Chapter 3
Current and Emerging Trends in the Design of Digital-to-Analog Converters

Sidharth Balasubramanian, Vipul J. Patel and Waleed Khalil

3.1 Introduction

The advent of digital computing, coupled with the continued scaling of CMOS devices, has led to the rapid growth in theory and applications of digital signal processing (DSP). In order to leverage the available increase in speed, complexity, and integration, high-performance analog-to-digital converters (ADCs) and digital-to-analog converters (DACs) are needed to ensure the highest signal fidelity when moving between the analog and digital domains. Traditionally, high-speed ADCs have attracted more attention in the research and scientific community than their DAC counterparts, mainly due to the inherent shift towards receivers in the study and analysis of radio systems [1].

This chapter aims to describe some of the design challenges and emerging trends for high-speed and high-resolution digital-to-analog converters. Section 3.1 presents an overview of the digital-to-analog conversion process. Section 3.2 delves into DAC characterization by outlining different sources of error and metrics used to quantify the DAC performance. A summary of DAC topologies and circuit limitations in the context of current-steering (CS) DACs is provided in Sects. 3.3 and 3.4, respectively. Section 3.5 details four major considerations in the design space of CS DACs. Realizing that the segmented architecture is the de facto standard in high-resolution DACs, a supplemental approach to segmentation is presented in Sect. 3.6. An in-depth survey of current and emerging architectural trends in high-performance DACs is discussed in Sect. 3.7. Finally, Sect. 3.8 concludes with a summary and a brief discussion on future directions.

S. Balasubramanian · W. Khalil (✉)
Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH 43210, USA
e-mail: khalil@ece.osu.edu

V. J. Patel
Air Force Research Laboratory, Wright-Patterson AFB, Dayton, OH 45433, USA
e-mail: Vipul.Patel@us.af.mil

P. Carbone et al. (eds.), *Design, Modeling and Testing of Data Converters*, Signals and Communication Technology, DOI: 10.1007/978-3-642-39655-7_3,

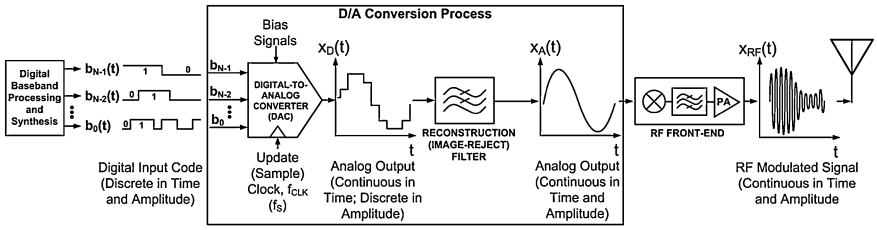

Fig. 3.1 Conventional transmitter block diagram

Figure 3.1 illustrates a conventional transmitter block diagram where the DAC is shown as the fundamental building block for waveform synthesis. Preceding the DAC, a digital baseband processor is used to generate N digital bits $(b_{N-1}(t), b_{N-2}(t), \ldots, b_0(t))$ that represent the waveform to be synthesized. The DAC and reconstruction filter then translate the digital input codes to an analog waveform, $x_A(t)$. Following the filter, $x_A(t)$ is upconverted to the desired RF band and amplified as represented by $x_{RF}(t)$.

The digital to analog translation process involves weighting and summing of voltages, charges, or currents derived from the input digital codes. A representative analog value, typically in the voltage domain, is then produced, where the full scale voltage (V_{FS}) is defined as the difference between the maximum and minimum voltage levels. Note that for current-steering DACs, the output current is converted to a voltage using a resistive load. The simplified representation of an N-bit DAC can be described as having N binary input bits defined by the following vector:

$$\hat{B} = \{b_{N-1}, b_{N-2}, b_{N-3}, \ldots, b_1, b_0\} \tag{3.1}$$

where $b_i \in \{0, 1\}$, b_{N-1} is the most significant bit (MSB), and b_0 is the least significant bit (LSB). The binary vector, $\hat{B}$, is converted to a corresponding decimal value, D:

$$D = \sum_{i=0}^{N-1} 2^i b_i \tag{3.2}$$

This weighted decimal value is then multiplied by a gain factor, such as V_{LSB} (where $V_{LSB} = V_{FS}/2^N$) for voltage-or charge-based DACs and I_{LSB} (where $I_{LSB} = I_{FS}/2^N$) for current-steering DACs, to yield the final analog voltage (or current):

$$V_{OUT}(D) = D \cdot V_{LSB} \tag{3.3}$$

$$I_{OUT}(D) = D \cdot I_{LSB} \tag{3.4}$$

The high-level architecture of a digital-to-analog converter can be conceptualized into two basic functions: (1) zero-order holding and (2) digital-to-analog translation, as illustrated in Fig. 3.2. Assuming an ideal sampling process, the

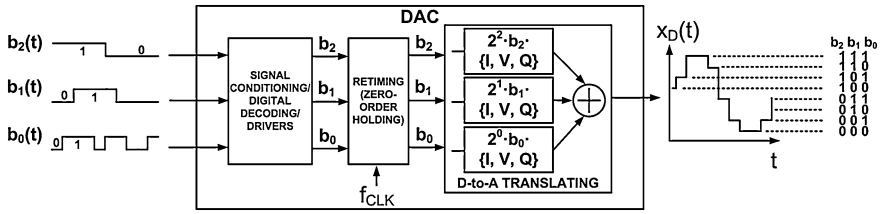

Fig. 3.2 Basic functions in a 3-bit digital-to-analog converter

DAC is fed with data in the form of impulse trains. However, finite switching time in real circuits makes it impossible to generate these impulses. Instead, the input signal is supplied as square-wave pulses, where a retiming register is often needed to ensure all digital input bits are aligned and synchronized to an update clock, f_{CLK} $(1/T_{CLK})$. This process of square waveform generation through holding the input for the duration of T_{CLK} is known as the zero-order hold (ZOH). The next step involves the actual translation from the digital to the analog domain. The converter circuit assigns analog (i.e. current, voltage, or charge) weights corresponding to the digital input code, and then sums them up to the final discrete (stair-step) output, $x_D(t)$. The reconstruction filter, also known as the image-reject filter, is typically an external component that smoothens the stair-step waveform by eliminating out-of-band frequencies. For signals generated at baseband, a low-pass filter is designed to eliminate frequencies greater than the Nyquist bandwidth (DC to $f_{CLK}/2$). However, in the case of intermediate signal generation beyond the first Nyquist zone, a bandpass filter is utilized.

3.2 Characterizing the DAC

In general, DACs are known to suffer from four inherent limitations: (1) quantization or truncation error, (2) image replicas, (3) nonlinear spurs, and (4) hold distortion. These limitations are illustrated in Fig. 3.3.

The finite resolution of the DAC results in inherent quantization noise that ultimately sets the minimum noise floor. Another inherent limitation in the DAC is attributed to its sampling nature. Assuming a DAC clocked at f_{CLK} is used to synthesize an output signal at f_0, image replicas are generated at $f_{CLK} \pm f_0$, $2f_{CLK} \pm f_0$, $3f_{CLK} \pm f_0$ and so on. Similar to an ADC, a DAC's output spectrum is divided into Nyquist zones defined at $nf_{CLK}/2$ where $n = 1, 2, 3, \ldots$, as illustrated in Fig. 3.3. When the generated analog signal approaches a Nyquist zone, the signal and its corresponding replica are close and comparable in magnitude. Hence, the replica acts as a strong interferer to the signal of interest. This requires a brickwall reconstruction filter and thus restricts the DAC signal generation to well below the Nyquist frequency. Another major limitation involves the nonlinear behavior in the DAC, which is well pronounced when the desired signal approaches the boundary of the first Nyquist zone. This results in high levels of harmonic and intermodulation

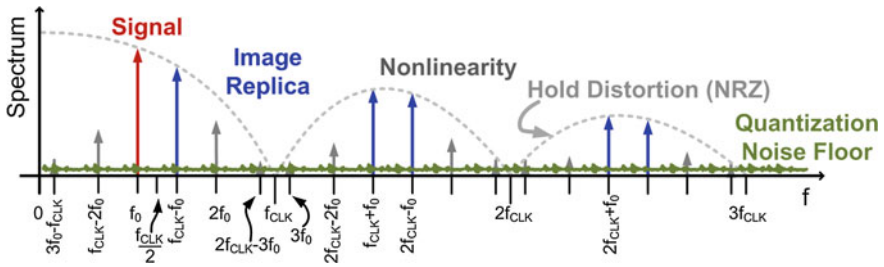

Fig. 3.3 Inherent DAC limitations

distortion products that can interfere with the desired band. Furthermore, the harmonic mixing amongst the DAC's update clock, the desired signal, and the distortion products, results in spurious components at $mf_{CLK} \pm nf_0$, for all integer m, n. Ideally, the signal and its image replicas are of equal magnitude at all points in the frequency spectrum. However, due to the imposed ZOH, the signal, image replicas, and distortion products are attenuated with a $\sin c$ response as described by (5), where D is the ON period during T_{CLK}. This effect is referred to as hold distortion. While there are several zero-order hold variations, the most often used is a non-return-to-zero (NRZ), where the DAC output is held for the entire duration of T_{CLK} (i.e. $D = 100$ %). However, the fast roll-off of the *sinc* response limits operation past the first Nyquist zone. Other hold techniques such as return-to-zero (RZ) can extend operation to multiple Nyquist zones. For instance, reducing the hold duty cycle to 50 % can extend the first *sinc* null to $2f_{CLK}$. Figure 3.4 illustrates both of these ZOH techniques in the frequency and time domains.

$$H(f) = \frac{\sin\left(\pi D \frac{f}{f_{CLK}}\right)}{\pi \frac{f}{f_{CLK}}} \exp\left(-j\pi D \frac{f}{f_{CLK}}\right) \tag{3.5}$$

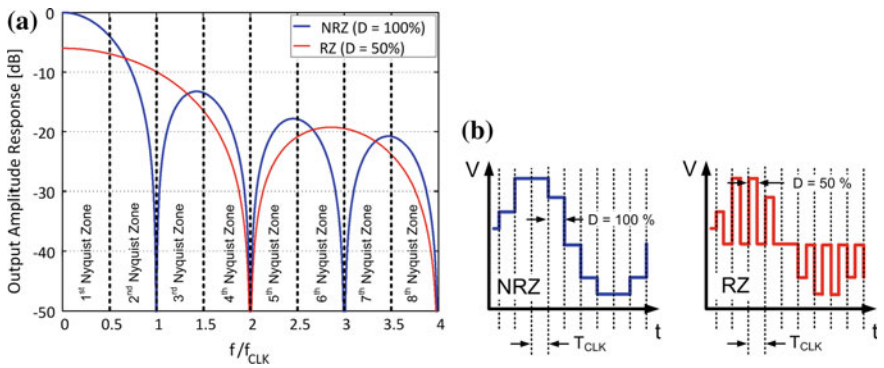

Fig. 3.4 **a** Common zero-order hold (ZOH) distortions segmented into Nyquist zones **b** NRZvsRZ Time domain ZOH for NRZ and RZ

3.2.1 Timing and Amplitude Errors

In addition to the aforementioned DAC limitations, there exist other numerous forms of DAC nonidealities which can be attributed to the physical circuit implementation. Specifically, these nonidealities can be lumped into two broad categories: timing-related errors and amplitude-related errors. While not seen as independent, each of the two categories can be further classified into static and dynamic errors. In basic terms, static refers to time-invariant errors that are induced by random or systematic mismatch effects. In contrast, dynamic refers to time-variant errors that can be attributed to code-dependent[1] transitions, jitter, glitches, and impedance variation. Figure 3.5 illustrates the four categories of error-static and dynamic timing as well as static and dynamic amplitude errors.

Examples of static timing errors Fig. 3.5a can be observed in delay mismatches amongst retiming flip-flops, clock skew between DAC circuit blocks, and delays attributed to switching pair mismatches. On the other hand, dynamic timing error

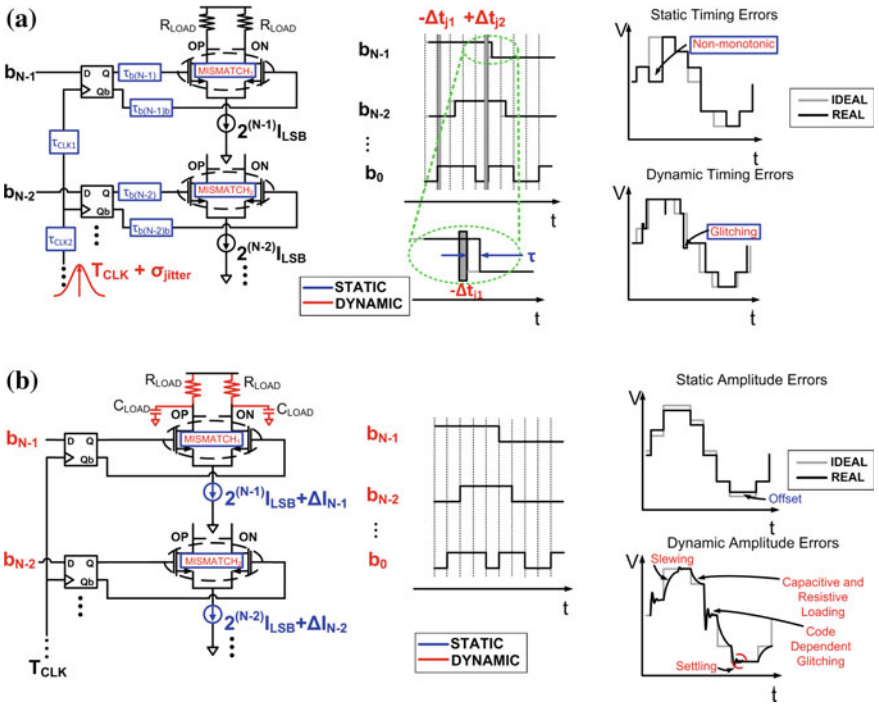

Fig. 3.5 **a** Static and dynamic timing errors **b** Static and dynamic amplitude errors

[1] Code-dependency is equivalently mentioned as signal dependency, where signal refers to the desired output waveform.

includes both random and deterministic clock jitter or phase noise. Similar to quantization noise, random jitter can increase the DAC's noise floor, while deterministic jitter is manifested as spurs in the DAC's output spectrum. Examples of amplitude-related errors are illustrated in Fig. 3.5b. The relative mismatch between the weighted current sources can induce static amplitude errors in the form of differential and integral nonlinearities as well as offset error. Whereas dynamic errors caused by large code transitions, parasitic loading, finite slewing, and finite settling times, further degrade the output amplitude accuracy. Timing-related errors can also be induced concurrently with amplitude-related errors resulting in varying settling times, slewing rates, and glitches.

3.2.2 Performance Metrics

In general, the performance of any given DAC can be quantified using a set of static and dynamic metrics. The choice of metrics depends on the desired application ranging from high precision systems such as potentiometers and medical instrumentation to waveform synthesis in high-speed communication systems. This section briefly introduces the commonly used metrics to characterize the DAC performance [2–5].

3.2.2.1 Static Metrics

Unlike the ADC's stair-step transfer function, the DAC's response is represented by discrete points, which maps a specific digital input code to a discrete analog value. The transfer function of the real DAC and its comparison to an ideal transfer function are used to evaluate the static (i.e. near DC) performance metrics. Generating the transfer function plot from a real DAC is a straightforward process by simply applying a digital input code and observing the output using a high-precision voltmeter. For simplicity, the transfer function of a real 3-bit DAC is overlaid on the ideal response as depicted in Fig. 3.6. The DAC static performance metrics, including offset, gain, monotonicity, differential nonlinearity, and integral nonlinearity errors can be extracted from the transfer function plots.

Offset and Gain Errors

A typical DAC transfer function is depicted in Fig. 3.6a. The DAC's offset and gain errors can be extracted from the transfer function using various methods. These include dividing the full scale voltage range by the number of quantization levels, using the end-points to generate a linear fit, or employing the best fit line. Owing to its simplicity, the end-point fit is the most preferred method to measure the DAC's offset [2]. The straightforward method to determine the DAC offset is

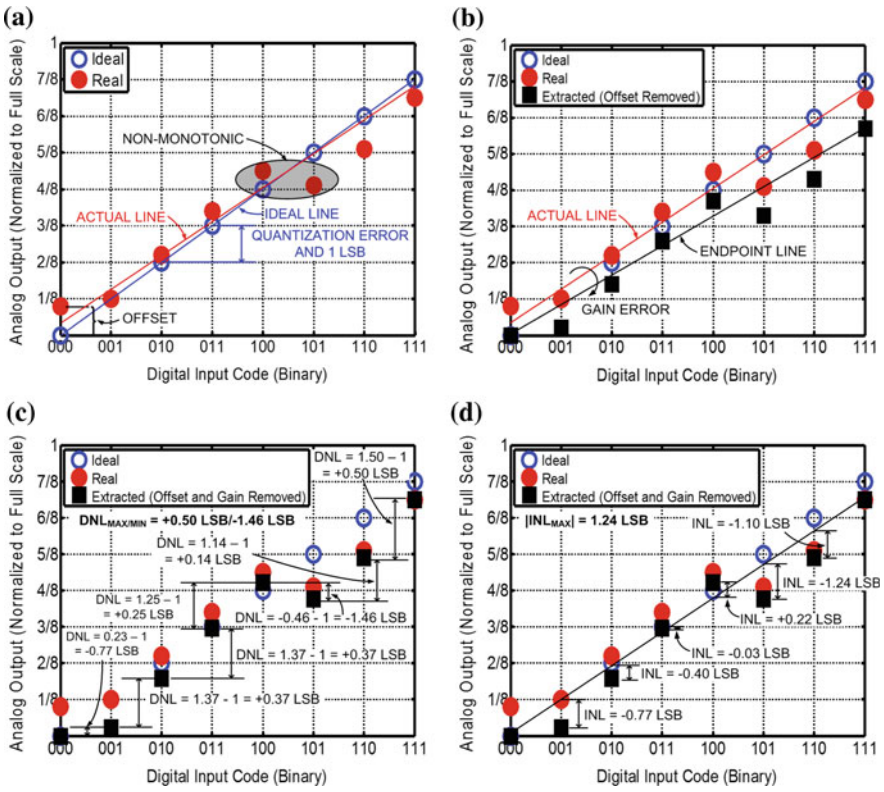

Fig. 3.6 3-bit DAC Transfer Function: **a** Offset error **b** Gain error **c** DNL **d** INL

by calculating the deviation between the real and ideal transfer functions when the binary input is all zeros. As illustrated in Fig. 3.6a, the y-intercept of the transfer function denotes the offset error. For target applications such as waveform synthesis, the DC offset can result in large carrier feedthrough, when upconverted in an RF transmitter.

After removal of the offset, the gain error is extracted from the deviation of the slope of the extracted transfer function versus the slope of the ideal transfer function ($y = x$), as depicted in Fig. 3.6b [5]. The gain error is seen as less critical, since it is often calibrated out by adjusting the input digital code. It is worth noting that both offset and gain errors need to be removed before extracting any further static metrics such as differential or integral nonlinearities.

Differential Nonlinearity (DNL)

The DNL error measures the step distance between the code i and the code $i - 1$ for the extracted DAC transfer function (after removal of offset and gain errors), as

illustrated in Fig. 3.6c. The difference is then compared to the ideal LSB value. The DNL for a given code i can be calculated as

$$DNL(i)[LSB] = (Code_i[LSB] - Code_{i-1}[LSB]) - 1[LSB] \tag{3.6}$$

For a given DAC, the minimum and maximum DNL values are typically reported to summarize its performance. The DAC is considered monotonic when the output signal is increasing (decreasing) as the digital input code is increasing (decreasing). Monotonicity is guaranteed if the minimum value of the DNL is greater than −1 LSB.

Integral Nonlinearity (INL)

The INL error can be quantified as the deviation of the extracted DAC transfer curve to the end-point line, as depicted in Fig. 3.6d. The INL for a code i can be calculated from the DNL as

$$INL(i)[LSB] = \sum_{k=0}^{i} DNL(k)[LSB] \tag{3.7}$$

In order to summarize the INL performance of a DAC, the absolute maximum INL is reported.

3.2.2.2 Dynamic Metrics

The DAC's dynamic performance can be inferred from its output spectrum. The most often used metrics to characterize the DAC's dynamic performance are SNR, SINAD, ENOB, SFDR, and IMD. It is worth mentioning that all of these metrics are typically measured within the desired Nyquist band of operation. A single-tone test is employed when measuring SNR, SINAD, ENOB, and SFDR, while a two-tone test is used to characterize IMD. A simulation of a 12-bit DAC with intrinsic nonlinearity is used to illustrate the extraction of these metrics in Fig. 3.7.

Signal-to-Noise Ratio (SNR)

Signal-to-noise ratio is defined as the ratio of the desired signal power (P_{Signal}) to the integrated noise power, excluding harmonics and DC offset. Typically, the specified noise power includes quantization noise (N_q), DNL error (N_{DNL}), thermal noise ($N_{Thermal}$), and random jitter (N_j).

$$SNR[dB] = 10\log_{10}\left(\frac{P_{Signal}}{N_q + N_{DNL} + N_{Thermal} + N_j}\right) \tag{3.8}$$

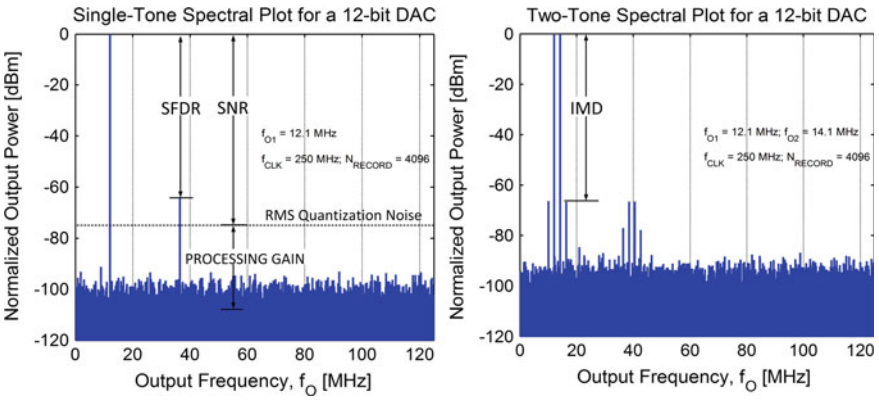

Fig. 3.7 Spectral plots for a 12-bit DAC

The SNR is generally specified over the entire Nyquist bandwidth. However, in some applications a narrow band filter is used following the DAC, and thus sets the integrated noise bandwidth well below its Nyquist. This process is otherwise known as oversampling and can effectively enhance the DAC resolution beyond its quantization limit.

Harmonic Distortion (HD*n*)

The *nth* order harmonic distortion is defined as the ratio between the power of the desired signal and the power of the *nth* harmonic, where $n = 1, 2, 3, \ldots$, and expressed as,

$$HDn[dBc] = 10\log_{10}\left(\frac{P_{Signal}}{n\text{th Harmonic Power}}\right) \tag{3.9}$$

When generating a single output tone at f_0, the *nth* harmonic component is observed at the $|nf_0 \pm kf_{CLK}|$ frequency where k is chosen to fold the harmonic term into the desired Nyquist zone.

Signal-to-Noise-and-Distortion Ratio (SINAD, SNDR)

Signal-to-noise-and-distortion ratio measures the ratio of the power of the desired signal to the power of the total noise, including harmonic distortion products ($P_{Distortion}$). The measurement does not include the DC component.

$$SINAD\,[dB] = 10\log_{10}\left(\frac{P_{Signal}}{N_q + N_{DNL} + N_{Thermal} + N_j + P_{Distortion}}\right) \tag{3.10}$$

Effective Number of Bits (ENOB)

ENOB is used to represent the effective resolution of the converter including all sources of noise and/or distortion. ENOB can be calculated from either SNR or SINAD and is represented as

$$ENOB[bits] = \frac{(SNR \text{ or } SINAD)[dB] - 1.76}{6.02} \tag{3.11}$$

Spurious-Free Dynamic Range (SFDR)

Spurious-free dynamic range measures the relative power of the desired signal to the power of the highest spur component generated within the targeted bandwidth.[2]

$$SFDR\,[dBc] = 10\log_{10}\left(\frac{P_{Signal}}{\text{Highest Spur Power}}\right) \tag{3.12}$$

This metric is considered the most critical in frequency synthesis since it determines the spectral purity of the output waveform with or without the presence of harmonics.

Intermodulation Distortion (IMD*n*)

In the presence of two or more input signals, inter-tone harmonic mixing can result in intermodulation distortion (IMD) products, located close to or further apart from the desired signals. IMD*n* measures the ratio of the power of the desired signals (P_{Signal}) to the power of the *nth*-order intermodulation product (P_{IMD_n}). The IMD products for two signals at f_{01} and f_{02} can span across $|(mf_{01} \pm nf_{02})|$ where $m, n = 1, 2, 3, \ldots$ Furthermore, the DAC sampling and aliasing processs can result in the folding of the IMD products to multiple Nyquist zones, which can be expressed as $|(mf_{01} \pm nf_{02}) \pm kf_{CLK}|$, where $m, n, k = 1, 2, 3, \ldots$

$$IMD_n = 10\log_{10}\left(\frac{P_{signal}}{P_{IMD_n}}\right) \tag{3.13}$$

In general, the third order intermodulation (IMD_3) is of most concern, as it generates the highest in-band distortion levels.

[2] Care should be taken in choosing appropriate FFT resolution bandwidth (or bin spacing) to set the minimum detectable power level.

3.2.3 INL Induced Distortion

Even though the previously described metrics can be divided into static and dynamic categories, their effects are not entirely independent. For instance, there is a clear link between the DAC static INL performance and its low frequency distortion behavior. A high INL error represents a large deviation of the DAC transfer curve from the straight line ($y = x$). This places an upper bound on SFDR at frequencies near DC [5, 6]. Thus, the INL can provide an estimate of the maximum SFDR before further degradation due to timing- and amplitude-related errors [2].

Continuing with the example of a 12-bit, 250 MS/s DAC, we estimate the prominent harmonic distortion order and its magnitude from the shape and maximum value of the INL, respectively. Expanding on the works of [5, 6], the INL is modeled using correlated second- and third-order polynomials. The second-order model is defined as $y = a \cdot x^2 + x - a$, while the third-order model is expressed as $y = a \cdot x^3 + (1 - a) \cdot x$. Here a is chosen such that the desired INL $_{MAX}$ is satisfied. Figure 3.8 illustrates the second- and third-order INL curves and their respective spectra with $\mathrm{INL}_{MAX} = 0.5$, 1, and 2 LSB. For each example of INL, the magnitudes of HD2 and HD3 are comparable. From the analyses of these two models, a heuristic approximation for SFDR as a function of the maximum INL can be expressed as

$$SFDR\,[dBc] \cong 20\log_{10}\left(\frac{2^N}{INL_{MAX}}\right) \tag{3.14}$$

3.3 DAC Implementation

The most straightforward implementation of the DAC involves an array of binary-weighted passive (capacitors and/or resistors) or active (current sources) components. Assuming a binary-weighted current-steering DAC with an LSB current of

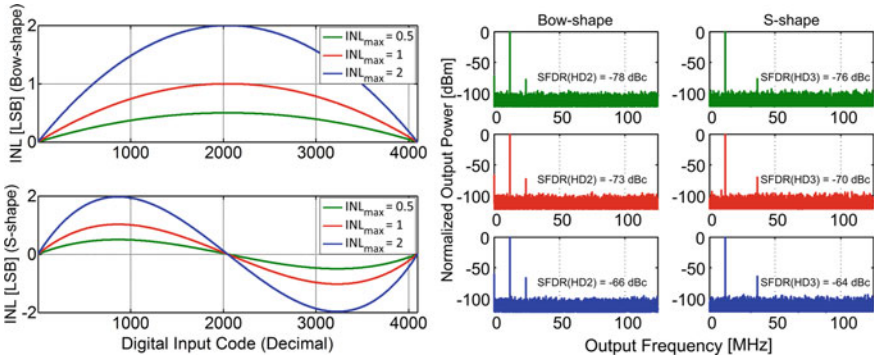

Fig. 3.8 Static INL shaping and the resulting effects on power spectral density

I_{LSB}, and denoting b_i[3] as the *ith* binary bit of a digital code, the output of the N-bit binary-weighted DAC can be expressed as

$$I_{OUT} = I_{LSB} \sum_{i=0}^{N-1} 2^i b_i \tag{3.15}$$

Alternatively, in a unary-weighted DAC, the current sources are all equal in magnitude; Thus, an N-bit DAC comprises $2^N - 1$ unary current sources. Denoting t_i[4] as the *ith* thermometer bit of a digital word, the effective analog current in response to the digital thermometer code is expressed as

$$I_{OUT} = I_{LSB} \sum_{i=0}^{2^N-2} t_i \tag{3.16}$$

Both the aforementioned architectures have their advantages and disadvantages. The binary architecture carries the benefit of using fewer control signals than the unary architecture. However, the accuracy requirements of the MSB cell versus LSB cell increases exponentially with the resolution. This results in the potential to exhibit code-transition glitches and loss of monotonic behavior. Such nonidealities can be mitigated by the unary architecture at the expense of increased chip area. A compromise between the two architectures is often made by segmenting the DAC, i.e. the MSBs and LSBs are represented by unary and binary structures, respectively. For an N-bit DAC segmented as $k : m$, such that the first k bits (MSBs) are realized as a unary structure, and the lower m bits are represented in binary, the effective analog output current is given by

$$I_{OUT} = I_{LSB} \sum_{i=0}^{m-1} 2^i b_i + 2^m I_{LSB} \sum_{i=0}^{2^k-2} t_i \tag{3.17}$$

For instance, a 12-bit DAC built using a 9-bit thermometer array and a 3-bit binary array is said to be 75 % segmented. The unary and binary architectures are two extreme cases of segmentation: a unary-weighted DAC is referred to as 100 % segmentation, while an all-binary DAC is 0 % segmented.

3.4 High-Speed DAC: Circuits and Limitations

A majority of high-speed DACs use the current-steering architecture, which offers faster switching and wider bandwidths compared to voltage- or charge-based DACs. This is primarily because the active devices are well known to switch faster in current

[3] b_i takes discrete values of 0 or 1 and referred in little-endian format.

[4] t_i takes discrete values of 0 or 1 and referred in little-endian format.

than in voltage mode. In addition, attempts to linearize the output buffer amplifier in voltage and/or charge mode DACs using feedback techniques, limit their speed of operation. This can be contrasted with using a simple load resistor in CS DACs.

An illustration of the current-steering architecture on a high-level abstraction is seen in Fig. 3.9. The DAC core comprises an array of binary and/or unary weighted current sources. The binary-switched current source array is scaled in units of $2^k \times I_{LSB}$, where $k = 0, 1, 2, \ldots, (N - m - 1)$ for an N-bit DAC segmented with m thermometer bits. The unary current source array, which is only used in thermometer or segmented DAC architectures, comprises 2^{m-1} current sources, each of magnitude $2^{N-m} \times I_{LSB}$. A corresponding array of current-commutating switch-pairs steer the direction of the current into one of the differential legs[5] of the DAC's output based on the input digital code. The switching pair cells can also be used to implement various hold operations at the output. Two load resistor cells, typically 50 Ω each, are used to convert the DAC's differential output current to a voltage signal (I–V). A binary-to-thermometer encoder maps the m most significant binary bits to a thermometer code that feeds the unary current source array.

In the context of high-speed DACs, it is important to examine the role and limitations of each component in the current-steering cell. A simple circuit schematic of a DAC current-steering cell is illustrated in Fig. 3.10. Transistors M_1 and M_2 constitute the current source and are normally biased from a current mirror reference cell. Transistor M_1 is the critical transistor that determines the magnitude of the cell current. However, the finite output impedance of M_1 affects the accuracy of the current source, thus forcing the need for the cascoding transistor, M_2. The source current is then steered to the positive or negative output leg in response to the input differential data signals, D and DB, by means of the commutating switch pair, ($M_{3,4}$). The size of the switch pair is scaled with the magnitude of the current to maintain the same source node voltages across all current cells. This is seen as critical to maintain an N-bit accuracy of the cascoded current source. If the output of the DAC is taken directly at the drain of the switching pair, there exists a data-to-output feedthrough resulting in high levels of switching glitches. In addition, the large aspect ratios and small channel lengths of the switching pair result in high load capacitance and low output impedance, respectively. Together, this limits the linearity performance of the DAC and hence cascode transistors ($M_{5,6}$) are added to isolate the output node from the gate of the switching pair.

3.5 DAC Design Space

In general, scaling of transistors, interconnect dimensions and power supply are quite favorable for digital designs. However, such trends are not entirely beneficial to analog and mixed-signal circuits. While the DAC in Fig. 3.9 exhibits a degree of repeatability that lends itself to an automated design methodology, the designer is

[5] Some books also use the term 'arm' as an equivalent to 'leg'.

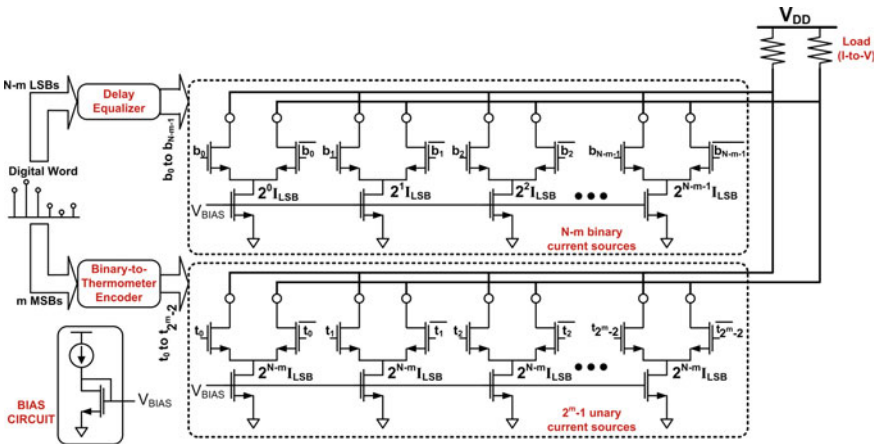

Fig. 3.9 A segmented current-steering DAC architecture

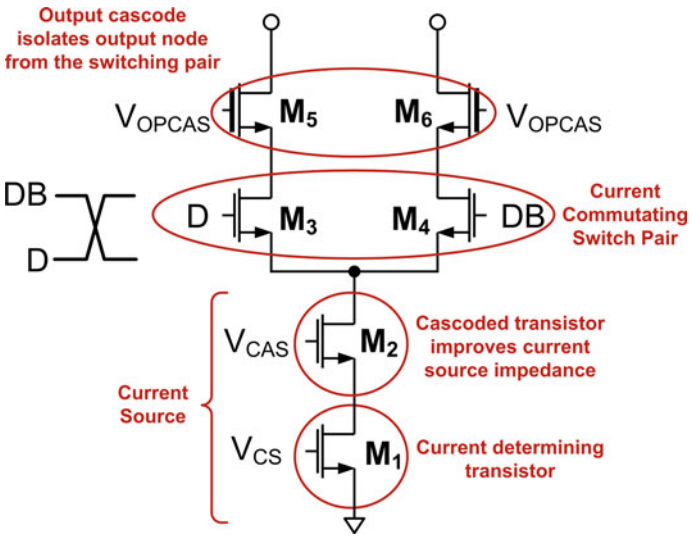

Fig. 3.10 A conventional current-steering cell

confronted with a complex design space and forced to resort to a custom design flow. To this end, a number of process-related limiting factors affect the performance of the DAC, resulting in failure to meet the target specifications. The design space of a DAC can be highlighted in terms of four major limitations: device noise, output impedance, signal swing and switching speed. A successful DAC design can only be achieved by carefully optimizing across this space to meet the desired specifications. The remainder of this section will address the DAC design space in detail.

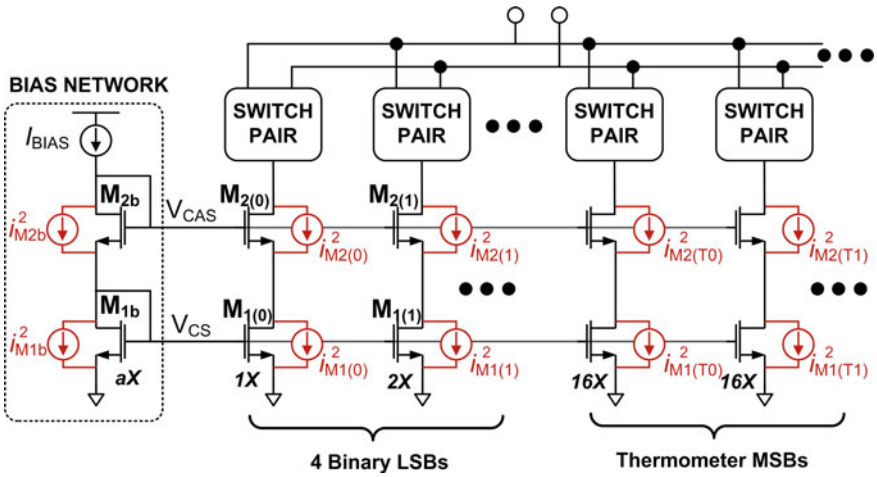

Fig. 3.11 Noise sources in a DAC

3.5.1 Device Noise

In addition to the intrinsic quantization noise, the DAC performance is also limited by the noise contribution from various circuit elements. The DAC's target resolution and desired bandwidth together set the maximum tolerable noise floor. In CS DACs, the current source array is the major contributor of noise. In addition to its own noise contribution, noise induced by the reference bias is further magnified by the current mirroring action, and thus can limit the overall noise performance. Figure 3.11 illustrates the noise sources in the DAC's circuit and its associated bias network. An $a:1$ current mirroring ratio is assumed between the bias network and the LSB cell. Accounting for the bias thermal noise contribution, the DAC's total output noise can be given by

$$i^2_{d_{DAC}} = 4kT\frac{\gamma}{\alpha}g_{m_{M1(0)}}2^N\left(1+\frac{2^N}{a}\right)\Delta f \tag{3.18}$$

where γ and α are process-defined noise parameters [7], and $g_{m_{M1(0)}}$ is the transconductance of the current source, $M_{1(0)}$.

Figure 3.12 illustrates the total output noise of a 12-bit DAC along with the isolated contribution of the bias network, as a function of a.[6] It is observed that the bias noise is the major contributor to the total output noise. In order to reduce the impact of the bias noise, a needs to be set much larger than 2^N. However, the power consumption specifications limit the maximum value of a that can be used; i.e. for a given LSB current and a bias mirroring ratio a, there is an expense of a times the LSB current in the bias cell.

[6] The flicker noise has been removed in the simulation.

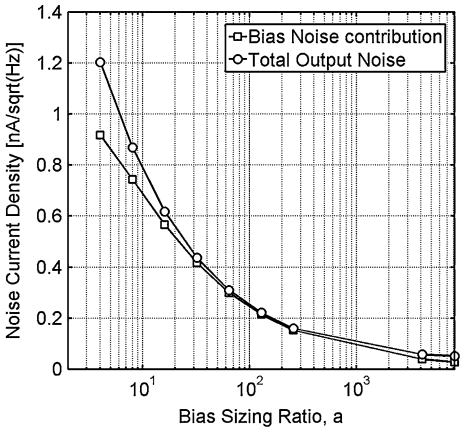

Fig. 3.12 Total output thermal noise contribution in a 12-bit DAC for various bias sizing ratios at 100 kHz

Assuming a full-scale output sinusoid current, the RMS signal power is given as

$$\text{Signal Power} = \frac{(2^N I_{LSB})^2}{2}. \tag{3.19}$$

Thus, the thermal SNR of the DAC over a bandwidth B is expressed as[7]

$$\text{Signal-to-Thermal Noise Ratio} \cong 10\log\left(\frac{aI_{LSB}(V_{GS}-V_T)\alpha}{16kT\gamma B}\right) \tag{3.20}$$

where $V_{GS} - V_T$ refers to the overdrive voltage of the current source transistor, $M_{1(0)}$. The SNR is seen to be independent of the resolution of the DAC, and can be only improved by increasing the bias mirroring ratio, a, or the LSB cell current, I_{LSB}. Let us consider a 12-bit DAC having an effective bandwidth of 100 MHz. The signal-to-quantization noise ratio is 74 dB. If the thermal noise floor is desired to be at least 10 dB below the quantization noise floor, and assuming $a = 1024$, $V_{GS} - V_T = 100$ mV and noise parameters, $\gamma = 2/3$, $\alpha = 1$, the minimum bound for LSB current is set to be 10.76 μA. The noise specification and bias sizing ratio together limit the minimum current (LSB cell) that may be used in the DAC. Along with the resolution and swing, it also sets the minimum achievable static current consumption of the DAC (including bias) in a given process technology.

3.5.2 Output Impedance

The static performance of the DAC is dependent on the intrinsic accuracy of the current sources.[8] Another element that can degrade the static performance is the

[7] The noise contribution of the DAC core is assumed negligible compared to the bias noise.

[8] This will be described later in detail in Sect. 3.6.

finite DC impedance ($Z_{CS}(0)$) of the current sources. In contrast, the DAC dynamic behavior is strongly dependent on the output impedance of the unit cell ($Z_{CELL}(s)$). In newer process technologies, channel length modulation effects are further exacerbated, significantly limiting the output impedance of a single transistor. Consider the current source in Fig. 3.13, comprised by transistors M_1 and M_2. Let us denote $Z_{DAC}(s)$ to be the effective parallel impedance looking into the DAC array. As the DAC cells are turned on, more cells are added in parallel, thus reducing $Z_{DAC}(s)$. Assuming an all-unary DAC, the total output impedance pertaining to the n^{th} code is given by

$$Z_{DAC}(s) = \frac{Z_{CELL}(s)}{n} \tag{3.21}$$

Accounting for the load impedance, R_L, the differential output voltage can be expressed as

$$V_{OUT} = I_{LSB} R_L \left\{ \frac{n}{1 + n\frac{R_L}{Z_{CELL}(s)}} - \frac{2^N - n - 1}{1 + (2^N - n)\frac{R_L}{Z_{CELL}(s)}} \right\} \tag{3.22}$$

It is observed that the total output impedance of the DAC changes as a function of the input code.

Such an effect is termed code-dependent impedance modulation, and is one of the fundamental factors limiting the distortion performance of a DAC. In the event of finite $|Z_{DAC}(0)|$, the static transfer characteristics become nonlinear. For instance,

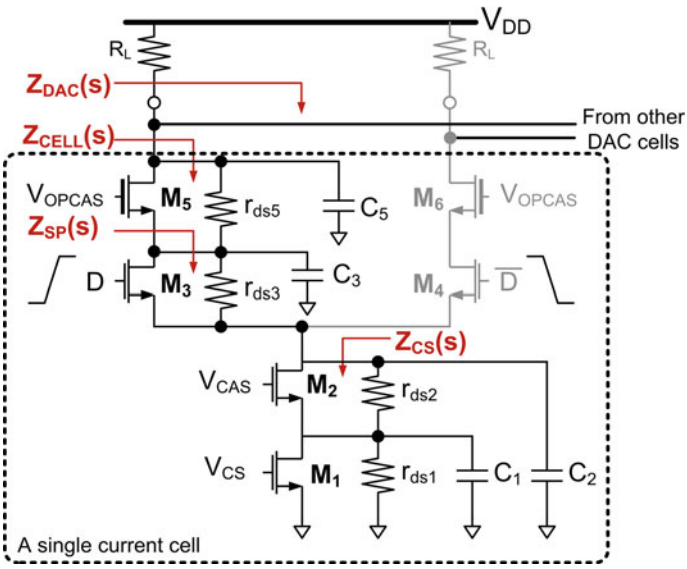

Fig. 3.13 Impedances in a current-steering cell

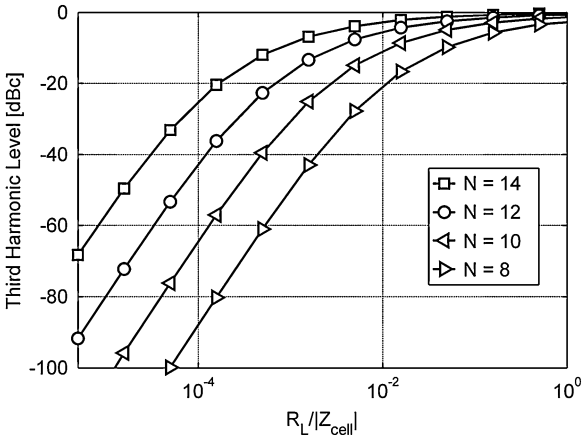

Fig. 3.14 HD_3 performance vs. $R_L/Z_{CELL}(0)$

a 12-bit DAC with an $R_L/|Z_{CELL}(0)|$ of 5×10^{-5} results in an HD_3 of about -53 dBc at low frequencies. This distortion level is significantly higher than the intrinsic performance of the DAC. Figure 3.14 illustrates the simulated HD_3 as a function of $R_L/|Z_{CELL}(0)|$ for various resolution. It is observed that the increase in resolution places more stringent requirement on the cell impedance.

The current cell in Fig. 3.13 is analyzed using a simple equivalent circuit model. In the absence of cascode transistors $M_{2,5}$, the DAC impedance is approximately $g_{m3}r_{ds3}r_{ds1}$. Since the switch pair is typically implemented using a minimum length device, it does not offer enough cascoding gain ($g_{m3}r_{ds3} \leq 1$). Thus, the DAC output impedance becomes a direct function of the current source impedance, r_{ds1}, and an R_L/r_{ds1} is not sufficiently small to attain high linearity. Cascoding is hence widely adopted to improve the impedance characteristics of the current source at the cost of voltage headroom. The cascoded current source impedance ($Z_{CS}(s)$) is analytically expressed as

$$Z_{CS}(s) = \frac{r_{ds1} + r_{ds2} + g_{m2}r_{ds1}r_{ds2} + sC_1r_{ds1}r_{ds2}}{1 + s(C_1 + C_2)r_{ds1} + sC_2r_{ds2} + sC_2g_{m2}r_{ds1}r_{ds2} + s^2C_1C_2r_{ds1}r_{ds2}} \quad (3.23)$$

Figure 3.15a illustrates the improvement in the low frequency impedance of the current source as a function of the ratio of the lengths of the transistors M_2 to M_1, for various M_1 lengths. While it is observed that a large channel length for M_1 improves the output impedance, increasing the length of M_2 further enhances the cascoding effect, resulting in output impedances in the order of tens of megohms. This significantly aids in achieving high static accuracy for the current sources. However, increasing transistor length, while maintaining its aspect ratio, increases the drain capacitances (C_1 and C_2), thus degrading the impedance at high frequencies (hundreds of megahertz). Figure 3.15b illustrates the impact of increasing channel lengths for M_1 and M_2 on the high frequency output impedance. Such opposing effects of increasing channel lengths call for an optimal choice to be

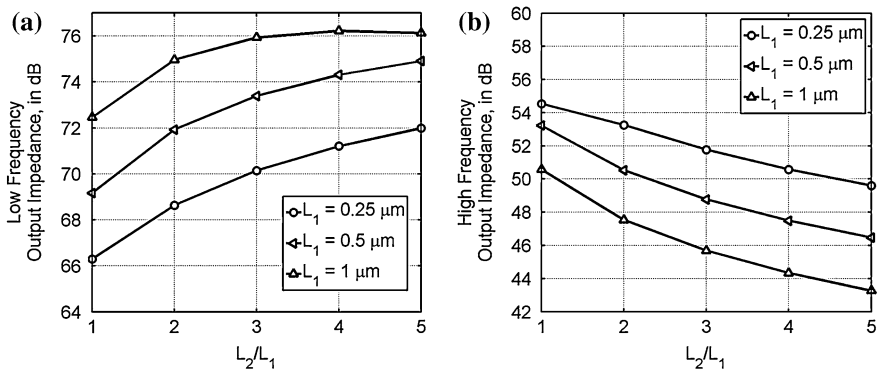

Fig. 3.15 **a** Low-frequency (DC) output impedance ($Z_{CS}(0)$) **b** High-frequency output impedance (Z_{CS} at 100 MHz)

considered when sizing the current sources, in order to strike a balance between the high and low frequency distortion limits.

On moving up the DAC cell, the impedance looking at the drain of the switch pair (when it is ON) is given by

$$Z_{SP}(s) = \frac{Z_{CS}(s) + r_{ds3} + g_{m3}r_{ds3}Z_{CS}(s)}{1 + sC_3(r_{ds3} + Z_{CS}(s)) + sC_3g_{m3}r_{ds3}Z_{CS}(s)} \tag{3.24}$$

In the case of high-resolution DACs (i.e. 10 bits or higher), Z_{SP} is not large enough to mitigate the effect of code-dependent impedance modulation. Furthermore, the large gate-drain capacitances of the switch pair ($M_{3,4}$) result in significant data feedthrough to the output. Cascoding transistors M_5 and M_6 are thus added to improve the isolation between the output node and the gates of the switching pair. Accounting for the cascode pair, the impedance of a single DAC cell (from Fig. 3.13) can be written as

$$Z_{CELL}(s) = \frac{Z_{SP}(s) + r_{ds5} + g_{m5}r_{ds5}Z_{SP}(s)}{1 + sC_5(r_{ds5} + Z_{SP}(s)) + sC_5g_{m5}r_{ds5}Z_{SP}(s)} \tag{3.25}$$

At low frequencies, the cell impedance reduces to

$$Z_{CELL}(0) \cong g_{m2}g_{m3}g_{m5}r_{ds1}r_{ds2}r_{ds3}r_{ds5} \tag{3.26}$$

As transistor feature size decreases, the channel length modulation parameter shows a dependence on the bias current and the overdrive voltage of the transistor [8]. Figure 3.16 illustrates the simulated impedance of the ON leg of a DAC cell, as a function of current in a 90 nm process technology. The use of cascoding in both the current source and the current cell, indeed improves the static linearity of the DAC by further increasing the output impedance $Z_{CELL}(0)$ enough to guarantee at least 12-bits of accuracy. The kink in impedance at low currents is attributed to the dependence of the switch pair channel length modulation parameter on the

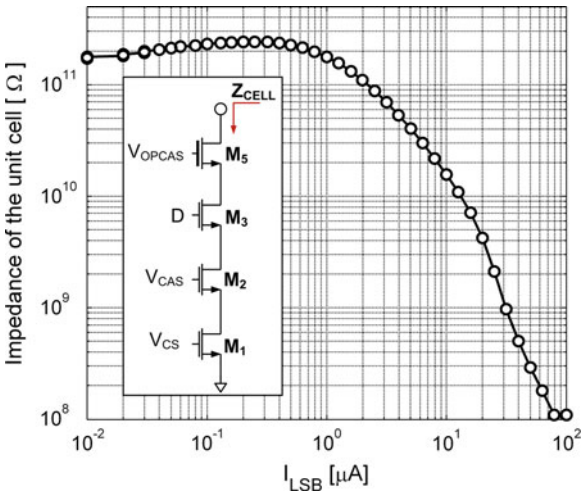

Fig. 3.16 DAC cell impedance as a function of current

overdrive voltage [8]. The distortion metrics in Fig. 3.14 can be used to determine the desired cell impedance. Figure 3.16 can then be used to estimate the maximum LSB current for the given process.

While the size of the current source array is made large enough to tolerate mismatches, it results in increased capacitance at the source node (V_S) of the switching pair. The manifestation of this capacitive effect is better understood in the temporal domain. As illustrated in Fig. 3.17a, when the data switch, transistors M_3 and M_4 shift from cut-off to saturation region or vice versa. However, this transition is not instantaneous; there exists a finite amount of time for which both switches are simultaneously on. During this period, the current source is immediately choked and the node V_S is discharged [9]. Once the switch pair's operating region transition is complete, the current source is forced to recharge the node, V_S, instead of delivering the desired current to the output node. The presence of a large capacitance at this node increases the recharge time constant. This effect is manifested as an output glitch that is proportional to the weight of the unit current source. In a DAC with multiple weighted cells connected together, the weighted glitches propagate to the output, creating a code-dependent glitch pulse.

Another form of dynamic distortion arises from the large aspect ratios of M_1 and M_2, thus resulting in large gate-drain capacitances and gate-source capacitances. These capacitances aid in the propagation of the switching behavior at node V_S to their bias nodes, as depicted in Fig. 3.17b. The bias fluctuation influences the magnitude of the i^{th} current source by a weighted coefficient α_i, such that

$$I_{OUT}(t) = I_{LSB}.\sum_{i=0}^{N-1} 2^i \alpha_i b_i(t) \tag{3.27}$$

It is seen that α_i is a function of several parameters such as the size of the switch pair, the impedance of the current source and the magnitude of the current. This

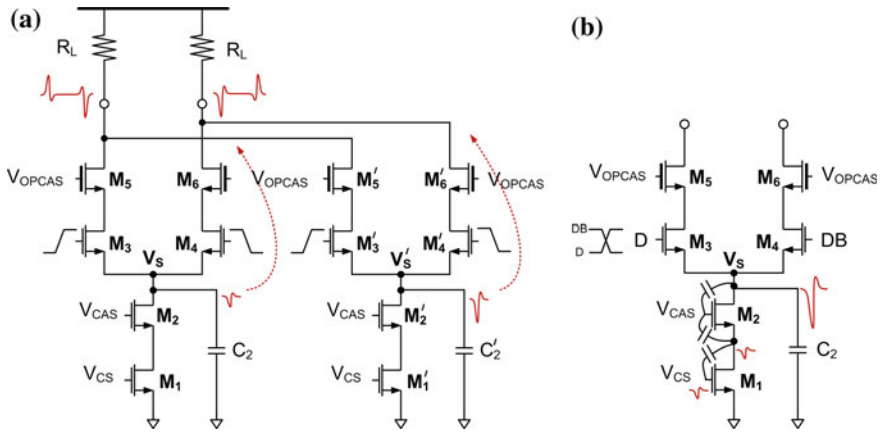

Fig. 3.17 **a** Glitch propagation from the source node **b** Modulation of the current source

distortion effect is mitigated by decreasing C_2 and increasing the gate capacitance of the bias nodes.[9] Care should also be taken to minimize the layout-induced coupling capacitance between the gates of M_1 and M_2.

3.5.3 Signal Swing

The noise-specified LSB current and resolution together determine the total current in a DAC. This eventually sets the output swing for a given load resistor, R_L. Thus, for a given supply voltage and transistor bias points, the maximum permissible signal swing is set to ensure all transistors are kept in saturation. For high-resolution DACs, the output swing can be allowed to increase further by raising the supply voltage. However, breakdown limits often mandate the use of thick-gate cascode devices ($M_{5,6}$) on top of the switch pair. Figure 3.18 illustrates the output swing as a function of the LSB current for different DAC resolutions. From the upper bound on swing limitations, the maximum permissible LSB current can be deduced.

Another major concern with large output swings is the linearity degradation due to the voltage-dependent drain capacitances of transistors, $M_{5,6}$. When the data switches, the output capacitance of the cell carrying no current is approximately given by

$$C_{OFF} = C_{gd5,6(OFF)}(V) + C_{db5,6}(V) \tag{3.28}$$

where C_{gd} is the gate-drain overlap capacitance of $M_{5,6}$, and C_{db} is the drain-bulk capacitance. In the ON state, the output capacitance of the cell is roughly given by

[9] Gate capacitances need to be relatively larger than the gate-drain capacitances of M_1 and M_2.

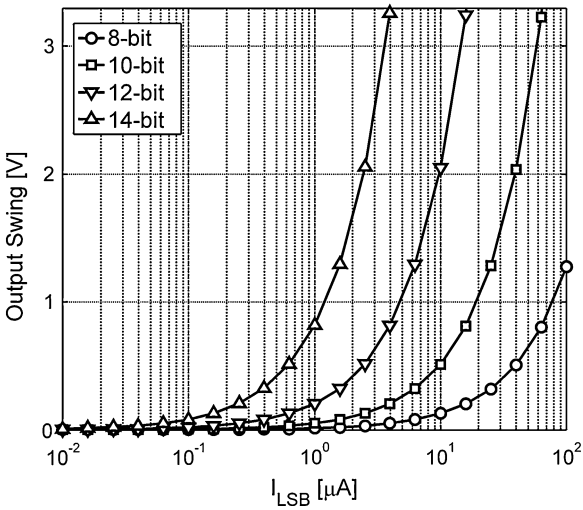

Fig. 3.18 Output swing as a function of the LSB current for various DAC resolution

$$C_{ON} = C_{gd5,6(ON)}(V) + C_{db5,6}(V) \tag{3.29}$$

This difference in capacitances in the ON and OFF states, along with the fact that the capacitance is voltage (output signal) dependent, results in code and amplitude dependent delays in addition to output load modulation. Together with the intrinsic transistor capacitances, interconnects can further increase the drain capacitance to the substrate. This eventually limits the maximum operation speed of the DAC cell. The parasitic capacitance at the output node can be minimized by layout techniques, while the mismatch in the ON and OFF impedances is alleviated by the use of leaker currents, as proposed in [10], and illustrated in Fig. 3.19b. It is shown that a leaker current of 1–2 % of the cell current is sufficient to maintain a fairly constant ON and OFF impedance [10]. While the use of leaker currents does not affect the differential output swing, it changes the single-ended swing by a constant DC value of $I_{LEAK} \times R_L$. The total sum of all leaker currents must be taken into consideration when estimating the lower bound of the single-ended swing, thus guaranteeing $M_{5,6}$ are maintained in saturation.

3.5.4 Switching Speed

The near-Nyquist performance of the DAC is highly dependent on the finite switching[10] and settling characteristics of the current cell [4]. This in turn is seen to be limited by the device transit or cut-off frequency (f_T) for a given process

[10] Switching refers to the action of turning a transistor from cut-off to saturation or vice versa.

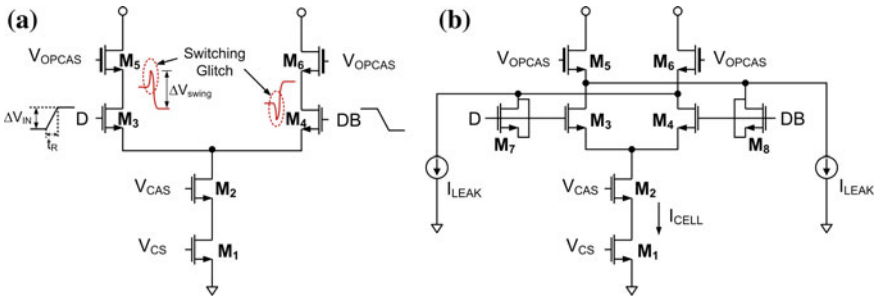

Fig. 3.19 **a** High-speed switching glitches in a DAC cell **b** A modified DAC cell with leakers and neutralization switches

technology. Figure 3.20 illustrates a hyperbolic increase in intrinsic device speeds with the reduction in gate lengths, as outlined by the International Technology Roadmap for Semiconductors (ITRS) [11]. In current-mode circuits, a general rule of thumb is to restrict the transistor switching speed to lesser than $f_T/20$ [12].

While the initial charging time of the DAC's output node is a function of the transistor's switching characteristics, the finite settling behavior is a function of the load capacitance and the rise time of the input signal. Figure 3.19a illustrates the switching action in a current-steering cell, where the data signal at the gate switches between voltage levels separated by ΔV_{IN}, with a rise time t_R. Using the model described in [12], the delay of a current-mode switching cell can be expressed as

$$\text{Delay} = k_{RC}\frac{\Delta V_{swing}}{\Delta I_{swing}}(C_{load}) + t_R \min\left(\frac{V_{OD}}{\Delta V_{IN}}, 1\right) \tag{3.30}$$

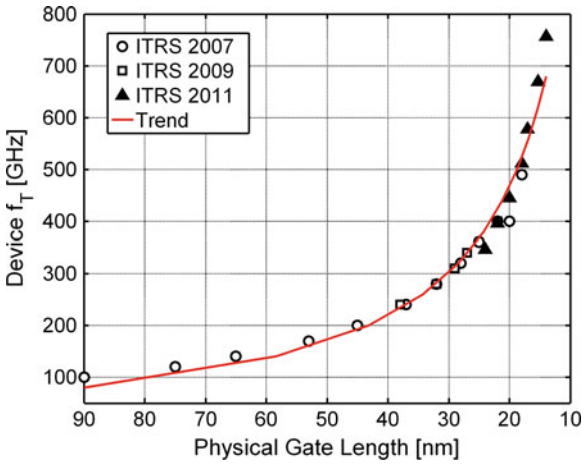

Fig. 3.20 Trend in device f_T as a function of feature length

where,

k_{RC}	$\cong \ln 2$
DV_{swing}	Single-ended voltage swing at drain of the switch pair
DI_{swing}	Single-ended current swing = Cell current
t_R	Rise time of the input at the gate of the switch pair
V_{OD}	Overdrive voltage of the differential switch pair = $V_{GS} - V_T$
DV_{IN}	Input voltage swing

All the above parameters are inter-dependent. For instance, a large value of ΔV_{IN} improves the switching characteristics. However, it also increases the swing at the drain of the switch pair (ΔV_{swing}), resulting in an increase in the overall cell delay. In addition, the maximum achievable ΔV_{IN} within a rise time t_R is limited by the process node. Further, the high-speed transition at the gate node increases the instantaneous voltage swing on the switch-pair drains, as well as those of the cascode transistors ($M_{5,6}$). In order to reduce these glitches, neutralization switches ($M_{7,8}$), are used to feed an equal and opposite glitch, as depicted in Figure 3.19b. Keeping the load capacitance fixed, the total LSB switch pair delay (sum of the switching and settling time to achieve 95 % of the final value of current) is simulated in a 90 nm CMOS process and the results are plotted with respect to the cell current in Fig. 3.21. The curves correspond to different widths for the switching pair, while the lengths are kept at minimum. From the settling time specifications, the minimum LSB current can be determined.

Hence, the DAC target speed and given process f_T together set the LSB current and the physical size of the switching pair. It is important to note that the current cells operating at lowest and highest current magnitudes have to switch simultaneously at the desired operation speed. In the event of mismatch in switching

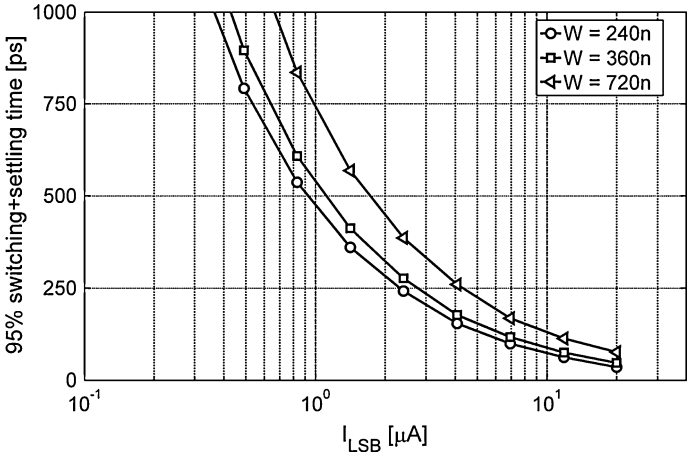

Fig. 3.21 Dependence of the transistor switching time on current

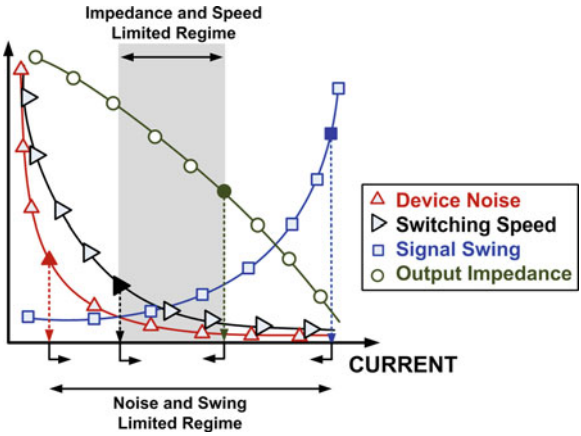

Fig. 3.22 The DAC design space: device noise, output impedance, signal swing and switching speed

speeds between the MSB and LSB cells, segmentation of the DAC is adopted as discussed in a later section.

Figure 3.22 illustrates the general trend for the four limiting parameters in the DAC design space as a function of the LSB current. The signal swing and output impedance together set the upper bound on LSB current, while the device noise and switching speed determine the smallest permissible current.

3.6 Segmenting the DAC

Although segmentation has been accepted as a mainstream option, it remains to be argued as to what is the optimal choice of the ratio of unary MSBs to binary LSBs. The limited accuracy of the MSB current sources can result in high levels of nonlinearity (i.e. large INL), which in turn degrades the dynamic performance of the DAC.

Based on the previous section, the DAC LSB current (I_{LSB}) is chosen to optimize across the design space. Subsequently, the transistor size and overdrive voltage need to be determined. An overdrive voltage of at least 100 mV is typically needed to ensure that the transistor is operating in strong inversion, and also to guarantee sufficient matching between the reference bias cell and the current mirrors. As discussed in Sect. 3.5.2, the lengths of the current source transistor and its cascoding device are set based on the output impedance requirement. As a result, the width of the transistor can be calculated. Another critical element that dictates the minimum size of the transistor is the mismatch accuracy between the MSB and LSB cell. One of the primary contributors to the variation between the two cells is the threshold voltage (V_T) mismatch between transistors [13, 14]. Let us denote the LSB and MSB currents for an N-bit DAC as

$$I_{LSB} = \frac{1}{2}\mu C_{OX}\frac{W}{L}V_{OD}^2 \tag{3.31}$$

$$I_{MSB} = \left(2^{N-1}\right)\frac{1}{2}\mu C_{OX}\frac{W}{L}\left(V_{OD} - \Delta V_T\right)^2 \tag{3.32}$$

where μ is the channel mobility, C_{OX} is the specific oxide capacitance, W (L) is the width (length) of the LSB current source transistor, V_{OD} is the overdrive voltage, and ΔV_T is the maximum mismatch error between MSB and LSB cell. In a binary cell array, the error in the MSB current source must be kept well below 0.5 LSB in order to maintain full accuracy. From (31) and (32), the maximum tolerable mismatch error can be expressed as

$$\Delta V_T \leq V_{OD}\left\{1 \pm \sqrt{1 + \frac{1}{2^N}}\right\} \tag{3.33}$$

This equation suggests that a lower V_{OD} implies a lower maximum tolerable V_T mismatch. On the other hand, if the V_T mismatch were to be fixed, we need a large V_{OD} to circumvent the mismatches. Figure 3.23a illustrates that the maximum tolerable V_T mismatch reduces as a function of the DAC resolution, further highlighting the issue of designing high-resolution DACs for a given mismatch constraint.

Reference [13] models the V_T mismatch as a Gaussian distribution with standard deviation, $\sigma_{V_T} = A_{V_{T0}}/\sqrt{WL}$, where $A_{V_{T0}}$ is a technology-dependent parameter. Therefore, in order to reduce the magnitude of σ_{V_T}, the transistor area can be increased, while maintaining a constant aspect ratio. Figure 3.23b illustrates the impact of increasing the area of the transistor on σ_{V_T} for a 90 nm CMOS process, assuming a V_{OD} of 100 mV. It is seen that even for a moderate resolution DAC, a large transistor area is required to minimize the impact of V_T mismatch. However, this results in reducing the high-frequency impedance of the current source, thus

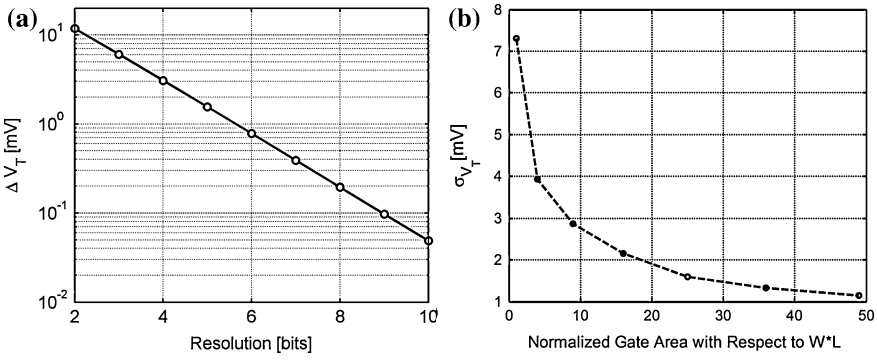

Fig. 3.23 **a** ΔV_T vs. Resolution **b** σ_{V_T} mismatch as a function of gate area

limiting the dynamic linearity of the DAC. The high-speed DAC designer is thus confronted with the challenge of meeting both static and dynamic linearity requirements. In the case of high-resolution DACs, the mismatch requirements dictate a transistor area that is prohibitive. In order to relax the transistor area requirements, a segmentation topology is adopted, such that the mismatch constraint is applied to the highest binary cell [14, 15].

Apart from mismatch-induced errors between the DAC current cells, fundamental circuit-level challenges, such as parasitic capacitance and finite output impedance, also influence the degree of segmentation. The degradation in transistor output impedance as channel lengths decrease, limits the maximum current that can flow through a transistor. As a result, a 100 % segmentation (unary DAC) is favored for low impedance processes. On the other hand, $2^N - 1$ current cells in a unary DAC increase the effective parasitic capacitance at the output node, thus limiting the speed of operation. In such cases, a 0 % segmentation (all-binary DAC) is preferred. In addition, a high degree of segmentation results in a significant increase in the DAC area (digital logic, analog current cells and interconnects) that makes timing compliance a challenge at the desired speed of operation. Thus, an optimal choice of segmentation needs to be made, with both process technology and circuit topology in mind.

3.6.1 The Segmentation Bound

As discussed in Sect. 3.5.2, a high current source impedance determines the accuracy of the current sources, while a high DAC cell impedance mitigates the effect of code-dependent impedance distortion. The extent by which a high Z_{CS} is required, is determined by the voltage fluctuation at the switch pair source node, V_S, relative to the LSB. It is noted that the resolution accuracy of the current sources needs to be maintained over the desired synthesis bandwidth of the DAC; i.e. both DC and AC impedances need to be sufficiently large. The LSB cell is designed to have the highest possible Z_{CS} over the synthesis bandwidth. As the current source is scaled in powers of two, Z_{CS} halves. The largest current cell that guarantees the desired output impedance across the synthesis bandwidth is where segmentation begins; all subsequent current cells are unary-weighted. This is the lower bound on segmentation (refers to the maximum number of binary cells that can be used) for the DAC. On the other hand, the desired DAC output bandwidth (computed from the load resistor and effective capacitance of all current cells) sets the upper bound on segmentation (or the maximum number of thermometer cells that can be used in the DAC).

Given the bounds on segmentation, the lower limit is more preferred as it implies fewer number of cells to be connected together, resulting in short routing lengths. This significantly aids in the reduction of the overall routing capacitance, thus decreasing clock skew mismatch and improving output bandwidth.

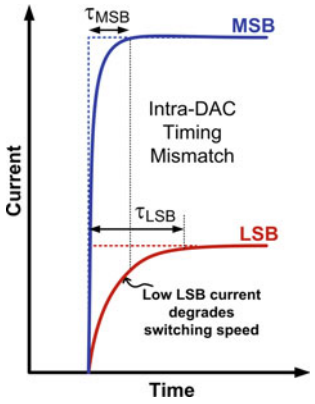

Fig. 3.24 Switching time mismatch in a DAC

Furthermore, a smaller effective area for the current sources helps decrease the ground line IR drop, which improves the cell matching. However, as discussed earlier, low segmentation designs demand high accuracy current sources. In addition, there exists a large ratio between the current magnitudes in the LSB and MSB cells in high-resolution DACs. As a result, the response times of the LSB cell (τ_{LSB}) and the MSB cell (τ_{MSB}) differ significantly, resulting in a mismatch in switching time instants, as depicted in Fig. 3.24. Such cell-by-cell timing mismatches result in the formation of output glitches that can limit the speed of operation, especially in gigahertz DACs. This creates a designer's paradox called the resolution-bandwidth trade-off. A high degree of segmentation helps alleviate this problem at the penalty of increased area and capacitance. A large chip area further results in being sensitive to process variations and IR drops that affect the matching between the current cells, both temporally and spatially. Consequently, the resolution-bandwidth trade-off comes into play.

3.7 Architectural Trends in High-performance DACs

Although BiCMOS technology is the predominant choice for high-speed operation, the issues of power, area and cost of integration have led to the wide adoption of CMOS-only processes. To this end, numerous circuit and architectural innovations have been proposed to improve the synthesis bandwidth and the linearity performance of DACs, enabling CMOS designs to compete with their BiCMOS counterparts [16]. In modern DACs, a high static linearity is obtained by using special layout techniques, trimming, calibration, dynamic element matching, etc. [15, 17, 18]. However, the dynamic performance of CS DACs have been known to fall rapidly with increase in signal frequency and clock rate. This section aims to introduce the reader to some of the well known techniques targeting high linearity and wideband signal synthesis.

3.7.1 Linearity Improvement

The dynamic performance of DACs is influenced by a number of factors such as MSB glitches, cell-to-cell timing skew, mismatch in settling time constants, low current source impedance and process gradients [14, 15]. Segmentation was one of the earliest approaches to enable high-speed DACs with good dynamic linearity [14]. It is seen as critical in minimizing the MSB glitch that occurs from delays in the switching cells. The DAC is split into segments that are thermometer or binary in nature. Each of these segments or sub-DACs can be implemented using either current dividers or resistor strings. The approach of combining current-steering with resistor strings was demonstrated in [19, 20]. In such architectures, great care needs to be taken to match time constants between the segments. Segmentation also increases the number of control signals to the DAC, resulting in timing skew between the cells. Such synchronization problems can occur from spatial variations or mismatch in the switch drivers. Spatial filtering using interleaving and inter-digitating help compensate for process gradients and thermal variations, thus improving intrinsic device matching accuracy [13]. Additionally, IR drops on the supply lines are reduced by the use of wide and thick metal lines, while power bus variations from the pins to the desired destination are addressed by binary tree structures [10]. However, these techniques often introduce large parasitic elements that limit the DAC's speed of operation.

The switching of currents also results in code-dependent glitches at the output. Deglitching or track-and-hold (T/H) circuits are typically employed to compensate for dynamic distortion, arising from switching and settling issues. A track-and-hold circuit holds the output constant while new data switches the DAC, and tracks only once the DAC output has settled. This enables the DAC glitches and settling errors to be mitigated. Nevertheless, the T/H circuit introduces its own errors such as pedestal, droop, clock feed-through and stepping errors [17]. When the input data switch, it must be guaranteed that the switch pair is not simultaneously off. Therefore, overlapping differential signals are used such that the transition point is optimal to achieve the best SFDR performance [21]. Ordinary switching also results in distortion from uneven pulse durations that is mitigated by the use of return-to-zero (RZ) switching. RZ implementation is realized by the use of an output switch, M_{RZ}, that shorts the output nodes together during the return phase, as illustrated in Fig. 3.25a. However at high frequencies, if RZ switching between two levels does not occur within the pulse duration, a memory effect of the input stream is manifested at the output in the form of code-dependent noise. Differential quad switching (DQS) Fig. 3.25b was proposed in [22] to mitigate the problems of RZ switching. In DQS, four logical signals obtained from the AND operation of data, clock and their inverted versions, result in a switching action at every clock edge; In other words, both high and low data signals are represented by rising and falling pulses, as illustrated in the timing diagrams in Fig. 3.25b. It is observed that the DQS scheme is equivalent to two RZ schemes operating in a staggered fashion. Owing to twice the frequency of switching than a single RZ, the code-dependent switching noise is pushed far out of the signal band. However, a major drawback

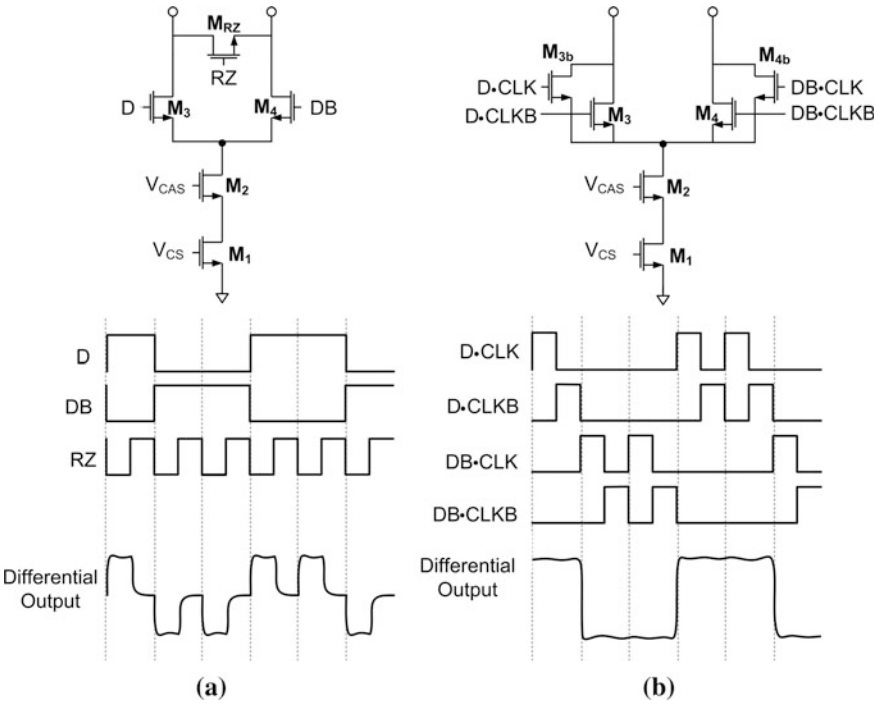

Fig. 3.25 DAC cell with **a** Return-to-zero (RZ) Switching **b** Differential-Quad Switching (DQS)

of the DQS approach is the increased dynamic power consumption due to the increased amounts of switching.

3.7.2 Towards RF Synthesis

One of the earliest and simplest techniques to synthesize high frequency signals beyond the Nyquist relied on band-pass filtering the inherent image replica components [23]. However, the availability of high selectivity filters limited the highest synthesizable frequency. In addition, the amplitude of the replicas in the higher Nyquist zones is reduced due to the *sinc* attenuation, translating to very low output power, or the need for a linear post-DAC amplification. RZ DACs [24] were used to push the *sinc* nulls to higher frequencies, thus improving the amplitude of the high-frequency spectral copies. The switch M_{RZ} in Fig. 3.25a is made large to have minimum ON resistance, so that the differential outputs are matched during the return phase. However, a large RZ switch loads the output node, thus degrading the bandwidth of operation. Furthermore, the rise-time requirements at the output node have to be met with respect to the effective ON

period of the DAC pulse, making the design of RZ DACs more cumbersome compared to their NRZ counterparts.

Another technique recently proposed to extend the DAC operation to near and beyond the Nyquist limit, is the use of partial-order hold (POH) [25]. A partial-order hold circuit, depicted in Fig. 3.26a, integrates the DAC output to realize a trapezoidal hold waveform. Such a technique has proven to result in image replica suppression over 40 dBc in the fourth Nyquist zone. However, this DAC architecture suffers from being sensitive to the POH period and demands tight control over the output signal rise time. In addition, both clock and POH jitter requirements are to be met, resulting in stringent specifications. Another means for high-frequency synthesis is the interpolation DAC, as depicted in Fig. 3.26b. It uses low sample-rate data that is fed into a digital interpolation filter and eventually fed to the DAC core. Although the speed of the digital interface is relaxed, the DAC still operates at the update rate. In the case of analog interpolation, the desired waveforms are created using microstepping methods [26] or using RZ DACs that perform an analog equivalent of zero padding.

Recent attempts have been made to combine the DAC and mixer functionality into a single topology to improve the overall system linearity and power consumption [27]. This DAC/mixer construct, referred to as the RF-DAC, is depicted in Fig. 3.27a. A mixer switch pair is placed on top of the DAC cells such that the local oscillator (LO) directly modulates the DAC output. The current-to-voltage-to-current conversion between the DAC and mixer, which introduces distortion, is completely removed. However, the up-conversion of the image-replicas and harmonic mixing of these replicas with the local oscillator, makes the post-RF-DAC filtering a daunting task. Reference [28] employs harmonic rejection mixers embedded into the DAC to suppress the harmonics caused by the mixer circuit, thus achieving greater than 70 dB of harmonic rejection. However, this technique greatly relies on the matching between the transistors that comprise the harmonic mixer.

Noise-shaping ($\Delta\Sigma$) techniques have also been employed in RF-DACs to improve the spectral quality of the DAC signal to obtain higher in-band SNRs, at the expense of large out-of-band noise [29, 30]. The delta-sigma modulator uses

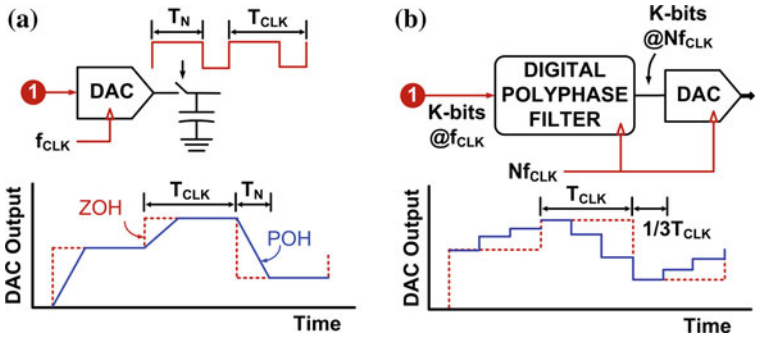

Fig. 3.26 DAC cell with **a** Partial-order hold DAC **b** Interpolation DAC

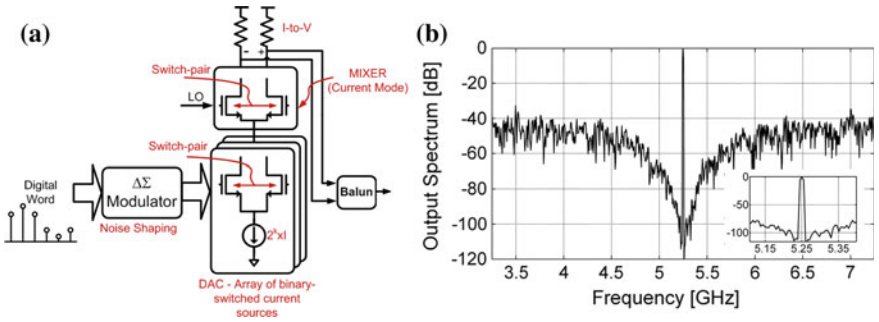

Fig. 3.27 **a** A simplified block diagram of a radio-frequency digital-to-analog converter **b** RF Output Spectrum using 2*nd*-order, 3-bit ΔΣ Modulator

lesser number of bits at the cost of increased rates of operation. However, the improvement in switching capabilities as processes emerge make this solution feasible to synthesize digital signals up to gigahertz frequencies. Figure 3.27b illustrates the output spectrum of a 5.25 GHz RF-DAC fed using a 2*nd*-order, 3-bit ΔΣ modulator. It is observed that the ΔΣ noise starts rising rapidly beyond the bandwidth of the modulator, eventually violating spurious emission requirements. The increased out-of-band ΔΣ noise and the spurious emission specifications together place stringent filtering requirements after the mixer, hence limiting the instantaneous bandwidth of operation to well below 100 MHz. Reference [29] integrates a high-Q passive LC bandpass filter to perform filtering of the out-of-band spurious and noise. However, the feasibility of this approach depends on the filtering requirements and the limited Q of on-chip passives. Alternatively, Reference [30] embeds a semi-digital FIR reconstruction filter in the digital-RF interface. The limitation of this approach lies in the need for large number of taps to obtain sufficient attenuation. Recently, a highly digital RF-DAC based transmitter exhibiting high linearity was proposed in Ref. [31]. The work demonstrated multi-band operation in 3G using a polar architecture, in which separate phase and amplitude paths were derived from the baseband digital signal. The phase signal modulates a digital controlled oscillator (DCO) and later acts as the LO signal, while a 14-bit amplitude signal is oversampled and then applied to the DAC current cells. A high dynamic range DAC with no noise-shaping was designed to relax the filtering requirements and obtain −160 dBc/Hz far-off noise specifications. However, there still exists the issue of image replicas in this structure, which limits the ability to further extend the bandwidth of operation.

The concept of a power DAC/mixer has recently evolved as an extension of the RF-DAC into the PA domain. In such a construct, as illustrated in Fig. 3.28, a high power transistor is used as the switching device and thus accomplishes both current commutating and current combining. However, the inherent trade-off between speed (f_T) and power capability (breakdown voltage) for semiconductor devices creates a

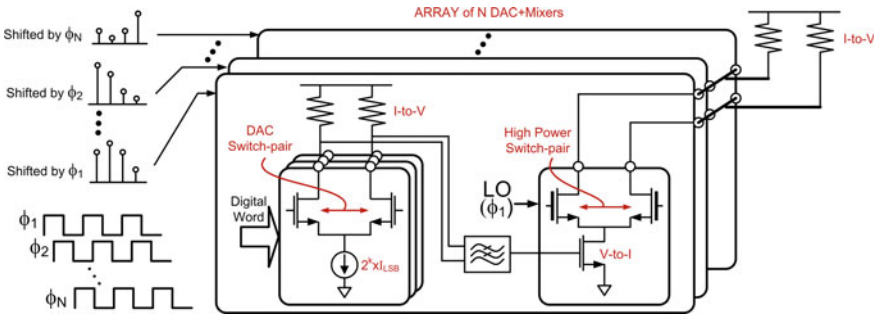

Fig. 3.28 Power Mixer Array

maximum achievable bandwidth that is power-limited. Parallel arrays of such DAC/mixers, proposed in [32, 33], was shown to cancel mixer nonlinearities by use of phase-shifted input signals and corresponding phase-shifted LO signals. Such a polyphase mixer has been shown to relax the mixer linearity requirements [34].

Another emerging architecture using time-interleaving topologies was proposed for high-frequency beyond-Nyquist synthesis [35, 36]. As illustrated in Fig. 3.29, a time-interleaved DAC comprises an array of DACs fed with interleaved signal samples and operating at interleaved instants of a clock period. It is also noted that the output of all DACs are connected together at all times; i.e. while one of the DACs is updating its output, the other DACs force their previously held values. This concept of hold and data interleaving was elaborated upon in Ref. [36] and proven to not hinder replica cancelation, while enabling beyond-Nyquist synthesis. The use of hold-interleaving was also shown to improve the replica suppression in the presence of gain and timing mismatches in [36].

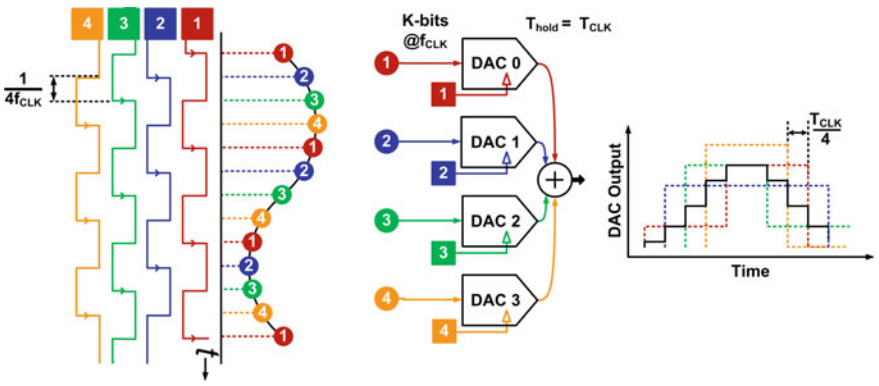

Fig. 3.29 DAC with data and hold interleaving [36]

3.8 Concluding Remarks

This chapter provided the reader with the holistic view of digital-to-analog conversion and associated challenges in CMOS process technologies. The DAC was first introduced as a system, and then abstracted to highlight some of the major limitations. A brief overview of performance metrics was then provided to quantify the DAC's static and dynamic performance. The complexity of the DAC design space, while perceived to be a simple array construct, forces a custom design flow. Attempts were made to simplify the DAC design space by breaking it into four parameters: device noise, output impedance, signal swing and switching speed. The reader was then introduced to the need for segmentation as a result of process variation and circuit limitations. Finally, a review of architectural and circuit techniques in the context of high-performance DACs to aid in circumventing the technological challenges, was presented. The need for higher resolution and higher speeds have kindled the interest in RF-DACs and interleaving, that are foreseen as promising for future ultra-wideband applications. The long-recognized fundamental limitations associated with process variations still remain a limiting factor in the implementation of high-performance DACs. However, the increasing gate densities in advanced CMOS processes allow for the realization of high-complexity mechanisms for self-calibration and self-compensation, which can effectively alleviate the various impairments suffered in the analog circuitry.

References

1. Balasubramanian, S., et al.: Ultimate transmission. IEEE Microwave Magazine. **13**(1), 64–82 (2012)
2. van den Bosch, A., Steyaert, M., Sansen, W.M.: Static and dynamic performance limitations for high speed D/A converters. Kluwer Academic Publishers, Dordrecht (2004)
3. Razavi, B.: Principles of data conversion system design. IEEE Press (1995)
4. Wikner, J.J.: Studies on CMOS digital-to-analog converters. Ph.D. Dissertation, Linkoping University, Linkoping (2001)
5. Maloberti, F.: Data converters. Springer, Dordrecht (2007)
6. Jenq, Y.C., Li, Q.: Differential non-linearity, integral non-linearity, and signal to noise ratio of an analog to digital converter. IMEKO International Measurement Confederation, Chicago (2002)
7. Tsividis, Y., McAndrew, C.: Operation and modeling of the MOS transistor. ISBN 978-0195170153, Oxford University Press, New York (2010)
8. Hiroki, A., Yamate, A., Yamada, M.: An analytical MOSFET model including gate voltage dependence of channel length modulation parameter for 20 nm CMOS. In: Proceedings of International Conference on Electrical and Computer Engineering. ICECE, Dec 2008, pp. 139–143
9. Chen, T., Gielen, G.G.E.: The analysis and improvement of a current-steering DACs dynamic SFDR-I: the cell-dependent delay differences. IEEE Trans. Circuits Syst. I, Regular Papers. **53**(1), (2006)

10. Lin, C.H., van der Goes, F., Westra, J., Mulder, J., Lin, Y., Arslan, E., Ayranci, E., Liu, X., Bult, X., abd, K.: A 12b 2.9GS/s DAC with IM3 $\ll -60$dBc beyond 1 GHz in 65 nm CMOS. IEEE J. Solid-State Circuits. **44**(12), 3285–3293 (2009)
11. Available online: www.itrs.net
12. Seckin, U., Ken Yang., C.: A Comprehensive delay model for CMOS CML circuits. IEEE Trans. Circuits Syst. I, Regular Papers. **55**(9), (2008)
13. Pelgrom, M.J.M., Duinmaijer, A.C.J., Welbers, A.P.G.: Matching properties of MOS transistors. IEEE J. Solid-State Circuits **24**(5), 1433–1439 (1989)
14. Lin, C.H., Bult, K.: A 10-bit, 500-MS/s CMOS DAC in 0.6 mm^2. IEEE J. Solid-State Circuits **33**(12), 1948–1958 (1998)
15. Bastos, J., Marques, A.M., Steyaert, M.S.J., Sansen, W.: A 12-bit intrinsic accuracy high speed CMOS DAC. IEEE J. Solid-State Circuits **33**(12), 1959–1969 (1998)
16. Balasubramanian, S., Khalil, W.: Architectural trends in GHz speed DACs. NORCHIP, Nov 2012
17. Bugeja, A.R., Song, B., Rakers, P.L., Gillig, S.F.: A 14-bit 100-MS/s CMOS DAC designed for spectral performance. IEEE J. Solid-State Circuits **34**(12), 1719–1732 (1999)
18. Oyama, B.,et al.: InP HBT/Si CMOS-based 13-bit 1.33Gsps digital-to-analog converter with $>$ 70 dB SFDR. IEEE Compound Semiconductor IC Symposium. Oct 2012
19. Jewett, B., Liu, J., Poulton, K.: A 1.2GS/s 15b DAC for precision signal generation. In: Proceedings of the IEEE International Solid-State Circuits Conference, Feb 2005 pp. 110–587
20. Halder, S., Gustat, H., Scheytt, C., Thiede, A.: A 20GS/s 8-Bit current steering DAC in 0.25 μm SiGe BiCMOS technology. In: European Microwave Integrated Circuit Conferenc. EuMIC, Oct 2008 pp. 147–150
21. Schofield, W., Mercer, D., Onge, L.S.: A 16 b 400 MS/s DAC with <-80 dBc IMD to 300 MHz and <-160 dBm/Hz noise power spectral density. In: Proceedings of the IEEE International Solid-State Circuits Conference, 2003 126–127
22. Park, S., Kim, G., Park, S., Kim, W.: A digital-to-Analog converter based on differential-quad switching. IEEE J. Solid-State Circuits **37**(10), 1335–1938 (2002)
23. Garceran, J.A.: Digital transmitter system and method. US Patent 6,944,238 (B2), Sep 2005
24. Choe, M-J.: Return-to-zero current switching digital-to-analog converter. US Patent 7,042,379 (B2), May 2006
25. Jha, A., Kinget, P.R.: Wideband signal synthesis using interleaved partial-order hold current-mode digital-to-analog converters. IEEE Trans. Circuits Syst. II, Exp. Briefs, **55**(11), 1109–1113 (2008)
26. Zhou, Y., Yuan, J.: An 8-bit 100-MHz CMOS linear interpolation DAC. IEEE J. Solid-State Circuits **38**(10), 1758–1761 (2003)
27. Luschas, S., Schreier, R., Lee, H.-S.: Radio frequency digital-to-analog converter. IEEE J. Solid-State Circuits. **39**(9), 1462–1467 (2004)
28. Maxim, A., et al.: A DDFS driven mixing-DAC with image and harmonic rejection capabilities. In: Proceedings of the IEEE International Solid-State Circuits Conference, Feb 2008 pp. 372–621
29. Jerng, A., Sodini, C.G.: A wideband $\Delta\Sigma$ digital-RF modulator for high data rate transmitters. IEEE J. Solid-State Circuits **42**(8), 1710–1722 (2007)
30. Taleie, S.M., Copani, T., Bakkaloglu, B., Kiaei, S.: A linear Σ-Δ digital IF to RF DAC transmitter with embedded mixer. IEEE Trans. Microw. Theory Tech. **56**(5), 1059–1068 (2008)
31. Boos, Z., et al.: A fully digital multimode polar transmitter employing 17b RF DAC in 3G mode. In: Proceedings of the IEEE International Solid-State Circuits Conference, Feb 2011 pp. 376–378
32. Shrestha, R., Eric, A., Klumperink, M., Mensink, E., Wienk, G.J.M., Nauta, B.: A polyphase multipath technique for software-defined radio transmitters. IEEE J. Solid-State Circuits. **41**(12), 2681–2692 (2006)

33. Kousai, S., Hajimiri, A.: An octave-range, watt-level, fully integrated CMOS switching power mixer array for linearization and back-off-efficiency improvement. IEEE J. Solid-State Circuits. **44**(12), (2009)
34. Xi, Y., et al.: Poly-harmonic modeling and predistortion linearization for software-defined radio upconverters. IEEE Trans. Microw. Theory Tech. **58**(8), 2125–2133 (2010)
35. Balasubramanian, S., Khalil, W.: Direct digital-to-RF digital-to-analogue converter using image replica and nonlinearity cancelling architecture. Electron. Lett. **46**(14), 1030–1032 (2010)
36. Balasubramanian, S., et al.: Systematic analysis of interleaved digital-to-analog converters. IEEE Trans. Circuits Syst. II, Exp. Briefs, **58**(12), 882–886 (2011)