Chapter 2
General Overview of Pipeline Analog-to-Digital Converters

Abstract This chapter provides a general background for the work carried out in this book. Therefore, its purpose is to cover all aspects of the developed work. First, some A/D converter (ADC) architectures will be briefly described. The common element of these architectures is the use of the multiplying-DAC (MDAC) circuit as their principal block. Advantages and limitations of the architectures will also be given. The MDAC circuit is one of the key elements of this book. Given that this work presents a prototype of a pipeline ADC, it is important to describe each of its building blocks. Besides detailing the function and importance of each block, related errors and performance limiting aspects will also be given. After the description of the pipeline converter sub-blocks, various static and dynamic performance parameters, and metrics that characterise ADCs are given. It will be the objective here to explain the parameters that fundamentally dictate the performance of ADCs. Finally, the chapter is completed with a state-of-the-art of medium-low resolution high-speed pipeline ADCs. Besides this overview, surveys of two key building blocks, namely, two-stage amplifiers and reference voltage circuits (in the context of A/D conversion), which deserved special attention in this work, are also presented.

2.1 MDAC-Based Analog-to-Digital Converter Architectures

There are many architectures of A/D converters, each with their own set of characteristics and capabilities to be used in different applications. Well known architectures are Full-Flash (or Parallel), Two-Step, Sub-Ranging, Folding, Integrating, Successive Approximation (SA), Algorithmic, Pipeline, Sigma-Delta modulators, and Time-to-Digital. It is also possible to find numerous combinations of the various existing topologies, such as: time-interleaving can be used as a means of increasing the sampling frequency (conversion rate) by arranging various converters of the same type in parallel; it is very frequent to find interpolation associated with flash and folding converters; the two-step topology can employ either flash or SA architecture in each

M. Figueiredo et al., *Reference-Free CMOS Pipeline Analog-to-Digital Converters*, Analog Circuits and Signal Processing, DOI: 10.1007/978-1-4614-3467-2_2,

step; the pipeline and algorithmic topologies usually employ flash converters in each step, etc.

Part of the work carried out in this book implements an ADC topology that employs MDAC circuits, consequently, the only converter topologies of interest, of the ones mentioned above, are those that use MDAC circuits as a means of obtaining a residue with amplification (i.e., simultaneous DAC, subtraction, and residue amplification functions). With this in mind the only architectures that will be discussed are the Two-Step (flash), the multi-step Algorithmic, and the Pipeline, targeting higher conversion rates. The time-interleaving technique will also be briefly described.

2.1.1 Two-Step Flash ADC

The Two-Step Flash architecture evolved from the Full-Flash converter. One of the main drawbacks of the latter is the number of necessary comparators, given by $N_{comp.} = 2^N - 1$, which scales exponentially with the resolution of the converter (N), making it, in some cases, impractical to implement due to the necessary die area. The Two-Step Flash topology alleviates the number of necessary comparators by quantizing the input in two steps, hence its name, as shown in Fig. 2.1a. The effective reduction factor in the number of comparators when compared to the Full-Flash ADC, is exponentially proportional to the converter's resolution, and is approximately given by[1] $2^{\frac{N}{2}-1}$. In other words, the higher the resolution the more area efficient it becomes to use a Two-Step topology.

As shown in Fig. 2.1a, each step (or stage) is composed of a quantizer, with a resolution (N_1 and N_2) inferior to the resolution (N) of the entire converter, thus requiring less reference voltages and comparators, and consequently, occupying less die area. Between the two steps an amplified residue voltage needs to be generated, which is achieved with a DAC, a subtraction operation block and a gain block. These three blocks constitute an MDAC circuit. The principle of operation is as follows: the input is sampled by the first quantizer during the sampling phase. During the residue amplification phase, the first quantizer decides the most significant bits (MSBs), which are then used to reconstruct a voltage (using the DAC), that is subtracted from the original sampled input and then amplified (by 2^{N_1}) to create an amplified residue voltage. Still during this phase, the second quantizer samples this residue. The objective of the amplification is to restore the residue to the full voltage range of the converter, thus facilitating the implementation of the second quantizer (Fig. 2.1b), or eventually, reusing the same quantizer in a cyclic way. During the final phase, the second quantizer (of resolution N_2) quantizes the residue to obtain the least significant bits (LSBs). The final digital output is assembled using digital logic, by adding the MSBs together with the LSBs.

Basically, this topology simplifies the quantization, by trading comparators with time. The Full-Flash converter achieves a quantization in two clock phases (one clock cycle), while the Two-Step needs at least three phases, i.e., one and a half clock

[1] Assuming that the same number of bits is extracted in both quantization steps, i.e. $N_1 = N_2 = N/2$.

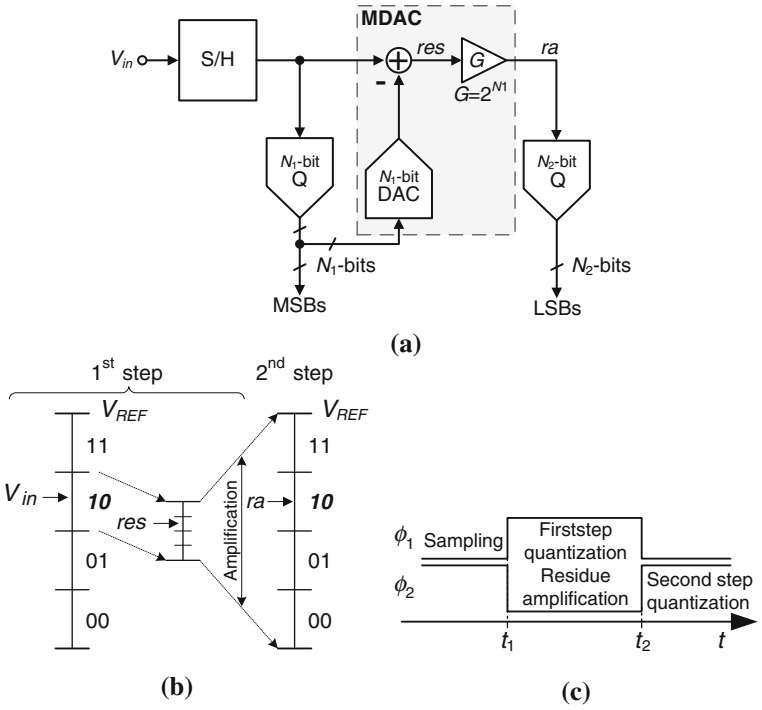


Fig. 2.1 The Two-Step A/D converter: **a** Block diagram. **b** Example of residue amplification. **c** Timing diagram

cycles (Fig. 2.1c). Although the throughput may be similar to that of the Full-Flash converter (one digital output per clock cycle), the Two-Step has higher latency.[2] If N_1 is made equal to N_2, then only one quantizer needs to be designed and, therefore, both quantizers may use the same reference voltages. The latter is also made possible by using a residue amplification gain of 2^{N_1}. The above assumptions consider that no digital redundancy is used.

If more steps (or stages) are added to the converter to simplify its implementation and relax the requirements of each step, which would eventually lead to a minimum resolution per step ($N = 1$), the resulting A/D architecture would be the Pipeline.

2.1.2 Pipeline ADC

The Pipeline converter's operation is basically the same as that of the Two-Step Flash. Each stage is responsible for quantizing N_j-bits, $j = 1, 2, \ldots, K$ ($N_j < N$) and generating an amplified residue for further quantization (performed by subsequent

[2] Latency is the number of initial clock cycles to produce the first digital output.

stages). However, the first stage does not have to wait for the residue of a specific sample to reach the end of the pipeline, to conclude its quantization. As soon as the first stage has performed its task, it may quantize the next input sample. This holds for all stages, which means that at any given time, except at the beginning when the converter starts its operation, all stages are processing data. Thus, the throughput of the Pipeline converter may be similar to that of the Full-Flash ADC, but its latency is high, even higher than that of the Two-Step Flash converter. The more stages there are in the pipeline chain, the higher the latency will be.

As shown in Fig. 2.2a, each stage of a Pipeline converter is composed of a flash quantizer and an MDAC. The flash quantizer quantizes the input sample (or residue) and generates N_j bits. In the literature it is common to find 1 to 4 bit quantizers (in half-bit intervals, 1.5-bit, 2.5-bit, etc.), the most common being 1.5-bit. The MDAC is responsible for reconstructing a residue voltage, determined by the N_j bits of the quantizer, subtracting it from the stage's input voltage and amplifying the result, to generate the residue voltage. This voltage is then held and passed onto the next stage where the stage's operation is repeated. The amplification of the residue, by 2^{N_j}, is justified for increasing its dynamic range to the full scale range of the converter, thus facilitating the implementation of subsequent quantizers. Each stage's operation is completed in two phases (one clock cycle): the first for sampling and quantizing, and the second for residue amplification (see the timing diagram of Fig. 2.2b).

Normally, all pipeline stages are designed with the same resolution to simplify the converter's layout and implementation, but, design trade-offs may determine that each stage have different resolutions. It is usual to find the first stage with a higher resolution. Another important reason for all stages to be equal (in resolution) is that the reference voltages are the same for all quantizers and DAC functions.

What concerns digital logic, the Pipeline converter employs digital synchronization logic to align (over time) all bits before producing the final digital output word. Aside from synchronizing, these converters usually employ digital correction, which corrects for nonidealities in the flash quantizers [97]. The Two-Step Flash converter, described above, may also employ digital correction logic. These digital blocks are detailed further on in this chapter.

2.1.3 Multi-Step Algorithmic ADC

The Algorithmic (or Cyclic) converter, as the name indicates, quantizes the input sample in an algorithmic or repetitive manner. Its principle of operation is basically the same as that of the Pipeline converter, in that an input is sampled, quantized, and a residue is amplified, and then the quantization and residue amplification process is repeated by subsequent stages. The main difference with the Algorithmic converter is that the conversion algorithm (sampling, quantization, and residue amplification) is repeated in the same physical space (reutilizing the same circuits) or area, while the Pipeline ADC repeats its operation over more area. In other words, the Algorithmic converter trades space for time, which means it has a longer conversion cycle.

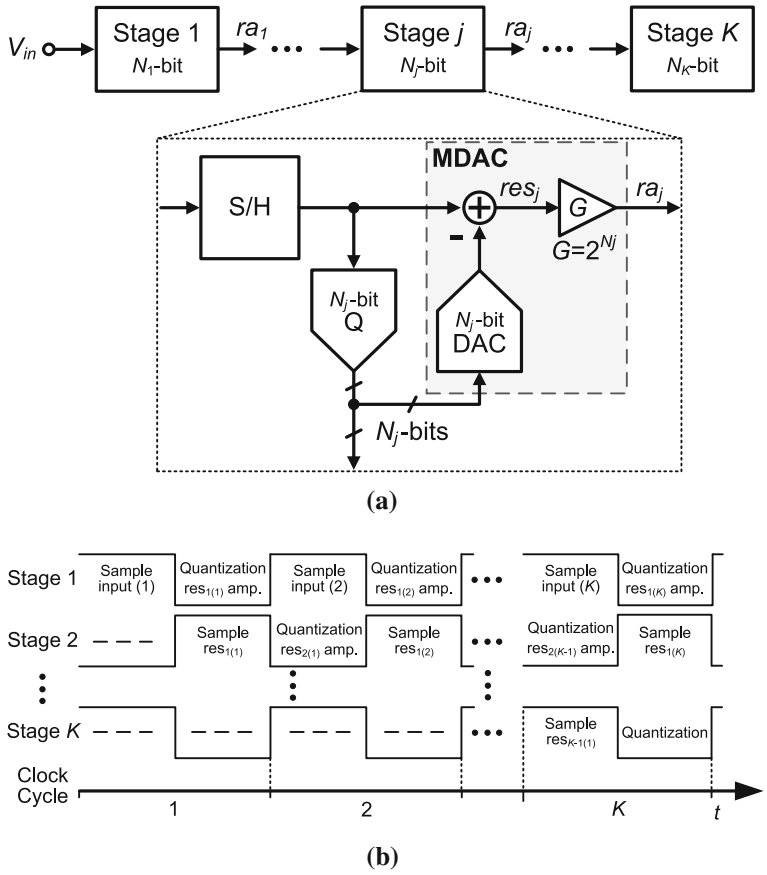


Fig. 2.2 The Pipeline A/D converter: **a** Block diagram of a stage, exemplifying its building blocks. **b** Timing diagram

The converter has a minimum number of stages (usually two), where the residue repeatedly passes through each stage, successively generating output bits, from the MSB to the LSB, as shown in Fig. 2.3a. After the LSB is generated, the process starts over again with a new sample.

The principle of operation is as follows: the input voltage is sampled by the first stage. It is then quantized to generate the MSBs. These bits are used to reconstruct a voltage (using a DAC and reference voltages) that is subtracted from the input voltage generating the residue voltage. This residue is then amplified (and held) to the full scale range of the converter and sampled by the second stage. This process repeats itself between the first and second stages until the LSB is generated. At this point a full digital output is ready, while the converter is sampling the next input sample (Fig. 2.3b).

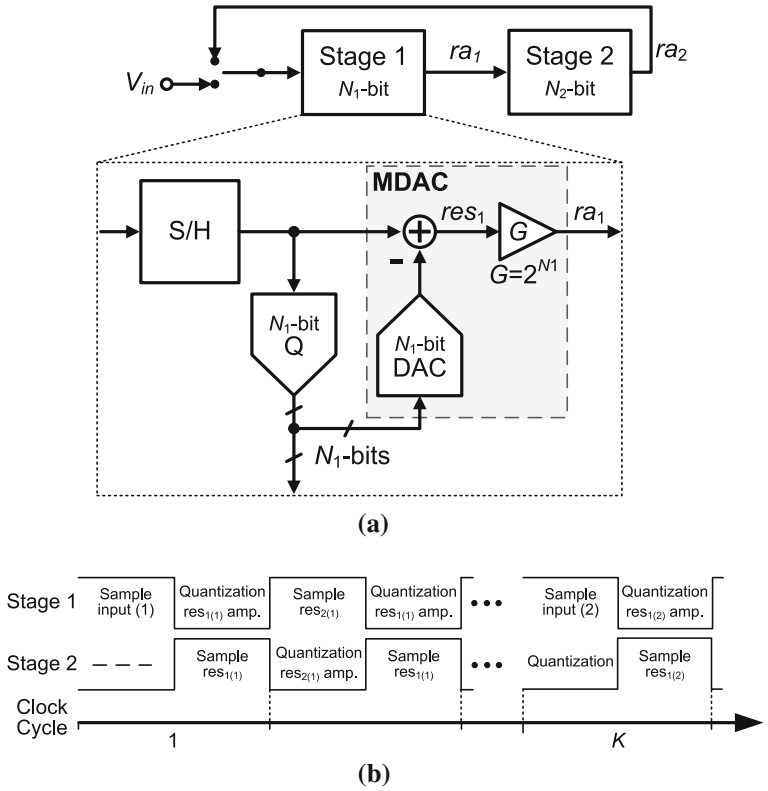


Fig. 2.3 The Algorithmic A/D converter: **a** Block diagram. **b** Timing diagram

Unlike the Pipeline, this converter is unable to process more than one sample at any given time, thus a converter of resolution N needs $N + 1$ clock cycles to quantize a single input sample. Compared to the Pipeline ADC, this converter's throughput is much lower, but has the same latency. It trades throughput with die area and facilitates layout and implementation due to the number of necessary blocks, which are much less than that used in a Pipeline converter.

The Algorithmic architecture must employ digital circuitry to hold and align the digital outputs of each stage before producing the final output word, and it usually employs digital correction logic.

In order to increase the throughput, the time-interleaved technique may be employed.

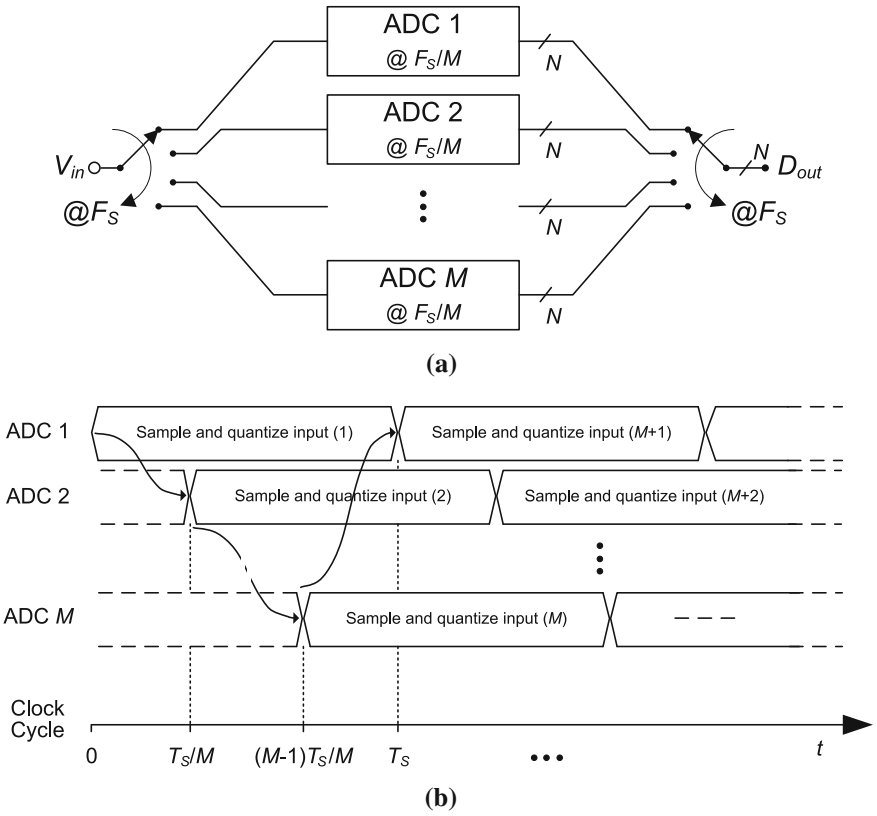


Fig. 2.4 Time-interleaving of A/D converters: **a** Block diagram. **b** Timing diagram

2.1.4 Time-Interleaving ADCs

Time-Interleaving [11] is a technique used to increase the throughput or conversion rate of a converter. This technique may be applied to all A/D converter topologies. It consists of using an array of M parallel converters multiplexed at the input and at the output as shown in Fig. 2.4a. Each converter operates at a conversion rate $F_{S/M}$ (where F_S is total conversion rate), making it easier to implement. The analog input multiplexer, adequately timed, is responsible for attributing an input sample to each of the converters over time, when it reaches the last converter in the array it starts over again. The digital output multiplexer guarantees that the timing sequence of digital outputs are in accordance with the sampled inputs. Each converter added to the array inevitably increases the die area and the power consumption of the overall A/D conversion system. The timing diagram of operation is shown in Fig. 2.4b.

Besides the limitations produced by each unit (multiplexed) ADC, the time-interleaved technique introduces its own limitations. These have mainly to do with

mismatches between the various unit ADCs that compose the converter [11, 39, 80, 91, 133, 172]. This topology is very sensitive to mismatches in the offset, gain, timing, and bandwidth of each unit ADC. All these extra errors (inherent to time-interleaving) cause a degradation of the converter's signal-to-noise ratio (SNR). No matter how large the error of a unit ADC is, as long as all other unit ADCs have the same error magnitude, no mismatch will exist. A brief description of these errors is given next and Table 2.1 presents a summary of the effects of time-interleaving mismatches [11, 39, 80, 91, 133, 172]. For this discussion, a two-channel time-interleaved ADC will be used as an illustrative example.

- **Offset mismatch** contributes with a component at DC (frequency, $f = 0$) in the output spectrum, which is already expected because all ADCs have an offset, but, it also adds a spurious tone at half the sampling frequency ($F_S/2$). The spectral locations and the magnitude of these components are independent of input signal amplitude and frequency.
- **Gain mismatch** contributes with a spurious tone at half the sampling frequency minus the input frequency ($F_S/2 - f_{in}$), thus its spectral location is dependent on the input signal frequency. The magnitude of the spurious tone is only dependent on the amplitude of the signal. The magnitude of the input signal is affected by this mismatch.
- **Timing mismatch** is due to variations in the sampling instant of each unit ADC, in other words, differences in the relative time between samples taken. This mismatch is similar to gain mismatch in the sense that it contributes with spurious tones at the same spectral locations ($F_S/2 - f_{in}$), but, the magnitude is dependent on the amplitude and the frequency of the signal. As with gain mismatch, timing mismatch affects the magnitude of the input signal.
- **Bandwidth mismatch** is due to differences in the sampling networks of each unit ADC. If each unit ADC has a dedicated sample-and-hold (S/H), and if each S/H has a different bandwidth, then bandwidth mismatch will affect the performance of the overall time-interleaved ADC [39, 91]. This mismatch adds similar contributions to that of gain and timing mismatches. The main differences are that the gain part of the bandwidth mismatch is now dependent on the input signal frequency and the timing part has a nonlinear dependency with the input signal frequency.

2.2 Building Blocks of Pipeline Analog-to-Digital Converters

The objective of this section is to briefly overview each of the constituent blocks of a Pipeline A/D converter. Besides an overview, errors related to each block will also be given. Various references are given throughout the section for a more detailed coverage and further reading. Note that, not all the blocks described below are necessary to build a Pipeline ADC. For example, the sample-and-hold (S/H) and decimation blocks are not strictly necessary. It should be equally noted that the blocks described

Table 2.1 Magnitude and spectral location of spurious tones due to mismatches and their effect on the signal ($A_{in}\sin(2\pi f_{in})$) component for a two-channel time-interleaved ADC. The offset and gain mismatches are given by o_i and g_i respectively ($i = 1, 2$). The relative timing mismatch is given by $r_i = \Delta t_i / T_S$, where Δt_i is the absolute timing mismatch and $T_S (= 1/F_S)$ is the sampling period

Mismatch type	Signal component (Magnitude)	Spurious component (Spectral location) (Magnitude)	
Offset	—	DC	$F_S/2$
		$\frac{o_1+o_2}{2}$	$\frac{o_1-o_2}{2}$
Gain	$A_{in}\frac{g_1+g_2}{4}$	$F_S/2 - f_{in}$	
		$A_{in}\frac{g_1-g_2}{4}$	
Timing	$A_{in}\cos(2\pi f_{in}\frac{r_1-r_2}{2})$	$F_S/2 - f_{in}$	
		$A_{in}\sin(2\pi f_{in}\frac{r_2-r_1}{2})$	
Bandwidth	see [91]	$F_S/2 - f_{in}$	
		see [91]	

below are generic, in the sense that they are found in most Pipeline ADCs. In recent developments some of these blocks have been substituted for more efficient ones and some have been eliminated (mostly for power and/or area savings).

2.2.1 Sample-and-Hold

The sample-and-hold (S/H) block is found at the very beginning of the converter. Its objective is to discretise, in time, i.e., to sample the input and hold the sampled input for the subsequent block to process it. The S/H converts a continuous-time signal into a discrete-time signal (the signal is still continuous in amplitude). Another circuit with similar functions is a track-and-hold (T/H), where the main difference to a S/H is that the output of the T/H tracks (follows) the input, then samples, and finally holds the sample. A simple version of a S/H and a T/H are shown in Fig. 2.5 with their respective timing diagram.

S/H circuits operate in two phases, the sampling and the holding phase, as shown in Fig. 2.5a. During the sampling phase, the switch (ϕ_S) is closed and the capacitor is charged to the input voltage. When the switch is opened, the input is sampled, and because the charge on the capacitor can not be destroyed, the sampled voltage is held on the capacitor. At this moment, the held voltage can only be sensed by a high input impedance block such as an amplifier (unity-buffer in this case).

There are numerous errors associated to sampling and holding an input signal [9, 84, 135]. Some are mentioned below:

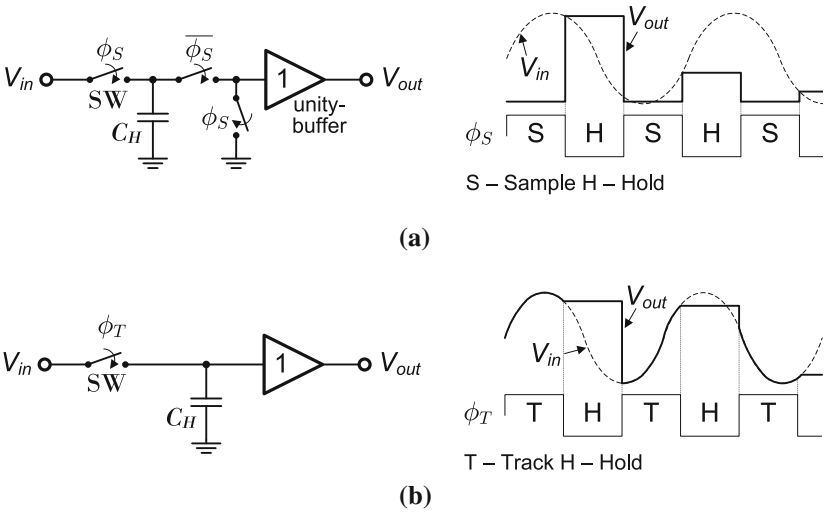


Fig. 2.5 Simple versions of (**a**) S/H and (**b**) T/H, with output buffer and output waveform

- **Finite sampling bandwidth** if the input signal's frequency is higher than the sampling bandwidth ($f_{-3\,\mathrm{dB}} = 1/(2\pi R_{SW} C_H)$), an output signal voltage with a phase difference is sampled.
- **Acquisition time** is the time it takes the amplifier (buffer) to settle. This error is associated with amplifiers which will be explained further on.
- **Sampling uncertainty (aperture error)** is the time uncertainty at the moment of sampling. This error has two possible origins: a long rise or fall time, or a sampling instant that changes from period to period. This error is particularly problematic in the presence of high frequency signals.
- **Sampling pedestal** is the error voltage added to the sampled voltage caused by the switch while it is turning off. This extra voltage is due to channel charge injection and clock feed-through. This error is particularly problematic when the error voltage is signal dependent, which adds distortion.

2.2.2 Multiplying-DAC

As seen in the previous section, the MDAC is a circuit which performs numerous functions. These functions include sampling the input signal (or residue voltage from a previous stage), reconstructing a voltage using a DAC, obtaining a residue (subtraction of the reconstructed voltage from the stage's sampled voltage), performing a gain to amplify the residue, and finally holding the amplified residue for the next stage. The block diagram of a generic MDAC is shown Fig. 2.6. Normally, a switched-capacitor (SC) network is employed to accomplish all these functions. Sampling is

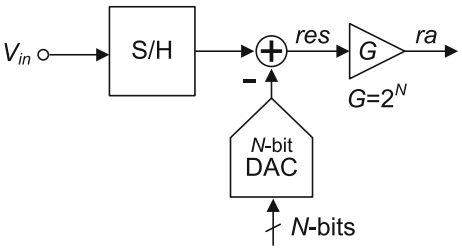


Fig. 2.6 Block diagram of a generic N-bit MDAC

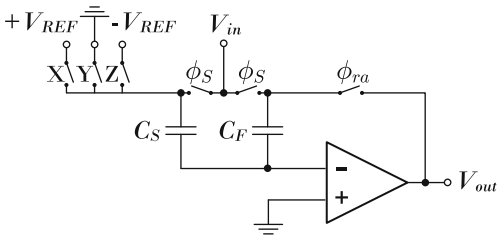


Fig. 2.7 Closed-loop switched-capacitor opamp-based 1.5-bit MDAC

achieved by means of switches and capacitors, similar to that shown in the previous subsection. The MDAC employs an operational amplifier (opamp) with a capacitive feedback network (closed-loop) to provide the DAC, subtraction, and amplification functions.

In the literature it is possible to find various techniques of implementing the MDAC function, either in open or closed-loop, such as, closed-loop switched-capacitor techniques, switched-current techniques [103], open-loop amplification [114], dynamic source follower amplification [76], MOS parametric amplification [123], and the substitution of the opamp for a comparator based circuit [15], among others. Each technique has its own advantages and limitations, not discussed here.

To understand the principle of operation of an MDAC circuit, a simple closed-loop switched-capacitor opamp-based 1.5-bit MDAC is used, as shown Fig. 2.7 (single-ended shown for simplicity) [98]. This resolution MDAC (1.5-bit) is chosen because it is one of the most widely used of all implemented stage resolutions. The input-output transfer characteristic is shown in Fig. 2.8a (ideal case). This characteristic can be described by the following expression

$$V_{out} = 2V_{in} + B \cdot V_{REF}, \tag{2.1}$$

where B represents the bit decisions made by the local quantizer (represented in Fig. 2.7 by X, Y, or Z which represent $B = +1$, $B = 0$, and $B = -1$, respectively) and V_{REF} represents the converters reference voltage. Equation 2.1 shows that the MDAC has a gain component ($2V_{in}$) and a reference shifting component ($B \cdot V_{REF}$).

Circuit operation is as follows: during ϕ_S the input (V_{in}) is sampled onto capacitors C_S and C_F. During the residue amplification phase (ϕ_{ra}), C_F is put in the opamp's feedback loop, and, due to charge conservation, the charge on C_S is transferred to

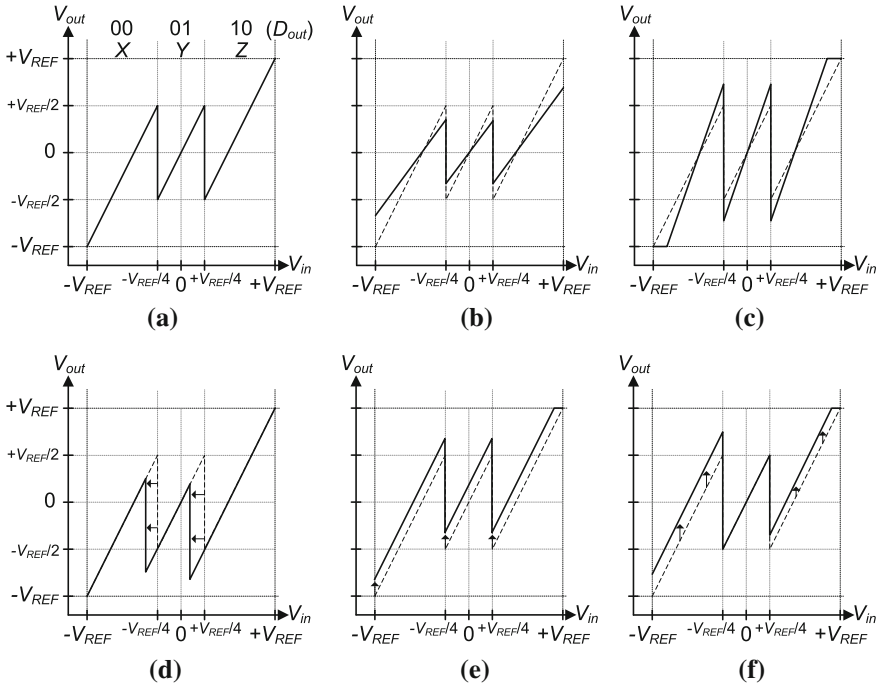


Fig. 2.8 Input–Output characteristics of the 1.5-bit MDAC and the effects of errors (*dashed line* represents the ideal situation): **a** Ideal. **b** Gain error: gain < 2. **c** Gain error: gain > 2. **d** Offset of the quantizer decision levels. **e** Offset in the MDAC due to charge injection or offset of the opamp. **f** DAC nonlinearity

C_F. The amount of transferred charge depends on the quantizer's decision (X, Y, or Z). If Y is high, i.e., if the Y switch is closed, the sampled input charge on C_S is transferred to C_F. In this case the output (V_{out}) will simply be $2V_{in}$ (parameter B of Eq. 2.1 is 0). If X or Z is enabled, then there will respectively be a charge addition or subtraction on C_S and the resultant charge transferred to C_F. In this single-ended version example, the DAC is only composed of capacitor C_S.

The limiting factors of closed-loop SC-MDAC circuits are given next. To aid the enumeration of these factors, the example of the 1.5-bit MDAC will be used as well as its transfer characteristics shown in Fig. 2.8.

- **Gain error** caused by capacitor mismatch (between C_S and C_F) and finite opamp DC gain [97, 102, 156]. Slopes of the characteristic vary from the ideal value of 2. See Fig. 2.8b, c.
- **Offset errors** can be caused by offset of the quantizer decision levels (Fig. 2.8d), by charge injection [86], or by opamp offset in the MDAC (Fig. 2.8e). Offset errors are of minor importance given that most errors can be corrected by the digital correction logic [97].

- **Nonlinearity errors** caused by DAC capacitor mismatch and nonlinearity errors present in the opamp [102] (Fig. 2.8f).
- **Thermal noise** comes from the ON-resistance of the switches which is sampled on the sampling capacitors and from the opamp [156].
- **Speed** conversion rate is limited by the opamp's closed-loop configuration, namely, the opamp's finite speed (GBW) and the closed-loop feedback factor [98] (slew rate may also be a limiting factor of the speed).

2.2.3 Local Flash Quantizer and Comparators

As the name indicates, this circuit is a quantizer based on the Full-Flash converter topology. In Fig. 2.9a, a generic N-bit flash quantizer is shown. It employs, in parallel, a number of comparators,[3] each with their own reference voltage, to be compared with the input signal. The output of each comparator is either 0 or 1, indicating that the input signal's voltage is lower or higher than the reference voltage, respectively. There are various methods of implementing the comparator: cascade of inverters (or simple gain stages), an opamp in open-loop, and the latched comparator [66, 84]. The first two circuits are designed to amplify the input signal or the difference between the input and reference signals to guarantee a logic output (0 or 1, or, in terms of voltage, the negative or positive saturation voltage, respectively). The latched comparator is composed of a pre-amplifier and a positive feedback latch. The pre-amplifier amplifies the small differential input signal and minimizes effects caused by the latch, while the latch guarantees logic levels at the comparator's output.

In Fig. 2.9b, the last block of the 1.5-bit flash quantizer is an XYZ encoder. This circuit is responsible for guaranteeing, depending on the decisions of the comparators, that either X, Y, or Z is one. As mentioned before, these signals decide the B parameter of the reference shifting of the MDAC (see Eq. 2.1). Output bits (b_i, b_j) are used for digital correction (discussed further on).

The reference voltages may be generated by a resistive or capacitive divider string, or by means of a SC network at the comparators' input. Normally for an N-bit flash quantizer, $2^N - 1$ comparators are necessary. For a half-bit quantizer (1.5-bit, 2.5-bit, etc.), only $2^N - 2$ comparators are necessary. N represents the nearest integer greater than the half-bit resolution, e.g., 1.5-bit corresponds to $N = 2$.

Factors that limit the performance of comparators and, thus, quantizers are given below:

- **Offset** the opamp and the pre-amplifier introduce an input-referred offset error which may be due to mismatches (between circuit components) or may be inherent to the comparator design [52, 84, 140].
- **Charge injection and clock feed-through** caused by channel charge and parasitics associated with the sampling switches while turning off [84, 135].

[3] The comparator is probably the most widely used component in A/D conversion, fundamental in practically all topologies.

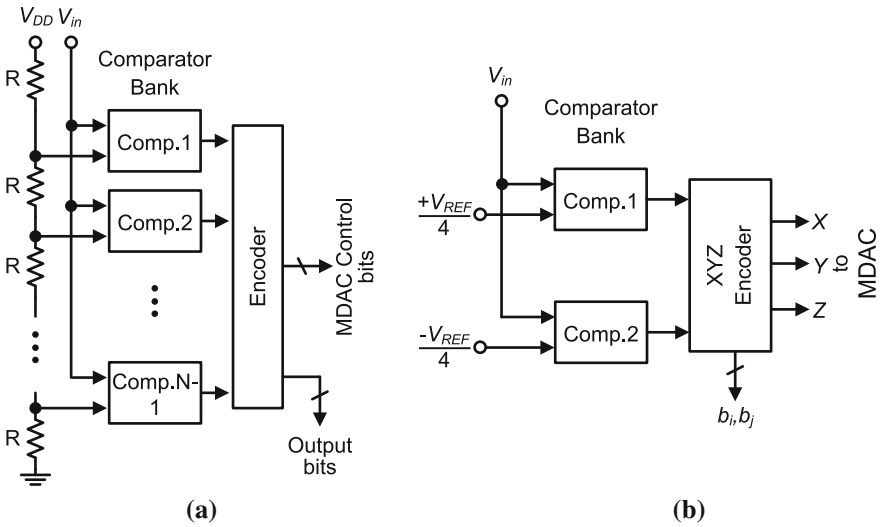


Fig. 2.9 Flash quantizer examples: **a** Generic N-bit quantizer. **b** 1.5-bit quantizer

- **Kickback noise** this error occurs during switching in latched comparators. When the comparator goes into latch mode, the high speed of the positive feedback latch causes high speed transients, which inject charge through parasitic capacitors back into the input signal, thus causing unwanted disturbances [52, 84, 140].
- **Comparison time** this is the time the comparator takes to produce a valid digital output. The worst case scenario is described by an overdrive recovery test, which defines the time the comparator takes to recover from a large input immediately followed by a very small input [104, 140].
- **Metastability** occurs when a very small input renders the comparator to be unable to produce a valid digital output. This error may be given as a probability of the occurrence of a metastable state [105, 140], as a number of metastable states per second [135], or, for the case of a flash quantizer with various comparators, as the mean time between failures (taking into account the number of comparators) [1].

2.2.4 Operational Amplifier and Common-Mode Feedback Circuitry

The operational amplifier, better known as opamp, is probably the most important and most used block in analog signal processing. The word operational comes from the fact that these blocks can be used to implement various functional operations. It is an active circuit which ideally has high gain, high input impedance, and low output impedance. Typically working in the voltage domain, a voltage is inputted and an

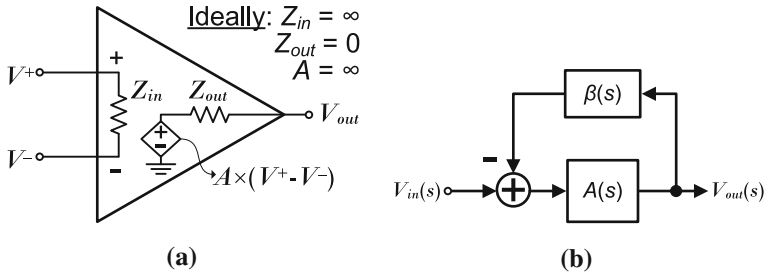


Fig. 2.10 Opamp: **a** Basic symbol. **b** Simplified block diagram of a feedback circuit

amplified voltage is provided at the output, i.e., the opamp is a voltage-controlled voltage source ($V_{out} = A \times (V^+ - V^-)$, where A is the opamp's open-loop gain), as shown in Fig. 2.10a. There are more types of amplifiers, but the one that deserves attention, given the work developed in this book, is the operational transconductance amplifier (OTA). This type of amplifier achieves a high open-loop gain (at low frequencies) at the expense of a high output impedance. They are widely used in SC circuits and do not need low output impedance because they usually only drive pure capacitive loads (and not resistive loads). Their gain and speed depend on the transconductance of specific transistors that compose the OTA.

Opamps usually operate in closed-loop form, inheriting all the associated benefits such as, less sensitivity to circuit and process, supply voltage, and temperature (PVT) variations, thus less distortion, higher input impedance, lower output impedance, and higher bandwidth. Figure 2.10b depicts a simplified block diagram of a negative closed-loop system, where $A(s)$ represents the opamp's open-loop gain (it represents Fig. 2.10a), $\beta(s)$ is the feedback network (with associated feedback factor), and s is the complex frequency.

As an example of the functionality and importance of the opamp, the SC circuit of Fig. 2.11a is used (this circuit is similar to Fig. 2.7 for $Y = 1$). For this example C_L represents a load capacitor and the objective is to double the input voltage (i.e., $V_{out} = 2\,V_{in}$). During ϕ_1, V_{in} is sampled onto C_S and C_F. During ϕ_2, the opamp (in closed-loop now) forces a virtual ground at its inverting input and, hence the voltage across C_S to zero. Due to charge conservation, the charge stored on C_S, $Q_S = C_S V_{in}$, is transferred to C_F. No charge goes to the amplifier because of its high input impedance. If C_F equals C_S then V_{out} becomes $2V_{in}$. Another way of exemplifying the importance of the opamp is: during ϕ_1, the voltages at the top and bottom plates of C_S are V_{in} and zero, respectively. Immediately at the beginning of ϕ_2, the top plate of C_S is connected to zero, which forces its bottom plate (and the opamp's inverting input) to $-V_{in}$. Simultaneously, the opamp (in closed-loop now) forces the virtual ground at its inverting input. In order to accomplish the virtual ground, the output voltage has to be increased by the same amount of voltage, i.e., V_{in}. Finally, the output will become $2V_{in}$ at the end of ϕ_2. If no opamp is used in the aforementioned explanations, neither charge conservation nor virtual ground would

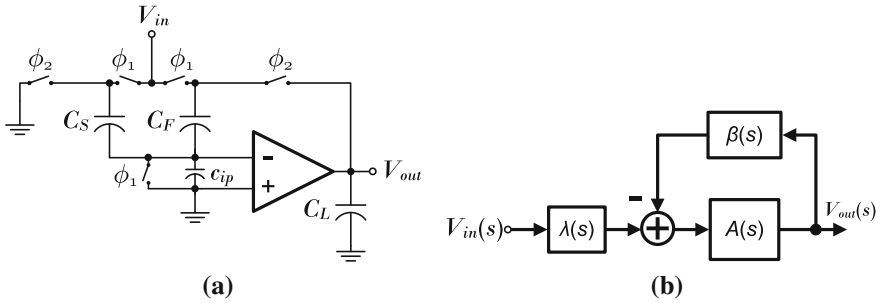


Fig. 2.11 Example using an opamp in a feedback loop: **a** SC circuit. **b** Simplified block diagram of the feedback scheme

occur. Consequently, the sampled charge would simply be distributed between C_S, C_F, and C_L, but V_{out} would not be $2V_{in}$ at the end of ϕ_2.

Figure 2.11b illustrates the equivalent block diagram of the SC circuit of Fig. 2.11a, where the feedback network is given by $\beta(s) = C_F/(C_F + C_S + c_{ip})$ and the input network is $\lambda(s) = (C_S + C_F)/(C_F + C_S + c_{ip})$ [28].

At the transistor level, many amplifier architectures exist. It will not be the objective of this work to describe any, but rather to give an idea of the common blocks used in most of them. The basic blocks are an input stage (where the input signal is connected), a gain or differential to single-ended conversion stage, and finally, an output driver stage (used to drive the load connected at the output) [66]. It is possible that, besides the input, the output stages be differential, thus inheriting the advantages of fully differential circuits, which will be given further on. When supply voltages were high (higher than 1.8 V in older CMOS technology nodes), gain could be achieved by cascoding (stacking) transistors. In this case enough gain was achieved with a single-stage amplifier, which also has the highest speed of operation. However, due to the power reduction necessity and low supply voltage (1.2 V and lower) tendency of modern nanoscale CMOS technologies (0.13 μm and beyond), it is no longer possible to cascode transistors, therefore, gain can only be achieved by multiple stages, i.e., by cascading stages. Besides gain, the output swing also decreases due to the supply voltage reduction, thus another reason to cascade stages. At this moment, we have arrived at a two-stage amplifier: the first stage mainly for gain and the second for output swing and speed with a small contribution to the overall gain. Due to compensation (to stabilize the amplifier), which is inevitable, the speed of the two-stage amplifier is reduced when compared to its single-stage counterpart. If more gain is needed, another stage can be added (three-stage amplifier), at the expense of speed and stability issues.

There are a number of parameters that characterize the performance of opamps. Most of these are limiting factors of their performance. Depending on the application the opamp is inserted in, some become more relevant than others. Here most of them will be described [65, 84, 92, 104, 141]. The parameters are:

- **Low-frequency gain**: There are two types of gains, differential-mode (DM) and common-mode (CM) gain. Usually an opamp is sized for a given DM gain while minimizing the CM gain. The former is fundamental as it determines the precision of the overall system where the opamp is used (usually in closed-loop). These gains, measured with the opamp in open-loop, are a function of frequency. At low frequencies they are called, the DC gain. An important sub-parameter of gain is its nonlinearity for different output voltages. Normally, for a small output voltage swing, the gain has its maximum value, but for larger voltage swings, the gain tends to be smaller.
- **Bandwidth**: There are two types of bandwidths, small-signal and large-signal (discussed in the following item), which characterize the high frequency performance of opamps. The former is related to the frequency dependence of the open-loop gain, which is caused by parasitic, compensation, and/or load capacitors present in the opamp. The open-loop gain versus frequency, has a constant value for low frequencies (DC gain, explained before) and then at a certain point in the frequency, the gain starts to decrease or roll-off (at a -20 dB/decade rate). This point is the opamp's bandwidth or the location of the dominant pole. The point at which the gain reaches unity is the unity gain bandwidth (f_u) or the gain-bandwidth product (GBW). Note that, GBW = f_u is only applicable either for a single pole opamp or for an adequately compensated multi-pole opamp [92, 104]. These two design parameters, f_u and GBW, determine the speed of the system the opamp is inserted in.
- **Slew rate** or large-signal bandwidth determines the rate at which the opamp can change the output voltage in the presence of large input signals. In large signal conditions, the opamp will try to provide current to charge the capacitors (compensation, load, etc.) of the system. The rate at which it does this is called the slew rate (SR). If the current is insufficient or the opamp does not have enough time (case of SC circuits) to charge the capacitors, the output will not reach the desired value and nonlinear distortion will arise.
- **Compensation and Phase Margin**: Opamps inserted in feedback loops can be potentially unstable, if not adequately compensated. A measure of this instability is called phase margin (PM). In single-stage opamps (which are less prone to instability), compensation is normally achieved (almost for free) by the load capacitor. In multi-stage opamps (unstable by nature), compensation capacitors and compensation schemes are inevitable.
- **Settling time**: This defines the time it takes the output to reach its final value within a given settling error (associated with the desired accuracy) when a step input is applied. This is probably the most important opamp parameter for SC circuits given that it is a time domain parameter, that includes the effect of gain (and its nonlinearity), small-signal bandwidth, slew rate, phase margin, and the closed-loop's feedback factor. Figure 2.12 [162] depicts an opamp's output response to an input step showing where each mentioned parameter plays its role. If any of them are not designed correctly, their effects will be pronounced in the output response and a longer settling time will probably occur.

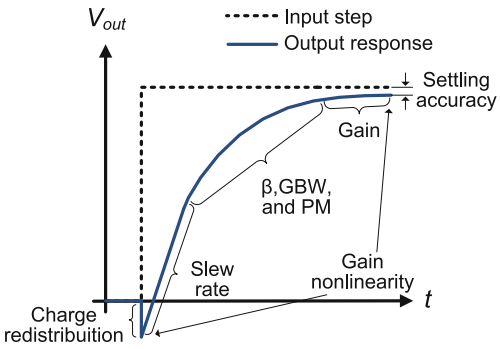


Fig. 2.12 Time-domain opamp output response (to an input step) and the role played by some opamp performance parameters

- **Output swing** is the maximum output voltage range possible that maintains the opamp functioning nominally, i.e., with the desired gain.
- **Common-mode input range** is the maximum input voltage range that guarantees negligible degradation of the opamp's performance.
- **Offset** is the output voltage when the input is zero. Ideally the output voltage should be zero, but will not be due to inherent design issues of the opamp (systematic offset) and mismatches between otherwise matched transistors (random offset). This offset voltage, if large enough, can limit the output swing.
- **Noise** is generated by devices with resistive components such as transistors (resistive channel) and, naturally, resistors. There are many types of noise, the two most commonly discussed are flicker (or $1/f$) and thermal noise. The latter is of more importance in high-speed or wideband opamps. Just as in the case of offset, the opamp may be designed for low noise but never for zero noise. Noise determines the smallest detectable input, the opamp may process.

Most parameters can be enhanced by using a fully differential structure for the opamp. In fact, opamp performance can double. This type of structure indicates that the opamp, besides having a differential input, also has a differential output. The advantages of the fully differential architecture are well known and are: larger gain, larger output swing, higher immunity to extrinsic noise, reduced distortion (suppression of even harmonics) and enhanced speed/power ratio [150]. Although the intrinsic noise level of fully differential circuits is higher, the signal-to-noise ratio (SNR) will still be higher than single-ended output circuits due to the larger output swing obtained.

An opamp designed with a fully differential structure needs to employ a special circuit to measure and control its output common-mode voltage. This circuit is known as the common-mode feedback (CMFB) circuit. Without this circuit the uncontrolled output CM voltage could drift to the supply rails. The CMFB circuit guarantees that this voltage stays approximately at midway between the supply rails. There are continuous-time and SC–CMFB circuits. Usually in SC circuits, the SC version of CMFB circuits is preferred.

2.2.5 Reference V/I and Buffering

Reference circuits are essential in analog and data converter systems. They generate reference voltages and currents that are used to bias circuits, to compare with other signals, for addition and subtraction operations, among others. In the specific case of data converters, reference circuits are determinant in defining the input and output full-scale ranges. Therefore it is necessary to guarantee a sufficient level of accuracy,[4] so that the overall performance of the data converter is not limited. To achieve this, they need to be independent of external conditions such as, process, supply voltage, temperature, and load disturbances. In the case where a reference voltage needs to drive a large capacitor or various capacitors amounting to a large capacitance (like in DAC circuits), or be used in a high-speed or high-accuracy SC circuit, an additional block needs to be added to the output of the reference circuit. Commonly known as a reference buffer, this block is used to maintain the reference voltage constant and to guarantee that it charges and discharges the capacitors it drives, in the available amount of time (particular case of SC circuits). In other words, the buffer must settle the reference voltage to within a given error, within a given time slot, which depends on the accuracy and speed of the converter.

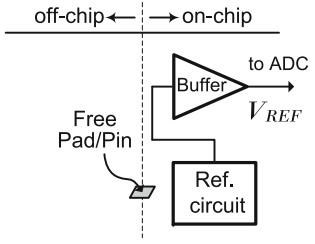
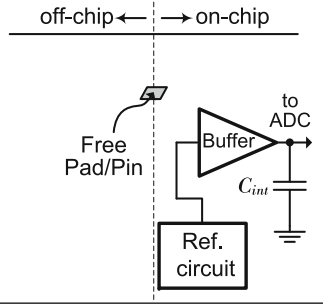
There are many forms of generating a reference voltage (resistive or capacitive ladders, bandgap circuits, etc.), but this is not the objective of this subsection. Instead, an overview of the issues and difficulties designers have to overcome to buffer and/or stabilize a reference voltage for SC converters, will be given.

It is possible to find many forms of reference voltage schemes in the literature. The options divide into on- and off-chip buffering with (or without) the use of on- and off-chip damping resistors and decoupling capacitors. The following table describes and analyses the most used forms of reference voltage circuitry and buffering schemes.

The conclusions extracted from Table 2.2 can be summarized as follows: the reference circuitry will occupy a large area, will dissipate a large amount of power, and/or will need at least one extra pin. Most of the currently employed solutions suffer from a combination of these drawbacks. From a system-level perspective, neither system-on-chip nor system-in-package designs are benefitted from voltage-domain reference circuitry. To avoid extra costs, no extra pins can be used, therefore, on-chip high-speed buffering must be employed, a trade between cost and power is made. If extra pins are available, then off-chip decoupling may be used (with a low bandwidth buffer), but then a penalty in area is paid for the off-chip decoupling capacitors. All this to avoid on-chip decoupling because silicon area is more expensive (valuable) than discrete components.

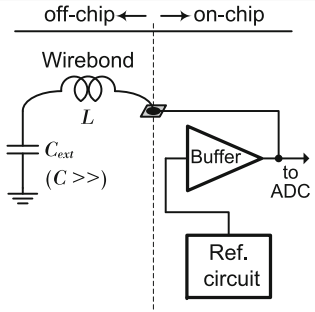
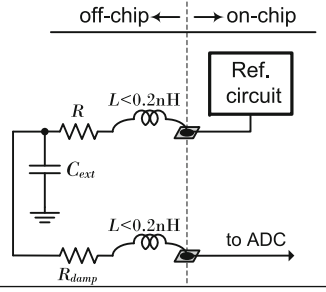
[4] In [113] it is shown that the reference circuit can have 1-bit lower accuracy than the resolution of the ADC.

Table 2.2 Description and analysis of the advantages and disadvantages of different reference voltage schemes

Description	Analysis	Diagram	Performance
On-chip buffer without decoupling [12, 14, 81, 158]	• High-speed buffer with wide bandwidth. • High power consumption. • Noise performance hard to achieve (as frequency rises, buffer output impedance rises). • [81] uses a deglitch circuit to minimize glitches during switching. • Due to large current peaks, the buffer may require dedicated supply pins.		• Accuracy ✢ • Power − • Speed ✢ • On-chip area ✢− • Off-chip area ✢
On-chip buffer with internal decoupling [20, 34, 78]	• Wirebond inductance too large for off-chip decoupling due to impractical amount of ringing on reference voltage. • Large on-chip capacitors, occupying large area, are unavoidable [20]. • On-chip RC filter [34].		• Accuracy ✢ • Power ✢− • Speed ✢− • On-chip area − • Off-chip area ✢

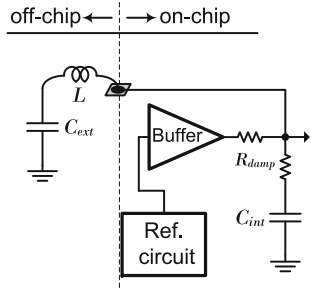
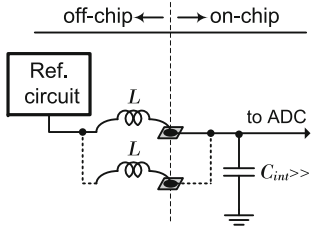
(Continued)

Table 2.2 (Continued)

Description	Analysis	Diagram	Performance
On-chip buffer with external decoupling [22, 31, 69, 109, 155]	• Low bandwidth buffer. • Low power consumption. • Noise performance and output impedance dependent on quality (ESL and ESR) of decoupling capacitors. • Wirebond inductance causes ringing of the internally generated reference voltage. Dampen ringing with large on-chip capacitors and resistors (these occupy large area). • Reduce inductance with special packaging [69]. • Additional pins.		• Accuracy +– • Power + • Speed – • On-chip area +– • Off-chip area –
On-chip reference without buffer with external decoupling [68]	• Reference is taken off-chip, RC filtered and dampened and brought on-chip again. • Two wirebonds in reference voltage path. • Special packaging unavoidable. • Two additional pins.		• Accuracy +– • Power + • Speed – • On-chip area + • Off-chip area –

(Continued)

Table 2.2 (Continued)

Description	Analysis	Diagram	Performance
On-chip buffer with external and internal decoupling [117]	• Use of external and internal capacitors and damping resistors. • Special internal decoupling scheme (low-V_T decoupling capacitors) [117]. • Additional pins.		• Accuracy +− • Power + • Speed +− • On-chip area − • Off-chip area −
Off-chip reference with internal decoupling [21, 88, 171]	• Use large on-chip capacitors to dampen the ringing caused by wirebond (occupying large area). • Internal decoupling capacitor may be larger than converter itself [171]. • For lower inductance more pins must be used [21]. • Additional pins.		• Accuracy − • Power + • Speed − • On-chip area − • Off-chip area −

Special packaging: chip-scale flip-chip with < 0.2 nH wirebond inductances

Parameters of reference circuits and buffers that may affect the overall performance of data converters are given next. They are:

- **Ringing** is caused by the inductance (L) of the wirebond and converter's capacitance (C) during switching in SC circuits. The ringing occurs at the natural frequency of the LC circuit. To reduce and dampen the ringing, large on- and/or off-chip decoupling capacitors and damping resistors must be used [53, 88, 113, 171].
- **Speed** of a reference circuit or buffer is determinant in the presence of perturbations [141]. During clocking of the capacitors in SC circuits, the reference voltage is constantly disturbed, and has to recover before the end of the phase to avoid incomplete reference settling which causes offset errors. This error may limit the converter's conversion rate.
- **Output impedance**: It is fundamental that the output impedance of the buffer circuit be as low as possible, to adequately feed the converter with the reference voltage, avoiding large voltage drops. To achieve low output impedance at high frequencies, large capacitors need to be used, which consequently occupy a large amount of area. Nevertheless, large capacitors help reduce the noise bandwidth and suppress external disturbances, if large enough [141].
- **Noise**: Directly couples to the stage's input signal during sampling and also to the output signals during the amplification phase. Because noise limits the smallest signal that may be converted, it directly influences the full-scale range of the converter [113].
- **PSR**: Power supply rejection determines the capacity of the buffer to reject noise of the supply (V_{DD} and ground) lines from coupling to the output buffered voltage. If noise couples to the output voltage, it contributes to the total output noise of the buffer.
- **Offset** errors limit the full-scale range of the converter. Negative offsets (lower reference voltage) saturate the residue voltage, thus reducing the conversion range, which causes distortion. Positive offsets (higher reference voltage) cause residue voltages to be smaller thus degrading the signal-to-noise ratio (SNR) of the converter [113]. These offset errors cause interstage gain errors [155] and overall converter gain errors.
- **Signal dependency**: When a reference voltage and input signal are connected to the capacitor's plates, any variation in one signal causes a modulation in the other. This is called signal-dependent modulation. Therefore, if a reference voltage does not settle adequately, it induces a variation in the input signal, which is then sampled. This modulation degrades the performance of the converter [113, 155].
- **Reference distribution** is not an easy task, particularly when references must be provided to widely separated locations across the die [142], or in converters with high clock rates or high resolutions [68]. The long lines cause voltage drops which alter the reference's original value. If each pipeline stage uses a reference voltage with a different value, the performance of the converter is highly degraded.

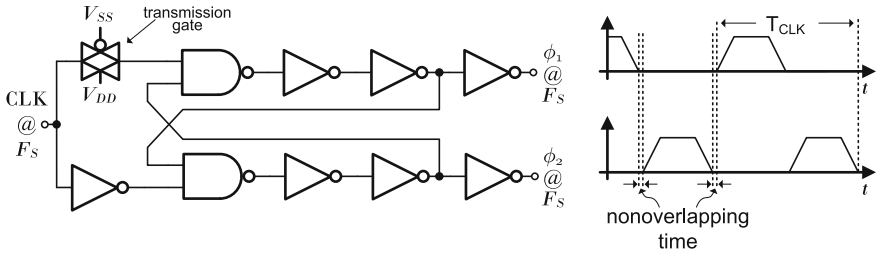


Fig. 2.13 Simple two-phase nonoverlapping clock generator and output waveforms

2.2.6 Clock Generation

Switched-capacitor circuits need clocking schemes to turn on/off its switches to achieve a specific function. Most data converters rely on a special scheme that permits the signal acquired during the sampling phase to be transferred and amplified without loss. In other words, it is imperative that no charge is destroyed or lost between the sampling and amplification phases. To achieve this, the overlapping time between phases must be zero. This is accomplished by a nonoverlapping clock generator. An example of a simple two-phase nonoverlapping clock generator, widely found and used in the literature, is shown in Fig. 2.13 [105]. This simple configuration has one clock input and two phases with 180° phase shift are provided at the output. The nonoverlapping time is controlled by the delay of the input NAND gate and the two inverters (before the feedback). To increase this time, more inverters (in an even number) can be added.

There are other schemes of generating nonoverlapped phases, namely using phase-locked loops [82] or delay-locked loops [37]. These circuits are much more complex than the simple two-phase nonoverlapping clock generator of Fig. 2.13. They are used in many applications such as in communication, data and clock recovery, FM demodulation, and high-speed or high-resolution data converters, among others. Given that they were not used in the work developed in this book, they will not be further discussed.

Clock generation or clock synthesis circuits limit the performance of data converters due to PVT variations, but mainly due to device and delay mismatches. These circuits contribute the following errors:

- **Clock jitter** is the random variation of the time the clock falls to zero (the falling edge is used here to define the sampling instant) around the ideal time. This error was briefly covered in the S/H subsection. The jittering of the timebase translates into an increase of the noise floor over all frequencies, and consequently, degrades the SNR of the converter. As mentioned before, this error is input signal dependent.
- **Clock skew** is a fundamental limitation inherent to time-interleaved converters. Each unit converter of a time-interleaved topology must sample the input at equally spaced instants in time, given by the M/F_S, where M is the number of unit converters and F_S is the total sampling frequency. Any deviation from these ideally

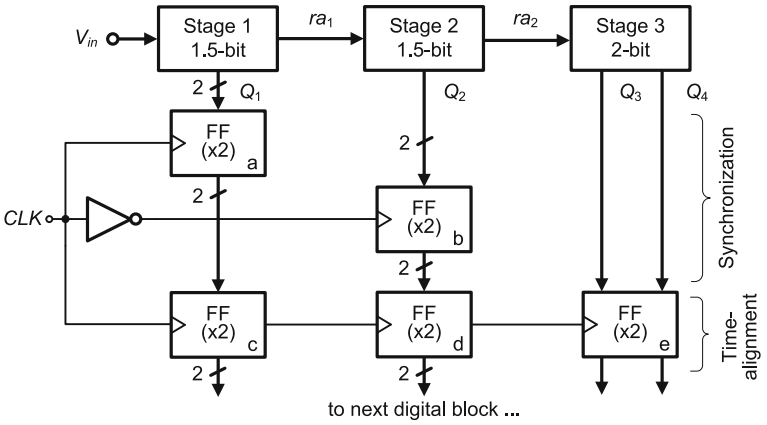
spaced sampling instants causes timing errors, which translate into deterministic spurious tones (visible in the converter's output spectrum), ultimately degrading the SNR of the converter. This error is also input signal dependent and was covered in Sect. 2.1.4 (timing mismatch).

2.2.7 Digital Backend and Decimation

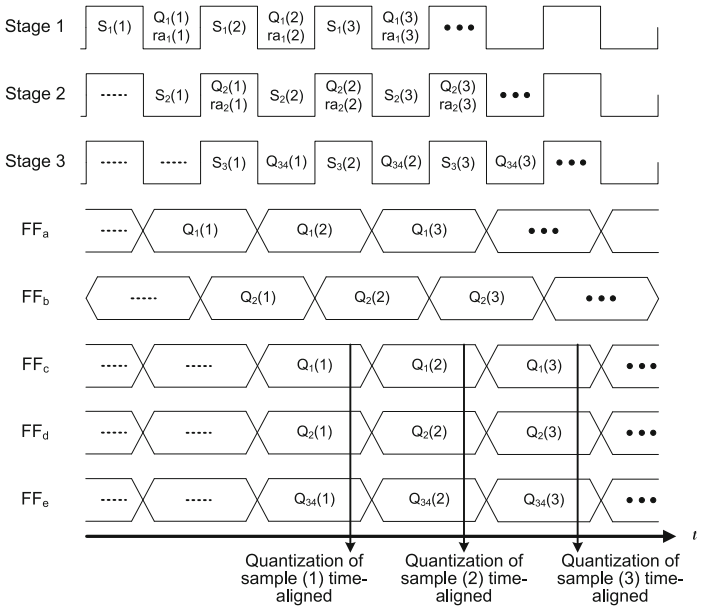
As mentioned in Sect. 2.1 for the various topologies, a backend only composed of digital circuits is necessary. These circuits store and align the digital outputs from each stage, which then move on to the correction stage, and are finally buffered to produce the final digital output word. Basically three digital blocks are necessary: synchronization logic, digital correction logic, and output buffers. The latter block (which is basically an even number of cascaded inverters) is optional and depends on where the digital outputs need to go next and if they need to be buffered or not. To exemplify the functionality of the circuits described in this subsection, a simple 4-bit Pipeline converter with two 1.5-bit stages and a final 2-bit stage will be used.

Synchronization logic is fundamental in converters that are unable to produce a digital word in one clock cycle, i.e., are unable to digitize the analog input at once (like Full-Flash converters). This logic is mainly composed of memory and shift circuits, such as flip-flops (FF) or shift registers. Taking the example of Fig. 2.14, when the analog input signal is processed by the first stage, its comparator produces digital outputs which need to be stored because the residue voltage will only be processed by the last stage of the converter at the end of one more cycle. Therefore, during this clock cycle, the bits from the first stage need to be stored. The same analysis can be made for the outputs of the second stage, which need to be stored for half a clock cycle. As can be seen in Fig. 2.14a, FF_a and FF_b synchronize the digital outputs, while FF_{c-e} align the outputs of a specific input sample for later processing. Figure 2.14b shows the timing diagram of operations and time-alignment, i.e., where the digital outputs of each of the sampled inputs are aligned at a specific instance in time. After all digital outputs are time-aligned they are ready for digital correction.

Digital correction logic is used to correct nonidealities and indecisions in the comparators used in the quantizers [97, 98]. Therefore, the quantizer's offsets can be as large as $\pm 1/2 V_{LSB} = V_{REF}/4$ for a 1.5-bit stage, which relaxes the specifications of the quantizer. This correction logic works on the basis that each stage has redundancy which is used for correction and eliminated by the correction logic itself. For example, a 1.5-bit stage has a true resolution of 1-bit and 0.5-bit redundancy. It has two digital output bits as shown by the input-output characteristic of Fig. 2.8a, where possible digital outputs are 00, 10, and 01. The 00 output indicates that the sampled input is certainly negative, while the 10 output indicates that it is certainly positive. The 01 output indicates indecision, i.e., the quantizer does not know if the sampled input is either positive or negative [105]. This decision is postponed to subsequent stages and the correction logic (with the digital outputs of these stages) will correct this indecision. All this is only possible if the quantizer's offsets are less than $\pm 1/2 V_{LSB}$.



(a)



(b)

Fig. 2.14 Simple example of synchronization: **a** Pipeline converter, synchronization and time-alignment logic. **b** Time scheme of operations

An example of how correction logic operates is shown in Fig. 2.15. In Fig. 2.15a, an ideal situation is depicted, i.e., the comparators do not have offsets. In Fig. 2.15b, one of the comparators of stage 2 has a positive offset (indicated by the curly arrow). Even though there is an offset, the final output is the same as in the situation with no offset. This example shows how digital correction corrects for nonidealities in the

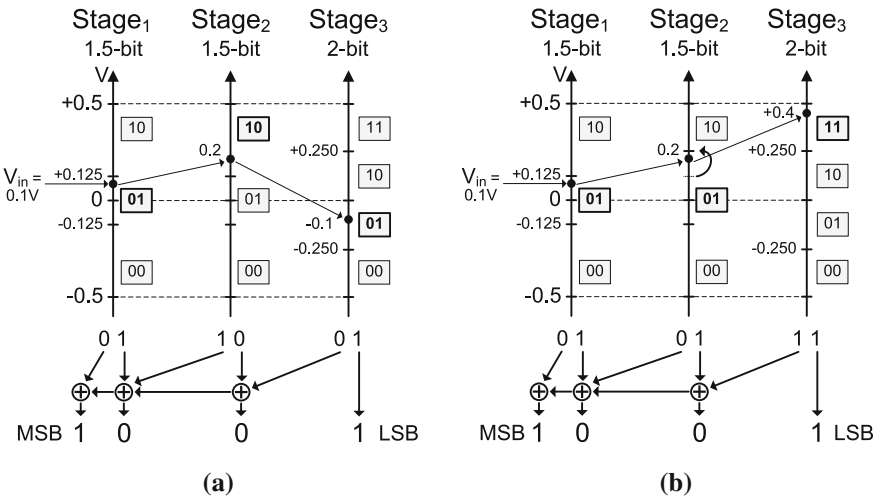


Fig. 2.15 Example of the operation of digital correction: **a** Ideal situation (and indecision corrected). **b** Offset error in stage 2: offset corrected

quantizers. Referring back to the example of Fig. 2.15a to demonstrate how digital correction also corrects indecisions. It is shown that the indecision of stage 1, i.e., stage 1 can not determine if the input is positive or negative (hence the 01 output), is corrected by the outputs of stages 2 and 3. This can be verified because the final result of the MSB (the sign bit of the digital output word) is 1, which indicates that the input signal is positive ($V_{in} = +0.1\,\text{V}$).

In high-speed converters, the digital output buffers and digital output pads are one of the main contributors to the increase in substrate (ground) noise. The constant switching, charging, and discharging of the output nodes, couple undesired digital noise into the substrate which affects the normal operation of the analog circuits. In order to reduce the digital noise, the speed of operation (clocking frequency) needs to be reduced. So, instead of acquiring all the outputs (at the maximum clock frequency), if only one output in every N is acquired, no information is lost and a lower clock frequency (now divided by N) permits reducing the digital noise. This technique is called decimation by a factor of N, and is explained more thoroughly further on in the book.

2.3 Performance Metrics of Analog-to-Digital Converters

It will be the objective of this section to present metrics and parameters that characterize the performance of ADCs (in general). These are divided into two groups: static and dynamic parameters. Static parameters discussed here will be offset, gain, differential nonlinearity (DNL) and integral nonlinearity (INL). These are usually

measured using DC, ramp, or low frequency signals. As for dynamic parameters, these are usually measured with high frequency signals, which stimulate these parameters. ADC dynamic performance metrics are signal-to-noise ratio (SNR), total harmonic distortion (THD), spurious free dynamic range (SFDR), signal-to-noise-and-distortion ratio (SNDR) and effective number of bits (ENOB). Although noise (represented by the SNR) does not depend on the input signal's frequency,[5] it will be discussed in the group of dynamic parameters.

2.3.1 Static Performance Parameters

The static performance of an ADC can be evaluated by its input-output conversion characteristic. An ideal 3-bit ADC situation is shown in Fig. 2.16, where the x-axis represents the analog input (normalized to the reference voltage, X_{REF},[6] or in LSBs) and the y-axis represents the quantization levels (D_{out}). As can be seen the ideal characteristic has a staircase waveform, where the width of each step is 1 LSB except for the first and last steps. The transition levels are taken in the middle of each analog input interval. The quantization (or amplitude discretisation) of an analog voltage, which intrinsically has infinite levels of quantization (zero error), by an ADC in eight quantization levels introduces an error. Graphically, this error is given by the difference between the staircase and the midpoint interpolating line (dashed line of the top graph of Fig. 2.16) and should ideally be limited between $\pm 1/2$ LSB as shown at the bottom of Fig. 2.16. This quantization error translates into an additive noise, i.e., the quantization noise. More on this noise is given in the dynamic performance parameters subsection.

With the understanding of Fig. 2.16, it is possible to describe most A/D converter static performance parameters.

2.3.1.1 Offset Error

Offset error is the horizontal difference between the first transition level of the real and the ideal ADC, as shown in Fig. 2.17a. It describes a shift for an analog input of 0 V, and graphically, the conversion characteristic is shifted horizontally. An expression for the offset error is given by

$$E_{offset} = \frac{x_{T_1} - x_{T_{1_{ideal}}}}{\Delta} \quad \text{(LSB)}, \tag{2.2}$$

[5] Intrinsic ADC noise is related with quantization and thermal noise. Other types of noise, such as, jitter and substrate noise (due to digital switching) are partially and often considered extrinsic to the A/D converter.

[6] X and x are used because they represent either a voltage or a current.

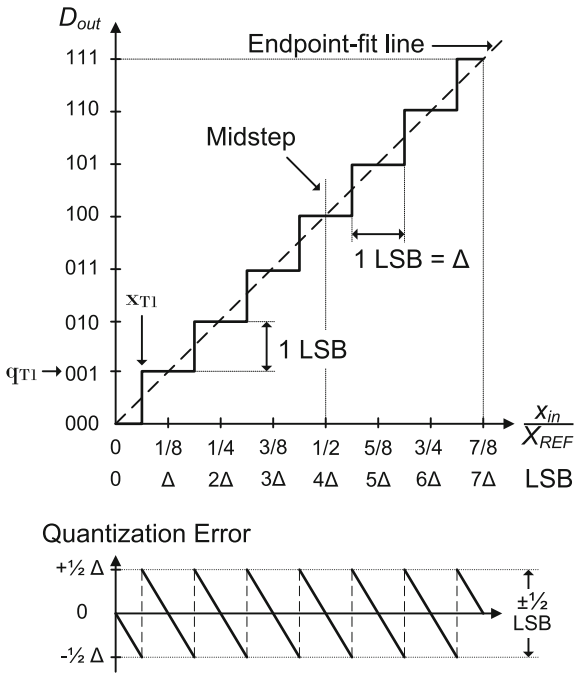


Fig. 2.16 Input–output conversion characteristic and quantization error of an ideal 3-bit A/D converter

where $\Delta = x_{LSB} = x_{REF}/2^N$ and N is the resolution of the ADC.

2.3.1.2 Gain Error

Gain error is the slope difference of the midpoint interpolating line of the real and ideal characteristics, as shown in Fig. 2.17b. For converter gains <1 and offset errors the output range of the ADC is limited. For gains > 1, a range of inputs have the same output (depicted in Fig. 2.17b). For the ideal case, this slope is unity. An expression for the gain error is given by

$$E_{gain} = \frac{q_{T_{2^N-1}} - q_{T_1}}{x_{T_{2^N-1}} - x_{T_1}} - 1 \quad \text{(LSB)}. \tag{2.3}$$

2.3.1.3 Differential Nonlinearity

In the ideal characteristic of Fig. 2.16, the horizontal difference between two consecutive transitions is exactly 1 LSB. In the characteristic of a real converter, any deviation from 1 LSB, between consecutive transitions, causes a differential

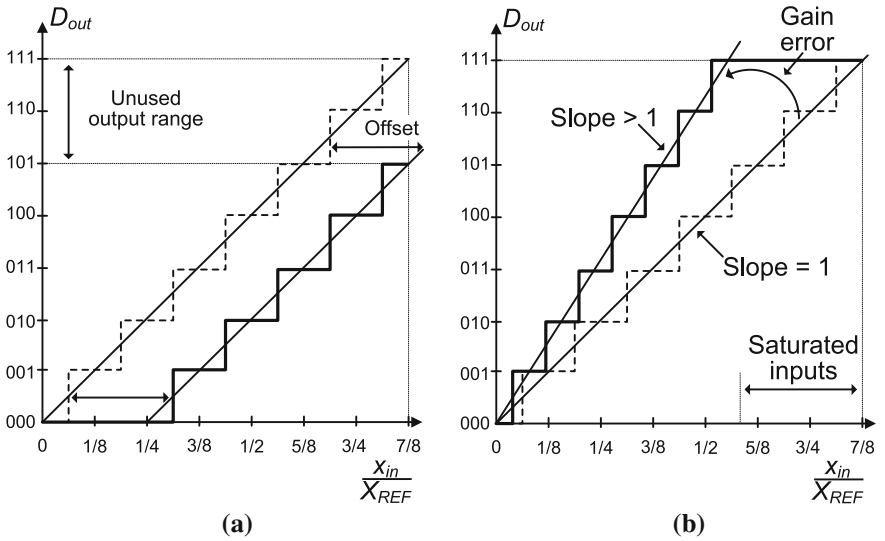


Fig. 2.17 Static A/D converter errors: **a** Offset error. **b** Gain error. Ideal characteristic represented by *dashed line*

nonlinearity (DNL) error. Figure 2.18 exemplifies some DNL errors. In an expression, the DNL can be given by,

$$\text{DNL}(i) = \frac{x_{T_{i+1}} - x_{T_i}}{\Delta} - 1 \quad \text{(LSB)}, \quad i = 1, \ldots, 2^N - 2. \tag{2.4}$$

The DNL is usually taken as the max ($|\text{DNL}(i)|$) for all i. Before determining the DNL for each code, offset and gain must be removed. This can be achieved using the endpoint-fit line, which creates a straight line from the first (origin) to the last (full-scale) code. The DNL profile is usually characterized by a graph with the digital output codes ($1 \ldots 2^N - 2$) for the x-axis and Eq. 2.4 for the y-axis. An example of this is shown at the bottom of Fig. 2.18a.

Observing Eq. 2.4, it is worth noting some special cases:

- $\text{DNL}(i) = 0$: two consecutive transitions are equal to 1 LSB. Also true for the first and last codes, due to the endpoint-fit line system.
- $\text{DNL}(i) = -1$: two consecutive transitions are equal, which means there is a missing quantization level, or missing code.
- $\text{DNL}(i) \geq +1$: two consecutive transitions are larger than 1 LSB. High probability of the existence of missing codes in the DNL profile [9].

The DNL characteristic provides information about the converter's behaviour code by code. This means that the way in which each output code is encoded (the encoding process), has an effect on the converter's linearity [77]. The encoding process depends on the DAC architecture (capacitor matching) used in the MDAC

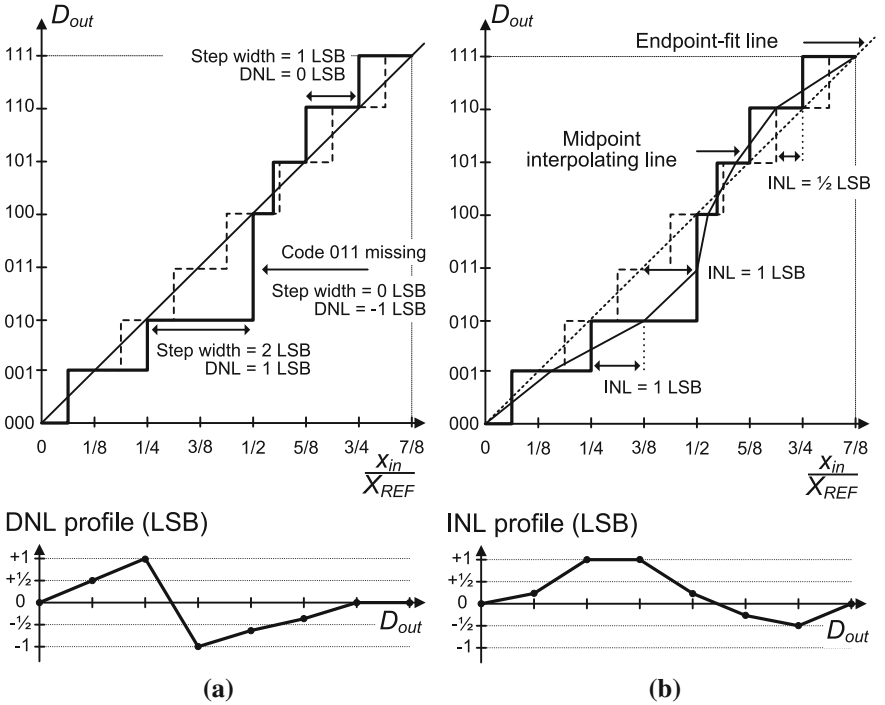


Fig. 2.18 Static A/D converter errors: **a** Differential nonlinearity (DNL). **b** Integral nonlinearity (INL). Ideal characteristic represented by the *dashed line*

circuit, and each architecture must be analysed separately [170]. Besides nonlinearity errors contributed by the MDAC circuits, the local stage quantizers also introduce errors with offsets. Fortunately, the latter can be corrected by the digital correction logic. Given that DNL errors cause the quantization characteristic to be different from the ideal one and quantization errors introduce additive noise, then DNL errors will also translate into an additive noise component called DNL noise, degrading the converter's SNR [77, 105].

2.3.1.4 Integral Nonlinearity

Integral Nonlinearity (INL) is a measure of the horizontal difference between each transition and its corresponding ideal transition. Using the endpoint-fit line to remove gain and offset errors, the INL becomes the difference from the midpoint interpolating line to the endpoint-fit line (straight line that connects first and last transition), as shown in Fig. 2.18b. INL can be defined by the following expression,

$$\mathrm{INL}(i) = \frac{x_{T_i} - \Delta(i-1) - x_{T_1}}{\Delta} \quad \text{(LSB)}, \quad i = 1, \ldots, 2^N - 1. \tag{2.5}$$

The INL can also be shown to be the cumulative sum of the DNL, given by,

$$\mathrm{INL}(i) = \sum_{j=1}^{i-1} \mathrm{DNL}(j), \quad i = 2, \ldots, 2^N - 1. \tag{2.6}$$

The INL is usually taken as the max ($|\mathrm{INL}(i)|$) for all i. Like the DNL profile, the INL profile is characterized by a graph with the digital output codes ($1 \,..\, 2^N - 1$) for the x-axis and Eq. 2.5 for the y-axis. INL errors are caused by capacitor mismatch in the DAC circuit of the MDAC and by finite gain of the opamp. There is a relationship between INL errors and harmonic distortion [42, 105, 130]. Therefore, a large INL indicates a large deviation of the conversion characteristic to the ideal one and could be an indication of a large amount of distortion.

2.3.2 Dynamic Performance Parameters

The dynamic performance of an ADC is usually characterised in the frequency domain, accomplished with the fast fourier transform (FFT). The FFT is a widely used algorithm to perform a Fourier transform on the output time data from the ADC. A simplified measuring process is as follows: a sinusoidal wave is used as the input to the ADC, which produces a quantized output, that is then fed to the FFT algorithm and, finally, a spectrum of the quantized input signal is produced. The output spectrum permits measuring all the dynamic parameters that will be described here. To aid the enumeration and description of these parameters, a hypothetical FFT of the output of a two-channel time-interleaved ADC (some errors exaggerated) shown in Fig. 2.19 will be used. For the equations presented below, the band of interest is considered the Nyquist bandwidth ($F_{S/2}$).

2.3.2.1 Signal-to-Noise Ratio

As the name indicates it is the ratio of the signal power to the noise power. Observing Fig. 2.19, the noise power excludes DC, signal, and harmonic components, but includes the spurious tones due to time-interleaving mismatches. The theoretical maximum signal-to-noise ratio (SNR) that an ADC can achieve is (only taking into account quantization noise),

$$\mathrm{SNR} = 6.02N + 1.76 \quad (\mathrm{dB}), \tag{2.7}$$

where N is the resolution of the converter. As already stated throughout this chapter, there are other noise sources which contribute to the total noise of the ADC (further degrading its SNR), which are clock jitter, DNL errors, and thermal noise. A more complete expression becomes,

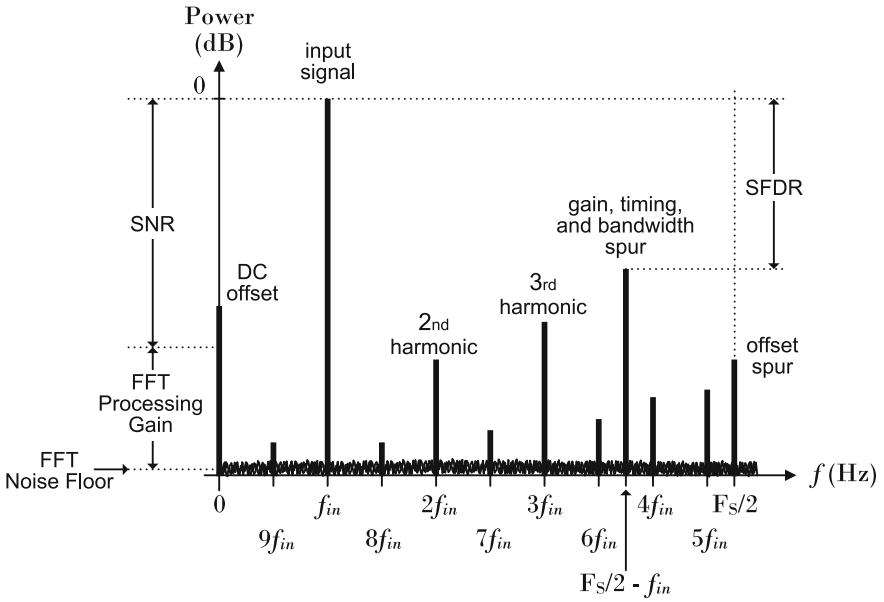


Fig. 2.19 Example of an FFT of a two-channel time-interleaved ADC

$$\text{SNR} = \frac{P_S}{P_j + P_{q+DNL} + P_{th}}, \tag{2.8}$$

where P_j, P_{q+DNL}, and P_{th} are the jitter, quantization plus DNL, and thermal noise power, respectively. Therefore, the overall SNR is input signal dependent (due to extrinsic effects): high signal frequency increases jitter and small signal amplitude reduces signal power. To obtain the result in decibels (dB), 10 log of the respective equation should be taken.

As shown in Fig. 2.19, the noise floor is situated such that each noise bin (discrete line) is, on average, below the full-scale by $6.02N+1.76$ plus the FFT processing gain (given by $10\log(n_{points}/2)$, where n_{points} is the number of points used to compute the FFT). This processing gain is particularly helpful when trying to distinguish harmonics from noise. The more points in the FFT, the lower the noise floor will be, but the harmonics will stay at their original magnitude.

2.3.2.2 Total Harmonic Distortion

When an input is quantized by a nonideal ADC, tones appear, in the output spectrum, at multiples of the signal's frequency. These tones are called harmonics, and the total harmonic distortion (THD) measures the ratio of the sum of the harmonics' power, P_H, to the signal's power (P_S). In an expression this is given by

$$\text{THD} = \frac{\sum_{i=2}^{h} P_H(i)}{P_S}, \tag{2.9}$$

where h represents the number of harmonics. THD is highly dependent on the input signal. At high frequencies and large amplitudes, distortion becomes more pronounced.

2.3.2.3 Spurious-Free Dynamic Range

This parameter measures the ratio between the signal power and the largest magnitude of any spectral component (excluding the DC component). This spectral component can be a harmonic of the input signal or a spurious tone, and is given by,

$$\text{SFDR} = \frac{P_S}{\max(P_{spectrum}(f))}, \quad f \in \{1, \ldots, F_S/2\} \backslash \{f_{in}\}, \tag{2.10}$$

where $P_{spectrum}(f)$ represents all spectral components except the DC ($f = 0$) and the signal components ($f = f_{in}$). At high input frequencies or large amplitudes, the limiting tone will probably be a harmonic, whereas at low amplitudes a spurious tone could limit the SFDR.

2.3.2.4 Signal-to-Noise-and-Distortion Ratio

This parameter is a complete indication of the overall dynamic performance of the converter. It is complete in the sense that it combines all performance degradation elements, such as, the noise sources of the SNR and the distortion components of the THD. In an expression the SNDR can be given by

$$\text{SNDR} = \frac{P_S}{\sum_{i=2}^{h} P_H(i) + P_j + P_{q+DNL} + P_{th}} = 10^{\frac{\text{THD}}{10}} + 10^{\frac{-\text{SNR}}{10}}. \tag{2.11}$$

2.3.2.5 Effective Number of Bits

Given that the SNDR of Eq. 2.11 will be less than the theoretical SNR limit given by Eq. 2.7, it becomes important to define a real resolution for the converter. This real resolution is known as the effective number of bits (ENOB) and an expression for it can be obtained by solving Eq. 2.7 for N. This is given by

$$\text{ENOB} = \frac{\text{SNDR} - 1.76}{6.02} \quad \text{bits}. \tag{2.12}$$

2.4 Overview and Comparison of Published Work

This section presents an overview of the state-of-the-art concerning the work carried out in this book. Given that this work presents results from silicon prototypes of two different circuits, namely an opamp and an ADC, an overview of published work related with these circuits is presented.

Besides the overviews of these two circuits, a review of published data concerning ADC reference voltage circuits will also be given. The objective of this review is to obtain some insight and criteria about the power dissipated and area occupied by these circuits in the context of A/D conversion. This review is fundamental because the ADC described in this book precludes all reference voltage circuits, therefore saving power and area. To be able to quantify, in average, the power and area saved by the proposed techniques, the results from this review will become useful.

2.4.1 Two-Stage Opamps

The following overview, summarized in Table 2.3, concerns class-A and class-AB two-stage opamps from the past ten years, simulated or fabricated in a CMOS technology with GBW > 30 MHz. In order to evaluate and compare the performance of the various opamps, a figure-of-merit (FoM) will be used [118]. This FoM basically evaluates the speed to current consumption ratio. The better the opamp, the smaller the FoM. However, this becomes a problem because the values of the FoM become too small, which in turn, makes it difficult to comprehend and compare, so the inverse of the FoM of [118] will be used instead, which is given by

$$\text{FoM}_{\text{OA}} = \frac{\text{GBW} \cdot C_L}{\text{Power}} \qquad [\text{MHz} \cdot \text{pF/mW}]. \tag{2.13}$$

Regarding Table 2.3, besides the technology and the supply voltage used, various AC and transient response performance parameters are given. The settling time and settling error are represented by T_S and $T_{S\varepsilon}$, respectively. Figure 2.20 depicts the FoM_{OA} of the opamps of Table 2.3 versus their respective GBW. A linear interpolation line (dashed line) in Fig. 2.20 demonstrates the average value of FoM_{OA}(GBW). The negative slope of this line, although small, demonstrates the difficulty in achieving high GBW with high energy efficiency. This challenge is well observed by the marked ([a], [b], and [c]) opamps. Opamp [b] [149] represents a very high GBW opamp with a given FoM_{OA}, while opamp [a] [73] achieves more than two times the FoM_{OA} but at the cost of a much lower GBW. The most energy efficient opamp ([c]) is reported in [73], and achieves a FoM_{OA} of 1213 MHz·pF/mW by employing a dynamic threshold technique and controlling the bulk voltage of critical transistors.

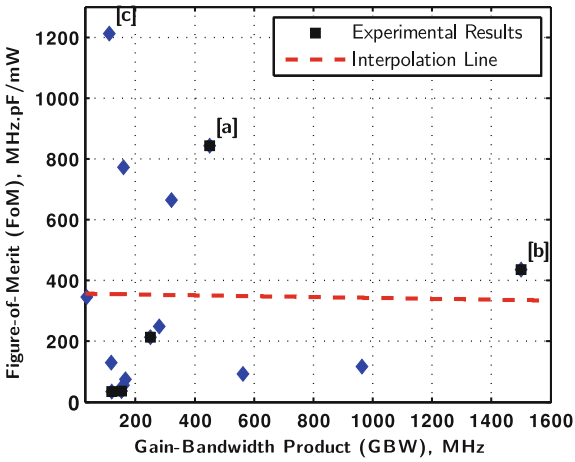


Fig. 2.20 State-of-the-art two-stage opamps with GBW > 30 MHz designed in CMOS technologies

Table 2.3 Overview of two-stage opamps from the past ten years with GBW > 30 MHz in CMOS technologies

Ref.	Tech. [μm]	V_{DD} [V]	DC Gain [dB]	$T_{S\varepsilon}$ [%]	T_S [ns@V_{pp}]	C_L [pF]	Power [mW]	GBW [MHz]	FoM_{OA} $\left[\frac{MHz \cdot pF}{mW}\right]$
[73]	0.13	0.5	51.0	N/A	N/A	6.5	0.6	112	**1213**
[159][a]	0.065	1.0	56.1	1	8@0.5	3	1.6	450	**844**
[132]	0.18	1.8	74	N/A	N/A	1.75	0.362	160	**773**
[147]	0.13	1.2	77	0.1	3.8@0.1	4	1.94	322	**664**
[149][a]	0.12	1.2	40.4	0.1	9.2@0.1	3.2	11	1500	**436**
[2]	0.18	1.0	64.4	N/A	N/A	5	0.522	36	**345**
[160]	0.25	1.5	78	0.24	7.2@2	8	9	280	**249**
[162][a]	0.5	2.5	90	1	15@0.7	12	14	250	**214**
[184]	0.25	1.5	69.5	0.01	20.1@0.5	6	5.5	119	**130**
[182]	0.25	2.5	72	1	2.1@0.6	3	25	963	**116**
[152]	0.5	1.8	75	N/A	N/A	1.8	11	562	**92**
[183]	0.25	1.5	80	0.01	7.1@0.5	4	8.9	167	**75**
[143]	0.13	1.2	68.7	N/A	N/A	1.57	4.51	160	**56**
[179][a]	0.5	3.0	84	0.1	47@1.2	10	43	153	**36**
[72][a]	0.5	3.5	90	0.1	10@1	1	3.5	120	**34**

[a]Experimental results

2.4.2 Medium-Low Resolution High-Speed MDAC-Based ADCs

The following state-of-the-art concerns medium-low resolution high-speed MDAC-based ADCs. The resolutions chosen are 6–8 bit, with ENOB > 5 bits and sampling frequencies, F_S > 200 MS/s. Only the MDAC-based architectures are chosen for a more fair comparison with the prototyped ADC described, as well as, the proposed techniques. In order to evaluate and compare the performance of the various ADCs, it

Table 2.4 Overview of 6–8 bit MDAC-based ADCs with ENOB > 5 bits and $F_S > 200$ MS/s designed in a CMOS technology

Ref.	Tech. [μm]	Resolution [bits]	F_S [MS/s]	ENOB [bits]	Power [mW]	Area [mm^2]	FoM_{ADC} $\left[\frac{\text{fJ}}{\text{conv.-step}}\right]$
[79]	0.09	8	320	7.3	12.8	0.53	**253**
[168]	0.065	8	800	7	30	0.12	**283**
[15]	0.18	8	200	6.4	8.5	0.05	**503**
[90]	0.18	8	200	7.7	30	0.15	**731**
[83]	0.18	8	200	7	22	0.32	**830**
[26]	0.13	6	1000	5.3	49	0.16	**1240**
[75]	0.09	7	550	5.7	60	0.37	**2045**
[75]	0.09	7	1100	5.7	92	0.37	**2206**
[134]	0.09	8	250	6.2	22.8	0.81	**2580**
[115]	0.09	6	10300	5.1	1600	N/A	**4699**
[136]	0.35	8	4000	6.1	4600	28.85	**33531**

is necessary to use an appropriate FoM. Although highly contested, the FoM mostly used and found in the literature is the Walden FoM [173] (or its inverted form), given by

$$\text{FoM}_{\text{ADC}} = \frac{\text{Power}}{2^{\text{ENOB}} \cdot \min\{F_S, 2BW\}} \quad [\text{J/conv.} - \text{step}], \tag{2.14}$$

where $\min\{F_S, 2BW\}$ represents the minimum between the sampling frequency and two times the ADC's bandwidth (BW). This FoM weighs the ADCs overall performance given by its speed (F_S) and linearity (ENOB, which includes SNR and THD), to its total power consumption. Unlike the FoM_{OA}, the smaller the FoM_{ADC}, the more energy efficient an ADC is.

Table 2.4 presents the state-of-the-art of the aforementioned medium-low resolution high-speed MDAC-based ADCs. Besides showing ENOB, F_S, power consumption, and FoM_{ADC}, the table also shows each ADC's resolution and occupied area, as well as the employed technology. The table only contemplates ADCs with experimental results.

Figure 2.21 shows the results of the FoM of the ADCs of Table 2.4 plotted against their respective sampling frequency (Fig. 2.21a) and ENOB (Fig. 2.21b). The latter, plots the FoM against normalized ENOB because the state-of-the-art consists of ADCs with different resolutions. Therefore, the various ADCs can be compared in terms of their linearity. The interpolation line of Fig. 2.21a clearly demonstrates the difficulty in achieving a good FoM at high sampling frequencies.

In Fig. 2.21 the ADC marked as [a] [79] represents the converter with the best FoM, while [b] [115]) marks the one with the highest sampling frequency. These two ADCs exemplify that for increasing sampling frequencies, the FoM degrades. Note however that, they were designed with different objectives. Concerning ADC [a], it achieves its FoM by using a multilevel power optimization algorithm based on geometric

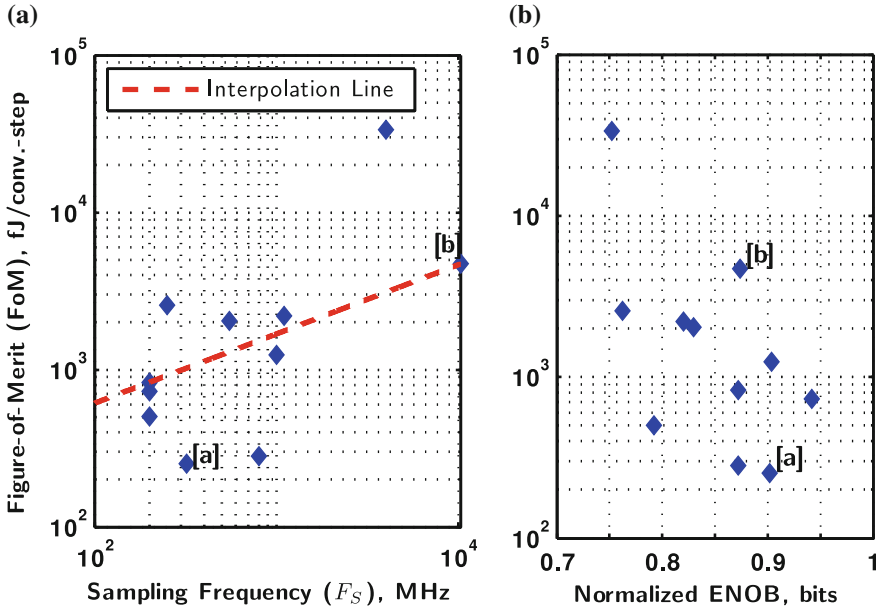


Fig. 2.21 The state-of-the-art of 6–8 bit MDAC-based ADCs with ENOB > 5 bits and F_S > 200 MS/s in CMOS technologies: **a** FoM versus F_S. **b** FoM versus normalized ENOB

programming. However, it needs two supply voltages, 1.2 and 2.1 V for digital and analog circuits, respectively. The total power of the ADC did not contemplate the power of the reference voltage buffers. ADC [b] uses digital calibration to correct for the nonidealities that arise from using open-loop amplifiers and mismatches between the interleaved channels. Besides calibration, each comparator uses a trimming circuit and each unit ADC is composed of two separate ADCs, one will be in operation while the other will be in calibration. Regarding calibration, only the ADCs with a high number of interleaved channels, namely, [115, 136] use digital calibration. What concerns reference voltage circuit power, only [79] indicates this power (but does not include it in the overall ADC power), while all others have omitted this value or have not indicated if it is included in the total power. Moreover, adding the reference power indicated in [79] to its total ADC power, degrades its FoM from 253 to 365 fJ/conv.-step.

2.4.3 ADC Reference Voltage Circuitry

The objective of this subsection is to obtain some knowledge and criteria concerning reference voltage circuitry related to ADCs. Reference voltage circuitry can be understood as the circuits used to generate, buffer, and decouple a reference voltage. As already mentioned, reference voltage buffers are one of the most power

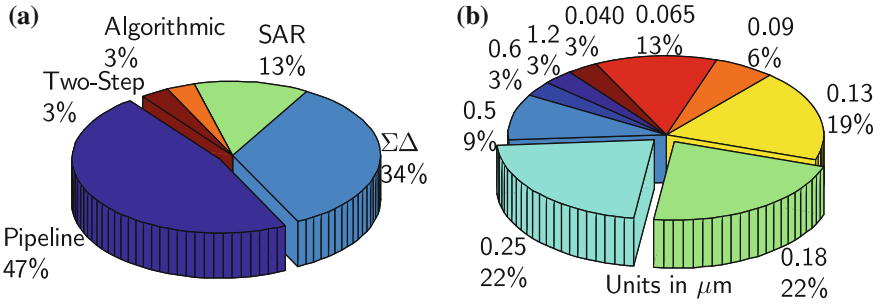


Fig. 2.22 Overview of the ADC reference voltage circuitry data set's characteristics: **a** Architectures. **b** Technology nodes

consuming blocks of ADCs. Besides consuming power, these circuits occupy a huge amount of silicon area due to the necessary decoupling capacitors (and damping resistors). In some cases these capacitors need to be so large and impractical to implement on-chip, they are placed off-chip, which brings other problems (explained in Sect. 2.2.5).

It is not an easy task compiling this overview because most articles found in the literature do not present the power consumption and area occupied by the reference voltage circuits. Nevertheless, the articles obtained that discriminate the power and area (of the various blocks that compose an ADC) will be enough to obtain some insight into this issue.

For this overview the experimental data from these references is used [4, 6, 8, 12–14, 16, 19, 20, 23, 24, 31–34, 36, 40, 53, 54, 70, 78, 81, 117, 125, 126, 129, 155, 157, 166, 168, 169, 187]. Figure 2.22 is used to characterize the data set for a better comprehension of the ADCs and their characteristics. Figure 2.22 depicts the architectures and the percentage of each architecture in the total population of the data set, while Fig. 2.22 shows the technologies used to implement the ADCs and their percentage of the total population. As mentioned before, the main objective of this overview is to obtain a rough estimate of the power and area of the reference voltage circuits compared to the ADC's overall power and area, respectively. From the mentioned data set, the percentage of the power consumed by reference voltage circuits (to the overall ADC power) is, in average, 29 %, while the area occupied by these circuits (to the overall ADC area) is, in average, 19 %. This area excludes off-chip decoupling capacitors.

From the data set it is possible to extract other interesting information, namely, the relationship between the reference circuit's power and: (a) its area, (b) the sampling frequency, and (c) the resolution of the ADC. These relationships are depicted in Fig. 2.23. The interpolation line of Fig. 2.23a clearly indicates a direct relationship between the power and area of the reference circuit. The interpolation line of Fig. 2.23b shows a smaller direct correlation with F_S, while Fig. 2.23c shows an even smaller relationship with the ADC's resolution.

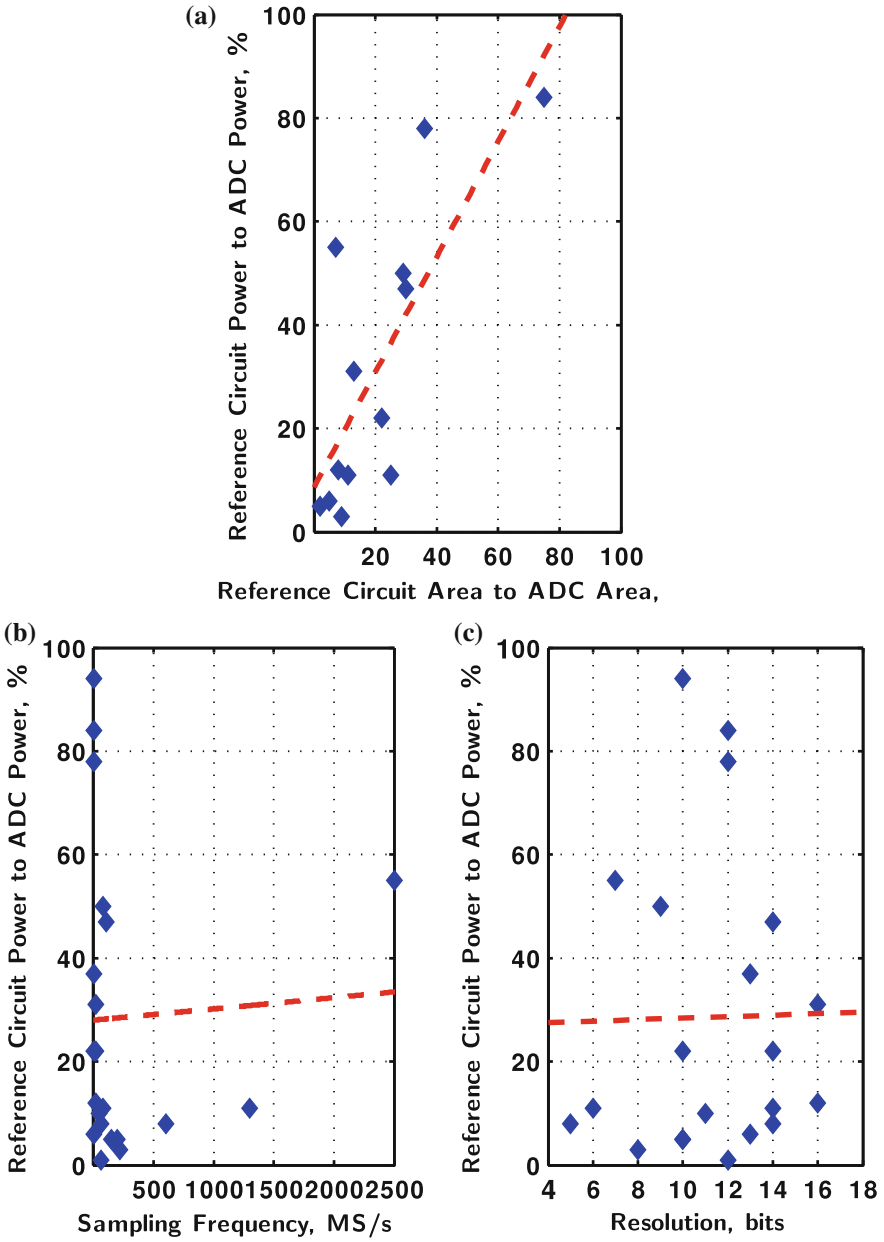


Fig. 2.23 Relationship between the reference circuit's power (in % of the total power) with: **a** Its area. **b** ADC's sampling frequency. **c** ADC's resolution. *Dashed line* represents the interpolation of the data points

Given such a small data set, it is important to note and remember that all these estimates and graphs correspond to a very rough representation regarding the power and area of reference voltage circuits in the context of A/D conversion.

This concludes the overview and comparison with published work. The following chapter presents the proposed MDAC circuits.

http://www.springer.com/978-1-4614-3466-5

Reference-Free CMOS Pipeline Analog-to-Digital Converters

Figueiredo, M.; Goes, J.; Evans, G.

2013, XVI, 184 p., Hardcover
ISBN: 978-1-4614-3466-5