitter follower Q3, will keep the base of Q1 lower than its emitter. Since Q1 is cut off, its collector voltage will be high (little current through R1) and that, through emitter follower Q4, keeps the base of Q2 high and Q2 conducting. The amount of both emitter current through Q4 and base current of Q2 is held constant by constant current source i_2. Note that any constant current source has *large voltage compliance* and very high source impedance; the voltage across such a current source may vary widely without changing the current.

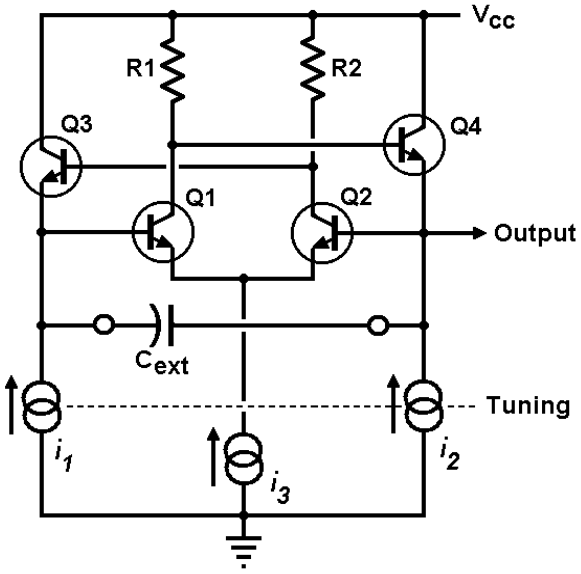


Figure 27-17 Simplified circuit of half of the MC4024 VCO to illustrate basic operation.

The tuning range is set by an external capacitor C_{ext} connected between the bases of Q1 and Q2. If Q1 base is low in voltage but Q2 base is high in voltage, C_{ext} will start charging to equalize its voltage, pulling charging current through i_1. Base voltage of Q1 will rise at approximately a linear rate. Eventually the Q1 base voltage will be high enough to cause it to conduct, dropping its collector voltage and the Q4-coupled voltage on Q2's base. Since both Q1 and Q2 emitters are to a common constant current source, Q1 conducting will take all current and Q2 will cut off. Q3 will conduct; its base going high from little voltage drop across R2. The switch over action is regenerative.

With C_{ext} voltage now reversed, it will begin charging to equalize voltage potential through i_2 until the base voltage on Q2 is high enough to cause it to conduct. The cycle repeats. If the transistors are reasonably matched and both constant-current sources also matched, the output will be close to a 50% duty-cycle squarewave. In applications, the MC4024 could easily span a 1:3 frequency change ratio by voltage control alone with a maximum operating frequency of 25 MHz at 25° C temperature. A 1:10, even 1:100 frequency ratio is possible although the maximum frequency will be lowered. Variable-capacitance voltage-tuned diodes can achieve only about a 1:2 frequency range tuning in L-C oscillators.[18]

The MC4024 and MC4324 had much more circuitry inside, including amplitude control and buffering to provide TTL compatible output. Figure 27-17 is a gross simplification. It does illustrate how a single external capacitance can set the oscillator's frequency range and not require a resonating inductor.

The TI Series of *TTL VCOs*

[18] Although a varactor diode capacitance has a capacity change ratio of about 1:5 maximum, the resonant frequency of an L-C circuit is determined by the *square root* of either L or C being variable.

These debuted in the 1970s as the 74x124 through 74x128. After some design changes those were discontinued and replaced by the 74x324 through 74x328 within a decade. By the end of the old millennium there were only four distinct types left in active production, mid-2006:

74LS624	Single VCO, optional internal 600 Ohm timing resistor, 20 MHz max.
74LS628	Single VCO, true and complementary outputs, 20 MHz maximum.
74LS629	Dual VCO, only true output state, 20 MHz maximum frequency.
74S124	Dual VCO, 40 to 85 MHz maximum (depending on control voltage)

These all have the practical advantage in not requiring any resonance-setting inductor, just a single fixed R and C with frequency tuning through a control voltage pin of 1.0 to 4.5 VDC. They can be designed-in just like a one-shot or lower-frequency astable. The voltage tuning range is 1:6 to 1:10 in frequency ratio.

Based on the *non-stock* information from distributors as of mid-2006, those are available only on large-quantity orders.[19] The expectation is that this series of TTL devices will soon become obsolete. With the greater availability of modular L-C-tuned VCOs, specified narrow tuning ranges, and newer monolithic ICs using *ring oscillators,* there is a lesser market need for the R-C-tuned VCOs having relatively wide tuning ratios at HF and above. This last series from TI is specified with low noise and good temperature stability, perhaps the ultimate in R-C voltage-tuned astables.

VCOs in *Function Generators*

A function generator IC is basically a free-running triangle generator with simultaneous squarewave and sinusoid waveform outputs, similar in functional interior circuitry to those in Figures 12, 16, and 17. Frequency is controlled by variable constant current sources alternately charging a single external capacitor. The squarewave output is derived from the internal flip-flop toggled by internal comparators. A sinewave is obtained by *shaping* the triangle waveform's peak excursions, rounding off the points in segments until the waveshape approaches a sinewave.[20]

As of mid-2006 available single-IC function generators were the XR2206, XR2211, and XR8038A from Exar, their maximum frequency at about 1 MHz.[21] Maxim-Dallas Semiconductor had the MAX038 topping out at 20 MHz but no longer *recommended for new designs.* All of those seem to have been used in ready-built and kit function generator test equipments for hobbyists and electronic service shops. These single-IC function generators have an advantage of automatically providing an oscilloscope sweep trigger at the start of any sine, triangle, or squarewave. The

---

[19] *Not in stock* is in reference to the device being listed in catalogs (along with prices) but the distributor does not normally stock them in their warehouse. With a large-enough order quantity in the near future they will then stock them for others ordering small quantities.

[20] Camenzind's *Designing Analog Chips* page 11-11 (which appears to be similar to Intersil's ICL8038 schematic in their 1998 datasheet) shows an array of bipolar transistor switches successively shunting a resistive voltage divider output leg to round off the triangle wave points. The Hewlett-Packard HP202 Function Generator (circa 1958) used biased diodes and successive resistive shunting in a similar manner, but with a higher-voltage triangular waveform derived from vacuum tube circuitry.

[21] The *8038* was originally from Harris Semiconductor, that acquired by Intersil, finally acquired by Exar.

sinewave output has a THD[22] or Total Harmonic Distortion of only about 1% up to 50 to 100 KHz, respectable purity. They can all tune over a 1:1000 frequency ratio by either a local voltage control or an external, applied voltage. Such a wide range allows *sweep frequency testing* or simulating FM.

Function generator ICs can function very well as voltage-controlled astable multivibrators with a choice of output waveshapes. Their only disadvantage is the maximum frequency for radio applications limiting their use to the final IF of 455 KHz or lower.

VCOs in PLL ICs

There are a number of specialty PLL ICs containing voltage-controlled oscillators, most of those having separate VCO outputs and phase-frequency-detector inputs. The most-available device was the descendant of the original *COS/MOS Micropower Phase Lock Loop*, the CD4046. In 74HC, 74HCT, 74AC, and 74ACT device types, the VCO tops out at about 18 MHz with a 5 VDC supply and includes a PFD or Phase Frequency Detector.

The Philips NE/SE 568A PLL combination has a VCO with a 150 MHz maximum frequency and the NE/SE 564 VCO has a 50 MHz maximum.[23] Both are useful as a 1st LO in an HF receiver or final mixing LO in a transmitter. The NE568A could be used as a VCO for a 6m transceiver but it would be pushing limits to use one at the amateur 2m band..

Return of the Ring Oscillator

The 1990s saw a resurgence of interest in the *ring oscillator* converted to monolithic form. A ring oscillator is a series of *odd* amplifier stages with the output connected back to the input for positive feedback. First used in radar transmitters of the later 1930s, their tuning was done by L-C resonances plus the delay due to *electron transit times* in vacuum tubes. In monolithic silicon form the delay is inherent in the junction structure (as it is with gates of logic devices) and that delay can be made variable by biasing the junctions from an external voltage source. Delay can be controlled in monolithic form and the external voltage control allows about a 1:4 frequency change ratio.

Most monolithic ring oscillators are a part of a PLL subsystem IC or complete transceiver on a chip. Three from Texas Instruments looked attractive in the mid-2006 time: TLC2932 (11 to 50 MHz tuning range), TLC2933 (37-100 MHz), and TLC2934 (36 to 130 MHz). All include a PFD and all have a *half-frequency* output available through an internal flip-flop. Except for the TLC2934 operating at a maximum supply voltage of 4 VDC, they will operate compatibly with 3.3 to 5.0 VDC supply voltages. As of mid-2006 TI had put up the notice that the TLC2934 was *NRND (Not Recommended for New Designs)*. The TLC2940, VCO only (28 to 75 MHz) has already become

---

[22] THD percentages indicate the sum total of **all** harmonics beyond the fundamental, referred to the RMS value of fundamental frequency. A good Wein Bridge audio oscillator has only about 0.25% THD.

[23] A circuit description of the insides of the NE564 is available in Philips appnote AN179, *Circuit Description of the NE564,* December 1991. This explains how Philips did their VCO control in detail, among other things.

obsolete.[24] The main drawback for hobbyists with these devices is the packaging, available only in the 0.050 inch lead spacing thin-shrink mountings.

ECL, PECL, and Others

The older Motorola MC1648 *VCO* design remains in production at ON Semiconductor, but only in a miniature, 0.050" lead-spacing package with part number MC100EL1648. While it was always advertised as a *VCO*, its frequency is variable ***only*** with an external voltage-variable-capacitance diode across the external L-C resonating circuit. This device is basically an assembly of transistors for good, low-noise oscillation up to 225 MHz plus a buffer to convert it to ECL logic levels. A stand-alone voltage-controlled-oscillator it is not.

Both ECL, Emitter Coupled Logic, and PECL, Positive (voltage supply) Emitter Coupled Logic, require some logic level translation to work with TTL. However, that is needed only for low frequencies. Higher frequencies can use capacitive coupling with Schmitt Trigger input gates to preserve the waveshape. The same thing can be done with discrete L-C oscillators, those allowing some flexibility in components and reducing dependency on continued logic device production.

Crystek carried a line of VCOs in SMD packages (only 4 connections) that run the gamut from 45 MHz to 5.5 GHz in 1:1.2 to 1:2.0 frequency range (depending on model). Most in their *CVCO55CL* series will operate at +5 VDC levels and the VCO control voltages are 1 to 10 VDC (again, depending on model). Their internal structure is not fully explained but these are assumed to be of the ring oscillator type. Their outputs are specified in dbm for a 50 Ohm system, typically less than +5 dbm. That requires amplification and shaping to work with any logic voltage levels; capacitive coupling will work since lower-than-minimum frequency components do not matter in any application. The *dash numbers* following the series specify minimum to maximum frequency of oscillation in MHz. The CVCO55CL-0045-0070 ranges 45 to 70 MHz and CVCO55CL-0100-0140 ranges 100 to 140 MHz. Cost may be prohibitive, ranging from $28 for the 45 to 70 MHz device to $41 for the 5.5 GHz device.[25] An obvious advantage is that they require ***no*** external discrete components to set frequency, only the control voltage.

---

[24] This can be a sign that a particular device series is on the way to obsolescence. The customer demand will eventually tell the story. Most of the ring oscillator VCOs were being used for the 315 to 1200 MHz low-power communications devices which were in a highly competitive consumer electronics market area in 1990s to 2000s.

[25] Mid-2006 distributor prices for quantities of 1 to 4.

Chapter 28

Position, Direction, and Conversion Logic

Digital circuitry is commonplace in contemporary radio tuning systems. Included is conversion of all-Binary to more human-understandable information, and vice-versa. As part of that are presented some rules on digital numbering systems and largely parallel forms of conversion using digital logic devices..

General

Position sensing is an important part of digital tuning. That is subdivided into two parts: Absolute position sensing and Sensing of Direction of Movement. Sensing direction of movement may need a counter to indicate the Up/increase and Down/decrease of movement. Absolute sensing of movement will probably need a form of digital interface to make it human-readable.

Most of the circuitry in here involves parallel hardware conversion along with the radix or numbering system of binary versus human decimal data. There is a lead-in to microprocessor or microcontroller conversion which is covered later.

Absolute Position Sensing

Absolute refers to ***all*** digital bits changing state relative to a position change of its input (or output) device. A problem with binary digital data is the mechanical sense of all possible bits changing to another state. An example is a 7 (binary 0111) changing to an 8 (binary 1000) or vice-versa. It is not reasonable to have three binary bit sensors all change at the same time to change the state of the 4th bit. An answer for that is ***Gray Code.***

Gray Code for the Sensor

Figure 28-1 shows a comparison of 4 bits of conventional Binary versus *Gray Code* Binary. Shading represents Logic 1 while clear represents Logic 0. The important thing about Gray Code is that ***one and only one bit changes state from any state position to any other state position.*** This is true for any number of digital bit positions, 2 through 64. It might be easier to read in the tabulation following, Table 28-1. In that Table the most significant bit is ***D*** and least-significant bit is ***A.***

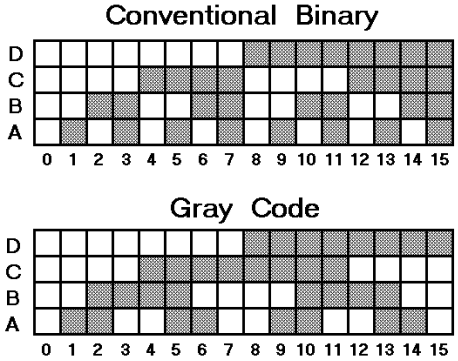
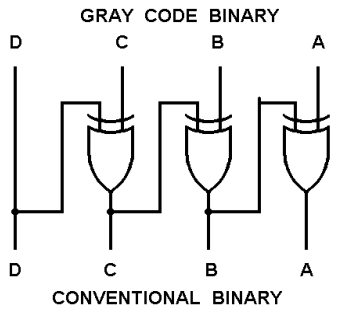


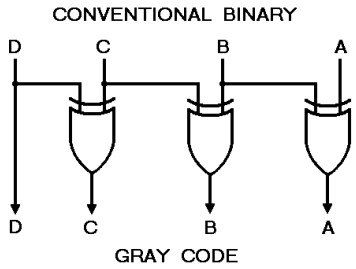
Figure 28-1 Pictorial Comparison of Binary versus Gray Code, 4 bits

Table 28-1 Tabulation of Conventional and Gray Code binary for Four Bits.

Conventional Binary				➜ ➜	Gray Code Binary			
D	C	B	A		D	C	B	A
0	0	0	0		0	0	0	0
0	0	0	1		0	0	0	1
0	0	1	0		0	0	1	1
0	0	1	1		0	0	1	0
0	1	0	0		0	1	1	0
0	1	0	1		0	1	1	1
0	1	1	0		0	1	0	1
0	1	1	1		0	1	0	0
1	0	0	0		1	1	0	0
1	0	0	1		1	1	0	1
1	0	1	0		1	1	1	1
1	0	1	1		1	1	1	0
1	1	0	0		1	0	1	0
1	1	0	1		1	0	1	1
1	1	1	0		1	0	0	1
1	1	1	1		1	0	0	0



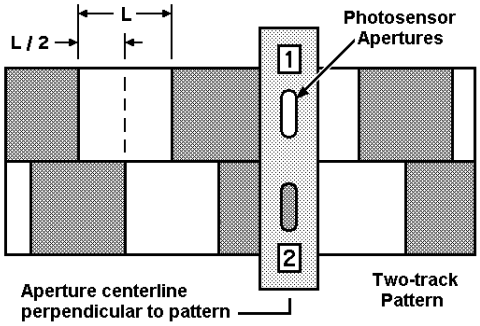
28-2 Gray Code to Conventional Binary Converter



28-3 Conventional to Gray Code Converter.

Conversion from Conventional to Gray or Gray to Conventional binary is quite easy using quad Exclusive-OR gates, a standard logic gate package. The only difference between the two types is that the Gray to Conventional takes one output from the preceding gate as an input. As to speed of response, ordinary 74HC family EXORs can switch in nanoSeconds; while there is a practical width converting from Gray to Conventional due to time build-up, it is still faster than a few micro-inches of mechanical dither possible in an all-conventional binary input-output.

Sensing Direction of Movement



28-4 A 2-track photosensor arrangement for direction sensing.

Direction sensing can be done with just two tracks in a modified Gray Code arrangement in an encoder, shown in 28-4. This is done in digital tuning for *up* and *down* incremental frequency changing, other circuitry doing the work of determining actual frequency. Two identical tracks can be arranged so that one track is exactly a quarter-period position apart from the other; in AC parlance, that is known as a 90-degree phase shift or one track in *quadrature* with the other.

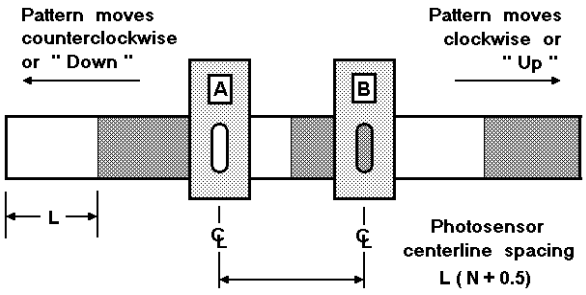
Two LED-photocell sensors, shielded from each other, pick up the track light and dark. This arrangement requires accurate alignment of the

photosensor apertures perpendicular to the track pattern. Modern rotary quadrature encoders dispense with two tracks in favor of the single-track arrangement shown in 28-5.

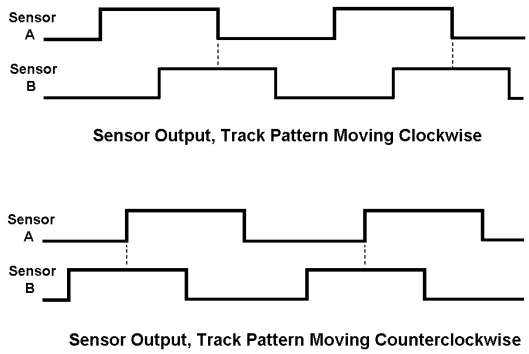
As a product, it is much easier to make a single track plus allowing some slight adjustment to position the two photosensors in quadrature.

What is even nicer, both the 2-track and 1-track encoders produce the same types of outputs when moved, shown in 28-6.

There are two ways to use the encoders. One way is to be able to differentiate which one came up to Logic 1 first, then remember it. It can be compared through an Exclusive-OR gate with the previous state from an internal memory. This will detect the ***edge of the first waveform*** to go up. Colloquially stated, such is called an ***edge detector.***



28-5 Evolution of 2-track encoder to the 1-track type, keeping photosensors apart.



28-6 Waveforms of photosensor output of either 2 or 1-track encoder.

Perhaps a simpler method is to use a dual D flip-flop as shown in Figure 28-7. This combines both the Exclusive-OR and memory features to output either direction of movement. Note the separate outputs for direction.

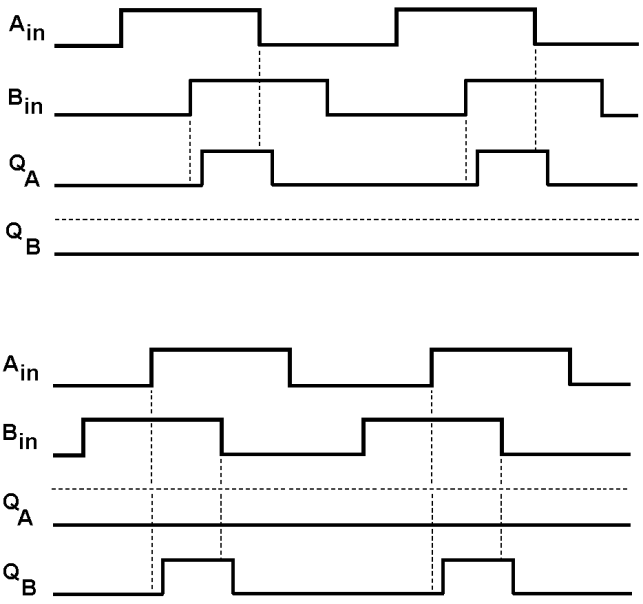


Figure 28-8 Waveforms of Figure 28-7 circuit for either direction of movement.

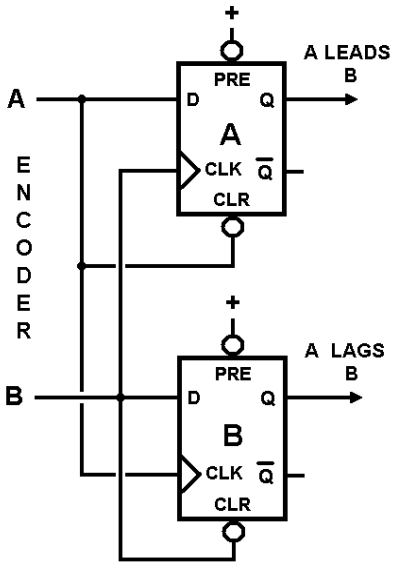


Figure 28-7 A *decoder* for either type of encoder using a single dual flip-flop.

Using Different Counter ICs

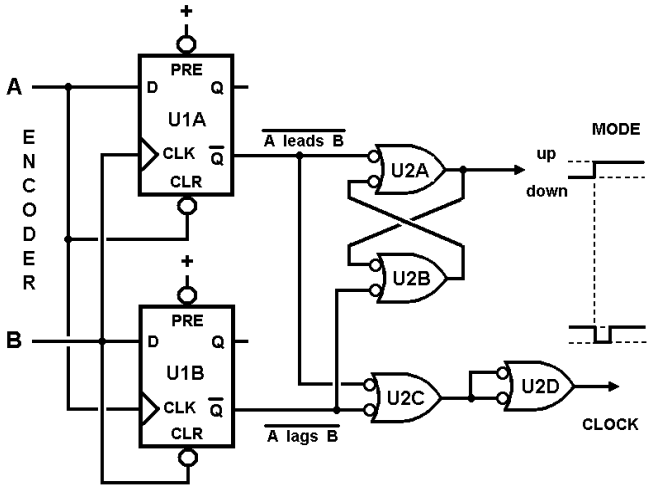


Figure 28-9 Modified output for certain other 4-bit counters such as the 74x190 or CD4510.

Any Up-Down counter will work with the simple encoder decoder of Figure 28-7. Where the Up and Down input is a 1-pin mode control requires a modification such as with Figure 28-9.

Figure 28-9 adds an R-S flip-flop (U2A and U2B) with input from the **Q-not** outputs of U1. U2 is a NAND gate, therefore an R-S FF would require active-low inputs. A single 2-input quad NAND gate IC can do the modified counter input mode and count control.

Increasing Resolution

The decoder of Figure 28-7 will produce one direction pulse for every cycle of the encode track. It's possible to provide a direction pulse for ***every*** transition; i.e., four transitions per cycle instead of just one. Consider the following table showing the A and B encoder output states for each direction of movement:

From→To States of AB Pairs During Quarter-Cycle Movements

Track A leading B	Track A lagging B
00 → 10	00 → 01
10 → 11	01 → 11
11 → 01	11 → 10
01 → 00	10 → 00

There are only four AB state combinations but there are eight combinations of from-to changes. In order to determine direction of track movement, the decoder must compare the just-prior AB state with the new AB state condition. That comparison requires some form of *memory* of states. In the circuit of Figure 28-10 the temporary memory is obtained by the R-C pair feeding the Schmitt trigger inverters, U1C and U1D. Exclusive-OR gates U2A and U2B effectively compare the state of encoder track B with the *previous* state of track A (U2A) and the state of encoder track A with the *previous* state of track B (U2B). That *comparison* is transitory, happening only during the delay caused by the R-C networks. That must happen before the Schmitt trigger inputs have reached their transition voltage levels. A transitory memory here is quite sufficient. The decoder need only act on the *beginning* of any track state change. No bistable memory devices are required.

The remaining part of the direction-determination is more subtle. U3A ORs the output of U2A and inverted output of U2B (from U1F) to obtain the negative-going output pulses for track A phase lagging phase B. U3B does the inverse for track A phase leading track B. The timing

waveform diagram of Figure 28-11 should be consulted to see this relationship.

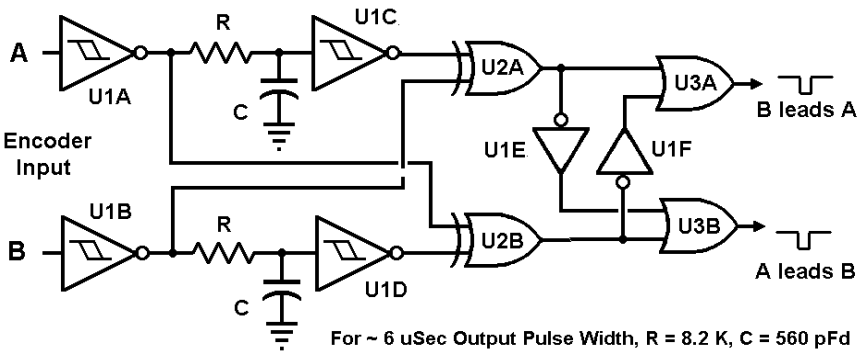


Figure 28-10 High-resolution decoder by Dr. Robert Dennis.

Negative-going pulses only appear at one or the other outputs, never both together. For any stable position of the encoder track, both outputs remain at a quiescent high state. NAND gates can be used for U3 rather than the NORs, no change in state or waveform conditions. Inverters U1E and U1F can be part of a hex Schmitt trigger input inverter package required for U1A through U1D or two inverters from another hex inverter package.

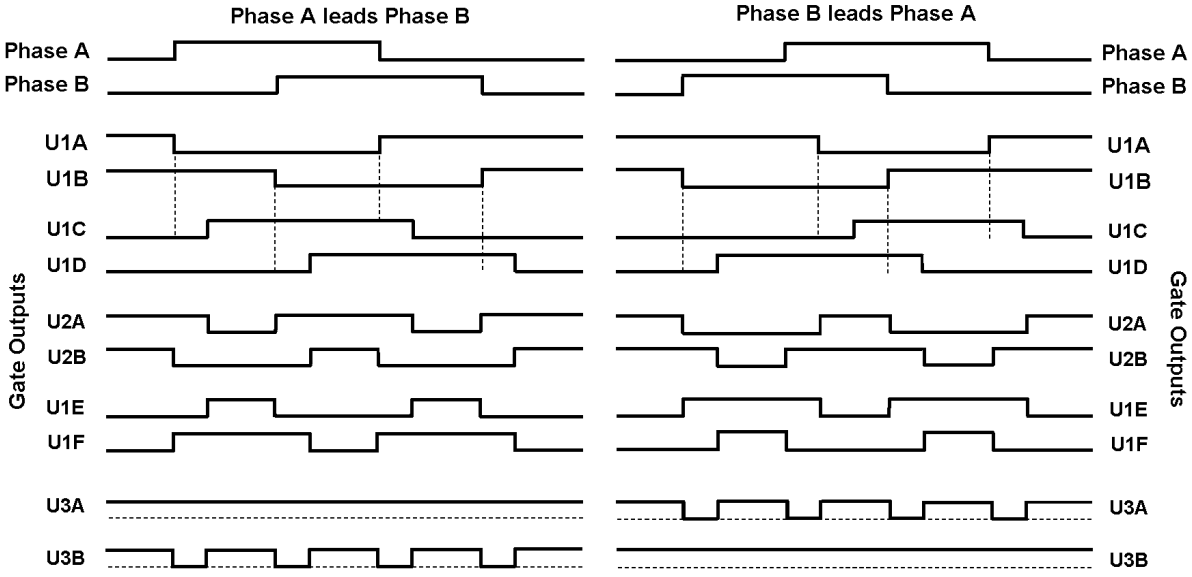


Figure 28-11 Waveforms of the Dennis Decoder circuit, each movement direction.

Adding the three-fourths of a quad 2-input NAND to the Dennis Decoder modifies the output to Mode-and-Clock counters, similar to Figure 28-9. U3C and U3D form an R-S flip-flop whose output will be the encoder track direction. If a slight delay in the clock pulses is desired, add the fourth 2-input NAND connected as an inverter to the clock pulse output. The positive-going edge of that clock pulse occurs at the end of the delay period of the R-C network in Figure 28-4. This allows a subsequent Up-Down counter to change count mode before any opposite-direction count begins.

Maximum Encoder Rates

Manual control encoders are essentially slow-speed devices compared to logic speeds. The

rate of change is dependent on the expected maximum rotation speed and the number of cycles on an encoder track. For a rotary control, an expected maximum rotation rate is probably about 300 RPM. The is equal to 5 revolutions per second. If the encoder track has 100 cycles on it, the rate of change of states is 500 Hz or a period of 2 mSec. A 10 or even 50 µSec delay in the Dennis Decoder would be quite acceptable.

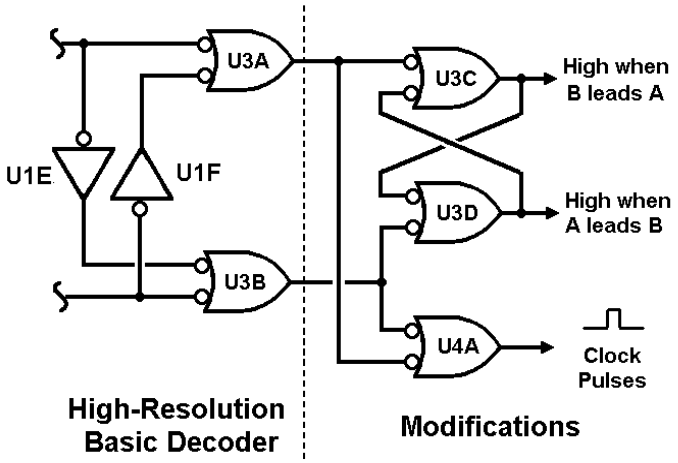


Figure 28-12 Dennis Decoder modifications

NUMERIC RADIX CONVERSION

Humans reckon in decimal numbers but digital logic works in binary. In order to work with binary logic, we humans need some ***radix or number base conversion*** circuits to relate decimals with binary bits. As a preface to that, consider again the *Bit Weights* in any binary word shown following.

Table 28-2 Weights of Bits in a Binary Word

Bit	Weight	Bit	Weight	Bit	Weight	Bit	Weight
1	1	9	256	17	65,536	25	16,777,216
2	2	10	512	18	131,072	26	33,554,432
3	4	11	1024	19	262,144	27	67,108,864
4	8	12	2048	20	524,288	28	134,217,728
5	16	13	4096	21	1,048,576	29	268,435,456
6	32	14	8192	22	2,097,152	30	536,870,912
7	64	15	16,384	23	4,194,304	31	1,073,741,824
8	128	16	32,768	24	8,388,608	32	2,147,483,648

The *weight* of any bit is that bit's decimal equivalent. To find the decimal value of any binary word, just add the *weights* of all bits at logic 1. For example, a binary 0111 would be 4 + 2 + 1 = 7. Another example is binary 110 0110 which would have a decimal value of 64 + 32 + 4 + 2 = 102.

In the beginning of computing, there were a great many different *weights* of bits. Many were selected or devised to fit the input-output devices of the day in the late 1940s to 1960. Eventually

those were all brought down to the simple power-of-twos incrementation shown following.

Table 28-3 Decimal to Binary Conversion

Decimal	V	U	T	S	R	Q	P	N	M	L	K	J	H	G	F	E	D	C	B	A
1																				1
2																			1	0
4																		1	0	0
8																	1	0	0	0
9																	1	0	0	1
10																	1	0	1	0
20																1	0	1	0	0
40															1	0	1	0	0	0
80														1	0	1	0	0	0	0
90														1	0	1	1	0	1	0
100														1	1	0	0	1	0	0
200													1	1	0	0	1	0	0	0
400												1	1	0	0	1	0	0	0	0
800											1	1	0	0	1	0	0	0	0	0
900											1	1	1	0	0	0	0	1	0	0
1000											1	1	1	1	1	0	1	0	0	0
2000										1	1	1	1	1	0	1	0	0	0	0
4000									1	1	1	1	1	0	1	0	0	0	0	0
8000								1	1	1	1	1	0	1	0	0	0	0	0	0
9000							1	0	0	0	1	1	0	0	1	0	1	0	0	0
10000							1	0	0	1	1	1	0	0	0	1	0	0	0	0
20000						1	0	0	1	1	1	0	0	0	1	0	0	0	0	0
40000					1	0	0	1	1	1	0	0	0	1	0	0	0	0	0	0
80000				1	0	0	1	1	1	0	0	0	1	0	0	0	0	0	0	0
90000				1	0	1	0	1	1	1	1	1	1	0	0	1	0	0	0	0
100000				1	1	0	0	0	0	1	1	0	1	0	1	0	0	0	0	0
200000			1	1	0	0	0	0	1	1	0	1	0	1	0	0	0	0	0	0
400000		1	1	0	0	0	0	1	1	0	1	0	1	0	0	0	0	0	0	0
800000	1	1	0	0	0	0	1	1	0	1	0	1	0	0	0	0	0	0	0	0
900000	1	1	0	1	1	0	1	1	1	0	1	1	1	0	1	0	0	0	0	0
	V	U	T	S	R	Q	P	N	M	L	K	J	H	G	F	E	D	C	B	A

Note: The alphabetic designation of binary bits is arbitrary; there is no standard practice on this.

Simple Decimal-to-BCD Conversion

The NAND gates of Figure 28-13 can convert a rotary switch decimal position to 4-bit parallel binary. Nine pull-up resistors are needed. The switch rotor contact grounding the NAND inputs will create the logic 1 states required in output (a NAND with active-low inputs is an OR).

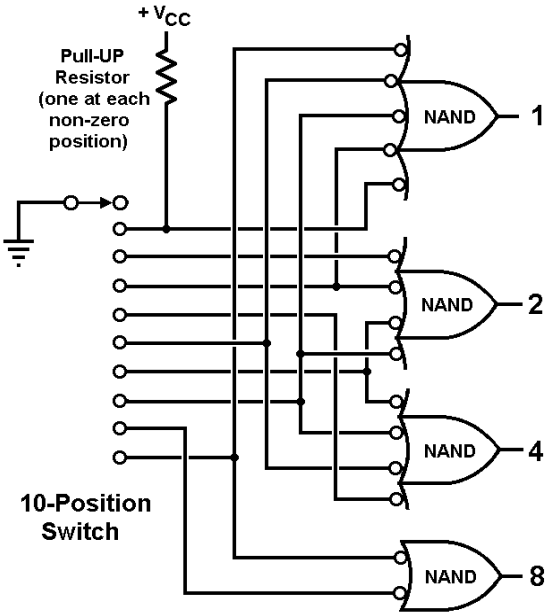


Figure 28-13 Simple 1-pole rotary switch controlling BCD output.

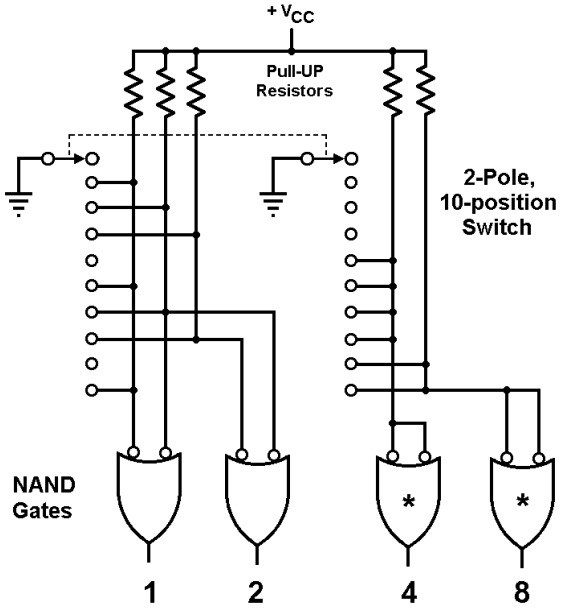


Figure 28-14 Two-Pole rotary switch to BCD output using less interwiring. Inverters can substitute for * markings.

If a two-pole switch is available, the simpler circuit of Figure 28-14 can be used. Note that the Figure 28-14 circuit needs only five pull-up resistors.

This use of NANDs as active-low OR gates can be adapted for almost any format of binary coding from rotary or linear switch contacts. Note that any logic 0 at any input of a NAND will cause the gate output to go logic 1. No software skills are needed for such innovations.

Multiple Decimal Digit to Binary Conversion in Hardware Logic

Parallel conversion of multiple BCD digits to a single binary word can be done with 4-bit Adders such as the 74x83, 74x283, or the CMOS CD4008. Those have four binary input pairs (for augend and addend), a *carry-in* input from the lesser 4-bit *carry-out*, and a carry-out to the next more significant 4-bit adder (or output).

Following Table 28-3, the bits can be ***added*** via binary adders to produce the final ***binary*** result. That is a rather simple process and can be worked out by pencil and paper, without any circuit breadboarding or power-on testing of any kind.[1] By illustration, the simple tabulation following allows binary addition to form the complete binary word for any tens and units inputs for a decimal equivalent of 0 to 99.

[1] Binary arithmetic rules need to be followed. That is fairly easy to learn.

	Decimal	Binary Bits G	F	E	D	C	B	A
Units selector	1	0	0	0	0	0	0	1
switch BCD	2	0	0	0	0	0	1	0
outputs	4	0	0	0	0	1	0	0
Logic 1	8	0	0	0	1	0	0	0
Tens selector	10	0	0	0	1	0	1	0
switch BCD	20	0	0	1	0	1	0	0
outputs at	40	0	1	0	1	0	0	0
Logic 1	80	1	0	1	0	0	0	0
Maximum switch state	99	1	1	0	0	0	1	1
Number of adder inputs, total		1	1	2	3	2	2	1

Note that binary bit A is never changed in conversion from BCD to binary (or vice-versa). It needs no hardware additions and can be wired around all devices. Note also that binary 10 is really a binary 5 (0101) that is *left-shifted* once. A left-shift of binary bits is equal to doubling the binary word value. A right-shift would be the same as halving the binary word value, division by two with the result remaining an integer (no fractional value, truncation).

To get any binary word value from 0 to 99, all that is required is to add the binary equivalents of the applicable four decimal values in any digit. A binary 7 is 0001 plus 0010 plus 0100 to equal 0111. A binary 70 would be 100 0110 from the addition of 000 1010, 001 0100, and 010 1000. Adding the binary 7 with binary 70 would result in decimal 77 which would be 100 1101 binary. This general scheme can be expanded to many decimal digits but will run into difficulty with the many bit additions necessary with large decimal numbers. Notice that bit D above will have three sources of addition, decimals 8, 10, and 40, while other have only 2 or 1. See Table 28-2.

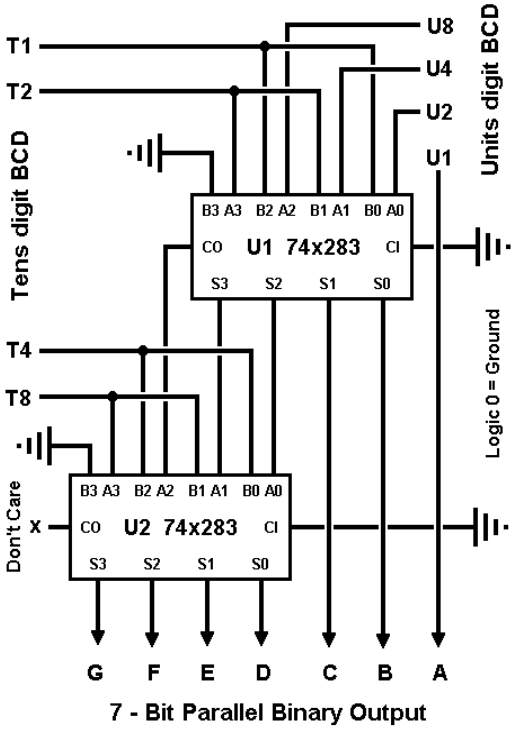


Figure 28-15 Two IC, two-digit BCD-binary Converter for decimal 0 to 99.

One way to convert a 2-decimal-digit input to 7-bit binary is shown in Figure 28-15. Note that the carry-output of the upper adder becomes an F-bit input to the lower adder. A maximum binary state from the Tens selector will have a 0-state F-bit if taken by itself. Adding a binary-9 from the Units selector results in 11 0111 (39) where a carry has resulted in an F-bit state of 1. There is no connection to the carry-out from the lower adder. No carries exist. The maximum decimal selection would be 99 and bit-G would be the highest bit set to a 1-state. Bit A does not need any conversion..

The tiered arrangement of adders can be expanded in width as desired. See the small table following. Note that bit-D in the truth table requires 3 inputs (U8, T1, T4) while bits B, C, E, and

F need only 2 inputs. A fourth bit-D input could be handled by the carry-in of the lower adder. No carries exist. The maximum decimal selection would be 99 and bit-G would be the highest bit set to a 1-state. Bit A does not need any conversion..

The tiered arrangement of adders can be expanded in width as desired. See the small table following. Note that bit-D in the truth table requires 3 inputs (U8, T1, T4) while bits B, C, E, and F need only 2 inputs. A fourth bit-D input could be handled by the carry-in of the lower adder.

		Binary Bits	
	Decimal	K J H G F E D C B A	
	99	0 0 0 1 1 0 0 0 1 1	← Maximum from Units and Tens
Hundreds	100	0 0 0 1 1 0 0 1 0 0	
switch BCD	200	0 0 1 1 0 0 1 0 0 0	
output states	400	0 1 1 0 0 1 0 0 0 0	
at logic 1:	800	1 1 0 0 1 0 0 0 0 0	
	999	1 1 1 1 1 0 0 1 1 1	← Maximum output binary state

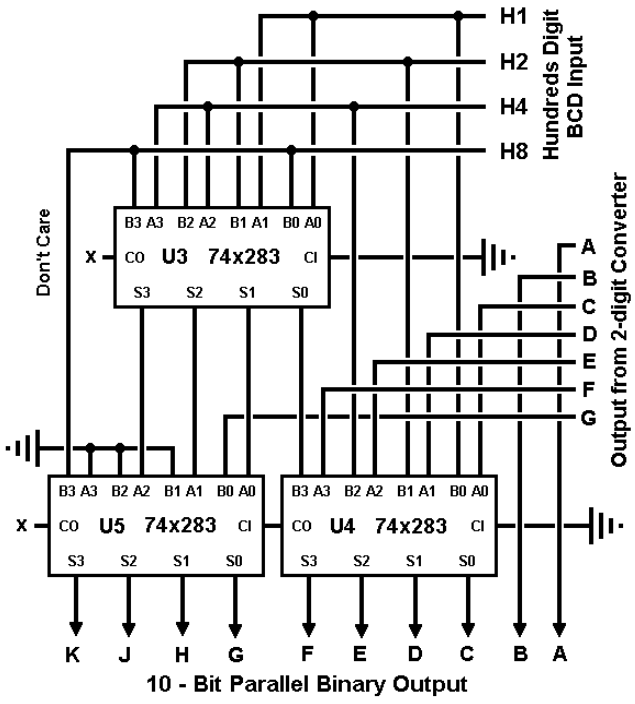


Figure 28-16 Adder expansion to 3 decimal digits

Rather than using a tiered structure, Figure 28-16 uses one 4-bit adder to form nearly all of the Hundreds selector binary states (U3). The output of U3 is then summed with the output of the adder-converter of Figure 28-15 in U4 and U5. A hidden advantage here is that the Units-Tens entire selector and logic circuit may be checked separately in case of a suspected problem.

Both of the carry-outs on the left are unconnected since there is no input combination state combination that results in a carry-out from either adder.. The maximum possible binary value equal to decimal 999 just fits ten bits width.

Binary bits A and B of the output are changed only by the Units-Tens encoder. Those bits are always 0 for any Hundreds BCD conversion. For that reason individual bit addition can begin at bit-C position and go higher.

Adding More BCD Inputs

It is possible to increase the adder size to handle 5 digits plus an offset. But, by then the hardware has grown cumbersome. The need for speed in parallel-handling logic is not always there. There should be a better way.

Binary to Decimal Conversion

This is a straightforward logic procedure if the number of binary bits is no greater than four. A conversion then consists of testing for bit state combinations equal to binary 10 (1010) through binary 15 (1111). If those exist a binary 6 (0110) is added; the result is the correct 8-4-2-1 BCD. The adder will produce a carry from that addition. That carry is a ***decimal carry*** and pertains only to the conversion from binary to decimal. To illustrate the conversion, assume a binary 15 state:

	1111	binary 15
	0110	+6
0001	0101	sum is a decimal 1 and a decimal 5 in separate 4-bit BCD digits

What slows most others down is getting binary inputs up to 8 bits. To handle all of those state combinations will require a decimal output up to 255 (3 digits). Handling 10 bits of binary input would yield a full 3-digit output but at a cost of rather a large amount of hardware. The obvious hardware alternative would be a microcontroller doing the conversion in software, the result placed in shift registers for static storage between conversion operations. While that is a time-dynamic conversion system, the time can be adjusted by software and the microcontroller clock rate.[2] An obvious advantage in hardware is that the microcontroller can be programmed to do many other tasks in between the few tens or hundreds of microseconds needed for a single conversion.

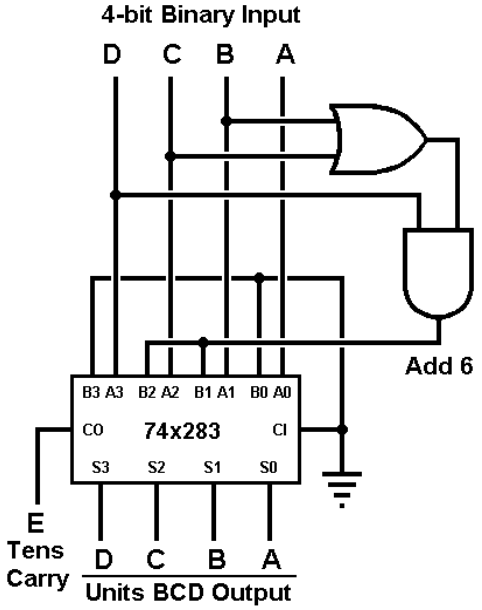


Figure 28-17 4-Bit Binary to BCD Converter using a 4-bit Adder and two gates.

Serial Radix Conversion

Both BCD-to-binary and binary-to-BCD conversion can be done *serially*. That means all input and output is handled sequentially for each bit or groups of four bits of both input and output data. While that can be done in available digital hardware, the low cost and small size (including ancillary components) of microprocessors and microcontrollers make their resulting tiny circuit board real estate requirements more attractive. While processor programming is a design area by itself, the input-output of particular processors will impact the other digital hardware and all must be tailored to fit. BCD-to-binary conversion is nearly the exact opposite in function of binary-to-BCD conversion.

---

[2] With microcontrollers (microprocessors with extra function circuitry on the same IC) of the new millennium operating at clock speeds of greater than 1 microsecond per single instruction cycle, conversion tasks can be done quickly. Software design is left up to the programmer's innovation. The obvious spectre for low-level-input receivers is the extra RFI generated by older microprocessors and their ancillary circuitry by periodic polling and input-output data handling.

Binary to BCD

The ***shift-left-and-add-three*** method is shown following using a maximum-value (65,535) 16-bit binary word. There must be exactly as many shifts as there are bits in the binary word. Any 4-bit group exceeding 4 (0100) gets 3 (0011) added to it.

BCD Register	Binary Register	
0000 0000 0000 0000 0000	1111 1111 1111 1111	Start
1	1111 1111 1111 1110	←Shift (1st)
11	1111 1111 1111 1100	←Shift (2nd)
111	1111 1111 1111 1000	←Shift (3rd)
+ 11		Add 3
1010	1111 1111 1111 1000	
1 0101	1111 1111 1111 0000	←Shift (4th)
+ 11		Add 3
1 1000	1111 1111 1111 0000	
11 0001	1111 1111 1110	←Shift (5th)
110 0011	1111 1111 1100	←Shift (6th)
+ 11		Add 3
1001 0011	1111 1111 1100	
1 0010 0111	1111 1111 1000	←Shift (7th)
+ 11		Add 3
1 0010 1010	1111 1111 1000	
10 0101 0101	1111 1111 0000	←Shift (8th)
+ 11 + 11		Add 3
10 1000 1000	1111 1111	
101 0001 0001	1111 1110	←Shift (9th)
+ 11		Add 3
1000 0001 0001	1111 1110	
1 0000 0010 0011	1111 1100	←Shift (10th)
10 0000 0100 0111	1111 1000	←Shift (11th)
+ 11		Add 3
10 0000 0100 1010	1111 1000	
100 0000 1001 0101	1111 0000	←Shift (12th)
+ 11 + 11		Add 3
100 0000 1100 1000	1111 0000	
1000 0001 1001 0001	1110	←Shift (13th)
+ 11 + 11		Add 3
1011 0001 1100 0001	1110	
1 0110 0011 1000 0011	1100	←Shift (14th)
+ 11 + 11		Add 3
1 1001 0011 1011 0011	1100	
11 0010 0111 0110 0111	1000	←Shift (15th)
+ 11 + 11 + 11		Add 3
11 0010 1010 1001 1010	1000	
110 0101 0101 0011 0101	0	←Shift (16th)
6 5 5 3 5	← BCD result	

BCD to Binary Conversion

BCD-to-Binary converts in the opposite way, ***shift-right-and-subtract-three.*** After each shift right, each four-bit group is examined. If it greater than binary 6 (0110), subtract binary 3 (0011) from it, ignoring any borrow (there will be none), then continue shifting. An example is the opposite of the previous, converting packed BCD of decimal 65,535 into 16-bit binary:

BCD Register					Binary Register	
6	5	5	3	5		**Start**
110	0101	0101	0011	0101	0000 0000 0000 0000	
11	0010	1010	1001	1010	1	**Shift➛ (1st)**
		- 11	- 11	- 11		**Subtract 3**
11	0010	0111	0110	0111	1	
1	1001	0011	1011	0011	11	**Shift→ (2nd)**
	- 11		- 11			**Subtract 3**
1	0110	0011	1000	0011	11	
	1011	0001	1100	0001	111	**Shift→ (3rd)**
	- 11		- 11			**Subtract 3**
	1000	0001	1001	0001	111	
	100	0000	1100	1000	1111	**Shift→ (4th)**
			- 11	- 11		**Subtract 3**
	100	0000	1001	0101	1111	
	10	0000	0100	1010	1111 1	**Shift→ (5th)**
				- 11		**Subtract 3**
	10	0000	0100	0111	1111 1	
	1	0000	0010	0011	1111 11	**Shift→ (6th)**
		1000	0001	0001	1111 111	**Shift→ (7th)**
		- 11				**Subtract 3**
		0101	0001	0001	1111 111	
		10	1000	1000	1111 1111	**Shift→ (8th)**
			- 11	- 11		**Subtract 3**
		10	0101	0101	1111 1111	
		1	0010	1010	1111 1111 1	**Shift→ (9th)**
				- 11		**Subtract 3**
		1	0010	0111	1111 1111 1	
			1001	0011	1111 1111 11	**Shift→ (10th)**
			- 11			**Subtract 3**
			0110	0011	1111 1111 11	
			11	0001	1111 1111 111	**Shift→ (11th)**
			1	1000	1111 1111 1111	**Shift→ (12th)**
				- 11		**Subtract 3**
			1	0101	1111 1111 1111	
				1010	1111 1111 1111 1	**Shift→ (13th)**
				- 11		**Subtract 3**
				0111	1111 1111 1111 1	
				0011	1111 1111 1111 11	**Shift→ (14th)**
				0001	1111 1111 1111 111	**Shift→ (15th)**
				0000	1111 1111 1111 1111	**Shift→ (16th)**
					Binary 65,535	

Cautions on Maximum Values

One has to stay ***within*** the maximum values of both BCD and Binary registers. As an example, suppose one has to convert 77,777 in BCD (0111 0111 0111 0111 0111) to Binary but has

only 16 shifts. In going through the exercise, after 16 shifts, the Binary register will contain a state of 0010 1111 1101 0001 but there is still one Logic 1 in the BCD register. The decimal value of the Binary register is 12,241. The 17th bit (not shifted out) has a weight of 65,536. When added to 12,241 the total decimal value of 17 bits is 77,777 which is correct.

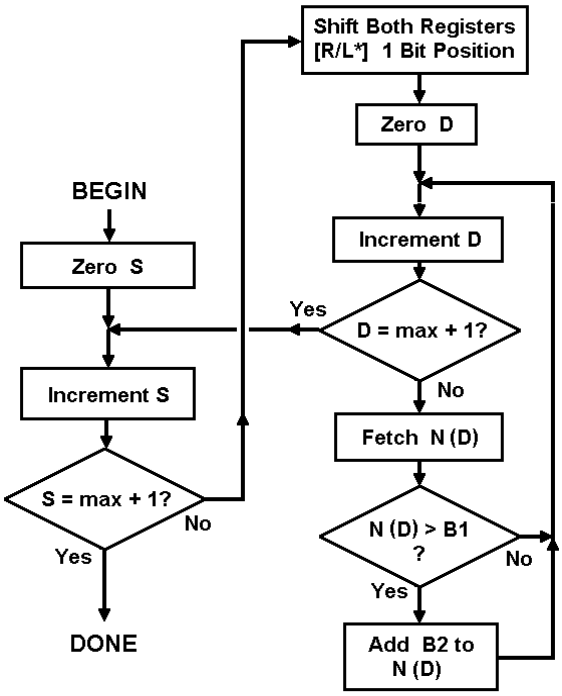
Rules apply. For either type of conversion, the number of stages in the Binary register will set the maximum number of shifts. At the end of those shifts the BCD register will be entirely Logic 0 for BCD-to-Binary conversion or the Binary register will be entirely Logic 0 for Binary-to-BCD conversion.

Microprocessor Conversion Method

Algorithms for either BCD-to-Binary or Binary-to-BCD conversion are so similar that one generalized flow chart Figure 28-20, serves for both. The process for either is a sequence of one-bit shift followed by examination of all the four-bit BCD digit binary values, that repeated until the maximum number of shifts is reached.

Figure 28-20 assumes two integer counters, S and D, to control the iterations. D is used as a *dimension* for the virtual array of N four-bit BCD numeric values in the BCD register.[3] The maximum value for S would be the number of bits in the Binary Register. The two binary constants, B1 and B2, assume addition in either conversion. *Two's complement* of binary 0011 allows the equivalent of subtraction by 3 in addition of the complement. The process follows the pencil-and-paper illustrations in the previous pages.

This flow chart is a general guide to writing the source code and it must be tailored to fit the various processor instruction sets. In early Texas Instrument microprocessors, the CPU (Central Processor Unit) could work directly with RAM locations and addition (iterative loop on the right) would result in a direct write back into memory when finished; other microprocessors have to use a separate *write to memory* instruction in order to keep the sum. Instead of incrementing S and D, they could be decremented (by one) after initial setting to their maximum-plus-one; decrementation would stop when the counters reached zero. In Figure 28-20 the incrementing follows the left-to-right convention of ordering the N ( ) BCD digit variables. In either conversion, the LSB of the BCD



S = Shift Counter D = Digit Counter
N (D) = 4-bit BCD Digit @ Position D

Constants Determining Type of Conversion

	BCD-to-Binary	Binary-to-BCD
R/L*	RIGHT	LEFT
B1 :	0111	0100
B2 :	1101	0011

Figure 28-18 A generic program flow chart for Binary/BCD conversion.

[3] *Dimension* is common to most high-level computer languages. In here the BCD Register is treated as the equivalent of an *array* of four-bit variables named *N*. When D = 1, N(D) is the most-significant BCD digit. The virtual array, N( ), should be in four-bit increments; for a 16-bit binary result, maximum D would be 5 and N(D) would then be the least-significant BCD digit.

Register must shift into the MSB position of the Binary Register (BCD-to-Binary) or the MSB of the Binary Register shifts into the LSB of the BCD Register (Binary-to-BCD).

Firmware[4] Replacement for Both Methods

EPROMs in the 1 Megabit capacity sizes were approximately $5.00 per Megabit in mid-2006 for single-quantity prices.[5] Rather than use a lot of digital logic devices taking up PCB real estate for a single purpose of conversion, a limited-range converter could be *burned into* an EPROM.[6] The address of the EPROM becomes the input data to be converted and the data output becomes the converted data. Range of necessary conversion is limited only by the various sizes and configurations of available EPROMs. The advantage is that conversion time of the EPROM data conversion is a single output read time, typically in the 40 to 250 nSec region. Devices range from the old *2716* 2048-word by 8-bit EPROM to the 27C1024 64K by 16-bit 40-pin package device.[7] In terms of PCB layout, the 40-pin 27C1024 takes up about six 16-pin DIP ICs.

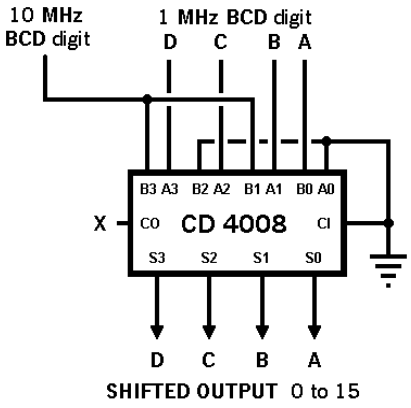


Figure 28-19 EPROM Address Shifter

In the author's case, a programmed 27C1024 replaced much conversion circuitry, dropping PCB space from 12 DIP spaces to only 6.[8] There was a problem to overcome and that lay in the ***17-bit input*** from the *4 ½-digit BCD* of the tuning counter. The tuning counter output states would range from 4900 to 15900 (4.9 to 15.9 MHz in 1 KHz increments) and there were 5 bits for the MHz digits. The 27C1024 EPROM addressing input is limited to 16 bits (65,536 binary words). The actual number of converted binary words needed were less than 11,000 and the binary data output would contain the 16-bit data correct for setting a PLL divider having an automatic IF offset programmed into it.

Some *Hidden* Capabilities of the EPROM Radix Converter

---

[4] *Firmware* is the colloquial expression for fixed memory values, such as from a PROM or Programmable Read-Only Memory.

[5] Most ultra-violet-erasable EPROMs from 16 Kbit to 1 Mbit capacity were about the same price for singles in that same year. Less applicable tax and shipping charges. Several manufacturers were selling all varieties.

[6] The correct term is *programmed into* rather than the colloquial *burned in.* The colloquial term came from some early programmable devices set by literally burning-open fusible links on the internal chip to set the data. Fusible links have become obsolete for *one-time-programmed* PROMs, replaced by ultra-violet-erasable EPROMs and *Flash* RAM. For hobbyists, the UV-erased EPROM is convenient in that the same device can be erased if the data needs to be changed.

[7] The venerable Intel 2716 became popular among computer hobbyists in 1980 and survived for better than a quarter century afterwards.

[8] Had that EPROM been available in a 0.3 inch pin row spacing package, the real estate would require only 3 DIP spaces. As it was, the 0.6 inch pin row spacing needed six spaces. More on that project is found in the finished project area of this book.

Keep in mind the fact that the LSB of the 16-bit output BCD-to-Binary converter doesn't change state. The LSB of the least-significant BCD digit still goes to the EPROM address but the EPROM data output LSB ***doesn't have to be used for the conversion.*** That EPROM data output LSB can remain at logic 1 except for other, specific uses. In the author's case, the data output LSB was programmed to logic 0 only at certain tuning frequencies for *presetting the tuning counter*. In this manner, the tuning could skip unwanted frequencies and go straight to other bands' edges, forming an automatic bandswitching scheme. The LSB of the EPROM *data output* could become a *mode bit* and used for the tuning counter's preset load input.[9] The normal data output to the PLL divider will see a fractional microSecond of disturbance but it will be perturbed anyway by the tuning's change to another frequency. The PLL is commanded to change frequency in any event.

In examining the *tuning counter's* frequency presetting, there is no need to preset to 1 KHz increments, at best, just 10 KHz increments. The least-significant four EPROM data outputs will have three bits free for other purposes (the LSB is the new *mode bit* and therefore used). One of those can be the band-change 10 MHz BCD digit bit. In this case, there isn't any need to do any *unshifting* of a four-bit 1 MHz preset data; that data output can go directly towards presetting the 1 MHz BCD digit without alteration.

There is another possibility of freeing up more EPROM data output bits on examining the range of binary bit ranges needed by the PLL divider. Looking at the extreme ends of the converted BCD input versus the necessary PLL binary divider states:

		PLL Divider Division Bits															
Tuning, KHz	PLL, KHz	R	Q	P	N	M	L	K	J	H	G	F	E	D	C	B	A
4980	26380	0	1	1	0	0	1	1	0	1	0	1	1	1	1	0	0
15840	37210	1	0	0	1	0	0	1	0	0	1	0	1	1	0	1	0

Bits R, Q, and P together have only two state combinations over the entire tuning range, 011 or 100. That data could be compressed into a single EPROM data output bit and that output hardwired with an inverter so that it produces the necessary divider division bits. That frees up two data output bits useful for other purposes, such as command lines for an external, relay-switched antenna matching network.[10] Those command lines would be active at any non-tuning-counter-preset frequency or essentially the vast majority of all frequencies tuned.[11]

Note that bit A in the tabulation is shown for completeness. The LSB of the least-significant tuning counter BCD digit would be wired directly to the PLL Divider. An examination of all the other bits in the necessary PLL Divider's division states did not indicate any bit combinations having similar state groups.

---

[9] Used in this manner, it is advisable to add a slight delay in the actual preset load active state. That allows the EPROM read-out time to stabilize and any necessary tuning counter preset input *setup time* to elapse. Once the tuning counter has been preset as desired, its outputs will change and therefore the EPROM address states will select a non-preset conversion data. That keeps the system from becoming an oscillator of sorts.

[10] Two data output bits allows four relay combinations, enough to switch in/out of as many matching networks. The consideration of relays (small, low-power types) is for low-resistance switching at relatively low characteristic impedances. Also, as time and nature dictate, the antenna(s) may change and matching changes.

[11] There are 3,140 one-KiloHertz increments in the actual tuning range of that receiver and only 14 of them have the auto bandswitching EPROM data output. 99.55% of the EPROM programmed data is BCD-to-Binary conversion.

Using an EPROM for Binary To BCD Conversion

This is, essentially the reverse of the preceding conversion. Binary would go to the EPROM address lines and the BCD digit data would be programmed in to appear at the EPROM data output. Again, the LSB of both binary and the least-significant BCD digit can go around the EPROM since those states are always the same. However, BCD digit data may have more bits than Binary. There might be some need to *pre-pack* the most-significant BCD digit data, then unpack it from the EPROM data output for display. That was thought of during the author's receiver design phase with the tuning counter all-binary, direct output to PLL divider (also binary input) and the converted BCD data used to drive an LED frequency display. EPROM programming would include subtraction of the IF Offset from the binary PLL states. That was discarded in favor of the tuning counter using BCD Up-Down counter ICs and no clear way to free up EPROM data output bits for ancillary purposes.[12]

With 16 binary bits there will be a need for 4+ times 4 bits for the BCD data output.[13] Increasing that to 20 bits would have 6 BCD digit outputs requiring 24 bits of converted data. With 24 binary bits, there would be 29 bits of BCD digit data. While some of that could be pre-converted in programming, the added *unpacking circuitry* needed begins to defeat the purpose of using an EPROM to reduce overall devices.

EPROM Programming

There are several ways to do that. The obvious way is to use a commercial programmer for a PC along with its software. Other than expense for a limited-use hardware device, the generation of a *program file* is an onerous task. The data in the finished EPROM is a giant *look-up table*[14] rather than a set of computer program instructions. In the author's example there would be slightly over three thousand entries to type in using acceptable EPROM data file formatting.[15] While the program entries would be linearly increasing from lowest address (tuning frequency in BCD) to highest, fourteen entries would be entirely different (the tuning counter auto bandswitching data). The finished file listing could be visually checked for accuracy but that is a tiresome task added on to the laborious typing-in task of generating the programming text file.

An alternative to the preceding is to use the partially-built BCD tuning counter with a temporary circuit using the binary PLL divider ICs wired up as a binary counter. The BCD tuning counter would still provide the address input for the EPROM and the binary states of the PLL divider temporary counter would provide the EPROM program data input. A slow-speed free-running

---

[12] The desired tuning band limits were, with one exception, on 100 or 50 KHz increments. Getting tuning counter (binary) presets required all EPROM data output bits (except bit A) to arrive at those frequencies. While other, farther-out-from-ends frequencies could have been used for auto bandswitching, such would not be appreciated by users.

[13] The amount of BCD converted data actually needed depends on the range necessary. To indicate all 65,535 states of 16-bit binary in BCD would require 19 data output bits total in the conversion, three extra for the most-significant BCD digit.

[14] *Look-up tables* in a computer program are constants embedded in the program code for read-only during program execution.

[15] Common-use EPROM data file formats are described in this chapter's Appendix C.

multivibrator would provide a common count input source for ***both*** counters. A block diagram of that one-use programmer is shown in Figure 28-20 on the next page.

The LED Display at the top is the intended tuning frequency display assembly with BCD digit input. The 16-bit Binary Counter sharing the common count clock input is preset to the Tuning Counter's start frequency in KHz plus the IF Offset and is wired to count up (the PLL divider would be wired to count down using 74AC191 ICs). The optional Binary Counter Display is a convenience in checking the count state.

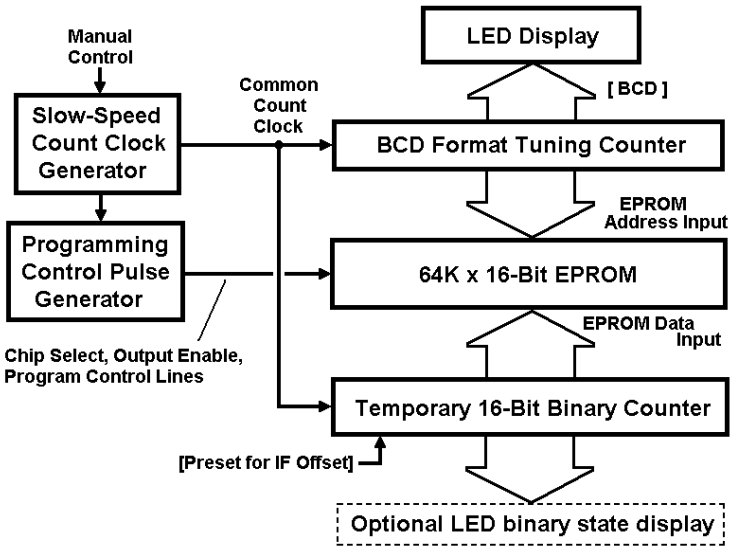


Figure 28-20 Single-Use EPROM Programmer Block Diagram for BCD-to-Binary Conversion programming.

Programming Control Pulses need some explanation. The AMD May 1998 datasheet on the 27C1024 device stated that PGM (Program input pin 39 in the 40-pin DIP) should go low for 100 µSec to program data input and that up to 25 pulses should be applied to insure proper programming at one address state...but no more than 25 program pulses after checking the data output after each program pulse.[16]

OE (Output Enable, pin 20) and CS (Chip Select, pin 2) should both be tied low (grounded). When PGM is high the Data Outputs reflect the program data and can be used to verify it. When PGM goes low the state of the Data Output pins are programmed into the EPROM. Trying to verify the Data Output after each PGM write normally requires the Data Input source to go into a high-impedance condition to avoid conflict in checking the Data Output state.[17]

For better isolation of data input source, something like a 74LS245 Octal Bus Transceiver IC (20-pin DIP) can be used. [18] Two would be required for 16-bit EPROMs. Another possibility is using four 74LS257 quad 2-input multiplexer with 3-state outputs, output connected to the EPROM data input and the multiplexer's output enable pin common with the active-low PGM pin. The multiplexer's OE pin at logic 1 would make the multiplexer's output pins high-impedance. The '257 device does not invert input states while the '258 does invert; both have the same OE and Select input pin control functions and states.

In the author's project, the already-built BCD-to-Binary converter of Figure 28-18 was used in place of a temporary binary counter and a relatively slow once-per-second controllable pulser made to the Up count clock input of the tuning counter. With each count clock pulse there would be

---

[16] This may be a bit severe but not unusual for manufacturers to state conditions that are quite conservative, perhaps to avoid too many returns from customers.

[17] Most manufactured EPROM programmers have a Verify feature that does not conflict with Data Input. Those are also of the *universal* kind in that they are designed for a large variety of programmable memory devices.

[18] Octal bus transceivers were a fairly common IC used in the first microcomputers for both the 8-bit data busses and the 16-bit address busses.

16 active-low PGM pulses via a couple of 555 Timer ICs and one binary counter IC. A manual push-button would allow single count clocks, with a toggle switch to select EPROM write or just increment the tuning counter. The latter was done to allow manual entry of the auto bandswitching data at one of the 14 bandswitch-sense frequencies. Four 74LS257s were used (with toggle switch selection) to have isolated EPROM program data input from the BCD-to-Binary converter or a pair of octal *DIP Switches* that could be set to the predetermined auto bandswitch preset data. A pair of toggle switches were used to set the two antenna matching network relay control line bits; the R, Q, P bits of the converter output were packed into a single bit P by an inverter. A pair of 8-light bar graph display LED assemblies driven by three hex inverters served as an indicator of the 16 bit EPROM data output. Since those inverter inputs were always connected to the EPROM data lines, they showed both write-in data and read-out data (for verification after writing). As long as the display didn't change between write and read, the EPROM data was assumed to be correct. The BCD tuning counter was preset to 4000 KHz to start the programming and manually stopped after reaching 15900 KHz, both as indicated on the intended Display assembly.[19] Note: The *stopping frequencies* (the auto bandswitch sense frequencies) would have to be predetermined ahead of time along with the manual data needed to write in the bandswitching data.

While the AMD datasheet stated that the V_{PP} (programming supply voltage) should be in a range of 12.5 to 13.0 VDC and the V_{CC} (operating voltage) should be 6.25 VDC, the author used 12.0 VDC and 5.00 VDC, respectively, with no problems. Had there been a problem with that, or of an unverified write, it was a matter of removing the EPROM and putting it in the UV *erase box* for a half hour, then doing it all again.[20] Total time of programming would have been about 4 hours but that was cut in half by skipping some of the tuning bandspace that would never be reached.. For a one-time-only effort that was reasonable. The prototype circuit board holding the Figure 28-18 converter and the extra EPROM programming circuits were kept in storage after the work was done.

---

[19] If the auto bandswitch feature worked correctly, the tuning counter would never go below 4980 KHz nor above 15810 KHz. TheFigure 28-18 circuit would provide the correct PLL Divider preset data with IF Offset from 3176 KHz to 15999 KHz. Any *extra* EPROM address states would be superfluous in actual use.

[20] The author has had success for a quarter century with a rather cheap 8-track tape cartridge box holding a UV lamp and its AC mains supply, the whole sold as a *finished product.* It had erased many a UV-erasable EPROM from old Intel 2708s to more modern devices such as the 27C1024.

Appendix 28-1

Various Bidirectional Counter Chains

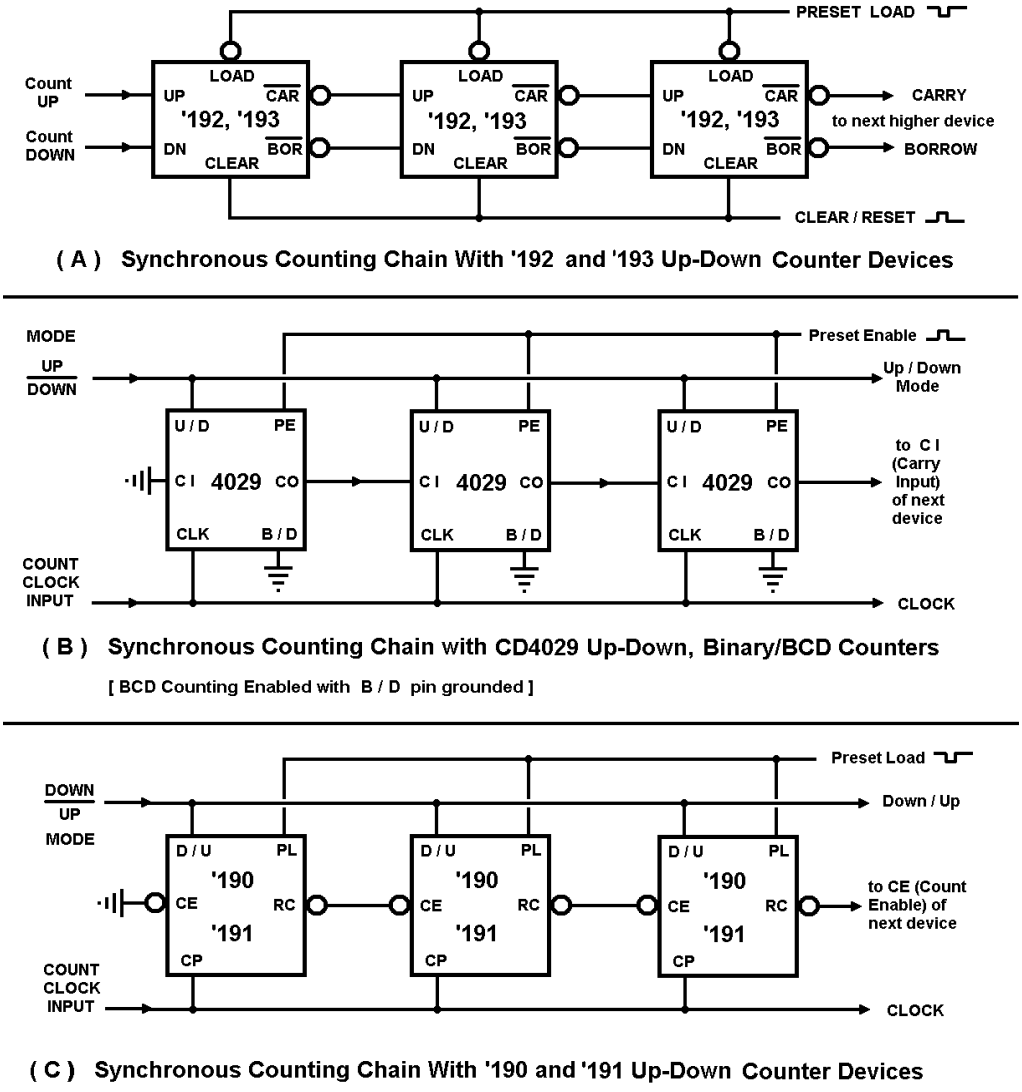


Figure 28-21 Chains of various bidirectional counter IC devices.

Figure 28-21 shows the difference between various bidirectional counter types to do the same counting tasks. The choice of synchronous counting is, for most uses, arbitrary. Using synchronous count devices insures that there are a minimum of extra little short-term spikes that might radiate a bit of RFI into sensitive stages. There is little cost difference between devices.

All shown in Figure 28-21 have parallel input preset capability. The CD4029 can be either binary or BCD count mode but has a maximum count-clock rate of about 4 MHz at a 5 Volt supply. Synchronous counters can operate at higher clock rates than *ripple-through-carry* devices. A 74AC191 with added gating can reliably preset to a state then count down to zero at rates greater than 40 MHz.

Chapter 29

OSCILLATORS

Oscillators are of self-repetitive-frequency AC energy sources. That frequency may be fixed or variable. This Chapter presents L-C tuned and quartz crystal unit oscillators with a general sinusoidal output. Sinusoids may be converted to rectangular waves by additional output stages but that can be done directly by multivibrators covered in Chapter 27.

General

Every **oscillator** is basically an active device with an input-to-output gain just slightly greater than unity. ***Positive*** feedback is applied through a frequency-selective network such that the output will be an oscillatory waveshape. Once oscillation is begun, it will seek equilibrium to keep on oscillating, depending on the amount of feedback, the voltage gain of the amplifier, and characteristics of the active device.

Figure 29-1 shows a *generic* oscillator using a linear-device amplifier. Assuming a voltage gain of A_V, the voltage divider of **Z1** and **Z2** is selected so that the feedback voltage is just slightly more than $(1/A_V)$. It is assumed that Z1 and Z2 are *perfect impedances* and have no intrinsic phase shift. Since the output is fed back in-phase, the input is enough to just slightly overload the amplifying element. Oscillation results.

Generally, oscillator circuits use L-C or quartz-crystal resonators in the feedback path such to make the oscillatory frequency a certain value. Output waveshape is then a function of:

1. Gain of the amplifying device.
2. Amplitude of the feedback.
3. Phase shift of the feedback.
4. Linearity of the amplifying device gain.
5. Stability of the amplifying device gain.
6. Stability of the feedback phase-shift.

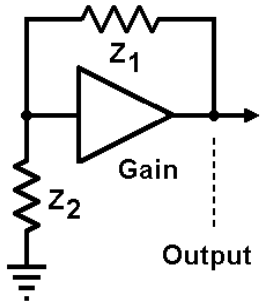


Figure 29-1 A generic oscillator.

In the beginning of radio as a technology, only L-C circuits were available to set both the value of frequency and the magnitude of feedback. Quartz-crystal resonators did not appear until about 1922. As a result, the basic oscillators were described by their innovators and those names became the names of oscillator circuit configurations. Those are shown in Figure 29-2 in a simplified form using NPN transistors with a positive supply voltage. Since transistors would not be available until after World War II, the original circuits used triode vacuum tubes.

The main point is that the base and emitter have a 180° phase-shift, thus satisfying the overall positive feedback situation at one frequency. In all of the basic circuits, oscillation

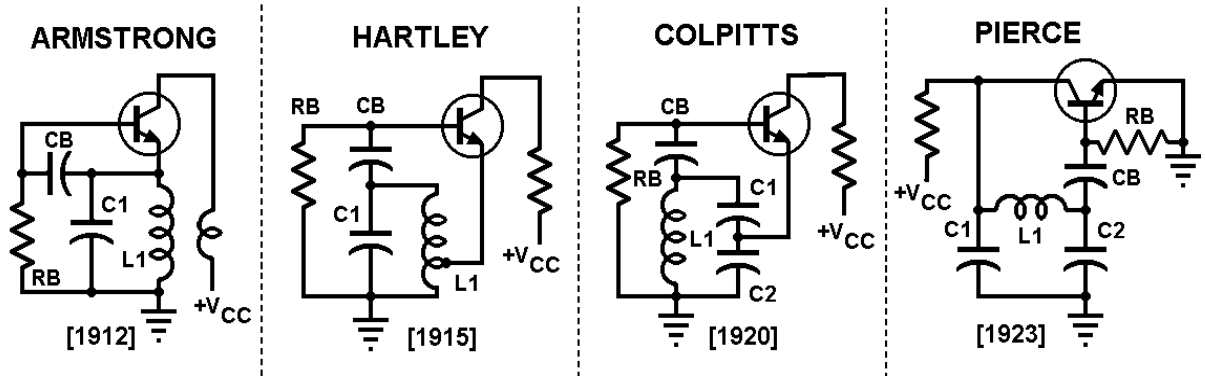


Figure 29-2 The four basic oscillator configuration. R_B and C_B have little to do with oscillation frequency, are there to show *base biasing.* The resistors in collector leads may represent a load impedance. Numbers in square brackets show year of first publication.

frequency is mainly due to resonant conditions of L1, C1, and C2. The ***Armstrong*** oscillator is also called a ***Meissner*** oscillator since they were first published at about the same time.[1] Armstrong would later invent the ***superheterodyne*** receiver and pioneer FM broadcasting. Ralph V. L. Hartley would later come up with the basic circuits for SSB AM modulation and demodulation. Edwin H. Colpitts' configuration would become the baseplate for the ***Clapp, Vackar,*** and ***Seiler*** modifications appearing after World War II. George W. Pierce's version would be the most-used form of IC crystal oscillator configuration by the 1980s.

All of these basic four forms could be modified to use quartz crystal units for very stable frequency sources...once the piezoelectric characteristics were better known.

Armstrong - Meissner Oscillator

Oscillation frequency is determined by L1 and C1 in parallel resonance. A small amount of output is fed back via a ***tickler*** winding physically added to L1.[2] RF polarity must be observed there since an opposite connection to the added inductor results in negative feedback and no oscillation.

Hartley Oscillator

Simply tapping down on L1, connecting the tap to the cathode (or emitter of common transistors now) would provide the positive feedback for oscillation. This became a favorite for HF and lower frequency receiver Local Oscillators prior to about 1960 since the tapped L1 could be made at lower cost and also work with the same value of variable tuning capacitor. That led to the (relatively) inexpensive multi-band receivers popular during the 1930s through 1970s.

---

[1] In Europe the ***Meissner*** name is preferred since Alexander Meissner (a German citizen) published at about the same time as a young Edwin H. Armstrong published his in the USA. Due to relative antiquity and the many civil court cases on this (then) new radio technology, it is unknown if there was any reciprocal publication in opposite countries at the time. Armstrong would use his oscillator circuit in a modified form to create his first ***regenerative detector,*** a triode circuit run just barely ***under*** oscillation to remarkably increase its sensitivity.

[2] ***Tickler*** is a jargon name. Since the technology was so new at the time, many such jargon names were invented and publicized largely by the media. Since academic technology was also new, jargon tended to become the de facto name before the correct technical terms were there. In the tube technology of the time, ***grid leak*** was the common name for the equivalent of ***Rb,*** something persisting until about the 1960s.

Colpitts Oscillator

If the Hartley could use a tapped inductor for feedback, then the ***Colpitts*** could use a ***capacitive divider*** via C1 and C2. Feedback polarity is the same as for the Hartley. The only drawback is needing an RF choke inductor to provide a DC path for the cathode or emitter. Oscillation frequency is determined by L1 and the ***series*** connection of C1 and C2. The Colpitts configuration was not used much for multi-band designs, being better suited for single-frequency or single-band operation.

Pierce Oscillator

The relative late-comer is the Pierce configuration, probably due to more attention paid to passive networks. Note that, in Figure 29-2, C1-L1-C2 form a familiar ***pi-network*** configuration. That allows better control of the amount of feedback from the source-impedance plate/collector to a lower-impedance base-emitter or higher-impedance grid. Note also that, at resonance, the phase difference of a pi-network is 180° input-to-output. That explains the configuration shown as essentially grounded-emitter.

L-C Oscillator Variations

Colpitts as a Baseplate

Figure 29-3 shows the major variations of the basic Colpitts circuit, all drawn in RF-simplified form. The Clapp was first published by James K. Clapp in 1948. Frequency is determined by resonance of L1 and the connections of C1 through C4. The Clapp works best with a higher-transconductance active device.

The Vackar was published in 1949 by Jiri Vackar. The Seiler was apparently a modification for high-HF to VHF. There are few references to the Seiler in publications.

Analysis is more complicated with transistors due to low base input impedance. Note that there were first done with tubes as active elements. Figure 29-3 is based on an illustration from the June, 1968, edition of *Ham Radio* magazine.

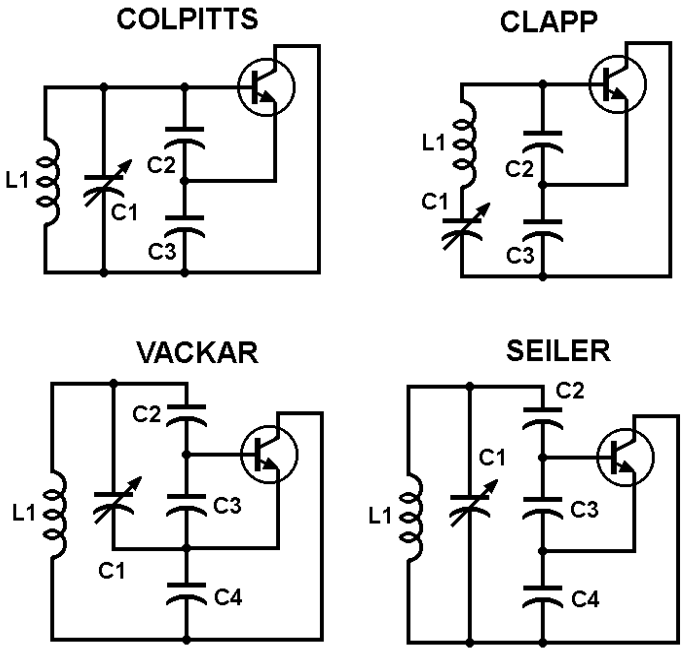


Figure 29-3 Simplified variations of the Colpitts.

The Vackar and Seiler configurations have (somewhat) exaggerated claims of stable performance, especially for amateur radio applications. It should be noted that some references cite the ***Clapp-Gouriet*** oscillators. Geoffrey Gouriet devised his circuit about 1938, similar to the Clapp published ten years later. Due to World War II secrecy, the Gouriet circuit was kept out of

publication until after that war. Applications of the Gouriet circuit are scarce.

Since the Clapp, Vackar, and Seiler all have improvements in frequency stability, and that they were initially done with tubes, it was important that ***physical*** frequency read-out was made stable. Considering that vacuum tube changes were once relatively frequent, this was important to eliminate changes in grid-to-ground capacity variations when changing tubes.

In more modern times, the ability to read frequency directly by counters obviated the need for physical frequency read-out. All that remained was standard environmental effects.

Stability and Environmental Effects

Temperature

This may be the biggest culprit in frequency variations. Capacitors, especially, have many grades of change per degree C depending on their construction. That includes variable capacitors for manually-tuned L-C circuits. Inductors have some manufacturing characteristics but that does not apply to powdered-iron or ferrite cores in home-built inductors. Such variations can be partly solved by applying fixed capacitors with an ***opposite*** temperature characteristic to that measured with a supposedly-stable L-C tuning circuit.

The old tale of ***waiting a half hour from first turn on*** is a left-over from the tube era. When vacuum tubes were used chassis structures were, by comparison to today, rather massive. Such relatively large structures took time to reach an equilibrium temperature and temperature changes reflected in slight variations in tuning of L-C circuits.

With quartz crystal oscillators, the quartz unit itself has some definite frequency-change characteristics, usually found in more extensive catalog information. Quartz crystal controlled oscillators tend to stay in their narrow frequency-change characteristics given by their makers. One can add small trim capacitors to align and calibrate such resonators. Such trimmers are fixed after alignment and seldom changed afterwards.

Supply Rail Changes

Unregulated supply voltages can change oscillator frequency by varying degrees. At one time the gas tube shunt-regulator was thought to be a cure, but it is only a stabilizer of sorts. Such wastes power, thus re-enforcing the old *wait a half hour after first turn-on* rule. The modern semi-conductor series regulator IC can do a much better regulation job.

Supply rail distribution can sometimes be a problem through sneak paths. Decoupling through R-C and L-C feeds may be necessary to keep changes induced by other, higher-current-drain blocks from affecting an oscillator frequency. That is part of the design process and can be figured ahead of time.

If necessary, small 100 mA series regulator versions can be added in place of larger, diamond-case ICs. Those allow a point-of-use regulation which is separated from other regulators by about 60 db from the common, unregulated input. Those are also good for stabilizing low-current, separate blocks powered from batteries. Battery voltages can change over their life, even in the same package size from having different electrochemical internal mixes. Such series regulation wastes battery life slightly, a small price to pay for more-stable frequency operation.

Mechanical

Vibration must be kept to a minimum. Vibration can cause FM-ing of RF output by small variations of capacitance and inductance changes. An oscillator should be stable enough to endure shock and other impact-derived mechanical disturbances.

Load Changes

This comes about primarily as Local Oscillator inputs to Mixers, with strong Signal input amplitudes. It is application dependent. If necessary, emitter-followers can be added to help *buffer* an oscillator from load effects.

The Electronic Breadboard

This is a PC with a ***SPICE*** program. Two sources are Linear Technology (LTSpice) and Texas Instruments (TINA-TI).[3] Both are download-free program sets. Both allow observation of oscillation waveforms, good for gross visual clues to harmonic content, and Fourier Analysis, for more precise harmonic content.[4] Any user can vary any component to see various effects.

One Saturday's efforts can equal a whole ***month's*** worth of effort handling the hardware on the bench. Also collecting a lot of parts that will not be used much now, may never be used. With LTSpice a schematic may be saved and any waveform or frequency-domain test can be saved to the PC for later review. LTSpice allows adding components to the program, includes a small subset program to add those. One can observe schematics (printable and storable) rather than lists of parts relative to their diagrams. This author recommends using a SPICE program, from whatever source, to do the first breadboard on any electronic circuit, simple or complex.

Start-Up Time Delay

When first energized, an oscillator circuit doesn't oscillate immediately. Reactances must charge up to their DC values. Feedback allows eventual oscillation. Start-up time may take a few hundreds of microSeconds to occur. Such start-up delay is related to many factors but it ***will occur.*** Generally, this delay time must be disregarded for any useful purpose.

It is wise to ***not use*** any keying of any oscillator, such as exemplified by older amateur radio articles using minimal parts. Given that a buffer amplifier can be built for less than $1 US with new parts, it is best to just let it run. Keying through an output circuit is much better.

---

[3] As of 2013, Linear Technology was from http://ltspice.linear.com/software/LTSpiceIV.exe. TINA is available from two firms: TINA-TI from www.ti.com (search for document SLOC241.ZIP); a saleable version, TINA 9.3 from DesignSoft in Hungary (www.DesignSoft.com) but costs money). TINA 9.3 includes a form of PCB layout software working directly from its internal netlist. Do not confuse the two programs even if they have very similar names.

[4] The author prefers LTSpice, primarily because he has used it for a long while and does not want to memorize all the commands necessary with TINA. That's a personal choice. LTSpice came out first with a ***free*** program download, primarily to sell its various switching power supply ICs. No (annoying) *netlists* need be typed in since a simple graphic schematic substitutes, leaving the netlist internal to LTSpice. ***Netlists*** refer to the typed-up lists of parts and their node connections in circuits, part of the jargon in such electronic programs.

Individual Circuits of the Basic Oscillator Family

General

These are all shown for a 5.0 to 5.5 MHz variable-frequency source.[5] One application was chosen to represent all instead of individual frequencies and components. The active device was picked to be a 2N3904 NPN BJT, again on purpose for standardization. Supply rail is fixed to +5 VDC. This is a fairly easy regulated-source voltage, used by the common digital logic device family. While higher supply voltages can be found in older oscillator texts, they are more a product of older, tube-structure thinking. Higher voltages dissipate more power; while such can provide more power, heat dissipation is greater than AC power output.

A Colpitts Oscillator

This is illustrated in Figure 29-4, along with a simple *buffer*[6] stage to help isolate the oscillator per se and its load. Q1 is the NPN version of the 2N3904-2N3906 complimentary pair and regarded as a general workhorse type of BJT. The diode is another workhorse, a 1N4148 general purpose silicon diode. Passive values correspond to ± 10 % types unless otherwise noted.

Diode D1 is there to catch negative swings of Q1 base. Since this is a Colpitts with a parallel-tuned base-emitter tuned circuit, higher DC supply voltages could allow a negative swing to possibly exceed reverse-biased base voltages on Q1.[7]

Base-to-ground AC voltage can be nearly twice the supply voltage peak-to-peak. The collector to ground AC voltage of Q1 is about 4.2 V peak-to-peak as a rounded square wave. R5 and inter-electrode capacity of Q2 makes the Q2 output nearly a triangular shape.

The tuned circuit is L1 and C1 in parallel-resonance plus base-emitter feed-back capacitors C3 and C4 determine operating frequency. Q1 base bias is provided by R1 and R2, C2 used as a bypass capacitor.

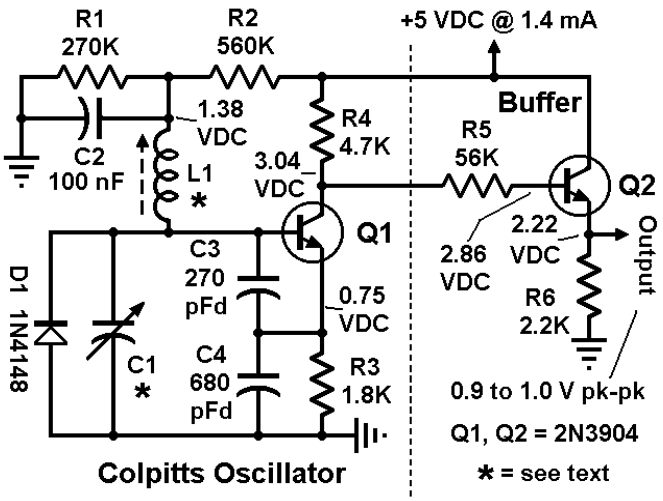


Figure 29-4 A Colpitts oscillator for about 5.0 to 5.5 MHz and a Buffer. *VDC* markings are for DC only. Output is 1.00 V peak-to-peak.

---

[5] Taken from the Main Tuning Oscillator assembly specifications of the earlier Heath SB Line of amateur band receivers, transmitters, and transceivers.

[6] ***Buffer*** is a colloquial term which may be varied as desired in waveshape and amplitude output. Their purpose is to isolate oscillator tuning from most load effects.

[7] Depending on the BJT construction, reverse-biased base-emitter junctions can exhibit a ***zener*** effect. For several years the author depended on 2N3638 BJTs for an approximate 6.5 VDC zener junction. Not all such BJTs have this as some have reverse breakdown values up to the limit of remaining electrodes.

As an example target, C1 is selected as 32 to 92 pFd, manually tunable, with a frequency range of 5.0 to 5.5 MHz. From Formula 15-1 of Chapter 15, a parallel capacitor is added to reduce the maximum-to-minimum range of C1. That makes **V** = 2.875 and **D** = 1.210. A parallel capacitor value would normally be 253.714 pFd. The series connection of C3 and C4 normally has a total capacity of 193.263 pFd. There is a difference of 60.451 pFd of C3 seriesed with C4 and $\mathbf{C_p}$.

Based on an LTSpice plot of output waveform, at best there would not be any parallel capacitance other than C3 and C4. Further, L1 would have a value of about 2.8 µHy to cover the frequency range. There is roughly a 60 to 70 pFd difference between Chapter 15 calculation results and the LTSpice analysis.[8] Based on the circuit and rather tough theory, the base of Q1 would appear to add more *equivalent capacity* than was expected.

Equations can be added to show where ***some*** of this extra capacitance (equivalence) comes from but the formulas become rather complicated. This is a ***hobby endeavor***, not a term paper, and time to work complicated formulas takes away from the hobby aspect. Given the free SPICE program, several trials can be run to zero-in on respectable component values. As it is, the real world would require alignment of an adjustable L1 and trimmable C1 to properly set the frequency range. As a hint, L1 would come very close to 2.78 µHy.

The Clapp Oscillator

The Figure 29-5 is is very similar to the Colpitts but the frequency is determined by the ***series resonance*** of the tuned circuit and feedback capacitors (C3, C4) are raised in value. There is no diode since oscillation occurs when the L1-C1-C2-C3-C4 tuned circuit is at a ***minimum*** impedance. Q1 base-to-ground voltage is at a minimum.

The higher value of C3 and C4 requires a high-β for Q1. C2 is there for alignment of C1. C2 adjusts the frequency span manual tuning of C1 while L1 sets the entire frequency of tuning. DC bias is provided by R1 and R2. No bypass capacitor is needed; values of R1 and R2 are high enough to have little effect on series resonance. Emitter resistor R3 is there for a DC return of Q1.

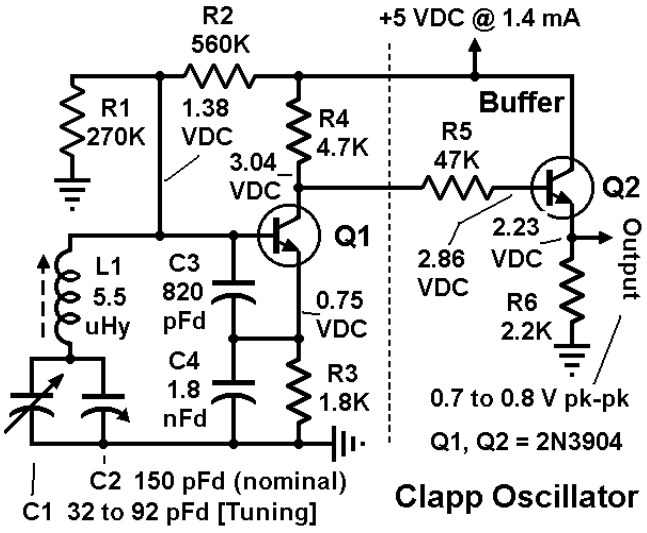


Figure 29-5 A Clapp Oscillator version, similar to that of Figure 29-4. Q1 is the oscillator and Q2 a load-isolating buffer.

Again, series resonance does not make an easy task of determining frequency. The procedure is to set the general value of C3 and C4 (in series) so that it is higher than usual and allows a parallel capacitor value across C1 to set the tuning span. As a direct calculation, C3 and C4 in series

[8] In LTSpice the output frequency is the inverse of the plotted repetition time. What was done was to count the total interval of 20 waveforms, then divide that by 20. That is a slight increase in accuracy although that is still rough indeed. Chapter 15 presumes that no other device other than a resistive load would affect results of adding series and/or parallel capacitors. Chapter 15 equations are a benign set that is not disturbed by non-resistive loads; SPICE programs show the circuitry as-is/

will be 563.359 pFd. This allows the parallel of C1 and C2 to be about 182 to 242 pFd. Total series-resonance capacity (C1 through C4) will be about 169 pFd at 5.0 MHz, requiring an inductance of about 5.9 µHy. High-end frequency is a bit over 5.5 MHz, therefore C2 is made a trimmer to adjust tuning span; about 140 to 180 pFd range must be used here..

In the first LTSpice analysis, proper oscillation frequency required the inductor to be about 5.5 µHy, not the idealized 5.9 µHy value. Much of that is due to the circuit, impedance of the base-emitter junctions of Q1 and R3 shunting effect. It was much easier to simply measure total ***time*** of 20 cycles, divide that by 20, then take the reciprocal of that for frequency.[9]

Q of the inductor has a great impact on oscillator start-up time. A Clapp series-resonant oscillator ***needs a high Q for stability.*** With a Q-equivalent of 60, start-up time increased to 1.5 to 2.0 mSec. The X_C of 5.9 µHy at 5.0 MHz is about 185 Ohms. For the effect of Q, divide that by Q to find the equivalent resistor for loss. 3.0 Ohms results in a Q of about 62. Compare that to the 300 µSec minimum oscillator start-up time with a Q of 300 using a Q-loss resistance of about 0.5 Ohms.

A Hartley Oscillator

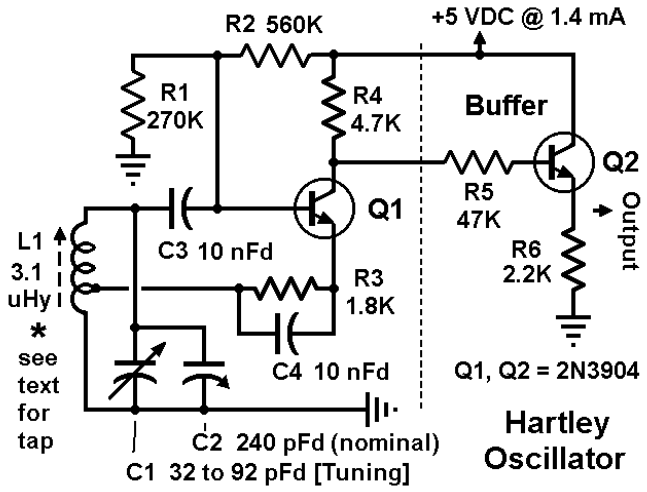


Figure 29-6 A Hartley oscillator configuration, again covering 5.0 to 5.5 MHz.

The circuit of Figure 29-6 is, essentially, much like the Colpitts and Clapp examples. DC conditions are the same but the feedback is from a tap on inductor L1. That tap is located about 1/4 of the way up from the ground end. L1 can be checked on the bench with an approximate 332 to 272 pFd capacitor.

Frequency is determined by parallel resonance of L1 and C1, C2. C2 here is a 220 pFd fixed value with an added parallel trimmer of 3 to 30 pFd. C2 aligns for the tuning span (width) while inductance of L1 sets the overall frequency range.

C3 and C4 both bypass the DC bias and need to be about 30 to 100 times higher in value than resonance capacity.

A Hartley configuration has some advantages in slight simplicity over a Colpitts and Clapp versions. Both L and C components for resonance have ground connections. The tap can be varied slightly to insure stable oscillation.[10]

A diode at Q1 base-to-ground is optional. The resonant circuit is parallel so the base voltage

---

[9] As with the Colpitts oscillator, the effect is off the ideal (passive) value of L and C with a rather large and complicated formula. Rather than slug through that, it was easier to see the entirety of oscillation and presume the LTSpice timing to be precise enough. It is difficult enough with low frequency ratios to be accurate with Chapter 15 formulas, good for non-oscillating, linear amplification. The technique of measuring time with LTSpice is optional; it may be about twice as accurate compared to using one oscillation cycle time and taking the reciprocal of that..

[10] The emitter tap can be anywhere from 1/5th to 1/3rd of the total turns from ground end. With toroidal coil forms it is more likely to be about 1/5th due to greater coupling possible. Cylindrical coil forms tend towards 1/4th to 1/3rd since there is less coupling. See Chapter 13 for modeling a coupled inductor for a compter.

can be driven negative. See the comments with the Figure 29-4 Colpitts oscillator for D1.

Vackar and Seiler Configurations

Referring to Figure 29-3, the Vackar and Seiler versions have a more-complex connection of resonant-circuit capacities. While those may fit vacuum-tube versions better, they do ***not*** seem to offer any special stabilities that make them worthwhile for semiconductor active devices. Indeed, those capacitor arrays can become a nightmare of complexity in trying to determine temperature-compensation necessary for wider environments.

A Small Summary of Configurations, So Far...

The Colpitts, Clapp, and Hartley examples are all at the same tuning frequency range. The oscillator sections per se draws only 2 mW maximum from the +5 VDC supply; the optional buffer takes another 5 mW. Each buffer can drive a 700 to 1 KOhm filter load to make its spectrum output more pure at its fundamental frequency. Another, following amplifier, can boost that filtered output. Note: Some of the older amateur radio texts boast of *minimum* stages and powering a final RF amplifier input directly. That is vacuum-tube thinking and says little about frequency stability. HF-range bipolar transistors are quite cheap at distributors, much less than their old tube counterparts.

The small power consumption of the examples and their lower supply voltages allow reasonable shielding, both conductive and thermal, to keep their operating temperature constant. The supply voltage can be higher, say from +9 to +12 VDC, allowing good R-C decoupling to further isolate the variable oscillator's frequency from outside interference.

Anti-Resonance

Peculiar to very old-school terminology, ***anti-resonance*** means ***Parallel-Resonance.***[11] Series resonance is also possible, especially so with quartz crystal devices.

Working With FETs As Active Devices

Allowing for the disparity between biases of *depletion-mode* and *enhancement-mode,* FETs can replace the BJTs in examples of Figures 32-4 to 32-6. Gate-Source impedances are higher with most FETs so there are some variations of the feedback capacitor values. At the same time, there is less effect on the tuned circuits. Inductor values approach the passive values of Chapter 15 more closely than with BJTs. If possible, a SPICE circuit model should be analyzed ***first*** to pin down the bias component values, start-up time delay, and waveshape of the output voltage. Bias resistor values for Gate and Drain junctions will be higher than with BJTs.

To restate, one Saturday afternoon spent with a SPICE model can easily equal 30 other days of the month working with just parts on a bench.

Fixed-Frequency Oscillators With Quartz Crystals

---

[11] To this author, ***anti-resonance*** denoted a ***LACK*** of resonance. Millions of parallel-tuned circuits later, this very old-school term should have been dropped, discarded, thrown-away. It seems peculiar to some instructors grounded in ancient thinking that goes back to 1900 or so, especially in regards to oscillators.

General

Crystalline silicon dioxide, better known as ***quartz***, is the premiere piezoelectric resonating element for hobbyist work. There are some other piezoelectric substances, such as lithium niobate, but ***quartz*** holds its resonant frequency at least 100 times better than all the known materials. As explained in Chapter 14, quartz crystal units can be used as series or parallel tuned circuits.

To better see the quartz crystal unit, consider the Pierce Oscillator conversion, as in Figure 29-7. The Pierce resonant circuit is shown at left. A ***model*** of a quartz crystal unit is in the middle, specific crystal unit ***component equivalences*** shown as darker lines. It should be noted that there exists an extra capacitance in parallel of the crystal unit, due mainly to its holder/container. This affects the slightly-higher ***parallel resonance*** frequency. ***Series resonant*** frequency is mainly from L, R, and C_{SERIES} of the model. Series resonance is slightly lower in frequency than parallel resonance.[12]

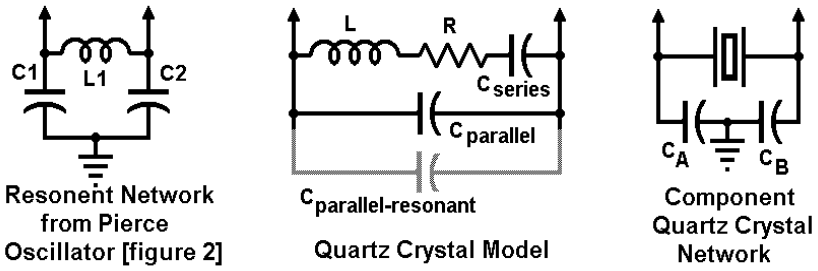


Figure 29-7 Conversion of Pierce Oscillator from an L-C resonance to a parallel-resonant quartz crystal unit.

At right in Figure 29-7 is the crystal unit with added external capacitors C_A and C_B whose total series capacitance value is equal to the ***manufacturers' specified parallel-resonance capacity*** minus stray capacities from the IC and PCB. If no capacity is specified by a manufacturer, then the crystal unit frequency is ***series resonant***.

After a decade into the new millennium, the Pierce Oscillator is probably the most used, being the clock oscillator of millions of microprocessors and microcontrollers. Any parallel-resonant circuit will have 180° of phase-shift. That is easy to apply to digital circuits using a single inverter gate. A series-resonant circuit has no phase shift at resonance; applied to digital circuit that would take two inverters in cascade. This is illustrated in Figure 29-8 using digital inverter gates.

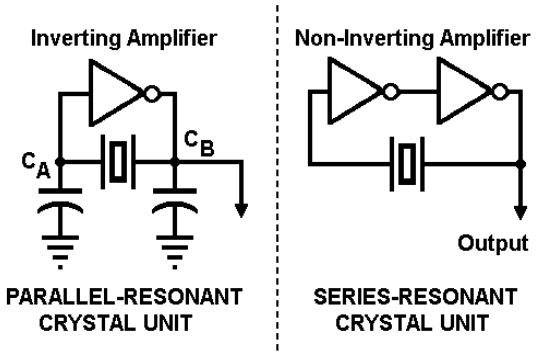


Figure 29-8 Crystal resonance selection using digital logic inverter gates.

Note that applying a parallel-resonant-mode crystal allows a slight amount of control of frequency. That is through making one of the two capacitors trimmers. This is not easily done with series-resonant-mode crystals since that frequency is ***intrinsically controlled*** by the quartz model.

Internal Quartz Equivalent Circuit Values

[12] There is little difference between series and parallel resonance in a quartz crystal unit, very roughly about 300 to 200 PPM or equivalent to (3/10,000 to 2/10,000).

These are difficult to obtain, both from being manufacturing-proprietary and from variations due to the ***crystal cut.***[13] There are, roughly, two dozen types of ***crystal cuts***, the ***AT-cut*** being, perhaps, the most available for the fundamental frequency range of about 400 KHz to 12 MHz. Unless a hobbyist has a crystallographic laboratory, it is useless to discuss more on actual cutting. The ***AT-cut*** has the most benign temperature coefficient versus frequency and is thus the most used.

Due to the great variation in quartz crystal models with their specific cut, a more precise measurement has to be done with a slow-tuning signal source and frequency counter measuring down to 10 Hz resolution (or better). Some of that was covered in Chapter 14, variations are found in the references. Measurement is precise work and takes considerable time to do carefully.

The fundamental frequency equivalent series-resonant arm can have inductances ranging from about 6 mHy at 400 KHz to 200 mHy at 12 MHz. Series-resonant capacitor equivalents will range from about 5 ***femto***Farads at 400 KHz to 30 femtoFarads at 12 MHz. Equivalent series resistance goes from about 1 KOhm at 400 KHz to 3.0 Ohms at 12 MHz. The resulting Q of the series-resonant equivalent arm can be 20,000 at 400 KHz to 100,000 at 12 MHz.[14] The equivalent parallel capacitor ranges from 5 pFd to 2 pFd but is subject to an added parallel capacity specified by the crystal unit maker for parallel-resonant frequencies.

Note that, while the parallel-resonant mode of a crystal unit is the most flexible for adjustment of frequency, a trimmer inductance can be added in series to slightly alter the series-resonant frequency. That requires some careful measurement of equivalent series inductance.

The difference between series-resonance and parallel-resonance can range from about 200 Hz at 400 KHz to about 15 KHz at a 12 MHz crystal unit. That attribute is one of the reasons for applying crystal units in place of L-C sections for very narrow bandpass filters.

Common Frequency Tolerances

This applies to purchased crystal units. The most common frequency specification is given as **±50 PPM** (Parts Per Million) or equal to ± 0.005 %. That specification is given over the stated temperature range of the crystal unit. A 50 PPM specification is equal to ± 500 Hz of a 10.0 MHz crystal unit. Better tolerances can be obtained but at an increased cost to the hobbyist. In all likelihood a frequency counter standard oscillator can be held to about ± 2 Hz frequency at 10.0 MHz in a hobby shop temperature environment of 70° F to 78° F.[15]

Controlling Oscillator Temperature

In older days, putting crystal units in small ***ovens*** was common to hold frequency constant. That was good for vacuum-tube circuitry which had considerable wasted heat dissipation. With solid-state oscillator and buffer circuitry, the ***entire circuit*** can be in the oven.

---

[13] Quartz crystal blanks begin by being ***cut*** from a much-larger crystalline ingot of quartz. The ***cut*** is done in three dimensions. The angle of cut determines a number of other factors, including the physical flexure of the quartz blank, its final shape, and a number of other factors. Many available texts devote much time to the different types of cut, as if the average hobbyist were also an expert on crystallography. Such data is not included here since it is out-of-league for even the expert hobbyist.

[14] These values come from a scattering of published information (when it is available) and personal experience. They aren't that quantifiable as to crystal cut since most were packaged in soldered metal cans.

[15] From personal experience over 30 years in a residence hobby shop environment, checked against NIST.

Using a box-within-a-box and small heater set for 120° to 130° F, a small heater control circuit can keep that to within ± 1" F. The author did that at 130° F in an older unit dissipating about 27 W at a room temperature of 82° F. Time-to-warm-up was about 15 minutes, time required for the oven to reach equilibrium. A 120° F oven temperature should fit more readily in residence home shops with built-in air conditioning and heating.

It should be noted that such as oven was constructed with a metal box ***inside*** a fiberglass-insulated outer box made of PCB stock. Fiberglass was from a furnace air filter, using two layers of the thin fiberglass material. An op-amp and thermistor (sensing heat) was mounted to one side of the crystal oscillator. The *heater* consisted of several carbon-composition resistors to the internal DC supply plus the small enclosed circuit dissipation.

Crystal Unit *Drive* Level

Most crystal unit manufacturers specify ***maximum drive levels*** of 5 mW to 10 mW to avoid any changes to a crystal unit frequency. This is normally from the equivalent series resistance of the crystal model. If that can't be measured, then the oscillator itself should be at a low power. The L-C examples of Figures 32-4 through 32-6 all take only about 2.1 mW in operation, quite good enough if converted to quartz crystal units.

Crystal Unit Miscellany

Ageing is the very slight drift in frequency over a long period of time. This applies to master oscillators, such as used in frequency counter time-bases. It is (very roughly) about 1/50th of the initial frequency tolerance.

Phase noise is a much-publicized specification which, actually, has little to do with crystal oscillators in radios. It is the short-term jitter in frequency. It will have a slight impact on FM and PM in radio communications at higher HF carriers, slightly more at VHF and UHF. Phase noise applies to clock synchronization such as in large communications installations. ***Phase noise*** has nothing to do with receiver front-end random noise that limits sensitivity in HF radios for AM, SSB, or CW although the implication exists. It matters for FM and PM on VHF and higher.

To simplify names, a crystal oscillator is sometimes abbreviated to ***XO.*** An Oven-Controlled crystal oscillator is abbreviated ***OCXO.*** A Temperature-Compensated crystal oscillator is abbreviated as ***TCXO.*** A crystal oscillator with Voltage Control (of a few PPM in tolerance) then becomes ***VCXO.*** ***VCXOs*** usually have a voltage-variable-capacitor diode in place of a trimmer capacitor to allow remote control of exact frequency.

Some makers specify frequency tolerance in terms of ***Pulling*** or ***Pushing.*** ***Pushing*** refers to frequency deviation due to power supply voltage changes. ***Pulling*** usually refers to ***all*** changes, voltage, temperature, (sometimes) ageing (in the case of Vectron units) causing a deviation from a specified center frequency.

Substituting Crystal Units for Oscillator L-C Tanks

Figure 29-9 shows the equivalence of quartz crystal fundamental units in place of the L-C resonant circuits of Figure 29-2. In the Colpitts version, C1 and C2 are usually smaller than with L-C resonators. A Colpitts crystal unit is parallel-resonant. The Clapp version of the Colpitts requires a series-resonant crystal unit. Both require a means of DC return of the emitter.

The Pierce was already discussed, along with its CA and CB. It has no DC blocking capacitor from Collector junction to the crystal; it is presumed the crystal unit is non-conducting.

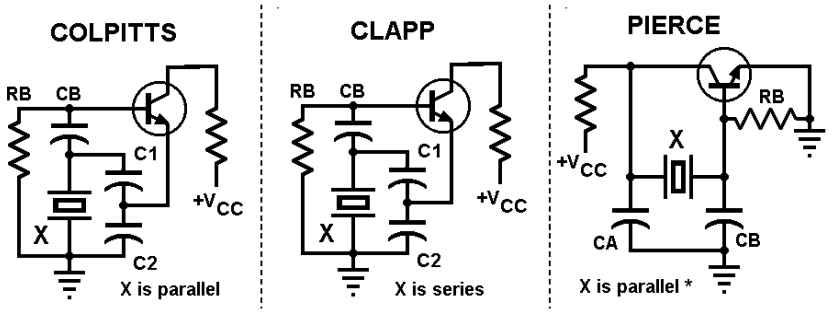


Figure 29-9 Simplified crystal unit basic oscillators.

The Armstrong-Meissner and Hartley versions are not shown since their feedback depends on inductor-coupling, not possible with quartz crystal units. The Vackar and Seiler versions of the Colpitts do not count here since their feedback comes from more-complex capacitor coupling that would add components having no usefulness with fixed-frequency sources.

Increasing Or Decreasing The Frequency

General

This can be done by heterodyning variable and fixed RF sources to create a newer band. This is common in multi-band transceivers, a band-switched set of fixed crystal-controlled oscillators mixing with a single variable RF source to create a set of same-frequency-span bands. The variable source will have the same tuning characteristic on each band. In some transceivers, there are two banks of bandswitched fixed RF sources, usually offset the same amount by the receiver IF. Transmit/Receive control selects which set of fixed RF sources.

Double heterodyning, at one time common with TV channel tuning systems, is good for a wide range of bands (channels) without requiring elaborate passive RF filtering for each band. An example is given in Chapters 41 and 42.

There are two major ways to create a high-HF, VHF, or UHF stable source based on a single crystal oscillator. The most obvious way, as done in older radios, is to ***multiply (by integers)*** a lower-frequency crystal upwards. Doublers and triplers were common. This was an advantage with World War II era VHF FM-PM radios where the crystal oscillator was about 400 to 900 KHz, modulated in PM, then that multiplied up to 96 times for *channelized* RF output at low-VHF.[16]

The next method was to multiply an ***overtone crystal*** in its output, the crystal unit usually cut for a higher harmonic content. That could be followed by more multipliers or direct amplifiers for VHF to low-microwave output. Output frequency was still controlled by a single quartz crystal.

Following that, an L-C controlled oscillator at the air frequency would then be ***divided,*** the divided frequency phase-locked to a single quartz crystal frequency. This began the ***PLL*** or *Phase Locked Loop.* With proper comparison frequency and some broad-banded design within the PLL, a number of individual frequencies could be generated, all of them controlled by a single crystal. With advances in IC production and design came the ***fractional-N PLL*** and, finally, the ***DDS*** or *Direct Digital Synthesis* system. The *fractional-N* and *DDS* could also create many crystal-

[16] AN/TRC-1, 3, 4 radio relay sets for 70 to 90 MHz carrying 4 simultaneous 3 KHz voice channels. Crystal units for those and the variations of FM vehicular mobile radios were the source for much of the surplus crystals available right after 1945 in the USA.

controlled frequencies, also by digital control.

In the two decades following the WWII period there appeared many different combinatorial heterodyne-and-divide systems, perhaps the largest being the Hewlett-Packard 5100 system which could generate any frequency 0 to 50 MHz in 0.1 Hz increments. That was fine as an accurate ***test instrument*** in the early 1960s but size, cost, and vacuum tube architecture would doom its life.[17]

Worthy of note as a much simpler method of accurate receiver tuning is the ***Wadley Loop*** designed by Trevor Wadley in the 1950s. See Appendix 29-1 for more on this Wadley system.

Multiplying The Basic Crystal Oscillator Output

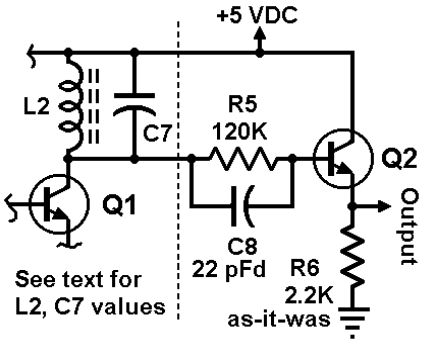


Figure 29-10 Modification of example oscillators for multiplying output 2x and 3x the fundamental frequency.

The basic oscillators of Figures 29-4 through 29-6 can be further modified to output second and third harmonics. A modification example is shown in Figure 29-10.

If the fundamental tank circuit is 5.0 MHz, then the modifications of Figure 29-10 will allow doubling to 10 MHz or tripling to 15 MHz at its output. This replaces R4 of Q1's collector with L2 and C7. Changes to R5 and addition of C8 help amplitude output for second and third harmonic.

Values for doubling and tripling:

Multiplication	L2	C7	Output
Tripling	1.1 µHy	100 pFd	0.3 V
Doubling	2.0 µHy	120 pFd	0.5 V

Output is peak-to-peak voltage

A word of caution on output: There is some PM on the waveform. The second and third multiples are still correct in the long-term, but there is some interaction with the base-emitter circuitry of Q1. Output should be followed by a bandpass filter centered around the doubled or tripled oscillator fundamental. The narrower the better to get rid of much of the PM.[18]

There is little fault if the multiplied output is applied to a PLL since the internal PLL counters will simply count over the PM so that its average frequency is a multiple of the fundamental. The PM may cause some spurious outputs if applied to a phase modulator for PM or FM RF output.

As it is, the modified oscillator circuit draws less than 2 mA. Decreasing the R5 value will improve multiplied power output amplitude about three times if changed downward to 27 KOhms. DC current drain can increase to 2.5 mA at +5 VDC with R5 at 27 K and a doubled output to 1.5 V peak-to-peak, tripled output to 1.2 V peak-to-peak. Most of that occurs in the Q2 buffer; Q1 current drain is about the same for L-C or crystal unit tank circuits.

---

[17] The HP-5105 arrived shortly thereafter for VHF-low-UHF to 500 MHz. Internal design rather similar to the HP-5100. An interesting system but too complex for normal homebrew laboratories. HP would replace this with simpler future instruments having more features.

[18] This PM was never mentioned in earlier tube-architecture articles, quite possibly because spectrum analysis of multiplier circuits was largely ignored. Phase Modulation still existed but following stages could damp out much of the PM by over-driving and narrow bandpass filtering in following stages..

Butler Oscillators

The so-called ***Butler Oscillator*** is one way to multiply a crystal oscillator frequency. A word of caution here: In the last half-century there has been a lot of confusion on ***naming.*** Some overtone oscillators are called *Butlers* while others are called *Overtone* or by their original configuration name such as Colpitts, Hartley, or Pierce. In the beginning it was common to call overtone oscillators by their original names while emphasizing that they had outputs tuned to an integral multiple of the crystal unit fundamental frequency. In the post-WWII era it was common to limit the number of stages (to keep production costs down) yet the origin of the ***Butler*** name remained obscure.[19]

Oscillator stages whose output was a multiple of the crystal unit fundamental frequency existed before there was any mention of a *Butler* name. They worked well enough. From then on the wordage got a bit extreme in some publications and names got a bit confused.

Butler/Overtone oscillators work best with quartz crystal ***series resonance.*** In addition, the crystal unit parallel capacitance can be ***tuned out.*** These are given in Figure 29-11. L1, C1, and C2 are tuned to the integral multiple of the series-parallel crystal unit fundamental. Feedback is supplied by the crystal to R_E. With BJTs or FETs, C2 would be about 2.5 to 3.0 times larger in value than C1.

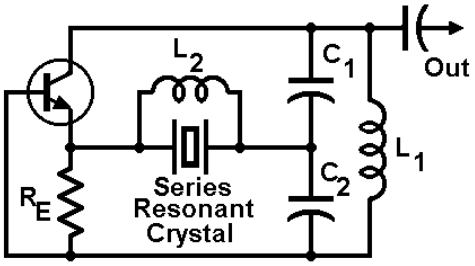


Figure 29-11 A simplified series-resonant crystal unit oscillator for overtone frequency generation.

The purpose of L2 is to minimize any parallel resonance with the crystal unit. This can be explained better by Figure 29-12 showing an equivalent model of a crystal unit at its fundamental and overtone frequencies.

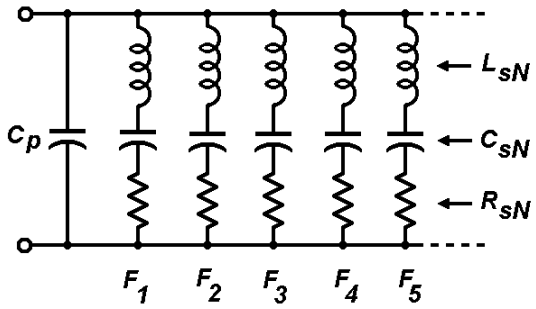


Figure 29-12 Equivalent circuit of an overtone crystal. Each series-resonant circuit is there for multiples of the fundamental frequency, F1.

In Figure 29-12 the overtone series-resonant circuits in parallel do not normally interfere with one another considering their very high Q. However, Cp, which can include a number of crystal unit parasitic capacitance can do so. Recall that there is very little frequency space between series and parallel resonant frequencies of the fundamental.

The purpose of L2 across the crystal unit is to effect a parallel resonance with Cp ***at the harmonic*** of fundamental crystal unit frequency**.** L2 can be a fixed inductor once the equivalent value has been found for Cp; it is not normally adjustable in-circuit.

The number of harmonics in a crystal unit depends on the crystal ***cut.***

PLLs or Phase-Locked Loops

[19] At least in a literature search. Much literature has been published but most of that is essentially repeats of previous work or specific boosts of a crystal unit manufacturer's model line. Many papers not produced by crystal unit manufacturers tended to emphasize so-called *simplified network analyses* using L-C equivalents, but for specific crystal unit cuts *as measured in their laboratory;* with different frequencies and different unit cuts, those might not apply, however their basic math was arranged. Some papers touted a particular method, devoting much to words in self-praise but little in detail of the method. It was most confusing to many hobbyists.

This combines digital electronics and voltage-tuned L-C oscillators to create a whole series of crystal-controlled frequencies which may be selected digitally. The typical block diagram is shown in Figure 29-13. The voltage-tuned L-C oscillator output is applied to a digital down-counter whose division is controlled by (some) digital control represented by ***N.*** The counter's output frequency is thus (***Fv / N***). That output is one of two inputs to a ***phase-frequency detector.*** The other input comes from a single-frequency crystal controlled oscillator represented by ***Fx***. The PFD (Phase-Frequency Detector) output produces a voltage to bring the voltage-controlled L-C oscillator frequency such that there is always a constant phase-difference, but always ***no*** frequency difference between the L-C oscillator and the crystal oscillator.

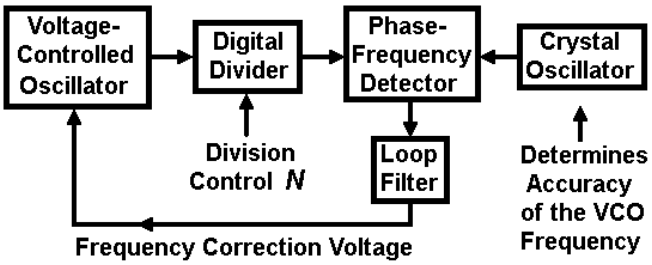


Figure 29-13 A generic PLL block diagram.

Output frequency is from the VCO (Voltage-Controlled Oscillator) and, when the Loop is ***locked***, its frequency is always (***N Fx).*** A single crystal-controlled frequency can determine the accuracy of several hundred individual VCO frequencies. All VCO frequencies can be selected by controlling the Digital Divider.

The Loop Filter converts the (normally) digital output of the PFD to analogue voltage for correction of the VCO frequency. Since Division Control is digital, many forms of manual control are possible, including selection by a microprocessor. There is more on the subject of PLLs and their Loop Filter calculation, plus DDS use, in Chapter 32.

The VCO should have a maximum to minimum frequency span of less than 2:1. This is to fit with the direct VCO control and to keep the Loop Filter within a normal range of values.

DDSs or Direct Digital Synthesis

DDSs are more complicated and, for IC devices, pack a few more things than the PLL circuitry within them. Their operation is generally the same as a PLL but they include a very fast digital-to-analogue converter for their output. Most include the VCO within, such as in the Analog Devices AD-9851, for output to 120 MHz. As with PLL topics, Chapter 32 also includes DDSs with more details of the AD-9851.

It should be noted that most DDS ICs also allow a ***greater detail of output frequencies.*** The AD-9851 can control frequencies down to ± 1.0 Hz increments at carrier frequencies of up to 100 MHz. As such it can be used to control frequency of a second or third LO in a receiver to control an incoming signal frequency for either USB or LSB or on-frequency (as with CW or AM) with clarity of demodulation. It can also be used for transmission the same way, or even as a signal generator sweep generator.

Some DDS ICs need a lowpass or bandpass filter of their output. This is the result of the IC internal design and noted in datasheets, partly as a result of unchangeable internal design. DDS ICs from Analog Devices have a software tool that both identified spurious outputs as well as the frequency control code in decimal, binary, and hexadecimal code.

Purchasing Complete Oscillator Units

There are a number of suppliers of individual crystal units plus complete oscillators, particularly in the HF to VHF range, available through the bigger distributors such as Allied, Newark, Digi-Key, and Mouser (to name a few). They come with several cautions:

For individual crystal units, you must ascertain whether their frequencies refer to ***series*** or ***parallel*** resonance. If your need is parallel, then what is the specified external parallel capacitor value? That may or may not affect a circuit.

For individual crystal units, you must either measure the equivalent circuits or take the manufacturer's data on that. Measuring an equivalent circuit can require more expensive test equipment and certainly more time than just measuring an individual capacitor or inductor.

Crystal units for overtone oscillator use require that a manufacturer specify the overtone frequency and, hopefully, supply an equivalent circuit for that.

You will need a ***tolerance*** on the frequency, whether crystal units or complete oscillators. *Normal* tolerance is ±50 PPM, equal to ±0.005 percent. Then there is the operating temperature range. Some have grading of different frequency tolerances for different operating temperatures. For difficult environmental conditions, you may need ***non-operating*** or storage temperatures.

For complete oscillator units, the choices are much larger. Output can be analogue or digital. Digital output is usually to standard +5 VDC logic but there is a growing trend towards lower digital supply voltages. If analog output, you need to find out the normal load impedance, at least that used in a manufacturer's production testing. Power supply voltage specifications are needed, although for many non-digital output units, they can usually operate at slightly lower supply voltages.

There is a great variety of packaging available, everything from somewhat large solder-in units to small plug-in types to SMT packages. Then there are pin-outs for the plug-in types, something not yet standardized. In short, buying a crystal unit or a complete oscillator requires a rather large mix of different specifications. An accurate frequency meter is necessary for either individual crystal units or purchased oscillator packages.

On the other hand, buying a complete crystal oscillator package can save a lot of time and trouble on the bench. If it can meet all of your required specifications, then it is ready to go. All you need then is normal safety precautions to make sure it remains good until after a project's last assembly.

References for L-C and Crystal Oscillators

There are, literally, hundreds of references available on the Internet. Most are re-writes of previous material and some concentrate too much on vacuum tube circuits or rather arcane devices which are not available now. There are a few which have *the right stuff* from the Internet in the author's opinion and are listed following. Due to the ephemeral nature of the Internet only the titles and other remarks are shown. As of 2013 these were all re-found via a browser's *search* feature.

[63]Any ***Tutorial*** by ***John R. Vig*** dated 1996 to 2008: Usually in a slide-show form detailing a great deal of information on quartz crystal oscillators, from crystal units themselves through environmental testing to application. Mr. Vig was with the U. S. Army, responsible for overseeing the ***SINCGARS*** family of secure field radios. *SINCGARS* is the current small-unit networkable, digital, frequency-hopping radio in the USA military inventory. There is lots of solid material in this Tutorial but some of it requires adapting into and digging out lots of material. Large.

[64] ***Principles of Quartz Oscillators,*** Hewlett-Packard Company, May 1997, document number 5985-7662H (may now be available from Agilent). Excellent tutorial, again concentrating on quartz crystals with relevance to frequency and time instruments. Many illustrations, graphs.

[65] ***Crystal Oscillators in CMOS*** by Bob Hay, ECE614 Project 2, Boise State University, College of Engineering. A slide-show of very concise oscillator circuit topics, including overtone oscillator circuits in addition to the Pierce crystal oscillator common in microcontrollers.

[66] ***Design of Crystal Oscillator Circuits*** by D. Neubig, DK1AG, approximately 1981. A large collection of circuits from 20 KHz to mid-VHF, apparently from several sources. It suffers from slight confusion in translation to English in places and includes some out-of-date semiconductors. It has some good information for all types of crystal oscillators.

[67] ***Crystal Characterization and Crystal Filter Design,*** by Nick Kennedy, WA5BDU, April 2008. Good collection of testing methods to derive equivalent crystal circuits.

[68] ***On Universal HF/VHF Low Noise Crystal Oscillators with Switching 4 Crystal Unit Possibility***, by Tasic Sinisa-Tasa, YU1LM/QRP (date unknown). Suffers some from translation to English but has considerable material on overtone crystal circuits and particular parts values.

[69] ***A Practical Test Set for Comprehensive Crystal Testing,*** by Chris Trask, N7ZWY, February 2008. Both a discussion of crystal characteristics and ways to test them plus a complete test set for measuring crystals.

Quite obviously there are many not included in this listing. The number of references turned out to be astonishing. Missing is a collection of Corning Frequency Control crystal units which had minimum to maximum ranges of crystal equivalent circuits for their 2001 *World Catalog.* Note: Corning Frequency Control was sold to Vectron International in 2004.

Appendix 29-1

A Single-Quartz Crystal Standard for a 0 to 30 MHz HF Receiver

The Wadley Loop

The time was the early third of the 1950s. A South African engineer, Trevor Lloyd Wadley, having worked for the British government during World War II, was now employed by a new communications company called *Racal.* Quartz crystal production had ramped down from its wartime peak and man-made quartz was not quite ready to fill the needs of the electronics industry. Wadley had an idea for an all-band, LF through HF receiver that could keep its accuracy dependent on a ***single quartz crystal oscillator.*** It became known as the ***Wadley Loop.*** It was first used in the Racal RA-17 from 1956 through 1962.

It would appear again as hand-carryable portable radio, the XCR-30, produced by the South African company called Barlow from 1969 to 1971. It then appeared in the Drake SSR-1 first produced in 1975. Then in the Yaesu FRG-7 in 1976 and, under the Realistic brand of Radio Shack, DX-300 and DX-302 by 1979. The Lowe SRX-30, produced in 1980, used it as well as the Standard C-6500 made in Japan.

How It Works

Figure 29-14 shows the block diagram of the Racal RA-17 front end. Manual Tuning over any 1.0 MHz span is controlled by the 2.0 to 3.0 MHz tunable IF, itself rather standard for its day. The ***VFO*** block is a continuously tunable L-C oscillator. More properly, that VFO can be thought of as a sort of analogue ***bandswitch without stops.***

In practice, the VFO is set approximately to integral MHz digits of the desired frequency. The ***Tunable IF***, in effect a conventional *monoband* receiver by itself, selects the remaining digits of the desired frequency.

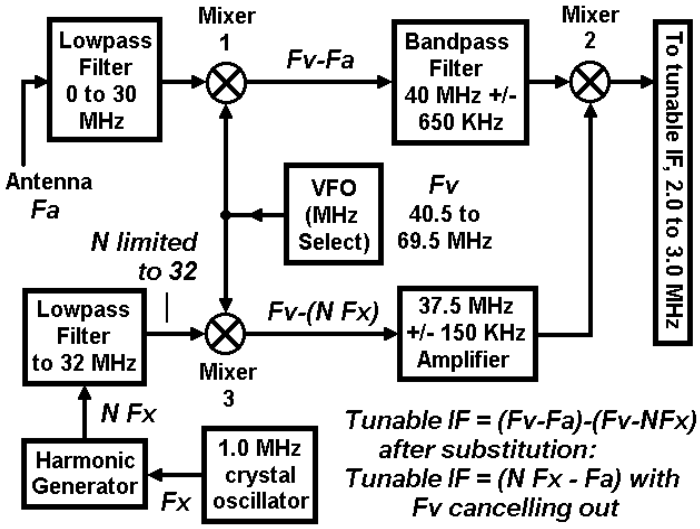


Figure 29-14 Front-end block diagram as used in the Racal R-17 with Wadley Loop to cancel drift.

The 40 MHz bandpass filter is fixed with a percentage bandwidth of 3.25% and the 37.5 MHz Amplifier BPF is also fixed with a percentage bandwidth of 0.80%. The system allows the ***same span*** of 1 MHz bands throughout all bands...all on the accuracy of a ***single quartz crystal oscillator.*** In this regard it could tune the same way as several other communications receivers of its time (notably from Collins Radio in their military line) but ***without*** nearly two dozen separate First LO crystal oscillator units. Remember that this was designed at an earlier time when quartz crystal units were not as plentiful nor were at the lower price of today.

Table 29-1 shows an abbreviated frequency table for the front-end of the RA-17. Note that the Tunable IF appears to *tune in reverse* to the Antenna input frequency band. That is really a minor matter for mechanical design or to procure a ganged tuning capacitor for the Tunable IF manual tuning. The various models since produced had general frequency range values to the original RA-17.

Table 29-1, Frequencies in the RA-17 Front-End, in MHz

Fa	Fv	N	Fv-NFx	Nfx-Fa
0 - 1	40.5	3	37.5	3.0-2.0
1 - 2	41.5	4	37.5	3.0-2.0
2 - 3	42.5	5	37.5	3.0-2.0
3 - 4	43.5	6	37.5	3.0-2.0
4 - 5	44.5	7	37.5	3.0-2.0
...	....	...	....	.....
25 - 26	65.5	28	37.5	3.0-2.0
26 - 27	66.5	29	37.5	3.0-2.0
27 - 28	67.5	30	37.5	3.0-2.0
28 - 29	68.5	31	37.5	3.0-2.0
29 - 30	69.5	32	37.5	3.0-2.0

Some of the Tough Design Problems

The 40 MHz bandpass filter had to have very good attenuation skirts to limit its acceptable percentage bandwidth to 3.5%. Racal used an *8-Resonator* type in their design, allowing for the losses inherent with L-C Qs. Similarly, the 37.5 MHz amplifier's bandpass filter, also 8-Resonator, added to the labor cost of initial alignment. Add to that the problem of maintaining a fairly even amplitude of 1.0 MHz harmonics, requiring more tube stages. While a single crystal unit saved one part of the cost, cost was increased in more stages and alignment labor.

The reason for specifications on the 40 MHz filter was to allow easier tuning at band edges. The VFO could be reset ± 1 MHz but the system allowed for maintaining reasonably good tuning. Lowest band operation came with a caution that limited low-frequency to about 0.5 MHz. There was a normal antenna problem at LF but there was little interest in LF and below at the time.

What didn't seem a problem was some of the Mixer spurious response effects that could result in some birdies in tuning.[20]

Future

Increase of solid-state technology, particularly having most of an LO on-chip, doomed the Wadley Loop HF tuning system. It was cheaper overall to use PLLs and DDSs and have frequencies digitally selected. It was produced for 24 years by 8 manufacturers. Just the same, the Wadley Loop was an excellent bit of innovative engineering.

[20] See Chapter 40 for the Multi-Band Converter project as well as Chapter 3 for spurious mixer products.

Chapter 30

VFAs, CFAs, and Related Circuits

The general term for *operational amplifier* has slowly been changing to *Voltage Feedback Amplifier* or *VFA*. There is more on VFAs here along with its follow-on, the *Current Feedback Amplifier* or *CFA*. This is an expanded version of original op-amps plus the introduction of closely-associated circuits used in and with VFAs and CFAs.

General

The ***operational amplifier*** came into being during WWII. Using vacuum tubes, its gain and frequency response could be set by *external components*, mostly resistors and capacitors. They worked well at slow speeds, such as for servo motor control in radar tracking units. In using tubes, they were limited to ***voltage feedback.*** In the two decades following, they picked up a new common term, ***op-amps.*** It persists. After the solid-state era had begun in the 1960s, there appeared the ***transimpedance*** or voltage-to-current converter. That was the birth of the lower-gain, lesser precision, but huge bandwidth ***current feedback*** IC block. Text material started to differentiate between ***VFAs*** (*Voltage Feedback amplifiers* or the op-amp-based circuits*)* and ***CFAs*** *(Current Feedback amplifiers).*

Operational/Voltage Feedback Amplifiers

Basic ***VFA*** or *Voltage Feedback Amplifiers* are shown in Figure 30-1. A voltage follower with a voltage gain of unity is in (A). Whatever voltage is on the positive input will appear at the output. Chief advantage of (A) is that its input impedance is moderate to high.

The VFA block itself has rather high voltage gain, usually about 80 db or so at DC. That block voltage gain is called ***open-loop.*** It is made to have a deliberate phase shift internally to prevent oscillation when feedback resistors are applied. Block voltage gain falls off at a steady logarithmic rate until it reaches the ***unity gain frequency.*** The most common unity gain frequency ia about 4 MHz. Below about 50

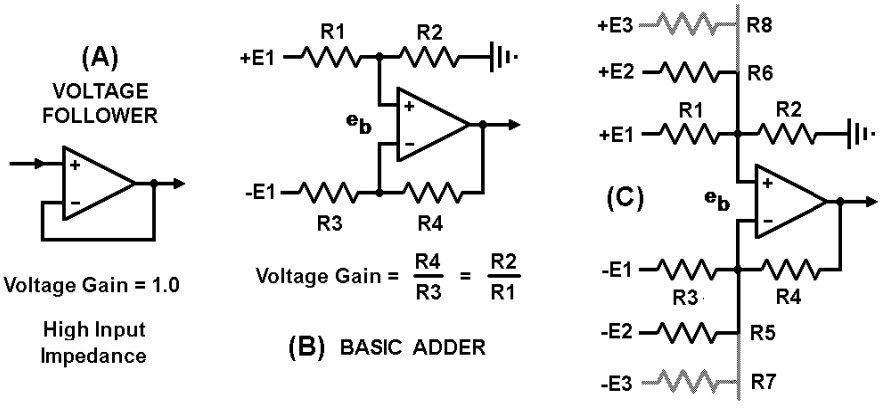


Figure 30-1 The basic VFA configurations

KHz it is still possible to achieve a specified gain under 20 db voltage with feedback as in (B).

In Figure 30-1 (B) the ***closed-loop voltage gain*** can be set by ratio of R4/R3 and R2/R1. Values of those resistors usually come from setting the e_B ***to ground*** as zero Volts and the differential potential of the positive and negative VFA inputs as equal. Note: Feedback will tend to force the equality of inputs' voltages.

R4/R3 must equal R2/R1 for both positive and negative-going inputs. If there is only one negative-going input, then R2 can be eliminated. R1 should be kept roughly the same as R3 so that the ***offset bias current*** has a minimal effect on ***input offset bias voltage.***[1]

Figure 30-1 (C) shows a multiple-input adder which can be expanded to many inputs of both polarities. Ratios of R4/R3 = R4/R5 = R4/R7 = (etc.) must still hold to establish the overall VFA voltage gain. Since R4/R3 also equals R2/R1, the ratios of R2/R6 = R2/R8 = (etc.) although R2 does not have to be exactly equal to R4. The ***ratios*** of R4/R3 must equal R2/R1 to keep the closed-loop voltage gain.

As long as the desired bandwidth is within the audio region, the small capacitance across the R2 or R4 will not affect the closed-loop gain. One way around that is to add capacitance across R1 and R3 such that their values are ***inverse*** of R2 and R4. That can be difficult with high gain since the feedback capacitance may be small and not easily measured. A saving grace is that added input capacity will be higher. It is best done with a lower-frequency sweep generator.

Suppose There Is Only One Active Input?

In the case of an ***amplifying inverter*** or a ***non-amplifier***, then Figure 30-2 would suffice. The reason for the third resistor has to do with ***offset current***, in particular with bipolar-input VFAs.

Bipolar-input VFAs use differential inputs, usually a variation of the dual transistor, common-emitter ***long-tailed pair.*** The common emitter is made to a constant current source. That raises the input impedance of bases of a bipolar transistor pair and also provides an opposite-polarity input to the other half of a pair.

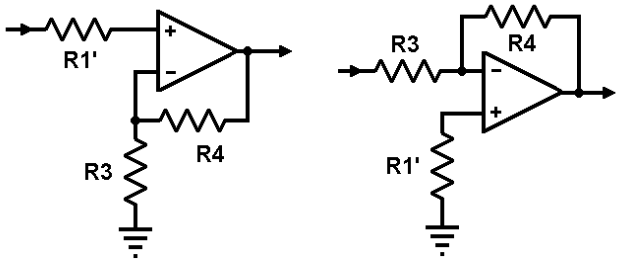


Figure 30-2 Using only one input of a VFA.

The R1' in Figure 30-2 is picked to be just slightly less than a normal R1 so as to equalize the resulting ***generated offset voltage.*** This generated offset is different than normal production tolerances of transistor pairs and may add-to or subtract-from any generated offset voltage. In cases of open-loop VFA voltage gains of 60 db or more, this offset voltage is usually less than 1% tolerance resistor values. But, it is there, and adding the third resistor will reduce such offset voltages.

Build-Up of Offset Voltages

[1] Datasheets for VFA op-amps will have those parameters specified. Those are important for those who are trying for high gain (> 20 db) over a wide temperature environment, but seldom for anything less stringent.

In a chain of VFAs offset voltage may reflect all the way to the eventual output depending on the overall gain. As an example, three stages, each with a voltage gain of 10, would add up to an overall voltage gain of 1000. A 5 mV offset in the first stage can result in a 5 Volt final output offset.

The cure there is to introduce an opposite-voltage offset at an earlier stage in the gain, using the *adder* configuration. Source for the added opposing-offset voltage should be from a regulated supply to keep it clean. Note that offsets can come from unequal bias currents in the VFA inputs dropping through external resistances. Manufacturers' datasheets almost always have both offset voltage and offset current maximums listed. Introducing an opposing offset to counter a built-in one is a simple matter of applying Kirkhoff's Voltage and Current Laws at DC.

For a high-gain chain of VFAs, an opposing offset voltage can be controlled by a trimmer resistor. This is not normally done in production since trimmers cost more plus the time spent by production technicians in aligning such conditions. Hobbyists have no such constraints, limited primarily by time available.

Differential or Instrument Amplifiers

In the circuit of Figure 30-3, inputs are at full differential connection. If FET inputs are available, the input impedances are very high. U1 and U2 form the input section with voltage gains of:

$$|A_{V\,(U1)}| = \frac{R_B + 2R_A}{R_B} = |A_{V\,(U2)}|$$

But, because of the connections to U3 inputs, inputs at E_1 and E_2 are opposed; i.e., they are differential. Note the vertical bars signifying *magnitude* of voltage gain.

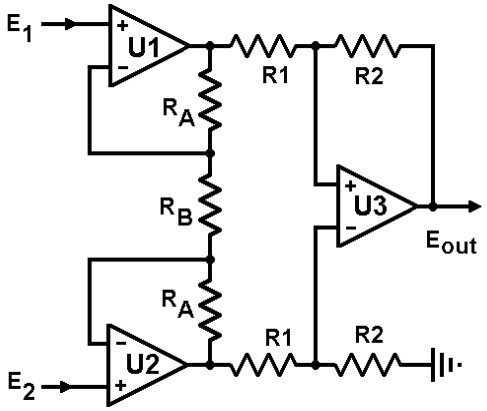


Figure 30-3 Fully differential amplifier suitable for instruments.

U3 is normally low-gain with additional voltage gain of (R2/R1) and most of the offset of U1 and U2 are reduced to a very low value. Using the same lot-number VFA blocks and low-tolerance resistors, the overall offset can be low even if the voltage gains are set very high. This is done in a later Chapter on a *Frequency Standard Comparator* project. In that circuit, total voltage gain from differential input to single-ended output is approximately 6300 times in the linear operating region.

Integrators

Integrators are usually described as forming a ***mathematical integral*** of an input voltage. Nearly all descriptions simply stop there. An integrator is basically a form of *lowpass filter.* Figure 30-4 shows such a typical integrator. At low frequencies the reactance of capacitor C is so great that overall circuit gain is as if it were not in-circuit. As frequency goes higher, the reactance drops and C becomes part of the feedback loop, dropping the gain.

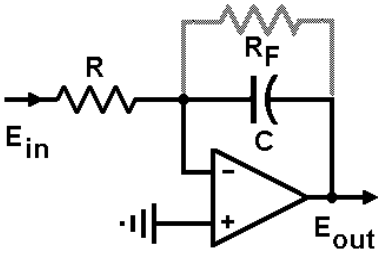


Figure 30-4 An integrator.

Frequency now becomes another part of the whole

circuit. Note the resistor R_F in Figure 30-4, in grey shading. It is there to set the overall gain at very low frequencies. Without that feedback resistor, overall gain would become the open-loop voltage gain at very low frequencies.

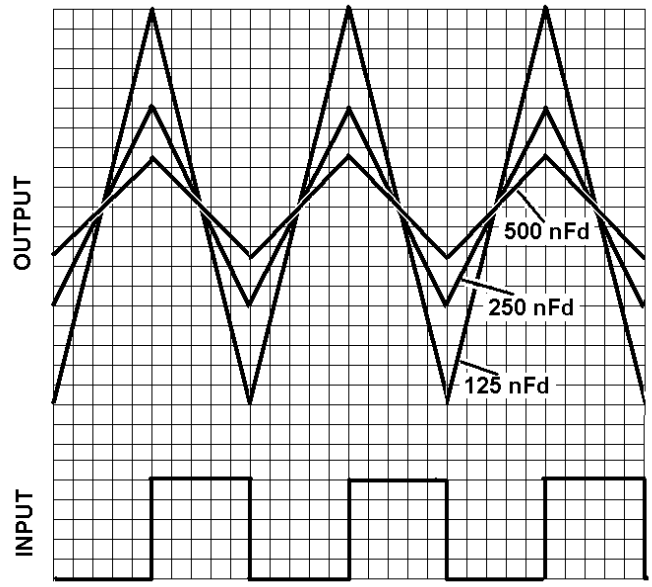


Figure 30-5 Waveforms of integrator in Figure 30-6 with constant low-level input plotted with same vertical scale as output. Period = 2 mSec.

To see the effect of differing capacitor values, examine Figure 30-5 for one frequency. In this case, the period is 2 mSec, R is 1 KOhms, C varied from 125 nFd to 500 nFd. Vertical scale is identical and the VFA block is operating in its linear region.

Note that output is inverted compared to the input. This is necessary since the integrator R and C both connect to the inverting input of the VFA block. DC gain is equal to or greater than 100 since R_F is equal or greater than 100 KOhms here.

Output peak-to-peak voltage is equal to the input when C equals 500 nFd at a frequency of 500 Hz. If the frequency is reduced, capacitance must increase proportionally. Output peak-to-peak voltage will increase if C is dropped in value, decrease if C is increased in value, both holding R constant. If R is increased (and R_F changed proportionally) then output will decrease with the same C value.

As R_F becomes lower and DC voltage gain decreases to less than 100, the linear rise and fall of the output will have a greater curve to it. That curve is barely visible to depart from linearity when DC voltage gain is about 10. The reason for that is the very high open-loop gain of a VFA block at DC. When the output voltage begins increasing, the feedback through C to the non-inverting input tends to make the output into a constant-voltage source. At high DC gains the whole circuit begins acting like the perfect lossless models depicted in many textbooks. Output voltage variation with C value is due to the amount of feedback from the output; a lower C value will have less negative feedback and AC voltage gain increases, hence the higher output voltage.

For operation within the linear region, the formula for a VFA integrator can use a modified integral expression of:

$$T = 2\,R\,C$$

Where:

T is time of Rise OR Fall in Seconds

R is series resistance in Ohms

C is series capacitance in Farads

As an example, pick R = 1000 Ohms and T = 1 x 10^{-3}, then solve for C:

$$T\ /\ 2\,R = C = 1 \cdot 10^{-3}\ /\ 2 \cdot 10^{3} = 500 \cdot 10^{-9} = 500\ \text{nFd}$$

changing its output. One way to do that is to add ***hysteresis*** to the triggering voltage. This is shown in Figure 30-8, using an inverting comparator powered by ± 5 VDC.

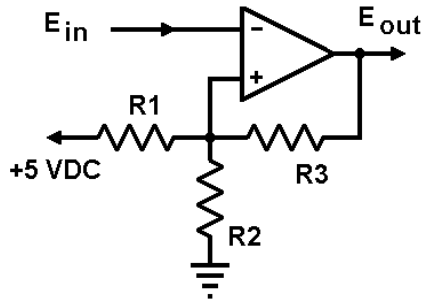


Figure 30-9 Hysteresis added to the comparator.

Hysteresis means that a voltage or current level going in one direction is different from the same signal going back to the same starting point. In this case it is advantageous to use it with a noisy signal to prevent bursts from noise at the trigger point.

Assume the circuit of Figure 30-9 and that the VFA output toggles between -5 and +5 VDC. Assume also that the input is at 0 VDC, R1 = R2 = 4.7 KOhms, R3 = 47 KOhms and, by connection, output voltage is high. Voltage at the non-inverting input is then:

$$E_0 = \frac{R_2R_3 \cdot V_{CC}}{R_1R_2 + R_1R_3 + R_2R_3}$$

Plugging in known values, E_0 equals 0.238 VDC. When the input exceeds this, output voltage falls and the new non-inverting input may be equated as:

$$E_1 = \frac{R_2(R_1 + R_3)\, V_{CC}}{R_1R_2 + R_1R_3 + R_2R_3}$$

Again, putting in known values will let E_1 equal 0.262 VDC. The input signal must fall below this new trigger level before the output changes state. In this example, the difference between the two trigger levels is 24 mV. That difference is often touted as the *noise immunity* of a comparator.

What happened here is a small amount of positive feedback added to the non-inverting input although not enough to make it oscillate; the VFA output will change its voltage abruptly from one limit to another. As a result the input signal will toggle on the lowest trigger voltage when VFA output is to go down, then toggle on the second trigger voltage when it is supposed to go up.

To increase the operating speed, a small amount of capacitance can be added in parallel with R3 to speed up switching. Most of the triggering rate is controlled by the VFA open-loop unity gain frequency. The higher the better for fast switching. In the real world, there are many variations on inserting this hysteresis. Several products are already on the market to use. One of the group is, surprisingly, the digital logic ***Schmitt Trigger***.[2]

Supply Rails: Double or Single-Ended?

Since most operational amplifiers were originally intended for multiple purposes, VFAs of today generally hold to that, preferring symmetrical ± supply voltages. That allows inputs to be referenced to chassis ground, to swing plus-minus from that ground level. That doesn't have to be

[2] Surprisingly, the Schmitt Trigger bears the name of its inventor, Otto H. Schmitt (1913-1998), a biophysics and zoology researcher. He simply had to ***make*** things in the pre-WWII years to further his research. The Schmitt Trigger was first conceived in 1934, using vacuum tubes. Schmitt did not develop it further, say to make it with semiconductors although many did after him.

fitm. Hobbyists can elevate inputs above ground and run VFAs from a single-ended supply rail. All that is required in design is that the (former) ground reference for inputs be moved to a voltage between chassis ground and the positive (or negative) supply rail voltage.

A few VFAs are designed to be either double-rail or single-rail. Those are an exception to the general rule. Usually such exceptions have notes in their datasheets to show how to use them. If not that, they will have Application Notes listed to show the same thing. The main thing to keep in mind is that, with differential inputs, linear operation is maintained when both non-inverting and inverting voltage inputs are near equal.

Outputs: Normal or Rail-to-Rail?

In the beginning of op-amps/VFAs, the output stages had limits on voltage swing. Those voltage swings could get within about 2 V of the supply rails. Any further voltage swing would result in distortion of the output. A cure for that was new versions of VFAs which had internal IC circuitry permitting output voltages within millivolts of supply rail connections. Naturally those *new rail-to-rail output* ICs cost more in the beginning. Either way, designers have to take into account the limits of output voltage swings in applications.

Current Feedback Amplifiers

General

Operational amplifier IC designers generally copied vacuum tube counterparts in the beginning. Customers were used to tube type designs. As a result, their unity-gain bandwidths tended to stay in the low-HF region for years. A few did pioneer IC design by taking radical steps on feedback, making such feedback ***current-dependent,*** hence the name of *Current Feedback Amplifier or* ***CFA.*** That allowed cleaning-up the IC inside plus allowing bandwidths of up to 100 times that of VFAs of the past.[3]

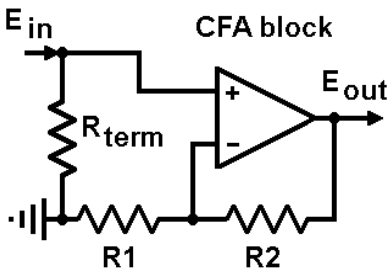


Figure 30-10 A general-use wideband CFA as described.

A CFA block has inverting inputs on the low side, closer to 100 Ohms, usually dissimilar in impedance. At the same time its ***gain-bandwidth*** is much higher than the common VFA block.[4] CFAs are excellent for video distribution, especially digital video. CFAs are used with closed-loop voltage gains of 1 to 10, seldom more.

Wideband CFAs

[3] Producible CFAs were first available about 1982.

[4] *Gain-bandwidth* refers to the useable bandwidth in Hz where the gain is approximately unity.

Shown in Figure 30-10, a general-purpose CFA block appears ***as if*** it was a VFA block.[5] The difference comes from the values of R1 and R2. ***R2 has been picked ahead of time and is found on datasheets.*** Try not to deviate much from that selected R2 value. Too much deviation will affect the bandwidth or can result in oscillation.

Input impedance of the non-inverting input can be described as *medium.* It is roughly on the order of older VFAs made entirely of BJT structures. R_{TERM} will probably be more like a shunt terminating resistor for an input of a wideband coaxial line. Input impedance of the CFA inverting input is, by contrast, ***low***. Just the same, using R2 as picked by the manufacturer, the voltage gain will be approximately (R2 / R1). R2 and R1 values are generally considered to be way too small for general-purpose VFA block amplifiers.

An Example of a Wideband Amplifier

This uses an Analog Devices AD8011, a low-power CFA block operating with +5 or ±5 VDC supplies. The objective is to make an amplifier good to 65 MHz with a gain of 10 db and operating with 50 Ohm source and load. The AD8011 made its debut in 2003 and was available in 8-pin DIP and SOIC.[6]

The -3 db bandwidth is 100 MHz for the non-inverting version and 130 MHz for the inverting version. Both meet the original criteria.

Note the series resistors on each output. Like many CFA blocks, the AD8011 is sensitive to load capacitance. About 39 Ohms for a 5 pFd load, 20 Ohms for 10 pFd, 14 Ohms for 15 pFd, and closing to 10 Ohms, all for less than 30 nanoSeconds settling to 0.1 percent on an output transient.

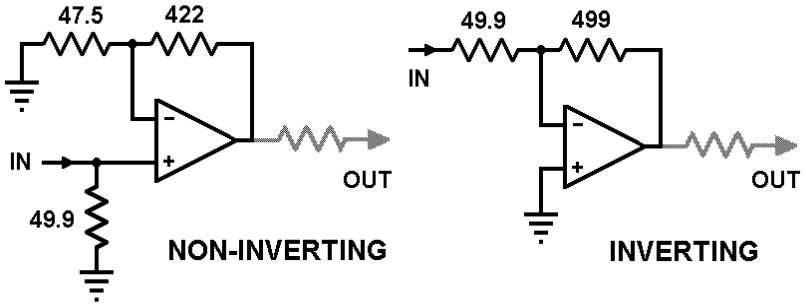


Figure 30-11 AD8011 as a 10 db wideband amplifier, run from ± 5 VDC supply rails with chip capacitors of 10 nFd and 10 µFd, in parallel, mounted close to the IC. [from Analog Devices]

Note the use of *settling time* as a specification. That is when all of the components in the IC will settle down after a step-change to an input. This must be measured with a very wideband oscilloscope, not always part of a hobbyist's test equipment.

Voltage gain of CFAs are seldom greater than 10 times the input. Just the same, the voltage gain generally follows the VFA scheme by dividing feedback resistance by input resistance at the inverting input.

Some CFAs intended for video distribution are characterized by source and load impedance

---

[5] Symbology for a CFA block as used in here is the same as for a VFA. Some of the semi-standardized symbology used by some makers is a bit too fixed and may not represent what actually happens inside. CFAs still have non-inverting, inverting inputs and (usually) just one output. Differences come from external components, not necessarily from the inside. Most of us can't get to insides of an IC to change it.

[6] This is not an endorsement of the AD8011. It may be obsolete when this is read. But, it is well-described at the Analog Devices website and becomes an object lesson for this part of the Chapter.

terminations, typically 75 Ohms as used in video broadcasting. Those CFA blocks will be tested and specified with both a series output resistor and termination resistor at the cable end.

Similarities and Differences of CFAs in regard to VFAs

Most of the VFA circuit configurations can be changed over to CFAs with due allegiance to the much-lower inverting input feedback resistances. That applies to Integrators and Differentiators as well. There is one caveat: Keep the inverting input stray capacitance to ground or common ***as low as possible.*** Remember that CFA block bandwidths are nearly 100 times wider in frequency than the average VFA; keep leads as short as possible.

CFA block non-inverting input impedance to ground is about *medium.* CFA inverting input impedance is generally ***low.*** CFA inputs are ***not differential*** as with most VFA blocks. Analog Devices AD8011 specification sheet indicates non-inverting input impedance as approximately 450 KOhms. It does not immediately state the inverting input impedance value.

Feedback resistor values supplied by CFA makers are hand-picked for a variety of specifications. One can use those values as a guide to change a CFA gain slightly, but only for a small change in value. For somewhat larger changes in gain, output should be checked on a very wideband oscilloscope (perhaps a spectrum analyzer) for oscillation or gross distortion.

Both VFAs and CFAs work well in their linear operating region. As such they might be used in transmitters, but only for relatively low power levels. In general, VFAs will simply clip the tops of waveforms that exceed linear-input levels. Some CFAs will exhibit the same action but that is more guess than fact. CFA use is fairly limited relative to VFA applications.

Transimpedance or Input-Output Converters

General

As can be guessed by now, the electronics industry uses its own jargon for certain actions. Vacuum tube basic operating and analysis principles used ***transconductance*** to refer to small changes in plate current to small changes in grid voltage. ***Transimpedance*** (particularly for CFAs) refers to small changes in collector current to small changes in base current.[7] One of the first uses was in photodiode signal amplification as shown in Figure 30-12.

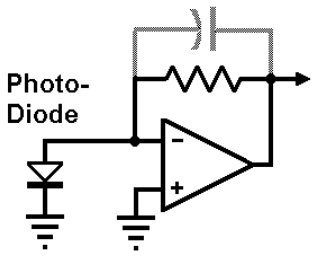


Figure 30-12 A photo-diode trans-impedance amplifier

The photodiode is back-biased relative to the VFA, is thus in its high source impedance state. Both VFA block inputs are at ground courtesy of the non-inverting input at ground. As the diode conducts more from photon input, the inverting input tends to go more negative while the VFA output goes more positive. Inverting input impedance is high, matching the reverse-biased photodiode impedance.

The capacitor (shown in grey) in Figure 30-12 tends to damp out any oscillation since it restricts the higher frequencies by reducing bandwidth. This doesn't work well with low-impedance CFA inputs.

Changing the photodiode into a capacitive sensor of high resistive impedance and keeping

[7] ***Transimpedance*** is a kind of catch-all word which can apply to practically any transistor junction.

the capacitor across the feedback resistor will make Figure 30-12 into a ***charge amplifier.*** Output voltage of the VFA block is the sensor capacitive charge divided by the feedback capacitance.

Sample-and-Hold Basic Circuit

This does not catagorize well but does manage to fit the bill for holding a charge as a voltage. There are many versions of this and all can be used with A-to-D (Analog to Digital) converter inputs, in the preamplifiers of low-end sampling oscilloscopes, and in recording music and voice. Main purpose is to ***sample*** a voltage level, then to hold that sample. A generic diagram is shown in Figure 30-13. This has voltage followers at input and output acting as voltage-gain-of-one buffers.

At rest, the switch armature is down and nothing appears on the output. If the switch is moved so that its armature is up, then C is charged by the voltage out of the left-hand voltage follower. The right-hand voltage follower repeats the capacitor voltage charge. When the armature returns to down position, the voltage charge on C remains and output continues to hold the C voltage charge. C will eventually leak off its charge and output drop to zero.

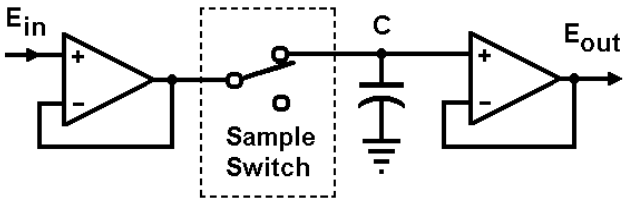


Figure 30-13 A generic sample-and-hold

The *Sample Switch* is always electronic. It can be a FET or a diode bridge or part of a functional block such as an 8-pin DIP National Semiconductor LF198 . Primarily it must have a very high impedance to the source when ***open*** and a very low impedance when the switch is ***closed.*** Depending on the application, the ***aperture time*** or time when switch is closed, will drive the electronic switch equivalent design. ***Hold time*** is the capability of the capacitor to hold a sampled voltage when the switch is closed. Sometimes ***droop*** is specified instead, referring to the held voltage dropping after the switch is opened.

Peak Voltmeter/Detector

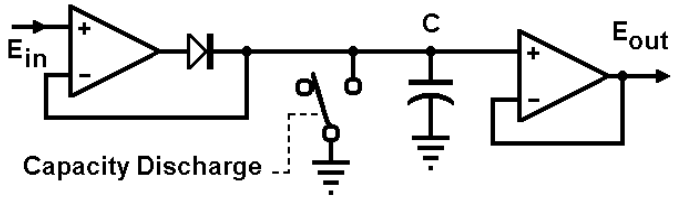


Figure 30-14 A generic peak voltmeter/detector

A close cousin to the sample-and-hold is the ***peak voltmeter*** or ***peak detector.*** Major difference between the two is that the input amplifier has a diode, either in series with the feedback or outside it. When the *switch* remains open in Figure 30-14, C will charge to the highest input voltage.

The output will remain pretty much as-is until the *switch* is closed to discharge capacitor C. The charge on C will remain depending on input impedance of the output amplifier, any losses of the input amplifier's output, and any other losses to ground, even the dielectric resistance of the PCB structure it is mounted on.

A monolithic sample-and-hold IC can be modified to be a peak voltmeter by inserting the charging diode and rewiring the open/close switch across C. Or it can be used as-is with a low-speed oscillator (manually adjustable) running the sample switch circuit.

Common to All Monolithic VFAs

Differential Amplifier or *the Long-Tailed Pair*

The first operational amplifiers used a dual-triode vacuum tube for the input stage, as shown in Figure 30-15 (A). When those became monolithic as semiconductor ICs, it took the form of (B).

The common name of *Long-Tailed Pair* grew out from schematics of tube type differential circuits. Basically this was for the high-resistance common cathode resistor that acted as a quasi-constant-current source.[8]

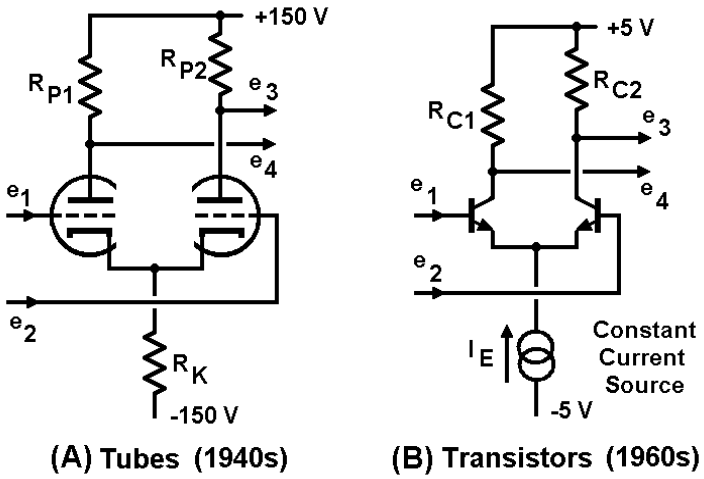


Figure 30-15 Differential circuits, vacuum tubes of 1940s (A) and (B) the first transistor work-alikes.

The common cathode resistor also allowed a ***phase split*** between the outputs and also to make the amplitudes at the plate equal. In Figure 30-15, the phase of e_3 is equal to e_1 but the phase of e_4 is opposite to e_3. This made a differential amplifier a preferred differential output driver for high-end music system amplifiers.

By the time of the monolithic semiconductor, constant-current sources could be made on-chip as shown in Figure 30-15 (B). This took only a maximum of five transistor junctions using a compensated *current mirror* or equivalent. The first monolithic VFA was a better performer overall compared to a tube circuit and used less DC power. Transistor junctions and (relatively small) resistors could be made on-chip and, with ingenuity and imagination, there could be capacitorless coupling from input to output. That, along with a good balance of characteristics allowed good ***common-mode rejection*** and maximum input voltages of nearly the full supply rail potentials.

Common-mode inputs refer to the same phase of signals applied to both inputs (e_1 and e_2 in Figure 30-15). Differential inputs should be isolated from one another to use them properly.

In the circuit of (A) there is no danger of any grid current changing the bias of either triode. There will be bias current with a BJT differential input. But, since transistor junctions can be made with FETs that have such low gate bias current, that can usually be neglected. What must be observed, especially in DC-coupled chains of VFAs, is ***offset voltage*** and ***offset current*** at VFA inputs. Those specifications are given on datasheets.

A Quick Trial of a BJT Differential Amplifier

[8] While a constant-current source could be built with tubes, the added heat, power dissipation, and cost tended to be prohibitive. The first plug-in operational amplifiers from *George A. Philbrick Researchers* used only two dual-triodes, the whole assembly slightly larger than the octal-base plug that was its base. Those Philbrick units depended on common-cathode resistance to substitute for constant-current sources. Philbrick grew and eventually became *Analog Devices.*

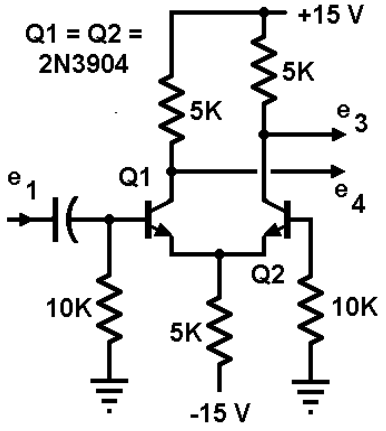


Figure 30-16 A quick trial model differential amplifier.

Largely from memory, the circuit of Figure 30-16 was modeled in LTSpice IV. Transistors are stock parts that were already becoming defacto standards in the 1970s.[9] There was no thought of optimizing this amplifier for any reason.

To begin, base voltage is very close to ground courtesy of the 10K resistors. Emitters would be -0.7 VDC relative to base voltage. The 5K common emitter resistor will drop about 14.3 VDC. That will make emitter currents to be 2.86 mA total or 1.43 mA per transistor.

For a β of 100 to 300 for Q1 and Q2, base currents will each be roughly 14 to 5 μA. The higher the β the better. Even with a β of 100, base-to-ground voltage will remain close to zero.

Collector voltages will be between +15 VDC and ground. In this case, using equal resistors of 5K and collector currents of 1.43 mA, there is a 7.15 V drop across the 5K collector resistors so the collector voltage to ground would be 7.85 VDC. There should be at least a ±4 V signal voltage swing at the collectors relative to ground, putting them at +3.85 to +11.85 VDC.

Power supply drain was 2.86 mA from each rail, close to 85.8 mW total at zero signal. Each collector resistor consumed 7.15 mW and the common cathode resistance took 40.9 mW. That left 30.6 mW dissipated in both Q1 and Q2 or 15.3 mW per transistor. That is very little compared to vacuum tube work-alikes but very important to IC designers.

The single signal input was picked as 10 mV peak-to-peak in analysis. That resulted in a collector voltage swing of 676 mV peak-to-peak at Q2 but 682 mV peak-to-peak at Q1. That is normal for this circuit. The 6 mV difference is only 0.86 percent but it exists. Reversing the single ended input to Q2 base resulted in a mirror effect. It is inherent in the circuit, not the model.

The reason that Q2 has a nearly-equal collector voltage (also in-phase with input) is due to the *long* common emitter resistance. It acts as a *pseudo-constant-current* device to keep the total emitter current the same for both transistors. As the input signal swings positive, Q1 total emitter current rises. But, with the large resistance in the common emitters, Q2 total emitter current must drop to equalize total current demand. In effect, Q2 emitter current drops, Q2 collector voltage to ground increases, all the while the Q1 emitter current was rising and Q1 collector voltage dropping. This circuit turns out to be a more perfect *phase inverter driver* circuit than its distant predecessor, a single triode with near-equal unbypassed cathode and plate resistors.

Finding the voltage gain of the circuit is more difficult. Analysis shows that voltage gain is about 68 times. At these emitter currents, the 2N3906 has a β of about 300. Input voltage excitation is rather solid at 10 mV peak-to-peak but the input current is only known mathematically through some laborious steps. Analysis with a SPICE program is much easier. With SPICE the AC base current is about 460 nA and AC collector current is about 138 μA, both peak-to-peak. AC β is also close to 300; at these low currents both DC and AC β should be close.

Input impedance is 10K in parallel with each base to ground or 21.6K measured. That would make it as 6.84K total. Again, this is easier to do with a SPICE program, using a series resistor and a 20 mV peak-to-peak excitation source and varying the series resistor until 10 mV is obtained at each base to ground. The series resistance determines the *magnitude* of input impedance.

---

[9] The 2N3906 is the PNP equivalent to the NPN 2N3904. Those are generally termed ***complimentary symmetry*** as are some other pairs such as the 2N4124 (NPN) and 2N4126 (PNP).

Constant Current Sources

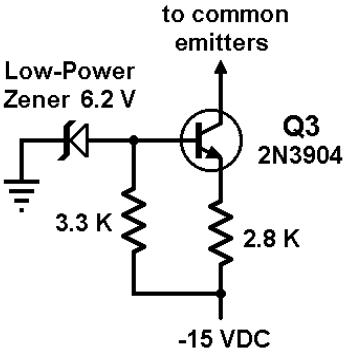


Figure 30-17 A simple constant-current source.

The circuit of Figure 30-17 can be used as a substitute for a single large resistor common to both emitters. It offers some improvement.

This circuit begins with picking a zener reference. Here a 6.2 V zener was selected. Base-to-ground voltage of Q3 would be fixed. A low-power zener diode here needs only a few mA supplied by the 3.3K resistor. Resistor voltage is 8.8 V and current is about 2.7 mA.

Higher-power zeners can be used but would need more current.

Assuming a β of roughly 200, Q3 base current will not upset the zener voltage much. Q3 base voltage to ground is -6.2 VDC and Q3 emitter voltage is then -6.9 VDC. Voltage across the emitter resistor is then 8.1 V and the emitter must conduct 2.86 mA. Resistance would then be 2.832 KOhms. That could be done by a series connection of a 1.8K and a 1K resistor for 2.8K. Collector current would be about 2.89 mA instead of 2.86 mA or about 1.14% high. Getting rather close to the required emitter resistance, a substitution of this constant current source does little in comparison to the original using a single large resistor.

When analyzed, replacing the common emitter resistance of 5K in Figure 30-16, this constant current source balanced out the Q1 and Q2 collector-to-ground voltages. It was better than the pseudo-constant-current single-resistor system. All other characteristics appeared unchanged.

Current Mirrors

Current mirror circuits are generally used in both VFAs and CFAs as constant-current sources, principally at the common emitter or common source of differential amplifiers. Unless there is a breakthrough in IC making for the home workshop, there is little point in describing them further for building by hobbyists.[10]

Remainders

Except for ***logarithmic detectors, A-to-D and D-to-A converters,*** and ***switch arrays***, the remainder of circuit blocks within a modern IC are essentially the same as can be built from discrete parts. The small die size and very small transistor junctions within an IC offer many features that make them operate better than their larger discrete component cousins. They are worth mentioning for the sake of completeness.

[10] It would be nice to explain them in detail, but the calculations are difficult and amount of temperature testing to ascertain operation is lengthy. There are limits in books intended for hobbyists and few have semi-conductor fabrication machinery for their own purposes.

Logarithmic Detectors

These are a chain of low-gain, limiting amplifier-detectors whose detector outputs are summed to the output. Response of each amplifier-detector is amplitude distorted so that the summation of their outputs equals a logarithmic response relative to the IC input. They are invaluable for presenting a wide dynamic range signal amplitude in a *compressed amplitude* form. That may be brought out as an analog signal for decibels or reconverted to Watts or anything else.

In conjunction with a microcontroller, these have been built into small instruments for power measurement that can be read from -100 dbm to +10 dbm without changing any front panel control. By using a 2-line alphanumeric display, the second line can also (mathematically) convert the dbm value to Watts, including adding the proper prefix to Wattage.

A-to-D and D-to-A Converters

While Analog-to-Digital is different from Digital-to-Analog in circuitry, their use is becoming widespread in consumer electronics and in radio all the way up into UHF. Consumer audio and video recordings are either in digital form or will likely standardize out to ***digital format*** in the near future. For consumers they are purity in themselves, free of clicks, scratches, and other noise of analog-medium recordings.

While in digital format, signals may be connected in many varied ways to demodulate a received signal. Rate of data will determine the circuit blocks used. USA DTV is presently about 6 MHz, but in compression from about an 18 MHz source. Time coding from NIST's WWVB 60 KHz transmitter is 1 bit per second. Digital sampling oscilloscopes have very high rates of digitization but the organization of data is more convoluted.

Amplitude sampling will determine the fidelity and amplitude of recorded music and voice. The *Nyquist criterion* for sampling says that the sample rate, which also determines the playback rate and fidelity, be ***two to three times the highest analog rate*** for best recovered audio. With voice and music compact discs, stereo could be captured with about a 100 KHz sampling rate, that including separate audio channels.

Switches in ICs

As a sidelight in IC devices, semiconductor switches were developed for various uses. One unlikely prospect was to use them as ***mixers for radio***, particularly for HF amateur radio. The *Tayloe Detector* (patented) used that for a so-called ***direct conversion*** receiver whose local oscillator was at or very near the carrier frequency of CW and SSB voice signals. Those would be nearly as good as multi-stage converter receivers although with some limitations on decoupling modes.