DIGITAL-TO-ANALOG CONVERSION

FUNDAMENTALS OF D/A CONVERSION

All electrical signals in the real world are analog in nature, and their waveforms are continuous in time. However, most signal processing is done numerically in a sample data form in discrete time. A device that converts a stream of discrete digital numbers into an analog waveform is called a digital-to-analog converter (DAC). One familiar example of a DAC application is the CD player. Digital bits, 1s and 0s, stored in the CD represent the electrical music waveform sampled in discrete time. DACs inside CD players convert the digital data stream read from optical discs into an audible sound waveform. DACs are used as stand-alone devices or as subblocks of other systems, such as analog-to-digital converters (ADCs), and cover the frequency spectrum from subsonic to microwave frequencies. Some examples of the systems in which DACs play an integral role are TV monitors, graphic display systems, digital voice/music/video systems, digital servo controllers, test systems, waveform generators, digital transmitters in modern digital communications systems, and multimedia systems (1).

DAC Resolution

The basic function of a DAC is the conversion of a digital number into an analog level. An N-bit DAC generates a discrete analog output level, either voltage or current, for every digital input word. The maximum range of the DAC is set by the reference voltage or current. In mathematical terms, the output of an ideal voltage DAC can be represented as

$$V_0(D_i) = \left(\frac{b_N}{2} + \frac{b_{N-1}}{2^2} + \cdots + \frac{b_2}{2^{N-1}} + \frac{b_1}{2^N}\right) V_r \qquad (1)$$

where V_r is a reference voltage setting the output range of the DAC and $b_N b_{N-1} \cdots b_1$ is the binary representation of the input digital word D_i. b_N is the most significant bit (MSB), and b_1 is the least significant bit (LSB). The digital-to-analog (D/A) conversion is a linear mapping of the input digital word to the analog output. Although purely current-output DACs are possible, voltage-output DACs are more common.

Because digital numbers are limited in length, the number of different possible DAC output levels depends on the number of bits used to form digital input word. For example, a DAC with two-bit digital word can have four possible outputs. A digital number of 00 represents one possible output, whereas 01, 10, and 11 each represents a different distinct output. That is, the total number of the output levels of a DAC using an N-bit long digital input word is 2^N. Therefore, the step size between the outputs created by two adjacent digital input words is $V_r/2^N$. This value represents the minimum voltage that can be resolved by an N-bit DAC. The resolution of a DAC is defined as the minimum resolvable output step, but the number of bits used for the digital input word is also quoted as resolution. In a strict sense, it is impossible to represent an analog waveform accurately with any limited number of discrete levels. However, real applications are limited either by noise or by other system requirements and need no finer resolution than necessary. For example, a video signal digitized with 8-bit resolution is sufficient for the current video standard, whereas CD players use 16 bit data to reproduce music.

Static and Dynamic Performance

An ideal N-bit voltage DAC generates 2^N discrete analog output voltages for digital inputs varying from 000 . . . 0 to 111 . . . 1 as illustrated in Fig. 1(a) for the ideal 4 bit DAC example. In the unipolar case, the reference point is the lowest output of the range when the digital input D_0 is 000 . . . 0. However, in the bipolar case or in differential DACs, the reference point is the midpoint of the full scale when the digital input is 100 . . . 0, whereas 000 . . . 0 and 111 . . . 1 represent the most negative and most positive DAC input ranges, respectively. An ideal DAC has a uniform step size and displays a linear transfer characteristic with a constant slope. However, in reality, the DAC transfer characteristic is far from being ideal as shown in Fig. 1(b). First, typical DACs have variance in step size and a curvature in the transfer characteristic measured in terms of differential and integral nonlinearities (DNL and INL), respectively. In addition, the overall transfer curve shifts up or down, and the slope is different from the ideal one. The former is defined as offset error, and the latter as gain error.

DAC performance is limited by two factors. One is the static linearity in representing digital numbers with a finite number of discrete output levels, and the other is the dynamic accuracy in the transition from one level to another. The D/A conversion process is illustrated in Fig. 2(a) with time-domain waveforms on the left and frequency spectrums on the right. Figure 2(b) shows the sampled digital waveform and its spectrum repeating at multiples of the sampling frequency f_s. Because the impulse response does not exist in the analog domain, the ideal D/A conversion is a sample-and-hold operation as shown in Fig. 2(c). It is still assumed that the DAC output is fast enough to make a step response. This sample-and-hold process exhibits a frequency response of a $(\sin x)/x$ function. If this sample-and-held waveform is low-

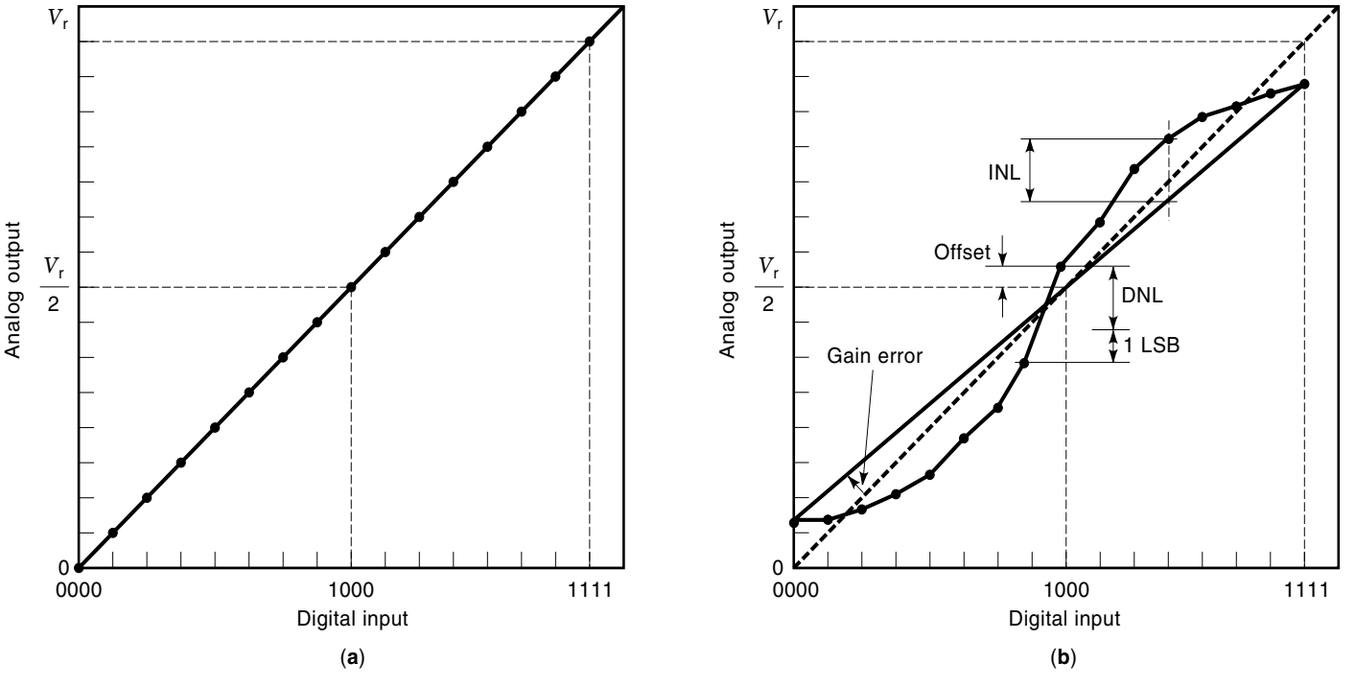

Figure 1. DAC transfer characteristics: (a) ideal case and (b) nonideal case.

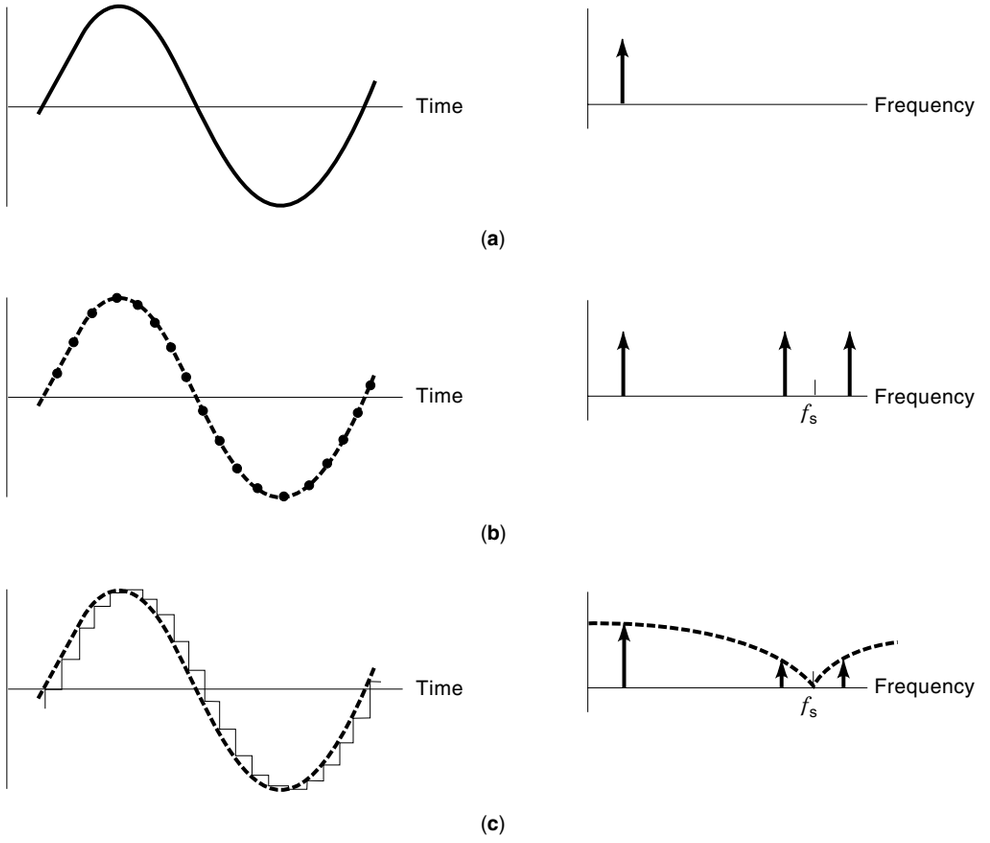

Figure 2. D/A conversion process: (a) input and its spectrum, (b) sampled data and its spectrum, and (c) sample-and-held DAC output and its spectrum.

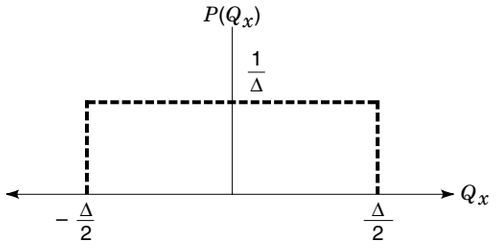

Figure 3. Density function of the quantization noise.

pass filtered, the original signal is restored except for a gain droop resulting from the $(\sin x)/x$ function. In practice, the DAC output cannot change instantaneously, and real DACs are designed to settle exponentially from one level to another. Therefore, the dynamic performance depends heavily on the actual implementation.

Signal-to-Noise Ratio

The lower bound in DAC resolution is set by the step size of the minimum incremental amount of the DAC output. Ideally this finite step size of the DAC output appears as a random noise in the D/A conversion. It is similar to the quantization noise of the A/D conversion process. The random noise can be modeled as an ideal DAC with an additive random noise source between values of $-V_r/2^{N+1}$ to $V_r/2^{N+1}$ as illustrated in the probability density function of Fig. 3, where Δ is $V_r/2^N$. Estimating the noise as the difference between the actual input and the nearest level, the noise $Q_x(n)$ lies between

$$-\frac{\Delta}{2} \le Q_x(n) \le \frac{\Delta}{2} \tag{2}$$

as shown in Fig. 3. Then the average noise power σ^2 can be calculated as

$$\sigma^2 = \int_{-\Delta/2}^{\Delta/2} \frac{x^2}{\Delta}\, dx = \frac{\Delta^2}{12} \tag{3}$$

The signal-to-noise ratio (SNR) is defined as the power ratio of the maximum signal to the inband uncorrelated noise. Because the maximum signal power is $V_r^2/8$, the well-known relation of the noise is derived using Eq. (3) as

$$\text{SNR} = 10 \times \log\left[\frac{3 \times 2^{2N}}{2}\right] \approx (6.02N + 1.76)\ \text{dB} \tag{4}$$

If the number of bits increases by one, noise is lowered by about 6 dB.

The SNR accounts for inband uncorrelated noise only. In reality, the DAC transfer characteristic is not linear, and the nonlinearity in the D/A conversion process appears as harmonics or intermodulation components in the DAC output. Considering nonlinearity, DAC performance is more accurately defined using the total signal-to-noise ratio (TSNR) or sometimes referred to as the signal-to-noise-and-distortion ratio (SNDR). For an ideal DAC without nonlinearity, noise is limited only by quantization, and SNDR should be identical to SNR. In some applications such as generating complex spectrums in wireless radio-frequency (RF) systems, the intermodulation from DAC nonlinearity is the most critical issue because strong interfering tones produce inband spurious components and degrade the inband SNR. The DAC performance for such applications is often measured as spurious-free dynamic range (SFDR), which is the ratio of the maximum signal component and the largest spurious component. The INL is the key design factor for low SFDR applications.

DNL and INL

The fundamental limit of the DAC performance is SNR, but it is more likely that DAC performance is limited by spurious and harmonic components. The DAC nonlinearity can be measured by DNL and INL. DNL is a measure of deviation of actual step size from the ideal size. Similarly, INL is defined as a measure of deviation of the midpoint of the actual step from the midpoint of the ideal step. DNL and INL are indicators of how ideal a DAC is and can be defined in an N-bit voltage DAC as

$$\text{DNL}(D_i) = \frac{V_0(D_{i+1}) - V_0(D_i) - V_r/2^N}{V_r/2^N} \tag{5}$$

$$\text{INL}(D_i) = \frac{V_0(D_i) - i \times V_r/2^N}{V_r/2^N} \tag{6}$$

for i = 0, 1, . . ., $2^N - 1$. DNL and INL are normalized to the ideal one LSB step, and the largest positive and negative numbers for both DNL and INL are usually quoted to specify the static performance of a DAC.

Several different definitions of INL may result depending on how two endpoints are defined. The two endpoints are not exactly 0 and V_r because of the offset and gain errors explained in Fig. 1(b). In most DAC applications, the offset and gain errors resulting from the nonideal endpoints do not matter, and the integral linearity can be better defined in a relative measure using the straight-line linearity concept rather than the endpoint linearity in the absolute measure. The straight line can be defined as two endpoints of the actual DAC output voltages or as a theoretical straight line adjusted to best fit the actual DAC output characteristic. The former definition is sometimes called endpoint linearity, whereas the latter is called best-straight-line linearity.

Monotonicity

A DAC is monotonic if its output increases as the digital input increases. The condition for monotonicity requires that the derivative of the transfer function never change sign and that the DNL be better than -1 LSB. Monotonicity of a DAC is important in all DAC applications, but it is a necessity in such applications as digital control and video. The worst DNL and INL in binary-weighted DACs usually appear at a major transition point. The major transition point is the point at which the MSB changes in the digital input such as between 011 . . . 1 and 100 . . . 0. If the MSB weight is smaller than the ideal value (1/2 of the full range), the analog output change can be smaller than the ideal step $V_r/2^N$ when the MSB changes. If the decrease in the output is larger than one LSB, the DAC becomes nonmonotonic. The similar nonmonotonicity can take place when switching the second or lower MSB bits in binary-weighted multibit DACs.

Monotonicity is inherently guaranteed if an N-bit DAC is made of 2^N elements for thermometer decoding. However, it is impractical to implement high-resolution DACs using 2^N elements because the number of elements grows exponentially as N increases. For high-resolution DACs, four different ways of achieving monotonicity exist. They are using the slope-type approach, the multilevel segmented DAC approach, calibration, and the interpolative oversampling technique. The first one is to use a linear voltage ramp and to control the time to stop digitally so that accurate voltage proportional to the digital word can be obtained. Oversampling interpolative DACs also achieve monotonicity by converting a pulse-density modulated bitstream into analog waveform. The slope-type DAC has a limited use in digital panel meters and in other slow measurement uses, and it is not covered here. However, the calibration and oversampling approaches are covered in separate sections.

DAC ARCHITECTURES

Resistor-String DAC

A resistor string made of 2^N identical resistors is a straightforward voltage divider. Switching the divided reference voltages to the output makes a DAC as shown in Fig. 4, which uses a 3-bit binary tree decoder. Because it requires a good switch, the stand-alone resistor-string DAC is easier to implement using metal-oxide-semiconductor (MOS) technologies. Resistor strings are widely used as an integral part of the flash ADC. One major drawback of using it as a stand-alone DAC is that the DAC output resistance depends on the digital input word and switch on-resistance. This nonuniform settling time constant problem can be alleviated either by adding low-resistance parallel resistors or by compensating for MOS switch overdrive voltages (2). The code-dependent settling has no effect on the DAC performance when used as an ADC subblock.

Film resistors such as Tantalum, Ni–Cr, or Cr–SiO used in bipolar process exhibit very good matching of above the 10-bit level with low voltage and temperature coefficients. However, in the MOS process, such high-quality resistors are not available. Either diffusion or undoped poly resistors are used, and the achievable matching accuracy is below 8-bit level with one-order higher voltage and temperature coefficients than those of film resistors in bipolar process. Although resistors are carefully laid out using large geometry, matching of resistors in integrated circuits is still limited by the mobility and resistor thickness variations. Differential resistor DACs with large feature sizes are reported to exhibit higher matching accuracy at the 11 to 12 bit level.

A resistor-string DAC is inherently monotonic and exhibits good DNL, but it suffers from poor INL. For higher INL, trimming or adjustment techniques are needed, but it is impractical to apply them to all 2^N resistor elements. A very practical method to improve the INL of the resistor-string DAC is to use on-chip unity-gain buffers and to adjust voltages at intermediate taps of the resistor string using conventional trimming techniques. For this, the buffers require high open-loop gain, low output resistance, large current driving capability, and wide bandwidth for accurate and fast setting. The more taps that are adjusted, the better integral linearity obtained. An added benefit of this INL trimming method is the reduced RC time constant due to the voltage sources applied to the adjustment taps.

Binary-Weighted Current DAC

Although the resistor-string DAC is simple to make for a small number of bits, the complexity grows exponentially as the number of bits increases. Binary-ratioed elements are simpler to use for a large number of bits. One of the simplest DAC architectures using the binary-weighted current sources is shown in Fig. 5 (3). It is the most popular stand-alone DAC architecture in use today. In bipolar technology, transistors and emitter resistors are ratioed with binary values as shown. In MOS technology, only ratioed transistors are used as in, for example, a video random-access-memory (RAM) DAC that is made up of simple PMOS differential pairs with

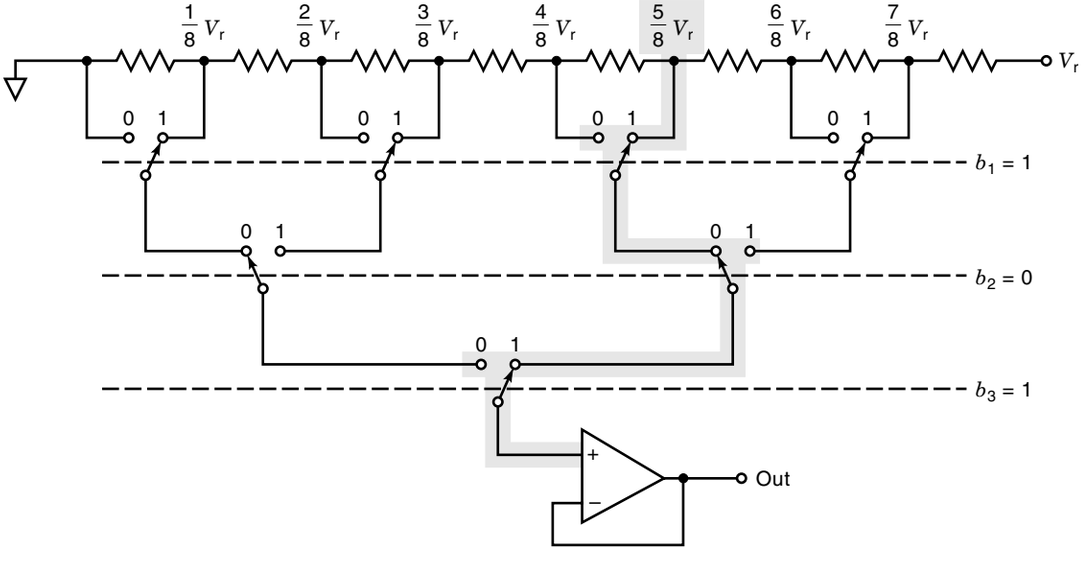

Figure 4. Resistor-string DAC.

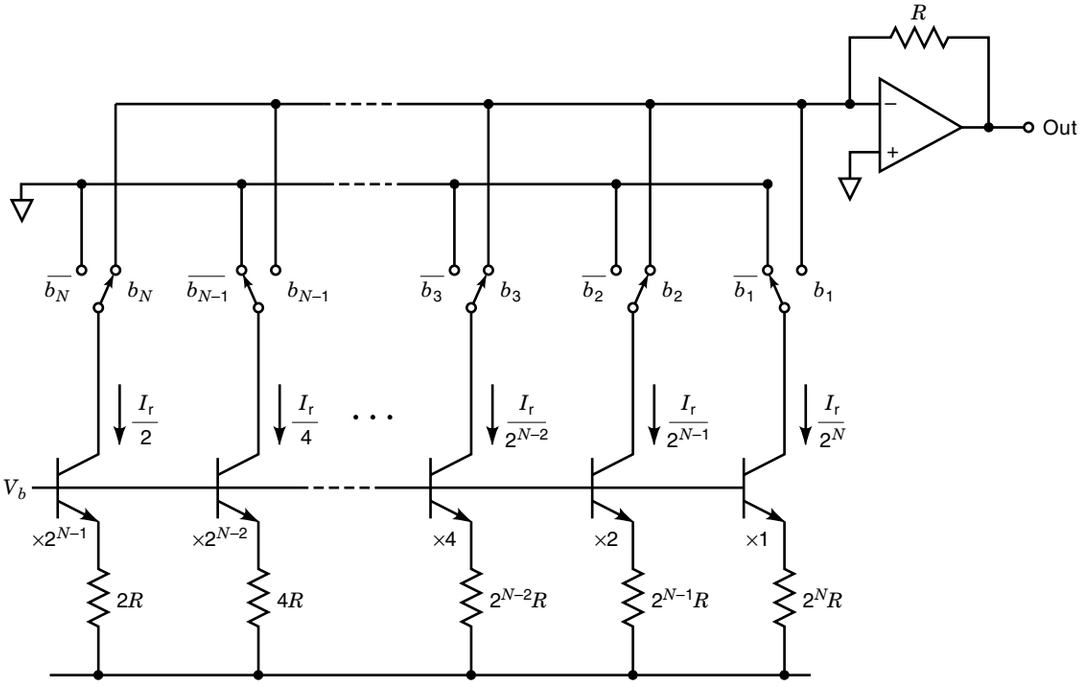

Figure 5. Binary-weighted current DAC.

binary-weighted tail currents. Although MOS DACs exhibit 8 bit level matching using 10 to 20 μm device size, bipolar DACs are known to have above 10-level matching using thin-film resistors. The current sources are switched on or off either by means of switching diodes or differential pairs as shown. The output current summing is done by a wideband transresistance amplifier, but in high-speed DACs, the output current is used directly to drive a resistor load.

Although binary weighting reduces circuit complexity, the number of transistors and resistors still grows exponentially because a unit component is repeatedly used for good matching. This complexity problem is alleviated using another type

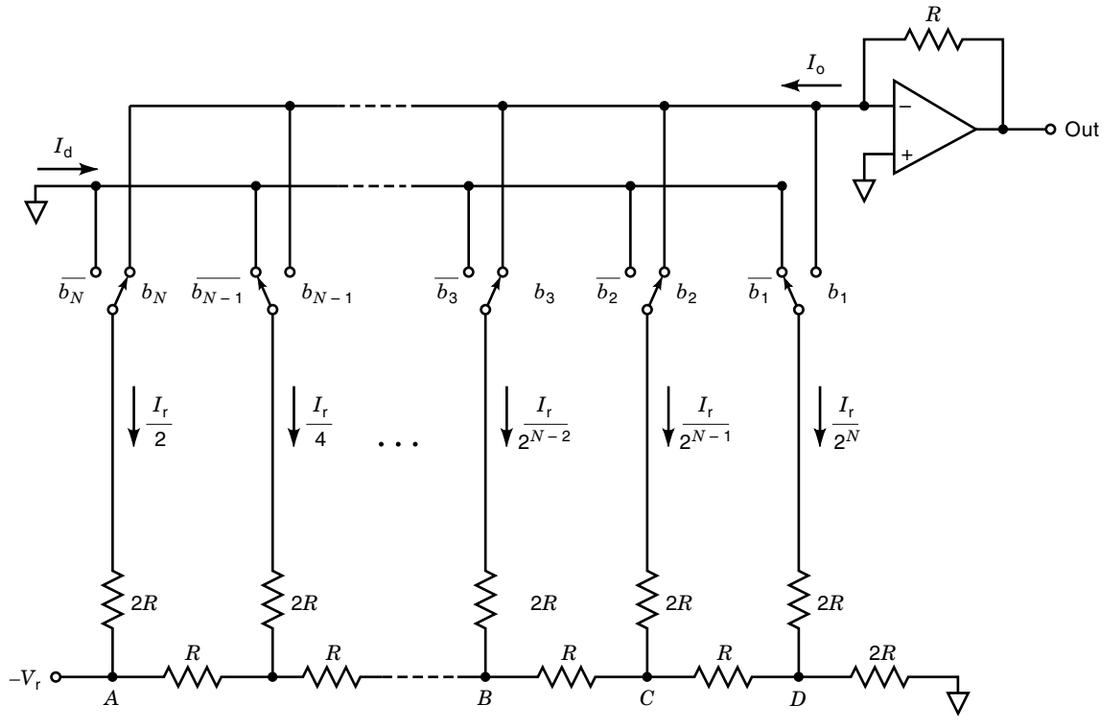

Figure 6. R–$2R$ DAC.

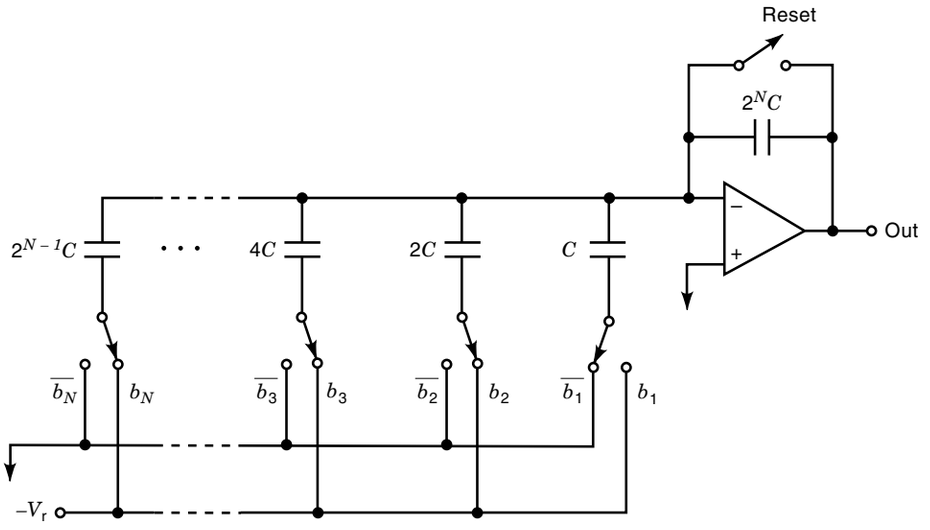

Figure 7. Binary-weighted capacitor-array DAC.

of current-ratioed DAC known as R–2R DAC shown in Fig. 6. The R–$2R$ network consists of series resistors of value R and shunt resistors of value $2R$. Each shunt resistor 2R has a single-pole double-throw electronic switch that connects the resistor either to ground or to the output current summing node. The operation of the R–$2R$ ladder network is based on the binary division of current as it flows down the ladder. At any junction of series resistor R, the resistance looking to the right side is $2R$. Therefore, the input resistance at any junction is R, and the current splits equally into two branches at the junction because it sees equal resistances in both directions. The advantage of the R–$2R$ ladder method is that only two values of resistors are used, greatly simplifying the task of matching or trimming and temperature tracking. Also, for high-speed applications, relatively low-valued resistance can be used for even more savings in the chip area. The major disadvantage of this architecture is the nonmonotonicity and high nonlinearity due to poor matching in fabricated resistors. The R–$2R$ DAC is a multiplying DAC (MDAC) that has an output proportional to the product of the reference voltage and the digital input word.

Binary-Weighted Capacitor-Array DAC

Capacitors made of double-poly or poly-diffusion in the MOS process are known to be very accurate passive components comparable to film resistors in the bipolar process both in the matching accuracy and voltage and temperature coefficients. The binary-weighted capacitor-array DAC is shown in Fig. 7. The DAC is made of a parallel capacitor array of N capacitors with a common top plate (4). Unlike DACs using resistors and currents, the capacitor-array DAC is based on a dynamic charge redistribution principle. Therefore, it is not convenient to use it in continuous-time applications, and for stand-alone applications, a resettable feedback amplifier periodically charging the top plate of the array and an output S/H or deglitcher are needed. The capacitor array is known to give a 10-bit level matching for this use. The matching accuracy of the capacitor in MOS technology depends on the geometry sizes of the capacitor width and length and the dielectric thickness.

As a stand-alone DAC, the top plate of the DAC is precharged either to the offset of the feedback amplifier or to the ground. One extra smallest C is not necessary for DAC as shown in Fig. 7, but as a subblock of an ADC, it is needed to make the total capacitance of 2^NC. Because the top plate is connected to the summing node, the top plate parasitic capacitance has a negligible effect on the DAC performance. The capacitor-array DAC requires two-phase nonoverlapping clocks for proper operation. Initially, all capacitors should be charged to ground. After initialization, depending on the digital input, the bottom plates are connected either to $-V_r$ or to ground. Then the output is the same as given by Eq. (1). Like the R-$2R$ DAC, the capacitor-array DAC can be used as an MDAC.

Monotonic Segmented DACs

As discussed in the previous section, a DAC system is greatly simplified using a binary-weighted DAC and binary input word, but monotonicity is not guaranteed. For monotonicity, the DAC needs to be thermometer-coded. Three examples of thermometer-coded DACs are shown in Fig. 8 for resistor, current, and capacitor-array DACs. Their operation is the same except for the thermometer-coded input. Because each segment is individually represented by a unit element, the DAC output should increase as a new unit element is switched in as the thermometer-code input increases by one. This guarantees monotonicity, but the problem is an exponential increase in the complexity as the number of bits increases.

Applying a two-step conversion concept, a monotonic DAC can be made using coarse and fine DACs. The fine DAC divides the next MSB segment into fine LSBs. Two examples of the monotonic-segmented approach exist. One is a combination DAC, and the other is the most widely used next-segment approach. Because it is difficult to achieve fine resolution using one DAC, both approaches use two separate DACs. The $R + C$ combination DAC uses two different DAC types to achieve two-level D/A conversion as shown in Fig. 9. The MSB resistor-string DAC supplies the reference voltages to the LSB capacitor-array DAC. When the top plate is initialized, all capacitor bottom plates are connected to the higher voltage of the next segment of the resistor-string DAC. During the next clock phase, the bottom plates of capacitors are

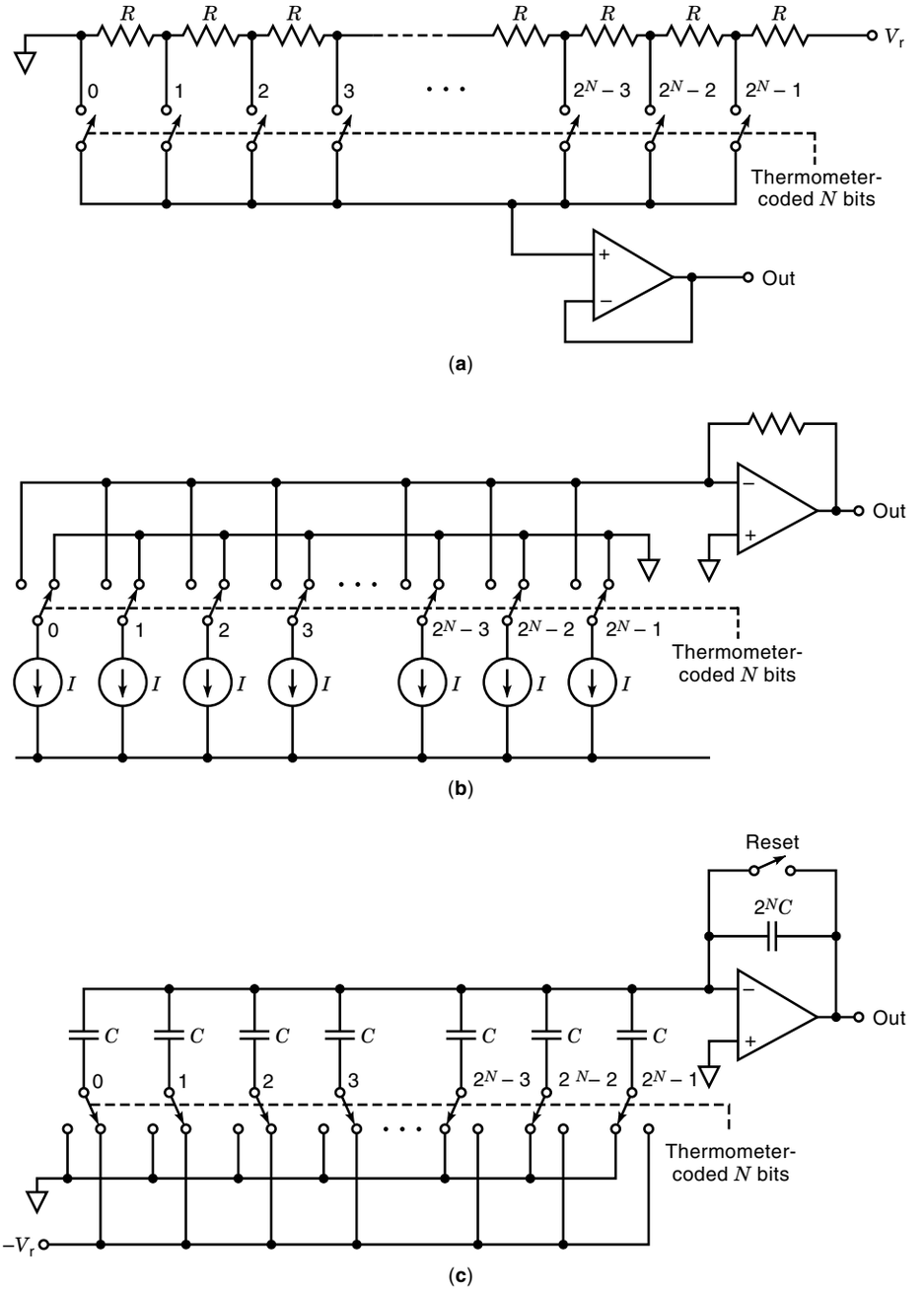

Figure 8. Thermometer-coded monotonic DACs: (a) resistor, (b) current, and (c) capacitor arrays.

selectively connected to the lower voltage of the segment if the digital bit is 1 but stay switched to the higher voltage if 0. This segmented DAC approach guarantees inherent monotonicity as far as the LSB DAC is monotonic within its resolution. However, INL is still poor because the MSB is set by the resistor string. Fully differential implementation of this architecture is known to benefit from the lack of the even-order nonlinearity and to achieve high INL.

The same idea used in the $R + C$ combination DAC can be implemented using two DACs of the same type. A segmented DAC with total N bits is shown in Fig. 10. For monotonicity, the MSB M bits are selected by a thermometer code, but one of the MSB current sources corresponding to the next segment of the thermometer code is divided by a current divider for the fine LSBs. As in the $R + C$ capacitor array, the fine DAC divider should have a monotonicity of $N - M$ bits. Although monotonicity is guaranteed with modest matching requirement, it is still limited in INL due to the limited matching accuracy in the MSB. This segmented current DAC, with and without trimming, is the most widely used DAC architecture.

DAC CONVERSION ACCURACY

DAC performance is directly affected by static linearity and dynamic settling. The former is limited by the matching accu-

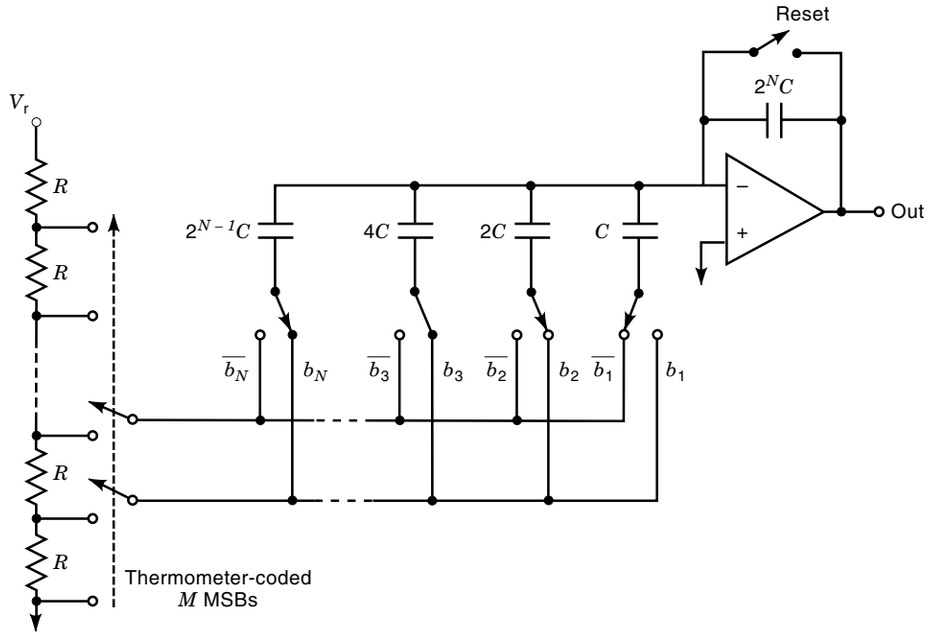

Figure 9. $R + C$ combination monotonic DAC.

racy of passive or active components, and the latter is limited by nonideal settling. When a DAC is used in the feedback path of an ADC, the DAC linearity also affects the ADC linearity. In ADC applications, the number of bits per stage is determined by how many bits are resolved in the DAC, depending on ADC architecture. However, the dynamic DAC settling requirement in ADCs is less stringent than those used as stand-alone devices. As explained in Fig. 2, stand-alone DACs should settle either infinitely fast or purely exponentially in making transitions from one level to another. Because fast settling is not possible except for extremely low-speed DACs, most DAC output should settle exponentially with a well-defined time constant. Main sources for nonideal dynamic settling errors are slew, glitch, and clock jitter.

Element Matching

Mismatch between elements occurs because of parasitic elements, uncertainties in the size of drawn shapes, and other varying process parameters. There is typically about $\pm 0.2\%$ mismatch between two resistors of the same value drawn with a length of 10 to 20 μm. For similar reasons, capacitors

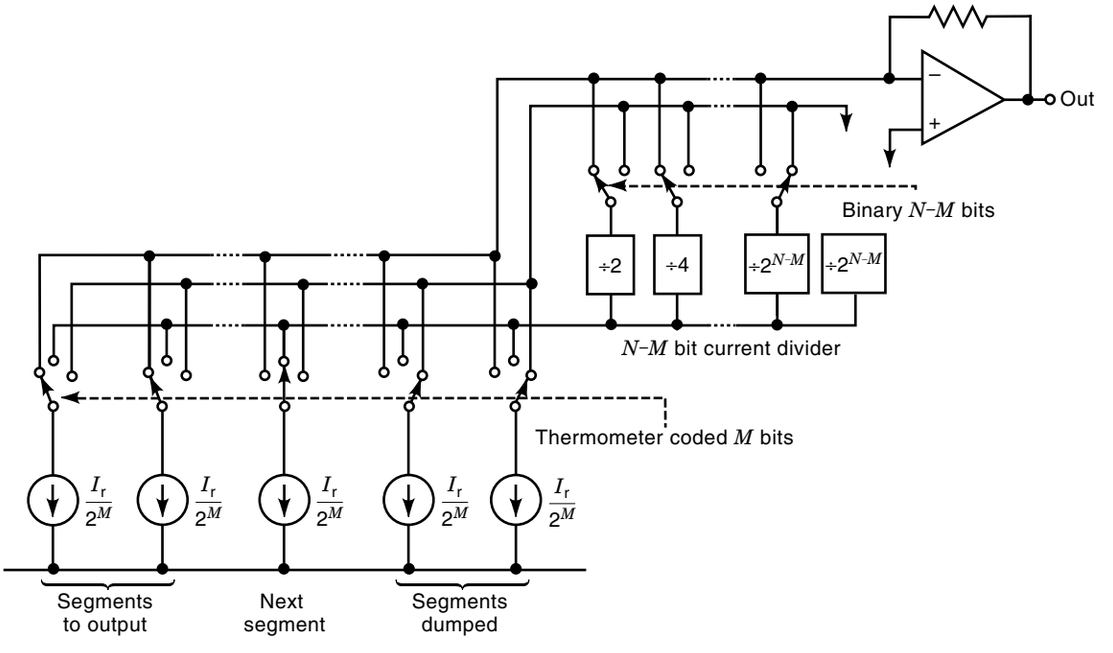

Figure 10. Two-level segmented monotonic DAC.

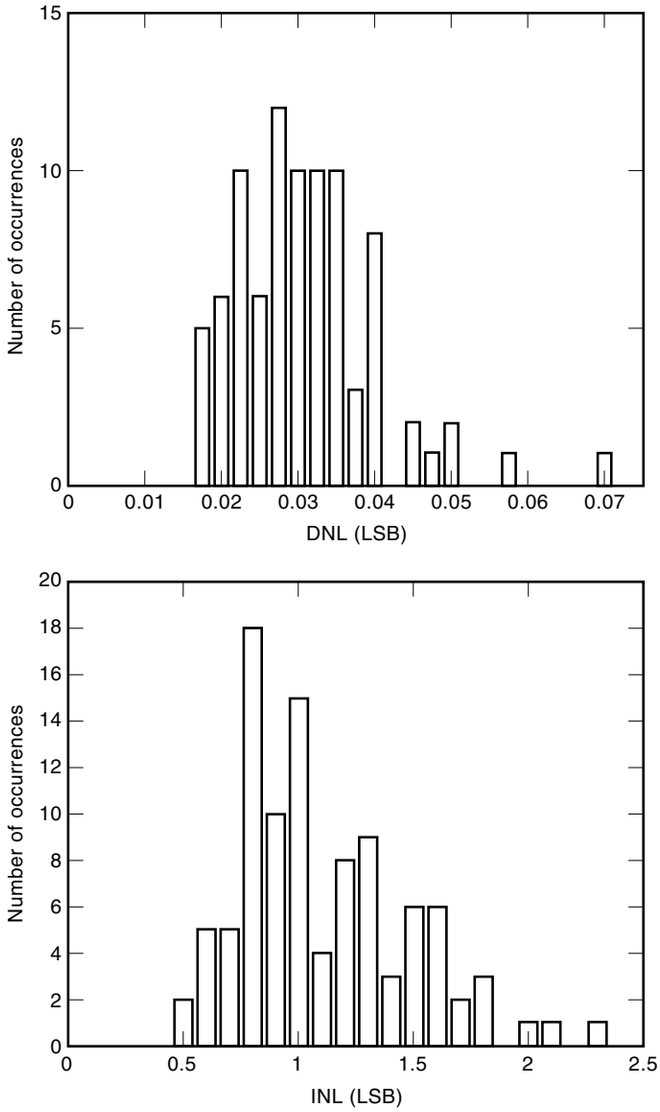

Figure 11. Simulated DNL and INL of a 12-bit DAC with 0.1% component matching.

and transistors also show mismatch due to process variations. The effect of component mismatch appears as static nonlinearity. Because of statistical randomness of the process variation, the mismatch effect on the DAC performance can be better understood using statistical Monte Carlo analysis. Figure 11 show the distribution characteristics of DNL and INL of a 12-bit segmented DAC with 1% random mismatch in current sources and the next-segment divider. The 12 bits are partitioned into four thermometer-coded MSBs and binary-weighted 8-bit LSBs. The x-axes for both graphs are in units of LSB, whereas the y-axes represent the total number of occurrences out of 100 samples. The segmented DAC architecture guarantees monotonicity regardless of the amount of mismatch in the elements. There exist many samples with INL greater than 1 LSB, but no DNL exceeds 1 LSB.

Slew and Settling

The output of a DAC is a sampled step waveform held constant during a word clock period. The ideal transition from

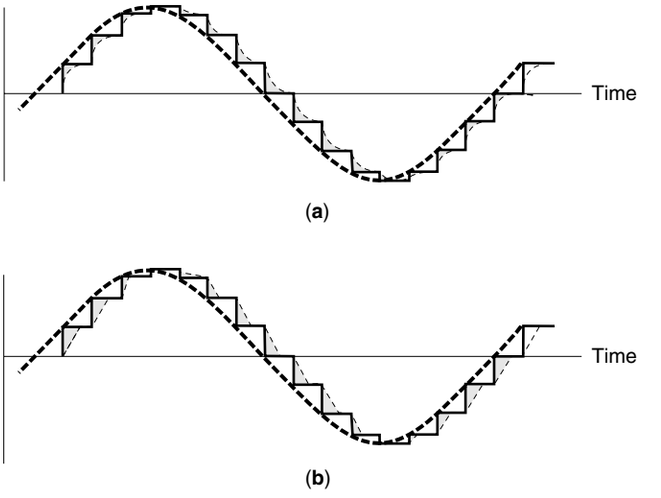

Figure 12. DAC transient errors: (a) settling and (b) slew.

one value to another is an exponential function. High-speed DACs usually have a current output, and the output is usually terminated with a 50 or 75 Ω low-impedance load. Therefore, the current output DAC can be designed so that the output may settle exponentially with a fixed time constant. However, for voltage output DACs, either a transconductance amplifier or a sample-and-hold amplifier is used as a buffer amplifier. In general, all amplifiers slew to a certain degree if a large transient signal is suddenly applied to the input. Figure 12 shows two waveforms with transient errors resulting from exponential settling and slew limiting.

In the exponential-settling case, the transient error defined by the shaded area is proportional to the height of the jump. This implies that any single time-constant settling does not produce any nonlinearity error. In reality, amplifiers settle with multiple poles although one of them is dominant. To see this effect, consider a two-pole settling case. The transient error can be calculated as follows:

$$\text{Transient error} = h\left(\frac{\tau_1^2}{\tau_1 - \tau_2} + \frac{\tau_2^2}{\tau_2 - \tau_1}\right) \tag{7}$$

where τ_1 and τ_2 are the two time constants. As in the single time-constant case, the transient error is a linear function of

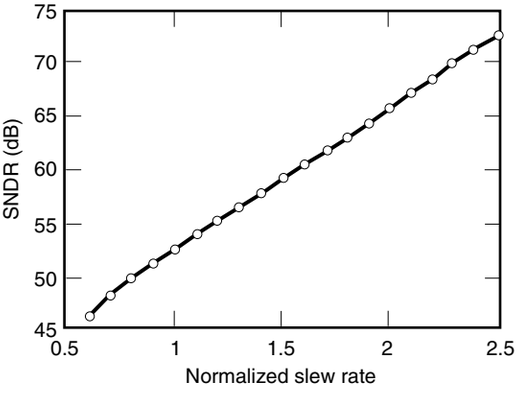

Figure 13. SNDR degradation of a 12-bit DAC due to limited slew rate.

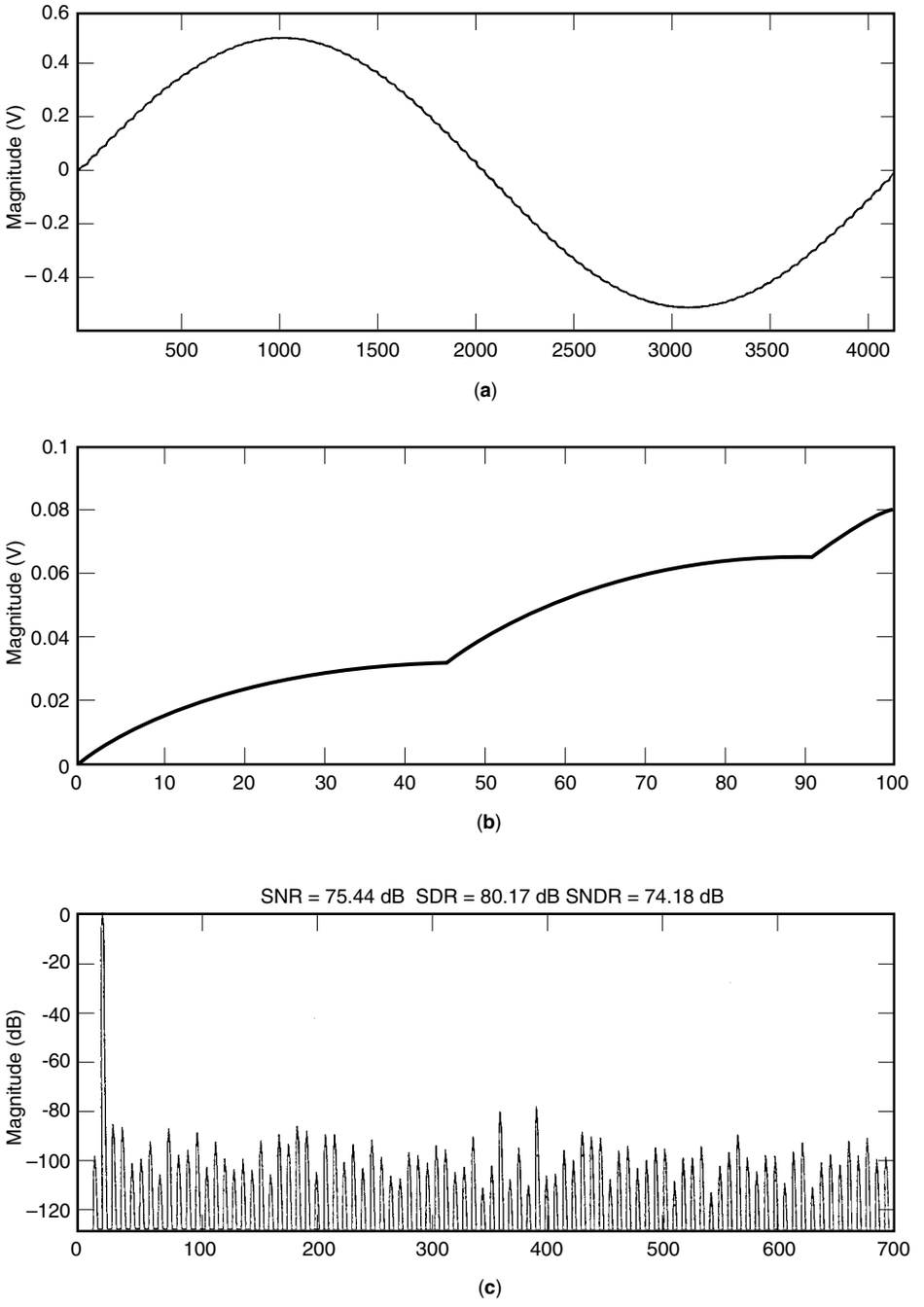

Figure 14. 12 bit DAC output with two poles: (a) time waveform, (b) its detail, and (c) its FFT.

h, and the preceding result implies that the two-pole settling generates only linear errors and will not affect the DAC linearity.

However, in the latter slew-limited case, the area is proportional to h^2. From the simple slewing distortion model, the worst case harmonic distortion (HD) when generating a sinusoidal signal with a magnitude V_0 with a limited slew rate of S is

$$\mathrm{HD}_k = 8\frac{\sin^2\dfrac{\omega T_c}{2}}{\pi k(k^2-4)} \times \frac{V_0}{ST_c}, \quad k = 1,\ 3,\ 5,\ 7, \ldots \tag{8}$$

where T_c is the clock period (5). For a given distortion level, the minimum slew rate is given. Any sinusoidal waveform has a maximum slew rate of $\omega_0 V_0$ at the zero-crossing point. Therefore, if a DAC has a slew rate much higher than this maximum slew rate, the DAC output will exhibit negligible slew-limited distortion. The summary plot of the SNDR vs. slew rate of a 12 bit DAC example is shown in Fig. 13, where the x-axis unit is normalized to the maximum slew rate of the sinusoidal waveform. As expected from Eq. (8), the SNDR is proportional to the slew rate. Figures 14 and 15 show the simulated output spectrums of a 12-bit DAC for two cases. One is the two-pole settling case, and the other is the slew-

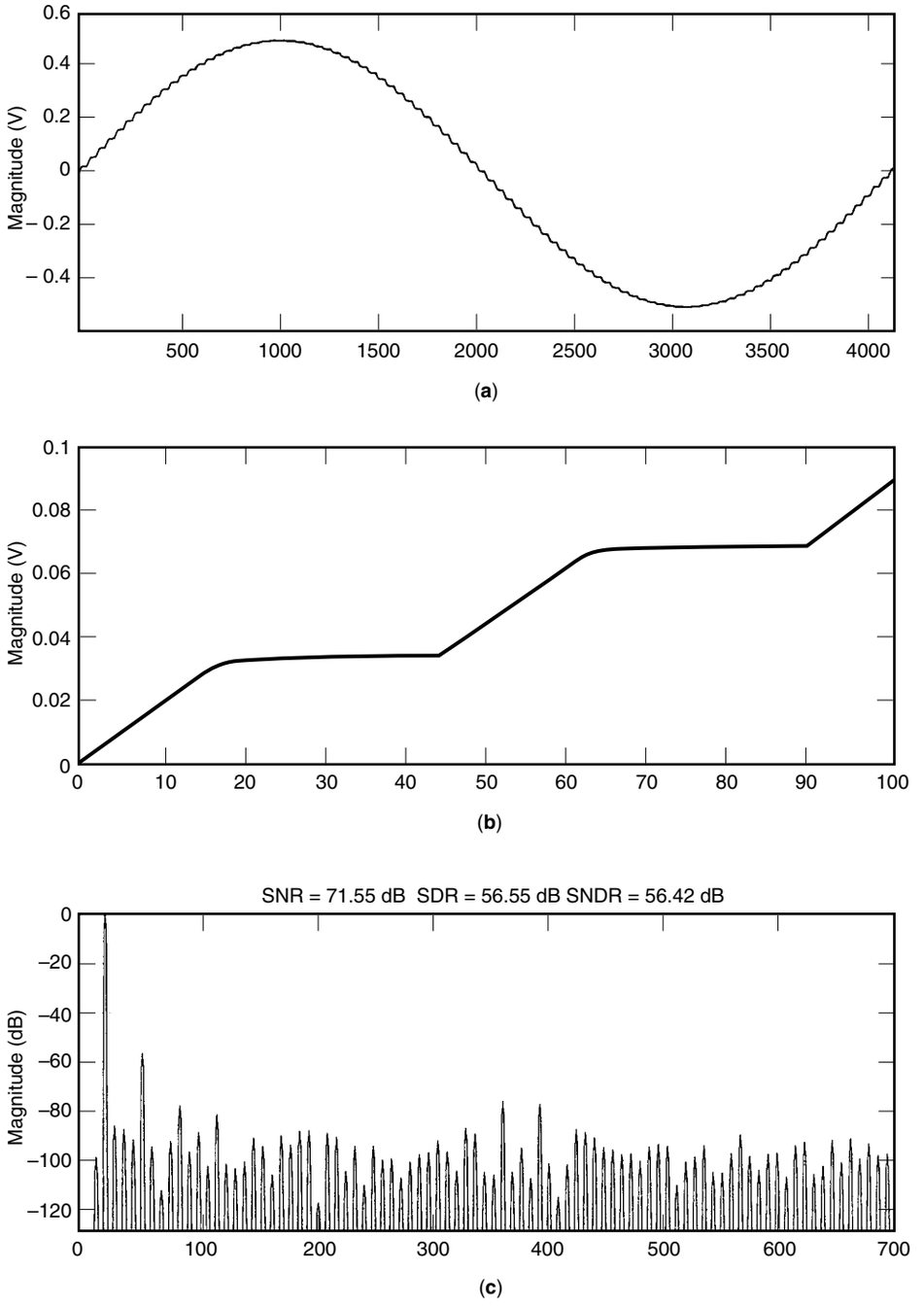

Figure 15. 12 bit DAC output with slew limit: (a) time waveform, (b) its detail, and (c) its FFT.

limited case. As predicted in Fig. 13, the SNDR of the former case stays high at 74 dB, whereas the latter case suffers from lower SNDR of 56 dB. The harmonics of the input increase drastically in the slew-limited case.

One more nonideal factor that contributes to the settling of a DAC is the parasitic capacitance. The parasitic capacitance of the current sources shifts the poles to display a code-dependent time constant. As a result, the dominant pole of a DAC varies according to input digital words. Because the parasitic capacitances are additive, the higher output levels where more current sources are switched on can exhibit a slower settling time than that of lower output levels.

Glitch and Clock Jitter

Ideally, all bits of the digital input word should switch at the same time so that any change in the digital word can be reflected in the analog output waveform simultaneously. However, even with the use of an input data latch to synchronize the digital word, different data paths cause a slight variation

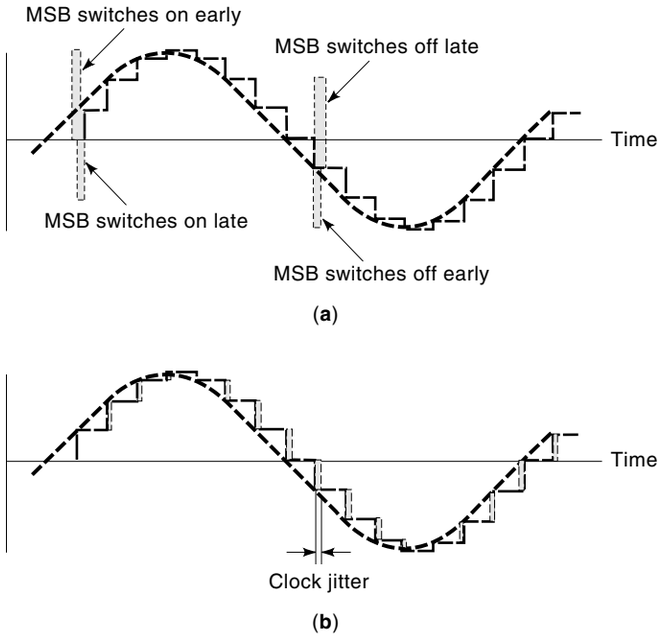

Figure 16. Effects of (a) MSB glitch and (b) clock jitter.

in the switching time of the digital bits. This phenomenon causes a glitch in the output and affects the performance of the DAC. For example, at the major code transition in a binary-weighted DAC, the digital word changes from 011 . . . 1 to 100 . . . 0. If all bits change simultaneously, then the output should increase exponentially by one LSB step. However, due to the slight time differences between switching of the digital bits, there may be an instance where the MSB bit switches on before the rest of the digital bits turn off, thus creating the appearance of the digital word 111 . . . 1 even momentarily. This will cause a positive spike in the output as shown in Fig. 16(a). Similarly, there may be an instance when the MSB switches on after the remaining digital bits turn off, thus creating the appearance of digital word 000 . . . 0. This will cause a downward trough in the output as shown. The same glitch occurs when the MSB switches off.

Obviously this situation is the worst case. A more realistic glitch spike and its effect on the output spectrum is shown in Fig. 17 where 5% glitch is simulated for the same 12-bit DAC with a sinusoidal input. The frequency spectrum of the output with glitches exhibits an elevated noise floor. Although a few techniques such as segmented antisymmetric switching have been proposed to alleviate the glitch problem, an easier solution is to employ a sample-and-hold amplifier at the DAC output as a deglitcher. The basic idea of deglitching is to keep the DAC in the hold mode until all the switching transients have settled. If the deglitcher is faster than the DAC, the slew rate limitation may improve. However, if the slew rate of the deglitcher is of the same order as the DAC, the slew distortion will still exist, now as an artifact of the deglitcher.

Another major source of conversion error is the nonideal clock jitter. The glitch results from the switching time difference of digital bits while the clock jitter results from the randomness of the clock edge itself. The clock jitter generates errors as explained in Fig. 16(b). The right signal at the wrong time is the same as the wrong signal at the right time. Even if the output waveform is correctly generated from the input word, the timing error will raise the noise floor of the DAC. If the jitter has a Gaussian distribution with a root-mean-square jitter of Δt, the worst-case SNR resulting from this random word clock is

$$\mathrm{SNR} = -20 \times \log \frac{2\pi f \Delta t}{\sqrt{M}} \tag{9}$$

where f is the signal frequency and M is the oversampling ratio. The timing jitter error is more critical in reproducing high-frequency components. In other words, to make an N-bit DAC, an upper limit for the tolerable word clock jitter is

$$\mathrm{Jitter} < \frac{1}{2\pi B 2^N}\sqrt{\frac{2M}{3}} \tag{10}$$

The simulated spectrum of a 12 bit DAC with 5% clock jitter is shown in Fig. 18. The effect is about the same as the glitch case. Figure 19 shows the SNDR of the same 12 bit DAC as functions of glitch and clock jitter percentages of the sampling period. It clearly demonstrates that the SNDR decreases as both glitch and clock jitter effects become more significant.

HIGH-RESOLUTION TECHNIQUES

The current trend of high resolution at high frequencies has stimulated many innovative DAC developments. Although there may be numerous approaches in making high-resolution DACs, only two distinctive approaches need special mentioning. They are the monotonic segmented approach and the oversampling interpolative approach. The former is exclusively used for high-speed DACs, but the oversampling approach is dominant at low-speed applications such as for instrumentations, digital audio, and digital communication channels. For high-resolution DACs, bipolar and bipolar-CMOS (BiCMOS) DACs are superior to bulk complementary MOS (CMOS) counterparts in speed. In nonoversampling DACs, linearity of more than 12 b requires resolution-enhancing methods such as trimming, dynamic matching, and electronic calibration.

Trimming

Component matching is limited by many variations in physical dimensions and process parameters. The effect of random size variation is reduced using physically large dimension. Careful layout using common centroid or geometric averaging can also reduce the process gradient effect. However, it is possible to further improve component matching by trimming on the wafer in a post-fabrication process. Resistor value can be either laser trimmed (6), Zener-zapped (7), or electronically controlled by switches using programmable read-only-memory (PROM) and erasable PROM (EPROM). The laser-trimming and the Zener-zapping processes are nonreversible. The long-term stability of trimmed resistors is a major concern, although electronical trimming using EPROM can be repeated.

The laser-trimming method involves using a focused laser beam to melt away part of the resistors to change their values. The precision required in the process and the irreversibility of the process make this option expensive. Because laser

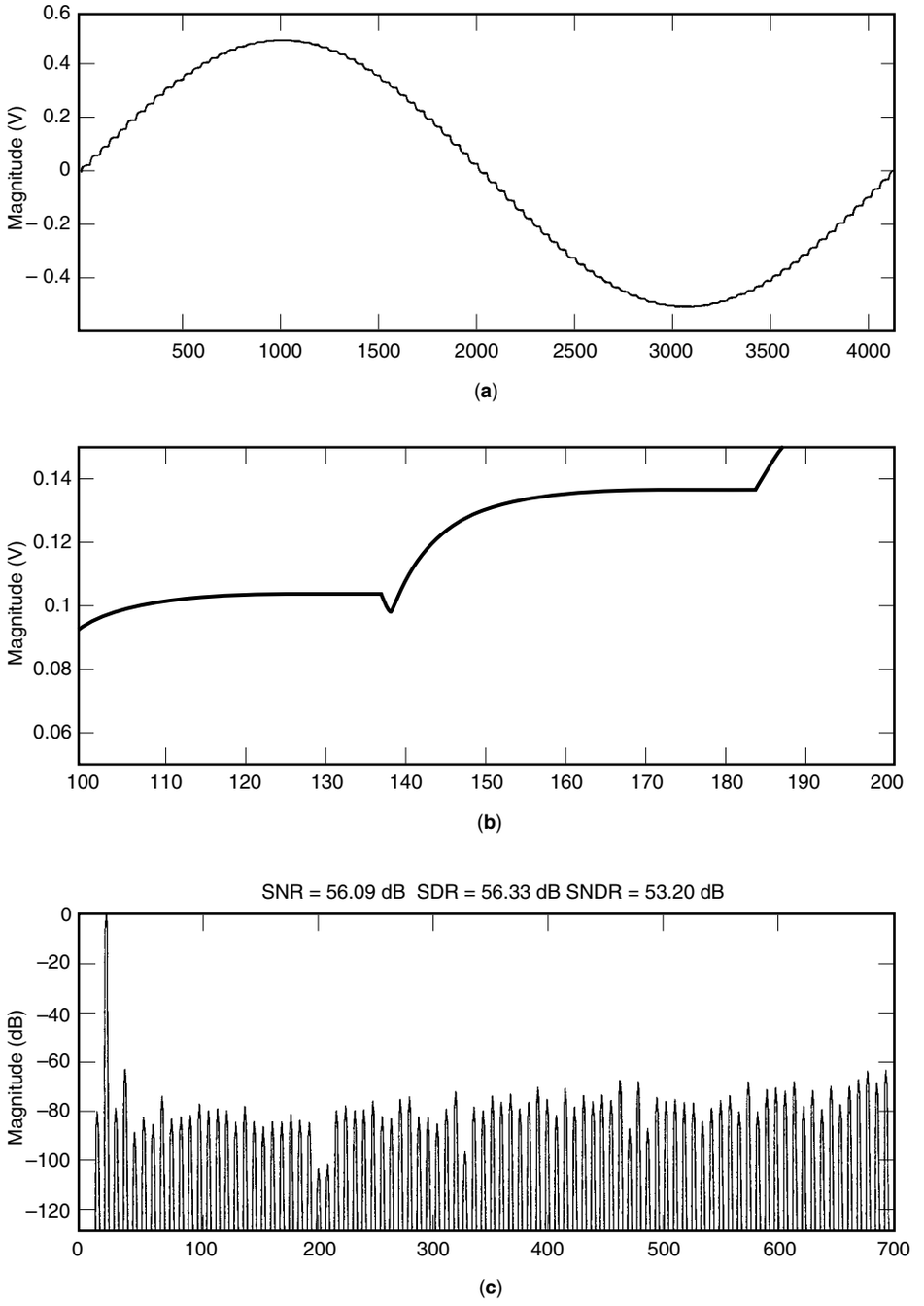

Figure 17. 12 bit DAC output with 5% glitch: (a) time waveform, (b) its details, and (c) its FFT.

trimming is continuous, very accurate measurement tools are needed to detect low levels of mismatch. The Zener-zapping method creates a resistor of desired value by inserting a series of small incremental resistors and by selectively bypassing them. Zener diodes connected in parallel with resistors are melted with short current pulses over hundreds of milliamperes. Fusible aluminum links are also used in this discrete trimming technique. Like laser trimming, the Zener-zapping method is also irreversible and requires precision equipment to measure matching errors. Furthermore, the stability of the trimmed values over a long period of time is still in question. The new trend is toward a more intelligent solution that involves the ability to recalibrate on demand by moving the error measurement process on-chip.

Dynamic Matching

The idea of dynamic matching is to improve the matching accuracy by time-averaging two component values (8). Figure 20 explains how the dynamic-matching technique can be applied to current dividers. Assume that two current sources are mismatched by Δ. By commuting the two currents I_1 and I_2 at

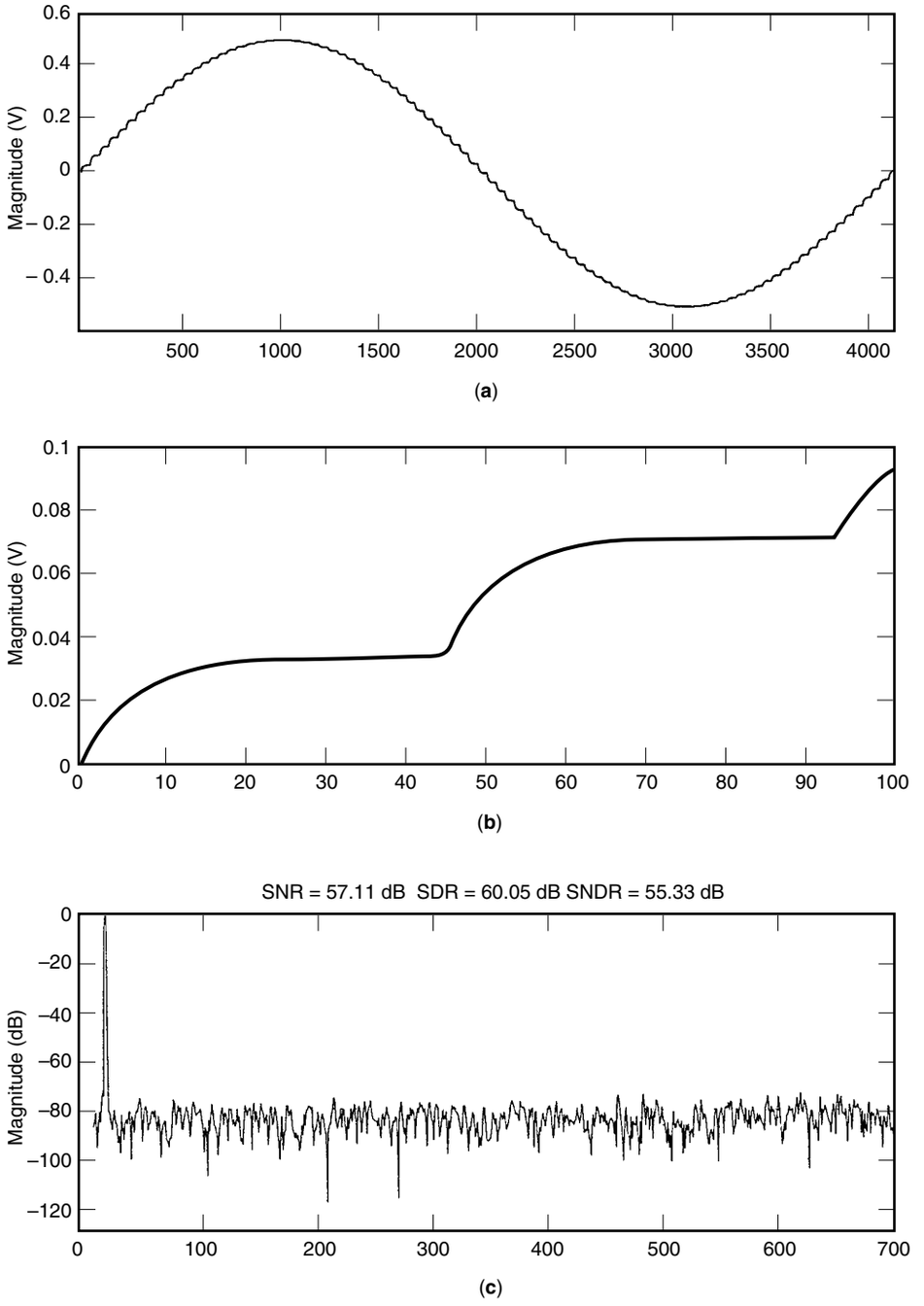

Figure 18. 12 bit DAC output with 5% jitter: (q) time waveform, (b) its detail, and (c) its FFT.

high speed to two new output ports with a 50% duty, the mismatch error $\Delta/2$ will be modulated by the high clock frequency. If these are lowpass filtered, the average current I will come out of the two terminals. This dynamic matching can be generalized for a large number of matching components using the butterfly-type randomizer as shown in Fig. 21 for the four-element matching case. The switches are controlled so that the four outputs can be the average of the four currents. In this case, the modulating frequency is half of the clock frequency. For matching a large number of elements, this fixed pattern noise moves to lower frequencies. To reduce this pattern noise, the clock frequency should be raised or the lowpass filters should have a sharper cutoff slope. A more sophisticated randomizer selects the switching path randomly with an equal probability to each output terminal using a pseudorandom number generator. This general randomizer can distribute mismatch errors over a wider frequency range. At least an order of magnitude improvement is achieved using the dynamic matching element method.

An alternative concept to dynamic matching is to replicate current or voltage unit element. The idea is to copy one master voltage or current repeatedly on voltage or current sam-

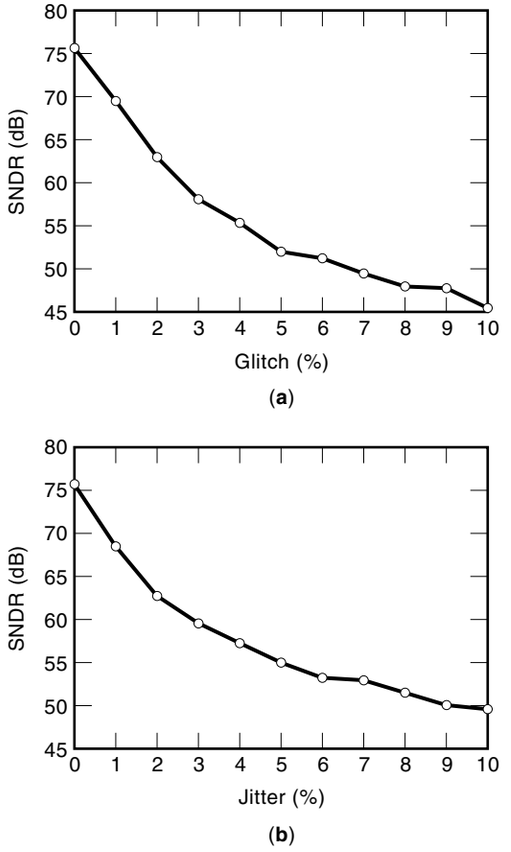

Figure 19. SNDR vs. (a) glitch and (b) clock jitter.

plers so that the sampled ones can be used to ratio elements. The voltage is usually sampled on the holding capacitor of sample-and-hold amplifier, and the current is sampled on the gate of the MOS transistor (9). This sampling method is ultimately limited by sampling errors. If the sampling accuracy

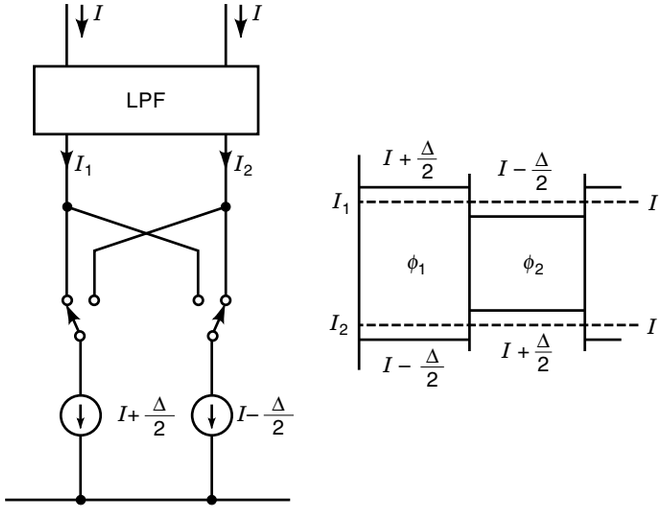

Figure 20. Dynamic matching of two currents and modulated two currents.

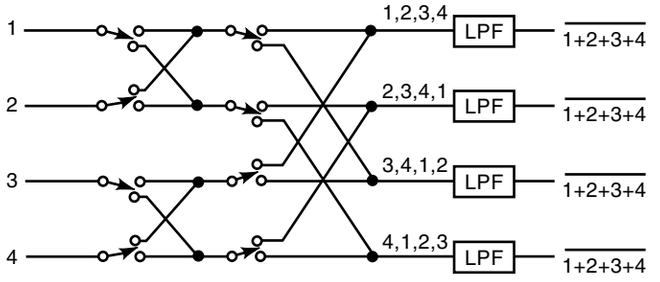

Figure 21. Butterfly-type randomizer.

is limited to N bits, it is equivalent to having passive components matching of N bits. Note that voltage and current sampling schemes alone are not sufficient enough to make a high-resolution DAC. This approach is generally limited to creating MSBs for segmented DACs or for multistep ADCs. The amount of time required to copy one master source repeatedly but accurately makes it impractical for high-speed DAC applications.

Electronic Calibration

Calibration is another alternative to the trimming and dynamic matching methods. It has become an intelligent electronic solution preferred to the factory trimming process. The electronic calibration predistorts the DAC transfer characteristic so that the DAC linearity can be improved. DAC nonlinearity errors are measured and stored in memory. Later during normal operation, these errors are subtracted from the DAC output. Error subtraction can be done either in the analog domain or in the digital domain. Calibration methods differ from one another in the measurement of nonlinearity errors. The most straightforward method is a direct code-mapping technique. All DAC code errors must be measured and stored in ROM. This method is limited because it requires precision measurements and large digital memory. A more robust way of calibrating a DAC electronically is self-calibration. This incorporates all the calibration mechanisms and hardware on the DAC as a built-in function so that users can recalibrate whenever calibration is necessary.

Self-calibration is based on the assumption that the segmented DAC linearity is limited by the MSB matching so that only errors of the MSBs can be measured, stored in memory, and recalled during normal operation. There are two different methods of measuring the MSB errors. In one method, individual binary bit errors, usually appearing as component mismatch errors, are measured digitally (10). The total error, which is called a code error, is computed from individual-bit errors depending on the digital code during normal conversion. The other method measures segment errors and accumulates them to obtain code errors. Because code errors are stored in memory, there is no need for the digital code-error computation during normal operation (11). The former requires less digital memory, whereas the latter requires fewer digital computations.

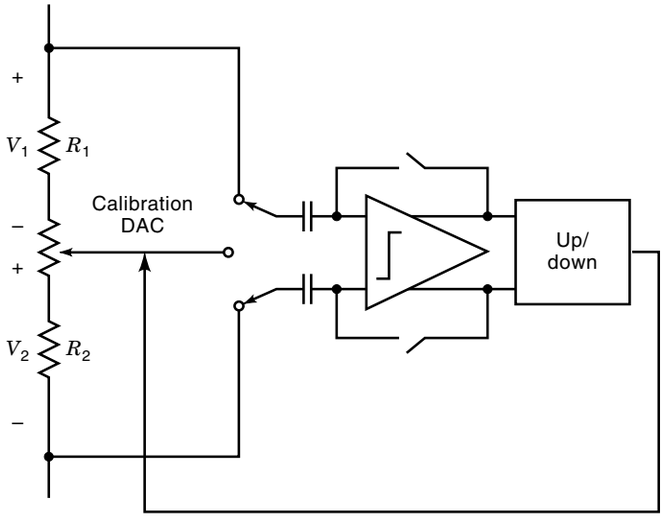

Figure 22. Resistor ratio calibration scheme.

Two examples of DAC ratio measurements are conceptually explained in Figs. 22 and 23. In the resistor ratio measurement of Fig. 22, the voltage across the two resistors is sampled on the capacitors, and the comparator is nulled. During the next cycle, the capacitor bottom plates are connected to the center point to sense the difference of V_1 and V_2. The comparator decision will move the center tap connected to the fine calibration DAC. This capacitive sensing is limited by the capacitor-matching accuracy between C_1 and C_2. To average out this capacitor mismatch error, the same measurement can be repeated with C_1 and C_2 swapped. The final center tap is set to the average of the two tap values obtained with the two measurements. The same principle can be applied to measure the current difference as shown in Fig. 23. The calibration DAC measures I_1 first using the up/down converter and moves on to measure I_2. The difference between the two measurements is the current mismatch error.

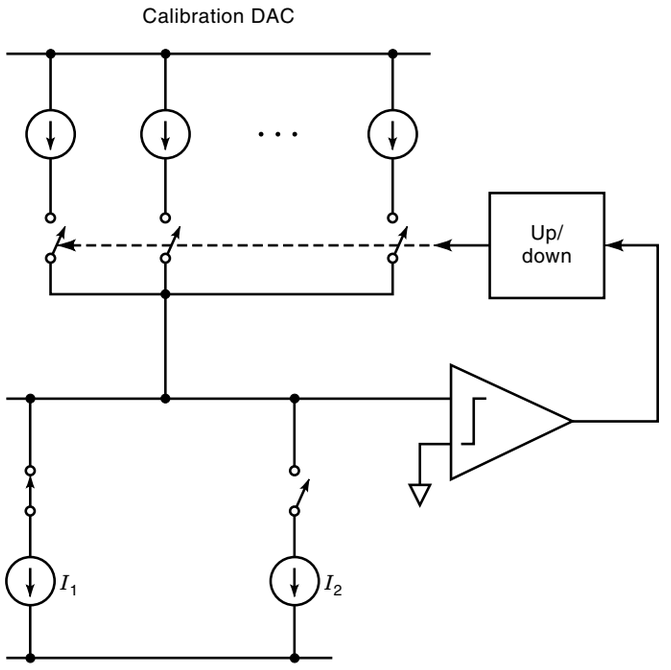

Figure 23. Current ratio calibration scheme.

INTERPOLATIVE OVERSAMPLING TECHNIQUE

All DACs have a discrete output level for every digital input word applied to their input. Although digital numbers can grow easily, generating a large number of distinct analog output levels is a difficult task. The oversampling interpolative DAC achieves fine resolution by covering the signal range with a few widely spaced levels and interpolating values between them. Rapid oscillation between coarse output levels is controlled so that the average output may represent the applied digital word with reduced noise in the signal band (12). This process is a tradeoff between speed and resolution. The oversampling idea is to achieve high resolution with less accurate component matching and has been most widely used to make DACs that need high resolution at low frequencies.

The interpolative oversampling DAC architecture is shown in Fig. 24. A digital interpolator raises the word rate to a frequency well above the Nyquist rate for oversampling. The interpolated data stream is applied to a digital truncator to shorten the word length. This data stream of shorter words, usually a one-bit stream, is converted into analog waveform at the oversampling rate. This oversampled output has a low truncation error in the signal band, and the out-of-band truncation noise is filtered out using analog low-pass filter (LPF). Figure 25 illustrates this oversampling D/A conversion process.

The sampling rate upconversion for this is usually done using two upsampling digital filters. The first filter, usually a two to four times oversampling FIR filter, shapes the signal band for sampling rate upconversion and equalizes the passband droop resulting from the $(\sin x)/x$ filter for higher-rate oversampling. The truncator is made of a feedback system called a delta-sigma modulator that pushes the truncation error out to high frequencies while passing the signal band unattenuated. Using a linearized model, the z-domain transfer function of the general digital truncator shown in Fig. 24 is

$$V_0(z) = \frac{H(z)}{1+H(z)}V_i(z) + \frac{1}{1+H(z)}E(z) \qquad (11)$$

where $E(z)$ is the truncation error. The loop filter $H(z)$ is chosen so that the truncation error may be high-pass filtered while the input signal is low-pass filtered. The order of noise shaping depends on the order of $H(z)$.

The oversampling effect begins to appear when the oversampling ratio is larger than 2. The noise of the Nth-order loop is suppressed by $6N + 3$ dB for every doubling of the sampling rate, providing $N + 0.5$ extra bits of resolution. Therefore, the dynamic range achievable by oversampling is

$$\text{Dynamic range} \approx (6N+3)(\log_2 M - 1)\ \text{dB} \qquad (12)$$

where M is the oversampling ratio between the sampling frequency and twice the signal bandwidth. For example, a sec-

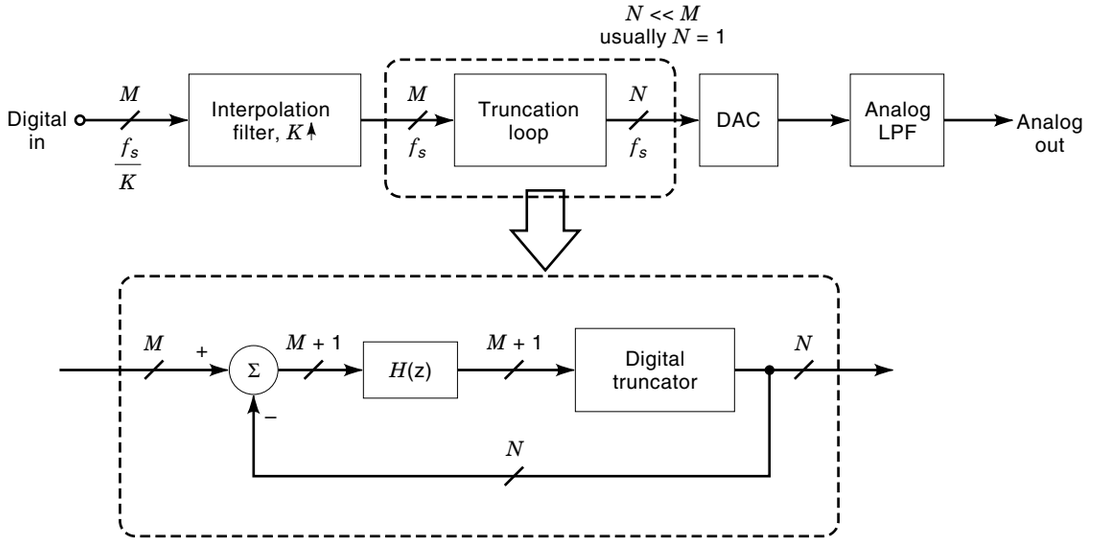

Figure 24. Interpolative oversampling DAC system.

ond-order loop with 256 times oversampling gives a dynamic range of over 105 dB, but the same dynamic range can be obtained using a third-order loop with only 64 times oversampling.

Digital Truncator

A noise-shaping delta-sigma modulator for digital truncation can be built in many different ways. The simplest three examples are shown in Fig. 26. The first-order loop shown in Fig. 26(a) outputs only N MSBs out of $M + 1$ bits from the integrator output. The remaining truncation error of $M - N + 1$ is fed back to the digital integrator along with the input. Therefore, the output of the first-order modulator becomes

$$Y(z) = z^{-1}X(z) - (1 - z^{-1})E(z) \tag{13}$$

This implies that the input appears at the output just delayed, but the truncation error in the integrator loop is high-pass filtered with one zero at dc.

In general, first-order designs tend to produce correlated idling patterns and require a high oversampling ratio to suppress inband truncation noise effectively. By increasing the order of the error transfer function, the inband truncation error can be suppressed with steeper-slope high-pass filter. The standard second-order modulator can be implemented as shown in Fig. 26(b) using a double integration loop. The output of the second-order modulator becomes

$$Y(z) = z^{-2}X(z) + (1 - z^{-1})^2E(z) \tag{14}$$

This second-order modulator is vastly superior to the first-order one both in terms of the oversampling ratio and the improved randomness of the idling patterns. However, even the second-order loop is not entirely free of correlated fixed patterns in the presence of small constant inputs. If loop order is higher than 3, fixed pattern noise does not exist. High-order modulators can be implemented using the same method as in the standard filter design (13).

Another variation of the high-order modulator is the cascading method (14). Rather than using error feedback, modulators can be cascaded to reduce the truncation noise in the same way. Figure 26(c) shows a second-order cascade modulator example. The truncation error E_1 of the first modulator is modulated again using another first-order modulator. If the truncation error of the second modulator is E_2, the outputs of the two modulators Y_1 and Y_2 become

$$Y_1(z) = z^{-1}X(z) - (1 - z^{-1})E_1(z) \tag{15}$$

$$Y_2(z) = z^{-1}E_1(z) - (1 - z^{-1})E_2(z) \tag{16}$$

respectively. The two outputs are added after they are multiplied by z^{-1} and $1 - z^{-1}$, respectively. Then the final output becomes

$$Y(z) = z^{-2}X(z) - (1 - z^{-1})^2E_2(z) \tag{17}$$

This is the same second-order modulator output given by Eq. (14). The problem with the cascading approach is that the output word is always longer than 2, even though two cascaded modulators have one-bit output ($N_1 = N_2 = 1$). That is, the higher-order noise shaping is possible with cascading, but error occurs in the output D/A conversion. The multibit output DAC requires the same accuracy as conventional DACs.

One-Bit Switched-Capacitor DAC/Filter

Digital processing for modulation and filtering is not limited in accuracy. However, all oversampling DACs are followed by an analog DAC and an analog low-pass filter, and DAC performance depends entirely on the rear-end analog components. The idea of oversampling to alleviate the component-matching requirement for high resolution does not apply to multibit DACs. Therefore, most oversampling DACs are designed with one-bit output. It is true that a continuous-time low-pass filter can convert the one-bit digital bitstream into an analog waveform, but it is difficult to construct an ideal

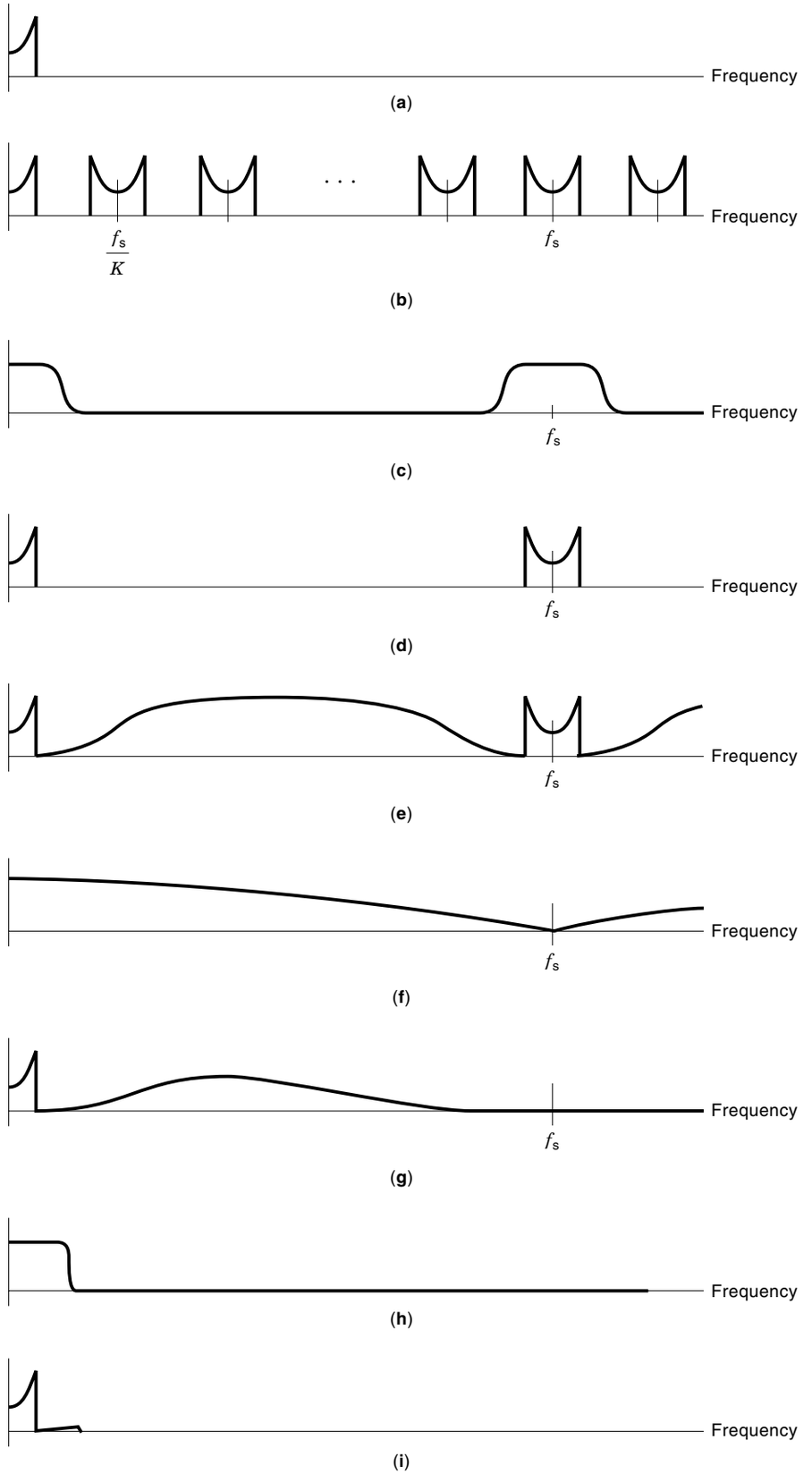

Figure 25. Oversampling D/A conversion: (a) input spectrum, (b) sampled spectrum, (c) interpolation filter, (d) interpolated spectrum, (e) noise-shaped spectrum, (f) sample-and-hold frequency response, (g) sample-and-held spectrum, (h) low-pass filter, and (i) DAC output.

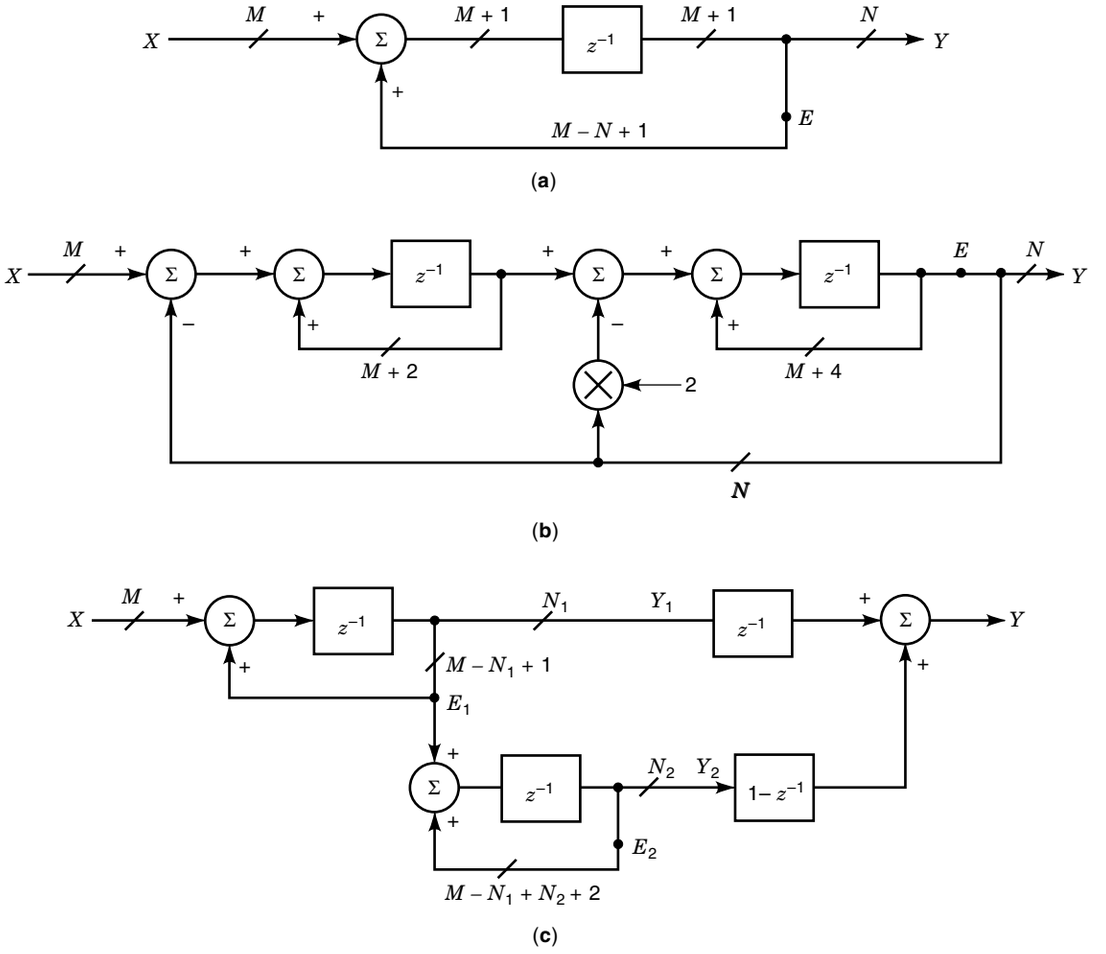

Figure 26. Digital truncators: (a) first-order loop, (b) second-order loop, and (c) cascaded architecture.

undistorted digital waveform without clock jitter. If the bitstream is converted into a charge packet, a high linearity is guaranteed as a result of the uniformity of the charge packets. For this reason, most high-resolution oversampling DACs are implemented using a switched-capacitor DAC/filter combination as the rear-end.

A one-bit switched-capacitor DAC with one-pole roll-off can be built as shown in Fig. 27(a) using two-phase nonoverlapping clocks ϕ_1 and ϕ_2. The digital signal is fed on the bottom plate of C_2 by sampling V_r or $-V_r$ depending on the digital bit. The end result is that a constant amount of charge C_2V_r is either added or subtracted from the integrator formed by C_1 and the operational amplifier. Because the digital signal is converted into a charge packet, the shape of the charge packet is not important as far as the incremental or decremental charge amount is constant. The use of the prime clocks is to make switch feedthrough errors constant by turning off the top plate switches slightly earlier than the bottom plate switches. The bandwidth of the filter is set to $f_sC_2/(C_1 + C_2)$. Operational amplifiers for this application should have high dc gain and fast slew rate. Operational amplifiers start to slew when an input voltage larger than its linear range is applied. When the charge packet of the sampled reference voltage is dumped onto the input summing node, it causes a sudden voltage step at the summing node. This one-bit DAC/filter configuration helps to reduce this voltage step because charge is directly dumped on the integrating capacitor.

The shaped noise is out of band and does not affect the inband performance directly. More precise filtering of the shaped noise can be done with higher-order switched-capacitor filters combined with the one-bit DAC. One second-order filter combined with a one-bit DAC is shown in Fig. 27(b). It is a second-order switched-capacitor biquad implementing the low-pass function. In oversampling applications, even a one-pole *RC* filter substantially attenuates high-frequency components around f_s, but higher-order continuous-time smoothing filters can be used. Analog filters for this application are often implemented using a cascade of Sallen–Key filters made of emitter follower unity-gain buffers.

DAC APPLICATIONS

DAC architectures are mainly determined by the operating speed. In general, D/A conversion rate is inversely proportional to DAC resolution. DAC applications can range from 20 bits at a few tens of hertz to 6 bits at a couple gigahertz. The most well-known applications are for 8 bit telephone voice,

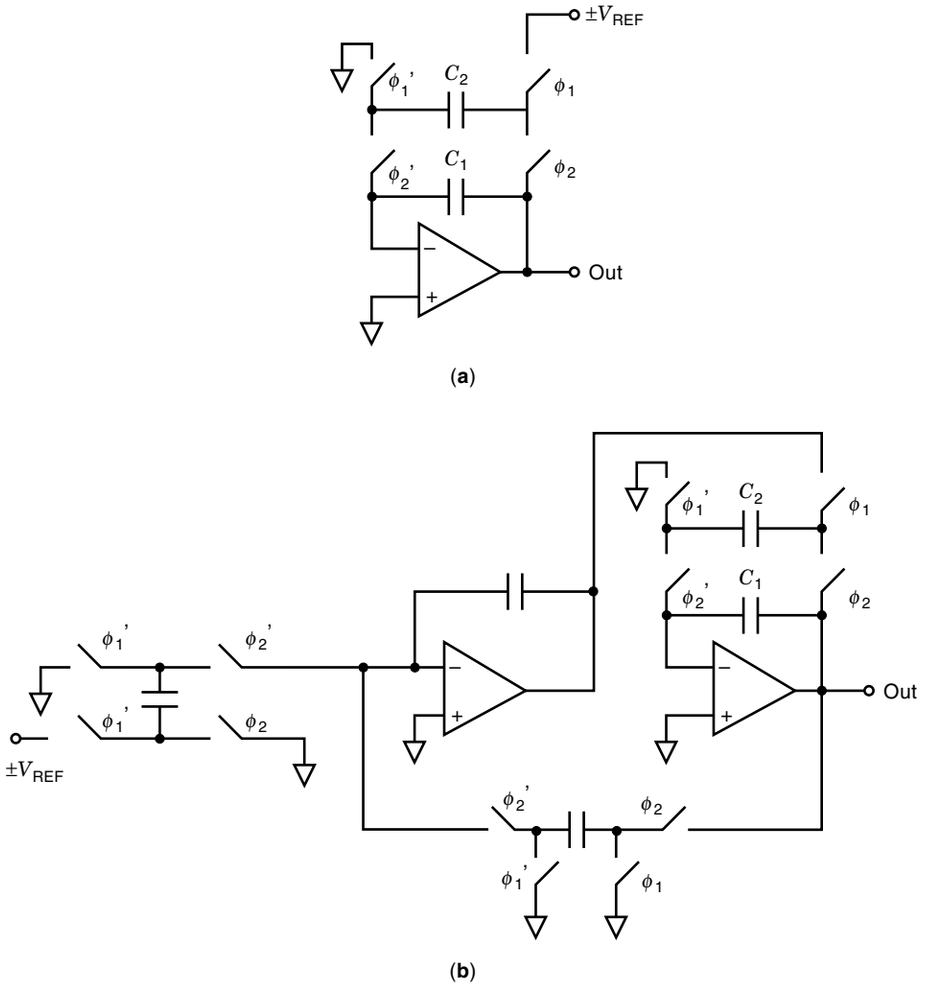

Figure 27. Time switched-capacitor DACs: (a) first order and (b) second order.

16 bit digital audio, 8 bit digital video, 8 to 12 bit waveform generation and high-resolution graphics, 12 bit communication channel, and 8 to 14 bit wireless radio and cell site. As discussed, two architectures stand out. They are oversampling DAC at low frequencies and segmented DAC at high frequencies. The former trades resolution for speed, but the latter needs resolution-enhancing techniques for high linearity because there is no room for speed to trade with resolution.

The 8 bit pulse-coded modulation (PCM) within a frequency band of 300 Hz to 3.4 kHz has been a standard since the 1970s for the telephone voice channel. One thing to note in the voiceband quantization is the use of the so-called μ-law coding scheme so that low-level signals can be quantized more finely than high-level signals. Because low-level resolution approaches that of a 14 bit system, it is possible to maintain constant SNR over a wide range of signal magnitude. Most voiceband DACs have been made using capacitor-array DACs and integrated in voiceband coder/decoder (CODEC) systems using MOS processes.

No field places a higher premium on DAC performance than digital audio. New multimedia applications have opened up new audio DAC applications as set forth in the audio CODEC standard AC'97 (1). The new standard covers bit rates of 8, 11.025, 16, 22.05, 32, 44.1, and 48 kHz. Even the 16 bit 44.1 kHz digital audio standard set for CD in the 1980s is now seriously challenged by many improved high-definition CD (HDCD) standards. The new DVD (Digital Versatile Disc) format has the capacity of high-resolution uncompressed 20 bit audio at a 98 kHz sampling rate. Video or movie systems also rely heavily on digital sound. Digital surround sound systems such as Dolby AC-3, DTS (Digital Theater System), and 5.1 digital surround systems have already become household names. R-2R or segmented current DACs with laser-trimmed film resistors are used to make 18 to 20 bit extremely linear DACs for digital audio. However, the oversampling technique can meet the demanding performance requirements of digital audio without trimming.

In the 1990s, digital video is leading the multimedia revolution in modern digital communications. Compressed digital video images are now accessible over computer networks from anywhere and at anytime. A 2 h MPEG-encoded (Motion Picture Experts Group) compressed video can now be stored on a 5 inch (12.7 cm) optical disc along with digital surround sound channels. Digital satellite networks provide quality digital video, and new HDTV (High-Definition TV) will eventually replace existing analog TV. All these are now possible through digitization of video signals. Standard NTSC video luminance signal covers about a 4 MHz bandwidth with color subcarrier at 3.58 MHz. The standard video luminance signal

is sampled with 8 bit resolution at 13.5 MHz, whereas color signals are sampled at 6.75 MHz or lower depending on the video formats. As more digital signal processing is used, the trend is to use 10 bits. In the new HDTV, the sampling rate is about five times higher at above 75 MHz.

DACs have become indispensable in the digital era. Digitization started from voiceband in the 1970s, digital audio in the 1980s, and digital video in the 1990s is about to move into intermediate frequency (IF) and RF for wireless telecommunications systems. Digitizing the IF in digital wireless systems makes it possible to perform data modulation and demodulation digitally in software. Low-spurious ADC and DAC are key components that will enable this software programmable radio environment. The architecture that can deliver both high-resolution and high-speed performance is limited to the segmented DAC. The current art estimated by recent works is projected to be 14 to 16 bits at low hundreds of megahertz range depending on process with trimming (15,16). However, the DAC performance envelope will be pushed out further as new architectures and device processes evolve.

BIBLIOGRAPHY

1. Analog Devices, Creative Labs, Intel, National Semiconductors, Yamaha, "Audio Codel '97," Rev. 1.03, Sept. 1996.
2. M. J. M. Pelgrom and M. Roorda, An algorithmic 15-Bit CMOS digital-to-analog converter. *IEEE J. Solid-State Circuits,* **SC-23**: 1402–1405, 1988.
3. J. J. Pastoriza, K. Krabbe, and A. A. Molinari, A High-Performance Monolithic D/A Converter Circuit. *IEEE Int. Solid-State Circuits Conf.,* Feb. 1970.
4. J. M. McCreary and P. R. Gray, All-MOS charge redistribution analog-to-digital conversion techniques—Part I. *IEEE J. Solid-State-Circuits,* **SC-10**: 371–379, 1975.
5. D. M. Freeman, Slewing distortion in digital-to-analog conversion. *J. Audio Eng. Soc.,* **25**: 178–183, 1977.
6. R. B. Craven, An Integrated Circuit 12-Bit D/A Converter. *Dig. IEEE Int. Solid-State Circuits Conf.,* Feb. 1975.
7. D. T. Comer, A monolithic 12-bit DAC. *IEEE Trans. Circuits Syst.,* **CAS-25**: 504–509, 1978.
8. R. J. Van de Plassche, Dynamic element matching for high accuracy monolithic D/A converters. *IEEE J. Solid-State Circuits,* **SC-11**: 795–800, 1976.
9. D. W. J. Groeneveld et al., A self-calibration technique for monolithic high-resolution D/A converters. *IEEE J. Solid-State Circuits,* **SC-24**: 1517–1522, 1989.
10. H. S. Lee, D. A. Hodges, and P. R. Gray, A self-calibrating 15-bit CMOS A/D converter. *IEEE J. Solid-State Circuits,* **SC-19**: 813–819, 1984.
11. S. H. Lee and B. S. Song, Digital-domain calibration of multistep analog-to-digital converters. *IEEE J. Solid-State Circuits,* **SC-27**: 1679–1688, 1992.
12. J. C. Candy and A. H. Huynh, Double interpolation for digital-to-analog conversion. *IEEE Trans. Commun.,* **34**: 77–81, 1986.
13. W. L. Lee and C. G. Sodini, A Topology for Higher-Order Interpolative Coder. *Proc. Int. Symp. Circuits Syst.,* May 1987.
14. T. Hayashi et al., A Multistage Delta-Sigma Modulator without Double Integration Loop. *IEEE Int. Solid-State Circuits Conf.,* Feb. 1986.
15. D. Mercer, A 16-b D/A Converter with Increased Spurious Free Dynamic Range, *IEEE J. Solid-State Circuits,* **SC-29**: 1180–1185, 1994.
16. B. J. Tesch and J. C. Garcia, A low-glitch 14-bit 100MHz D/A converter. *IEEE Bipolar / BiCMOS Circuits and Technology Meeting,* 1996.

BANG-SUP SONG
University of Illinois

DIGITAL VIDEO. See IMAGE SEQUENCES; MULTIMEDIA VIDEO.

DIGITAL WIRELESS COMMUNICATIONS. See CELLULAR RADIO.

DIODE LASERS, MANUFACTURING. See LASER DESKTOP MACHINING.

DIODE, RESONANT TUNNELING. See RESONANT TUNNELING DIODES.