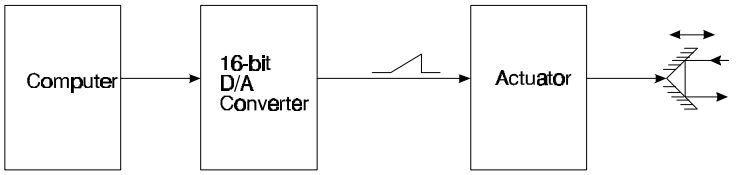
Digital-to-Analogue Conversion

The problem of digital-to-analogue conversion and vice versa is the problem of how to simulate a system which may take an infinite sequence of values of voltage at an infinite sequence of times, with a system which has a finite set of voltage values and may be updated only at a finite rate. This mismatch is at the heart of, and the cause of, most of the problems with conversion from the electronic world of volts, amps and time to the computer world of clocks and finite word lengths.

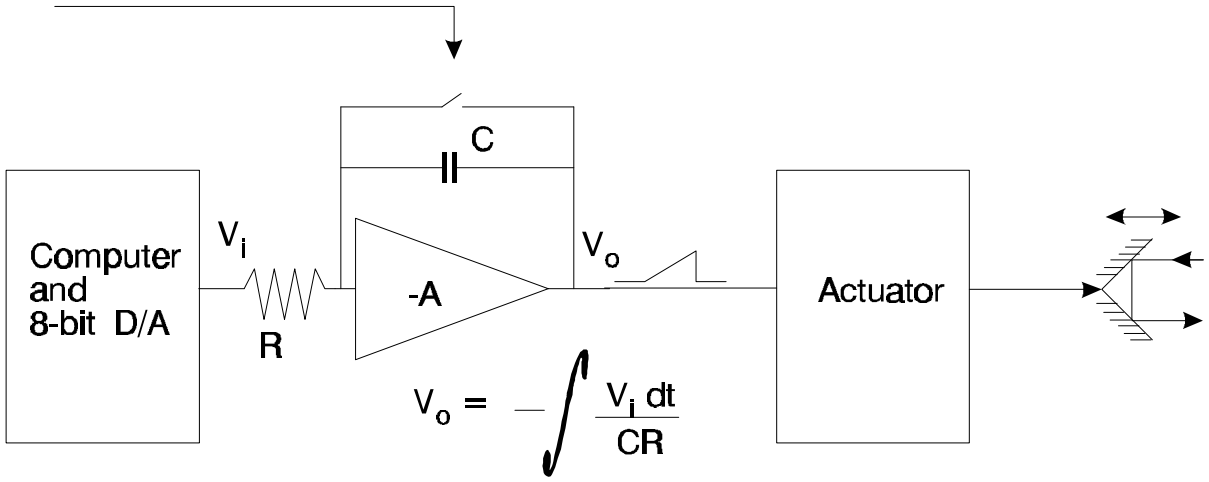
In general it may be said that the more money you pay, the closer to the real world you can get in terms of resolution in voltage and time. The question to be addressed is whether this is an altogether desirable state or if a little thought can produce a result which is as good, or in some cases better than, the expensive solution using only a few cheap parts.

To give a simple example: Suppose we wish to slowly move a mirror using a piezo-electric transducer and it is important that the motion be linear, i.e. not stepwise. We could spend a tremendous amount of money and get a very accurate Digital-to Analog Converter (DAC) and feed a count sequence to it as shown here.



Moving a Mirror - D/A Converter

As an alternative we could use a less accurate DAC and follow it by an integrator which will necessarily produce a smooth output as shown in the next diagram.



Moving a Mirror - Integrator

Now in order to reset the integrator we may need a couple of extra parts but the result is much nearer perfection than can be obtained reasonably. To give some figures: A very good DAC can produce 2^{16} states or a resolution of 0.2mV in a 10v swing. A simple integrator can easily better that. However we have lost some flexibility in that we cannot now arbitrarily choose the rate of the ramp, being limited to 256 discrete rates if we use a simple 8-bit DAC in our design. However often when moving a mirror in an optical system 256 rates are plenty[18]. In a practical system we also should give some thought to measuring where the mirror is and how to return the mirror to the rest position - but this is only an example.

A further advantage of the integrator system is that it relieves the computer of the ramp timing and therefore makes the time constraints of the program easier. Again with figures: If we ramp through 10v in 1 sec at 2^{16} resolution, we require updates every 15µS which is a substantial job for most micros (an 8088 can just about make it with optimised code) whereas our integrator solution is a "set and forget" thing where we only need to check at the limits - if at all.

---

[18] It is an observed fact that most controls are only used in 2-3 positions for any given problem. Since producing very wide range controls is an expensive proposition in general, careful attention to the required range is a very good and thrifty thing.

Thus a little, or much, thought given to the strategy of any converters will probably repay itself. However in order to have constructive thoughts it is necessary to know a bit about how things work. I therefore plan to run quickly over the main philosophies of DACs and then to discuss their relative strengths and weaknesses in a fair degree of detail - this is necessary because many a good experiment has been ruined, or been less than the best, because of inadequate thought at the planning stage.

Practical Devices

DACs come in a number of styles with various properties. The key parameters for the unit are:

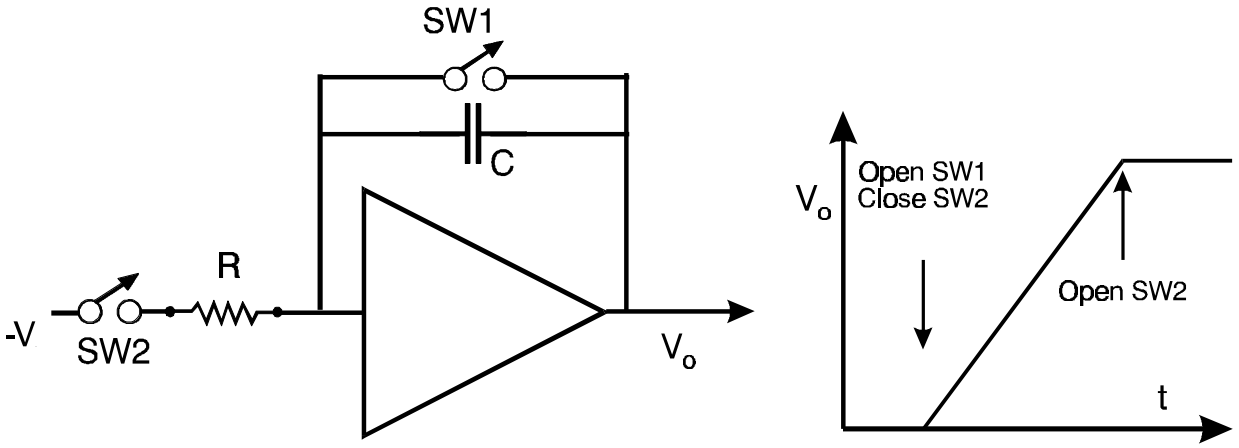
Resolution - what is the smallest average change in output voltage which can be effected?

Accuracy - what is the relationship between the output and some standard of voltage? This parameter needs several other parameters such as linearity to explain properly.

Speed - how fast can the output change? Important but less easy to quantify is: what does it do while changing?

Integrating Converters

The simplest illustration of the above is a converter relying on an integrator and time:

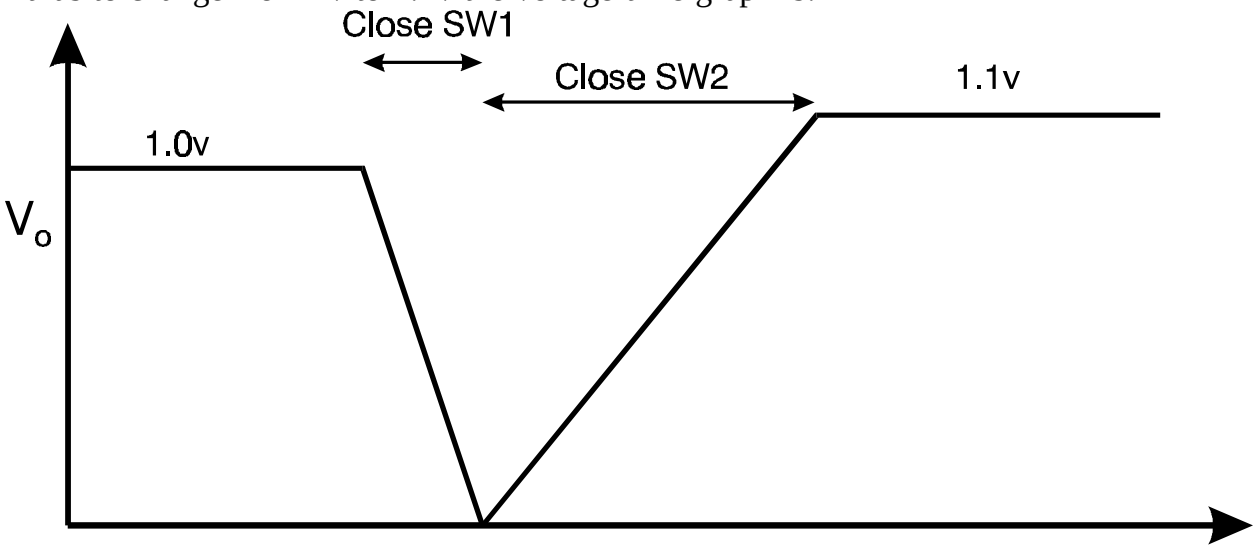


An Integrating D/A Converter

The computer closes SW1 for a long enough time to discharge the capacitor and then opens SW1 and closes SW2 for a finite time, t_1, after which the voltage at the output is:

$$V_o = \frac{V t_1}{C R}$$

Notice that the resolution is determined essentially by the timing resolution which can be made arbitrarily fine by the use of a short basic time interval counted for many counts. The accuracy is determined by the constancy, linearity and perfection of V,C and R. The speed is slow to very slow. Notice that the converter does very odd things while changing its value - thus to change from 1v to 1.1v the voltage-time graph is:



Integrating D/A - Changing the value

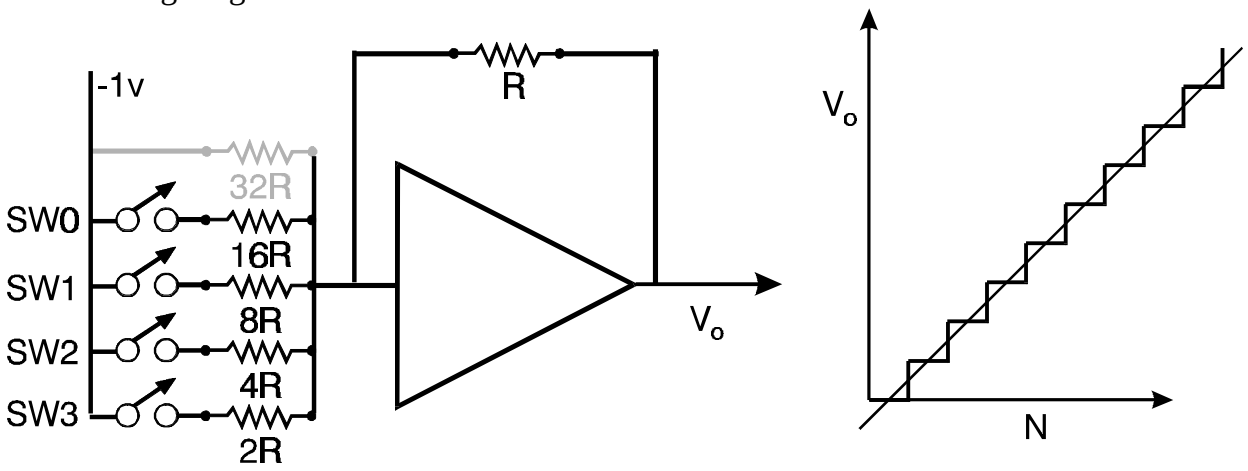
and just to finally confuse the issue, notice that since no op-amp is perfect and no capacitor is perfect there is a tendency for the voltage to drift from the set value after some period of inactivity.

Some of the problems can be alleviated by using two voltage sources of opposite polarity and then running the converter from the appropriate reference. However the value will still steadily drift from the correct one after some time. This may not be a problem if the system is in a servo-loop where the exact value is under some other control. Using bi-directional references the movement from 1 to 1.1 volt is a smooth ramp of defined slope.

Although it is the simplest converter and has some intrinsic merit - the counting converter is very little used.

Switched Converters

The commonest form of DAC converter is the switched sort which is illustrated by the following diagram:



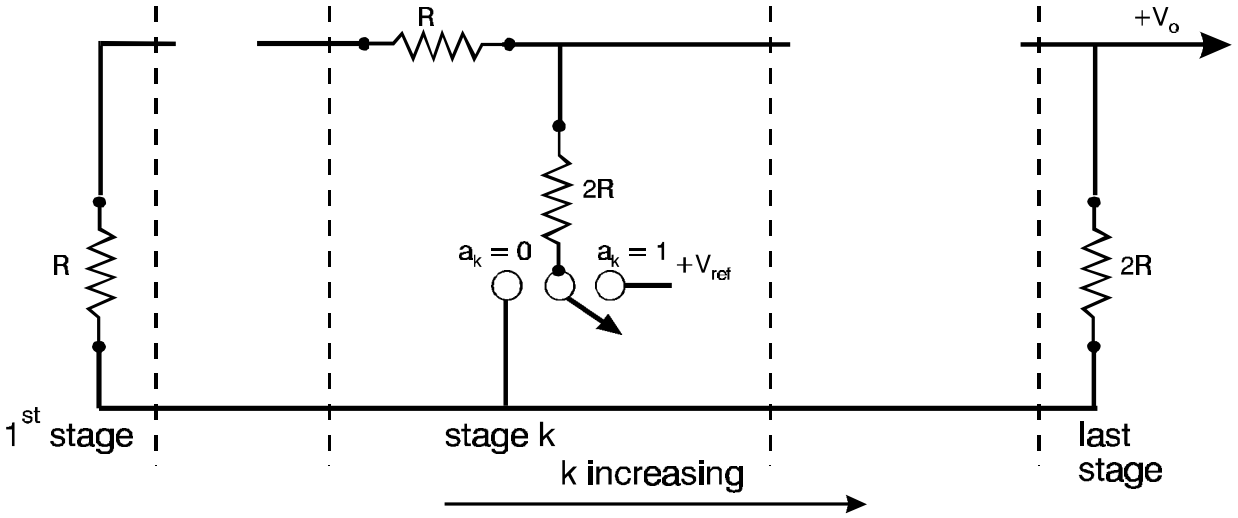
Summed Source D/A Converter

The setting of the switches SW3-SW0 sets up a voltage of 0 - 0.9375 v in increments of 0.0625 volts. The first thing that we can see which is a problem here is that if we plot a graph of V vs N where N is the number on the converter and then run a "best straight line" through the results, we find that the line does not pass through the origin. An alternative statement is that the fit to a line through the origin is within +0 or -1 LSB of the line. We can make a better fit by offsetting the whole line by +0.0313v with an extra resistor which gives a fit of ±1/2 LSB to a straight line through the origin[19]. We can now see clearly the

[19] An alternative approach to the problem of correct representation is to use a converter without any offsetting and then regard the number N as a real (ie floating point number). If we then round the floating point number as we apply it to the converter (such that fractions ≥0.5 are rounded up), then the regression line goes through the origin and the output for 0.000 input is 0V. This illustrates an important principle that *software* can replace *hardware* sometimes with great advantage. The reverse is also true - hardware can replace software at times, also to great advantage. The trick is to know which is better when!

problem of mapping an infinite set of voltages (the straight line) onto a finite set of values (the staircase) and the clincher is that an input of 0 comes out as +0.0313v and is within the accuracy statement of the device!

Even worse things happen if we use real resistors - say 5% accuracy. If the four top resistors are -5% then the output for state 7 is 0.4688*1.05 = 0.4922v. If the bottom resistor is +5% then the output for 8 is 0.5313*0.95 = 0.5047V. The difference between 7 and 8 on the output is then 0.0125V instead of the expected 0.0625V - a substantial, even worrying change. Some of the problems inherent in this design can be overcome by the use of more sophisticated networks and resistor chains instead of the simple scheme used above.

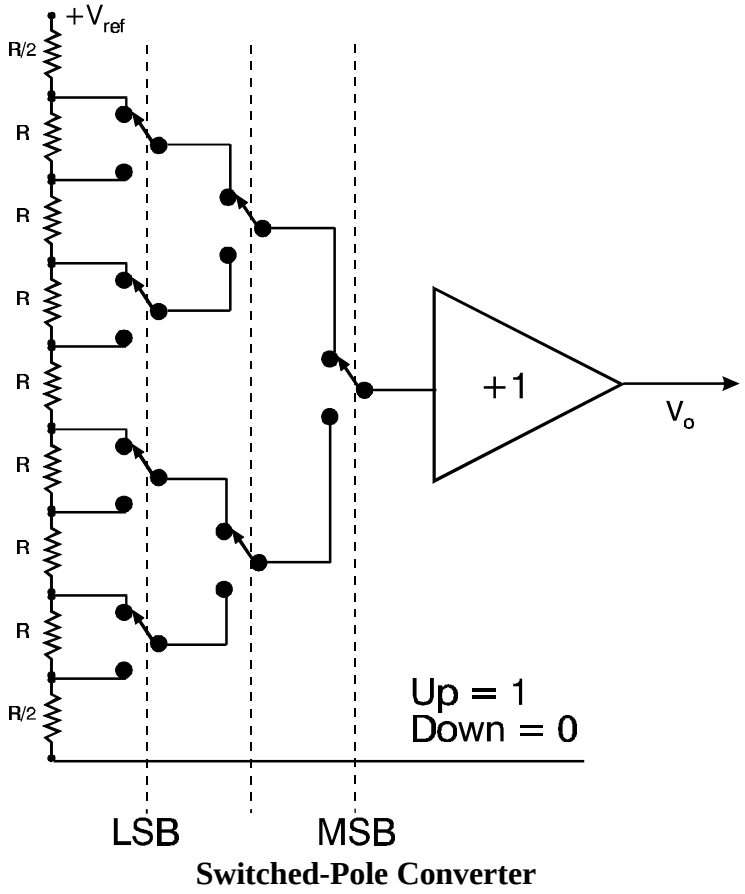


An R-2R Ladder

One common alternative network is the "R-2R ladder" which uses only resistors of value R and 2R (or just one value if 2R = R + R). The problems of matching a large number of resistors of the same value are much lower than the problems of matching resistors of disparate values such as in the scheme above. A sequence of stages is then set up so that each stage produces an appropriate weighting for the system (see accounts in Millman: Microelectronics for details).

$$V_o = \frac{V_{ref}}{3} \sum_{k=0}^{k=n-1} \frac{a_k}{2^k}, \qquad a_k = \{0,1\}$$

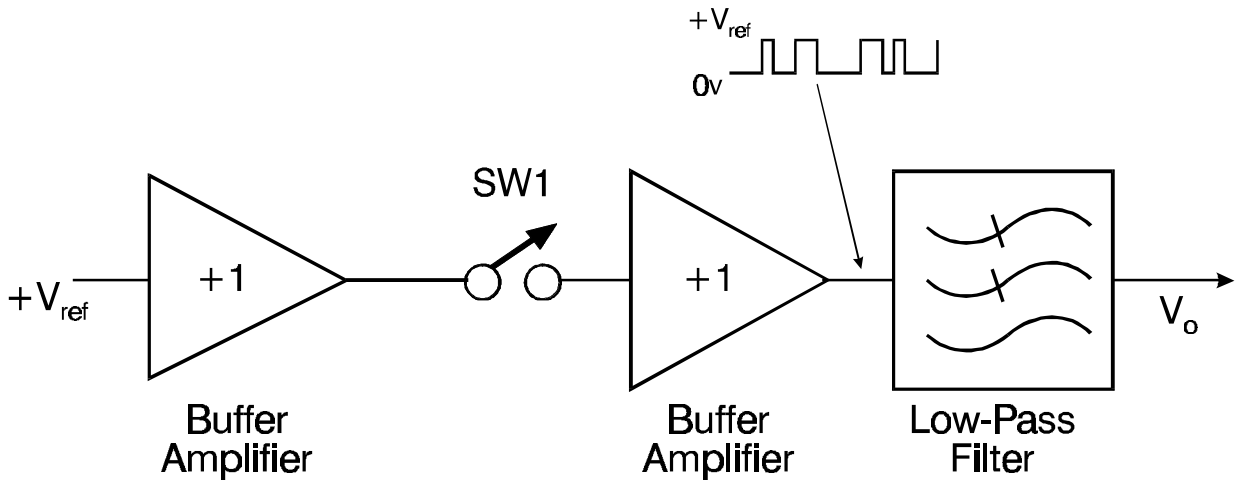
Another scheme which is gaining in popularity in integrated circuits is to use a resistor chain as shown here. This scheme simply involves the selection of the appropriate "tap" off the resistor chain with a set of switches. Its strength is that the output voltage must at least increase with increasing N - which the circuits above are not constrained to do - and it uses large numbers of identical parts which is terrible in discrete circuits but heaven for makers of integrated circuits. Notice the use of the end resistors to offset the converter output.



Switched-Pole Converter

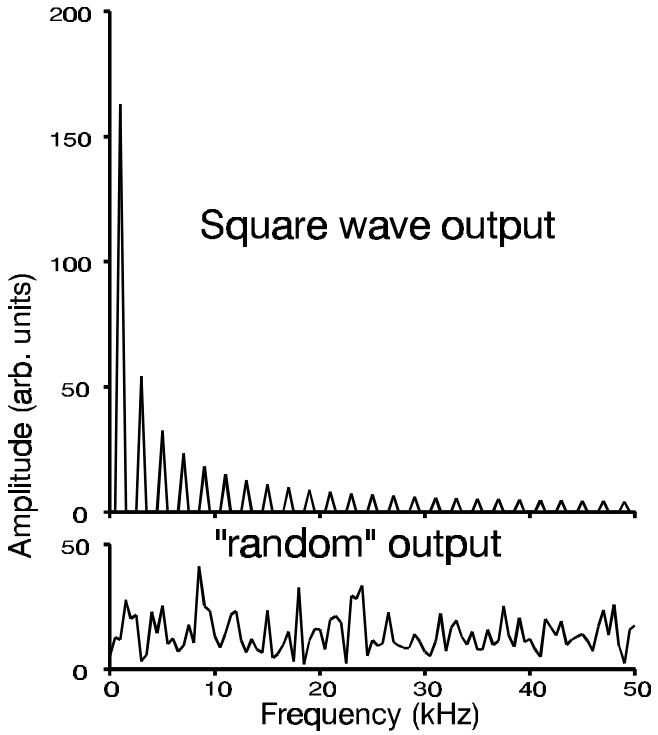
Stochastic Converters

Most of the practical DACs available use one of the above schemes with some refinements to salvage manufacturers pride in producing a different product. However several other schemes are possible and I will just illustrate them with one - the stochastic converter. This converter consists of a single switch followed by a low-pass filter.



Stochastic Converter

The method of operation is to operate the switch SW1 so that the fraction of time f which it spends in the "on" position is given by $f = V_o/V_{ref}$ when V_o is the required output. the simplest strategy for this scheme is to take a time period of n cycles of a clock ϕ and turn the switch on for m of them such that $f = m/n$ - note that again the result is a set of discrete values. However if the update time is t secs then the output will have a harmonic content at harmonics of $1/nt$. A better strategy therefore is to update at the same rate but have the update have a "probability" of being "on" equal to f. There are several well-known ways of doing this but the simplest is to pick a random number between 0 and 1 and turn the switch "on" if the number $< f$ and "off" otherwise. The output will now be a random sequence which will not have any particular frequencies embedded in it (hopefully).



A particular point to note is that if the random number takes a finite but variable time to generate then it will be necessary to use an external clock

(real-time clock) to fix the update rate.

In order to demonstrate the difference between the two approaches more thoroughly, here is an example: I used a sampling rate of 128kHz and then made a 7-bit converter by taking 128 time slices. I am going to try to set the output up for 1/2 scale - 1/2 the time on and half the time off. The simplest strategy would be to put the switch on for half the time and then off for the other half. This produces a square wave of frequency 1kHz whose frequency analysis is in the upper graph of the diagram[20]. An alternate strategy is to use a random number generator generating values between 0 and 1. For every time slot we generate a random number and if it is greater than 0.5 we put the switch on and if it is less we turn it off. The frequency analysis of the output is shown in the lower graph on the same vertical and horizontal scale. It is obviously easier to reduce the frequency content of the random sequence than the regular sequence because of the absence of peaks with high energy. Colloquially we can say that the energy in the first case is concentrated in the peaks whereas in the second case it is spread throughout the spectrum.

Non-Linear Systems

So far we have discussed the DAC as though it is always designed as a linear device. By linear we mean that the relationship between the input and the output is of the form

$$V_o = mN + c$$

where *m* and *c* are constants. However this is not always desirable and several converters have been designed to emulate curves rather than a straight line. One such system is a "companding" converter. These are used in telecommunications and often follow the "Bell μ-255" law which can be stated as:

$$V_o = \frac{V_{fs}}{255}\left(\exp\left[\frac{N/(2^m - 1)}{0.18}\right] - 1 \right)$$

where $V_{fs}$is the full-scale output. This can also be expressed in the alternative form relating the number N to the output V_o.

---

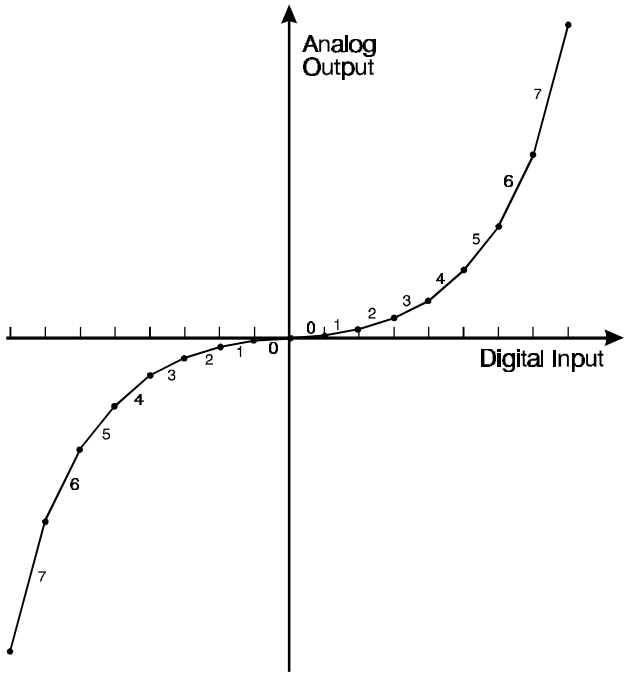
[20] The fact that the frequencies come out as "triangles" on the graph instead of just vertical lines is a consequence of my methodology for creating the graph - I generated the square wave as a time series and then fourier analysed it. Since the series length was finite - the frequencies were not perfectly defined.

$$N = \left(2^m - 1\right)\ 0.18\ \ln\left(1 + 255\frac{V_o}{V_{fs}}\right)$$

The advantage of these schemes is that they can accurately convey the size of a large signal without compromising the resolution of a small signal. This implies that they can handle signals with a large "dynamic range" The term refers to the ratio of the largest to the smallest expected useful values for the quantity. A voltage range from 0.1-1V has a dynamic range of 10x. A voltage range of 0-1V has infinite dynamic range. However if you would regard the smallest "significant" signal as 1mV then the dynamic range is the ratio of the maximum to the minimum signal - 10^3 in this case. It is often useful to restrict dynamic range in order to make things more easily measurable and sometimes necessary to cope with a very large dynamic range in a finite size of result, e.g. by using range switching or a log amplifier.

In digital-to-analogue terms the dynamic range of a linear converter is usually something like 2^n:1 where n is the number of bits in the converter. If we use a companding converter as we are discussing now the dynamic range can be larger. The Bell μ-255 law allows a range of about 4000:1 in 8 bits whereas a linear converter could only manage 256:1. (In fact our realisation of the μ-255 law doesn't do quite as well as the theory because of the straight line segment approach)

Dynamic range is often expressed in "dB" which is $20\log_{10}$(ratio). Thus 256:1 is 48dB.

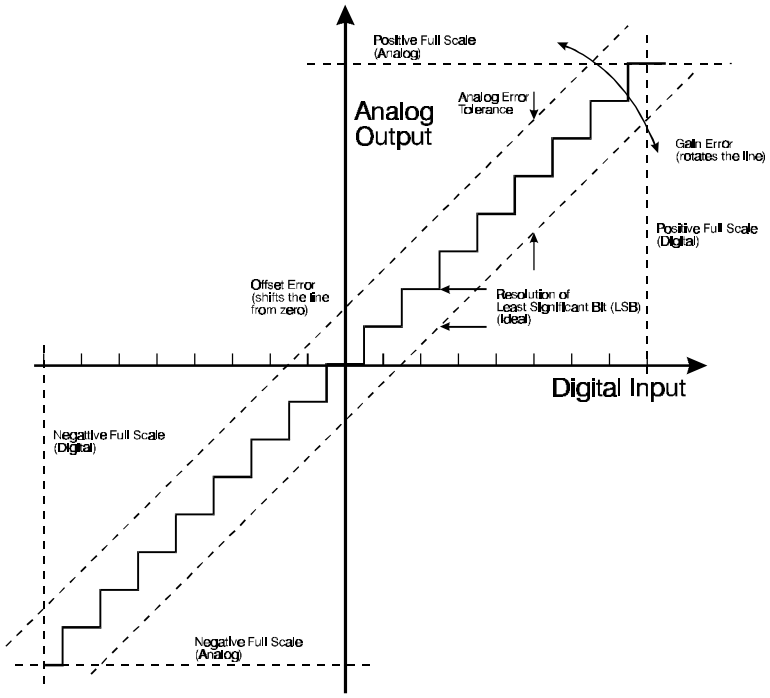


Bell System μ-255 Companding Approximation

A practical realisation of a companding DAC is shown in this figure where the most significant four bits of the input number define which segment (chord) of the converter is used, including the sign, and the last four bits determine where on the chord the output is. The curved characteristic is therefore approximated by 16 straight line segments.

Terms and Definitions

At this point we introduce some of the terms that one reads in specification sheets. Manufacturers specify a large number of parameters for DACs. Some will be relevant in a given situation - some will not. This list cannot be exhaustive, but at least it tries to set down some of the main terms and discuss some of the issues associated with them.



Specifications of a Digital-to-Analog Converter

Full Scale

Essentially a DAC can produce an output which is a scaled version of an input (although the input may simply be a reference voltage). If this is a unipolar output we could expect zero input code to correspond to zero output and 2^n-1 to correspond to the maximum output which would be full-scale. A bipolar output would be approximately symmetrical although note that since the number of codes is even and zero output is one of the codes, there is going to be a slight asymmetry in the output range. The mapping of the input codes to the bipolar output may be as an offset, where 0 input corresponds to maximum negative output and 2^n-1 to maximum positive output, 2s-complement, or by magnitude and a separate sign bit.

Multiplying DACs

As mentioned above, the action of a DAC can be considered to be the multiplication of an input by a fraction. Some DACs actually achieve this in practice and therefore become digital multipliers. Such DACs may be 1-quadrant (+ve input and fractions only), 2-quadrant (usually bipolar input and +ve fraction but could be the other way round) or 4-quadrant (bipolar input and multiplier)

Monotonicity

In an ideal DAC, the change in output from one input code to the next is the same. We have

already noted the difficulty with maintaining this relationship in practical cases. It is very desirable that the output stay the same or rise with each increment in the input code, even if the rise is not the theoretical step size. This can be accomplished if the converter steps are always within $\pm 1/2$LSB of the correct size. Such a converter is said to be "monotonic" and is achieved for converters with DNL (see below) of $< 1/2$LSB.

Gain Error
This is the difference between the actual analog output range and the ideal value from the data sheet. It speaks to the average slope of the characteristic. A gain error is only serious if it drifts or if you are relying on the theoretical value, ie no trimming. The gain most commonly changes with temperature and time. In addition the actual output is also dependent upon the stability of the reference input where one is used.

Zero Error
This is the difference from true zero when the appropriate code for zero output is given to the converter. Zero error is more likely to be a problem than gain error and can be particularly troublesome if it drifts.

Settling Time
The time taken from the change in the input code for the converter to come within some specified error band of the resulting output. Should be specified for a Full-scale change.

Glitch
A switching transient appearing on the output during a code transition. Its values can be expressed in volts or amps (as appropriate to voltage and current converters) and time or in terms of the charge transferred.

Power Supply Sensitivity
The change in output of the converter due to a change in the power supply voltage(s). Usually expressed as %full scale/% supply change.

Compliance and Output Resistance
These parameters (for current and voltage DACs respectively) determine how near they come to being ideal sources under differing load conditions. In practical design it is better to isolate any load changes from the DAC using a buffer amplifier and then, since the load is fixed, these parameters are folded into the gain error.

Accuracy

This is really a system specification since it actually specifies how the output relates to the absolute standard of voltage (or current). This involves the converter, the reference and the traceability of their various calibrations to an international standard. However even assuming that the full-scale is accurate, the accuracy of intermediate values depends upon the Integral Non-Linearity of the converter (see below).

Resolution

The resolution of the DAC is the smallest output change which can be made in the output. This can be a hard thing to describe in general, but for a monotonic DAC with reasonable Integral Non-Linearity it is bound up with the Differential Non-Linearity (see below).

Non-Linearity

Two types of non-linearity in the mapping of input code to output are recognised

Integral Non-Linearity (INL) is the degree to which the line deviates from the ideal straight line through all the states (or more usefully the line joining the zero and full-scale points). Good INL means that the converter has good accuracy and a low distortion of a digital waveform being convertered (ie if asked to convert a sine wave a converter with good INL will produce less harmonic content than one with a poor INL)

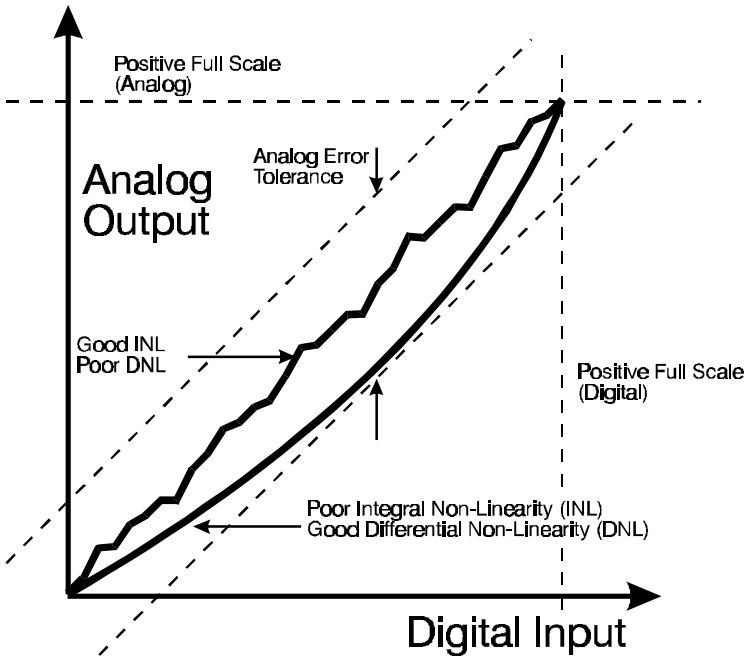


Illustration of Integral Non-Linearity and Differential Non-Linearity in a Digital-to-Analog converter

Differential Non-Linearity (DNL) is the degree to which the steps between the codes are uniform. Good DNL implies good resolution, because all the steps are almost exactly the same size, and a good noise performance[21].

[21] A discussion of noise is outside the scope of these notes, consult a good book on digital signal processing.

Notice that good DNL does not mean good INL because there could be a systematic variation in the steps which produces a curve. Even though the deviation for each step is small, the effect is cumulative. However good INL tends to imply good, but not necessarily excellent, DNL.

Crudely speaking INL controls the "straightness" of the transfer characteristic whereas DNL controls the "smoothness".

INL and DNL are becoming more important as converters become more sophisticated. However at the time of writing not all manufacturers are good about distinguishing between them and providing the necessary data to assist you in your choices.

Transition Issues

The various converters have very different reactions if you move from one value to another - and this is a point to watch if your equipment has a response time anything like the "settling time", i.e. the time that it takes a converter to settle to a new value. The stochastic and integrating (bidirectional) converters however are smooth in their transitions - if a bit slow.

One serious word of caution about reading specifications of converters - always remember that all manufacturers have one eye on the competition and are keen to show you how their product is better than anybody else's. Always be sure that you know what the manufacturer means by the terms in his specification, the words may not mean the same thing to both of you, and always beware of specifications which are important but not quoted. If in doubt - test it before committing yourself and even if not in doubt, test it.

Summary

There are three important parameters associated with a D/A converter: *Accuracy, Resolution and Speed* but that these were not exhaustive - i.e. you might well want to know more about the device than these three parameters.

The main thing to realise is that in general you get what you pay for and it is unlikely that you will get a Cadillac for the price of a Volkswagen. However if your aim is merely to get from A to B, either will do the job satisfactorily.