

Chapter 2

Analog to Digital Conversion

2.1 High-Speed High-Resolution A/D Converter Architectural Choices

Since the existence of digital signal processing, A/D converters have been playing a very important role to interface analog and digital worlds. They perform the digitalization of analog signals at a fixed time period, which is generally specified by the application. The A/D conversion process involves sampling the applied analog input signal and quantizing it to its digital representation by comparing it to reference voltages before further signal processing in subsequent digital systems. Depending on how these functions are combined, different A/D converter architectures can be implemented with different requirements on each function. To implement power-optimized A/D converter functions, it is important to understand the performance limitations of each function before discussing system issues. In this section, the concept of the basic A/D conversion process and the fundamental limitation to the power dissipation of each key building block are presented.

2.1.1 Multi-Step A/D Converters

Parallel (Flash) A/D conversion is by far the fastest and conceptually simplest conversion process [1–21], where an analog input is applied to one side of a comparator circuit and the other side is connected to the proper level of reference from zero to full scale. The threshold levels are usually generated by resistively dividing one or more references into a series of equally spaced voltages, which are applied to one input of each comparator. For n -bit resolution, $2^n - 1$ comparators simultaneously evaluate the analog input and generate the digital output as a thermometer code. Since flash converter needs only one clock cycle per conversion, it is often the fastest converter. On the other hand, the resolution of flash ADCs is limited by circuit complexity, high power dissipation, and comparator and reference mismatch. Its complexity grows

exponentially as the resolution bit increases. Consequently, the power dissipation and the chip area increase exponentially with the resolution. The component-matching requirements also double for every additional bit, which limits the useful resolution of a flash converter to 8–10 bits. The impact of various detrimental effects on flash A/D converter design will be discussed further in [Section 3.4.1](#).

To reduce hardware complexity, power dissipation and die area, and to increase the resolution but to maintain high conversion rates, flash converters can be extended to a two-step/multi-step [22–39] or sub-ranging architecture [40–53] (also called series-parallel converter). Conceptually, these types of converters need $m \times 2^n$ instead of $2^{m \times n}$ comparators for a full flash implementation assuming n_1, n_2, \dots, n_m are all equal to n . However, the conversion in sub-range, two-step/multi-step ADC does not occur instantaneously like a flash ADC, and the input has to be held constant until the sub-quantizer finishes its conversion. Therefore, a sample-and-hold circuit is required to improve performance. The conversion process is split into two steps as shown in [Fig. 2.1](#). The first A/D sub-converter performs a coarse conversion of the input signal. A D/A converter is used to convert the digital output of the A/D sub-converter back into the analog domain. The output of the D/A converter is then subtracted from the analog input. The resulting signal, called the residue, is amplified and fed into a second A/D sub-converter which takes over the fine conversion to full resolution of the converter. The amplification between the two stages is not strictly necessary but is carried out nevertheless in most of the cases. With the help of this amplifying stage, the second A/D sub-converter can work with the same signal levels as the first one, and therefore has the same accuracy requirements. At the end of the conversion the digital outputs of both A/D sub-converters are summed up.

By using concurrent processing, the throughput of this architecture can sustain the same rate as a flash A/D converter. However, the converted outputs have a latency of two clock cycles due to the extra stage to reduce the number of precision comparators. If the system can tolerate the latency of the converted signal, a two-step converter is a lower power, smaller area alternative.

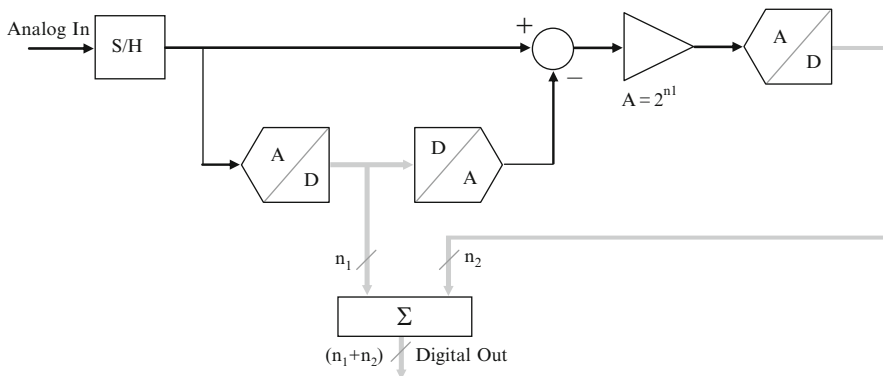


Fig. 2.1 Two-step A/D converter

2.1.2 Pipeline A/D Converters

The two-step architecture is equipped with a sample-and-hold (S/H) circuit in front of the converter (Fig. 2.1). This additional circuit is necessary because the input signal has to be kept constant until the entire conversion (coarse and fine) is completed. By adding a second S/H circuit between the two converter stages, the conversion speed of the two-step A/D converter can be significantly increased (Fig. 2.2). In a first clock cycle the input sample-and-hold circuit samples the analog input signal and holds the value until the first stage has finished its operation and the outputs of the subtraction circuit and the amplifier have settled. In the next clock cycle, the S/H circuit between the two stages holds the value of the amplified residue. Therefore, the second stage is able to operate on that residue independently of the first stage, which in turn can convert a new, more recent sample. The maximum sampling frequency of the pipelined two-step converter is determined by the settling time of the first stage only due to the independent operation of the two stages. To generate the digital output for one sample, the output of the first stage has to be delayed by one clock cycle by means of a shift register (SR) (Fig. 2.2). Although the sampling speed is increased by the pipelined operation, the delay between the sampling of the analog input and the output of the corresponding digital value is still two clock cycles. For most applications, however, latency does not play any role, only conversion speed is important. In all signal processing and telecommunications applications, the main delay is caused by digital signal processing, so a latency of even more than two clock cycles is not critical.

The architecture as described above is not limited to two stages. Because the inter-stage sample-and-hold circuit decouples the individual stages, there is no difference in conversion speed whether one single stage or an arbitrary number of stages follow the first one. This leads to the general pipelined A/D converter architecture, as depicted in Fig. 2.3 [54–89]. Each stage consists of an S/H, an N -bit flash A/D converter, a reconstruction D/A converter, a subtractor, and a residue amplifier. The

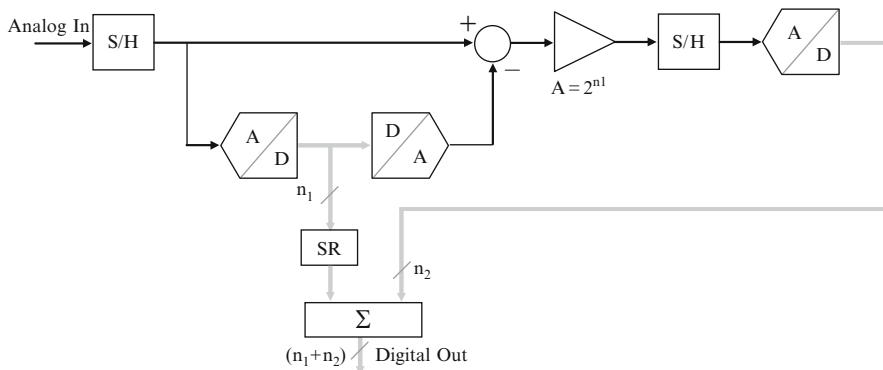


Fig. 2.2 Two-Step converter with an additional sample-and-hold circuit and a shift register (SR) to line up the stage output in time

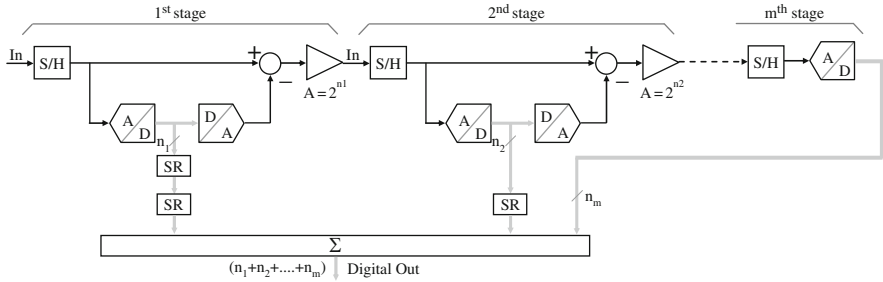


Fig. 2.3 Multi-stage pipeline A/D converter architecture

conversion mechanism is similar to that of sub-ranging conversion in each stage. Now the amplified residue is sampled by the next S/H, instead of being fed to the following stage.

All the n -bit digital outputs emerging from the quantizer are combined as a final code by using the proper number of delay registers, combination logic and digital error correction logic. Although this operation produces a latency corresponding to the sub-conversion stage before generating a valid output code, the conversion rate is determined by each stage's conversion time, which is dependent on the reconstruction D/A converter and residue amplifier settling time. The multi-stage pipeline structure combines the advantages of high throughput by flash converters with the low complexity, power dissipation, and input capacitance of sub-ranging/multi-step converters. The advantage of the pipelined A/D converter architecture over the two-step converter is the freedom in the choice of number of bits per stage. In principle, any number of bits per stage is possible, down to one single bit. It is even possible to implement a non-integer number of bits such as 1.5 bit per stage by omitting the top comparator of the flash A/D sub-converter used in the individual stages [59]. It is not necessary, although common, that the number of bits per stage is identical throughout the pipeline, but can be chosen individually for each stage [65–69]. The only real disadvantage of the pipelined architecture is the increased latency. For an A/D converter with m stages, the latency is m clock cycles. For architectures with a small number of bits per stage, the latency can thus be 10–14 clock cycles or even more.

2.1.3 Parallel Pipelined A/D Converters

The throughput rate can be increased further by using a parallel architecture [90–106] in a time-interleaved manner as shown in Fig. 2.4. The first converter channel processes the first input sample, the second converter channel the next one and so on until, after the last converter channel has processed its respective sample, the first converter has its turn again (see Section 3.1 for extensive discussion on timing-related issues in time-interleaved systems).

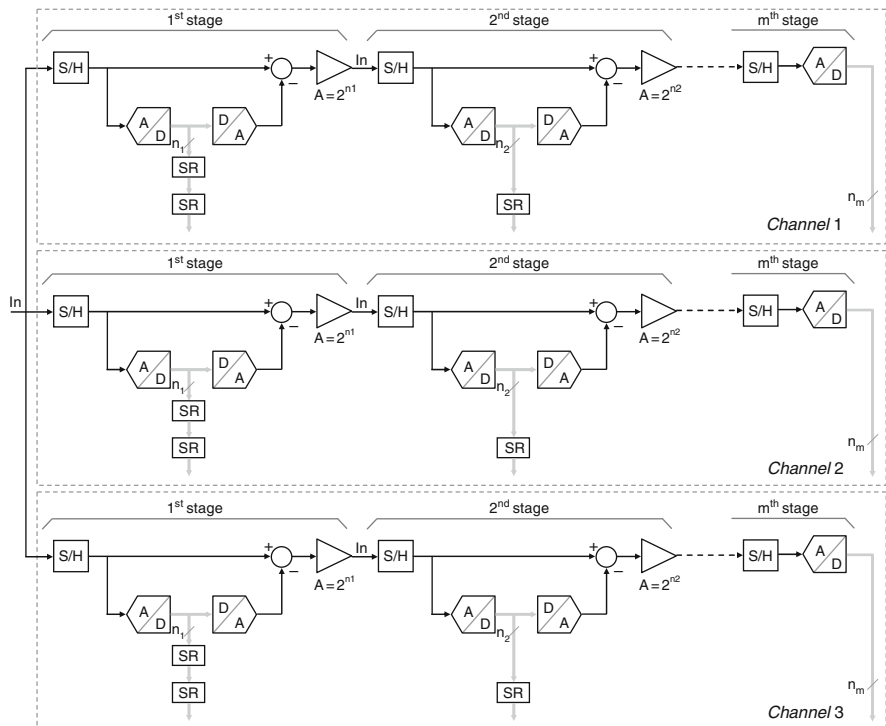


Fig. 2.4 Parallel pipeline A/D converter architecture

The individual A/D converters therefore operate on a much lower sampling rate than the entire converter, with the reduction in conversion speed for each individual converter equal to the number of A/D converters in parallel. The only building block that sees the full input signal bandwidth of the composite converter is the sample-and-hold circuit of each A/D converter. Theoretically, the conversion rate can be increased by the number of parallel paths, at the cost of a linear increase in power consumption and large silicon area requirement. A second problem associated with parallel A/D converters is path mismatch. During operation, the input signal has to pass different paths from the input to the digital output. If all A/D converters in parallel are identical, these paths are also identical. However, if offset, gain, bandwidth or time mismatch occur between the individual converters, the path for the input signal changes each time it is switched from one converter to another. This behavior gives rise to fixed-pattern noise at the output of the composite A/D converter which can be detected as spurious harmonics in the frequency domain [90]. How these errors are seen in the spectrum of the sampled signal will be discussed in conjunction with the time-interleaved S/H in Section 3.2. The parallel architecture is advantageous when high sampling rates are necessary which are difficult to achieve with single A/D converter. Although the architecture is straightforward, parallel A/D converters usually are not the best compromise when it

comes to increasing the conversion rate of medium speed converters. For the A/D converter family described in this book, it has therefore been decided in favor of two-step/multi-step converter to obtain higher speed.

2.1.4 A/D Converters Realization Comparison

In this section, a number of recently published high resolution analog to digital converters are compared. Tables 2.1, 2.2, 2.3, and 2.4 show the FoM of the realized A/D converters from 1998 until 2008 distributed over the categories: flash, two-step/multi-step/subranging, pipeline and parallel pipeline. Normalizing the dissipated power P to the effective resolution ENOB and to the effective resolution bandwidth ERBW, the figure of merit, $\text{FoM} = P / (2^{\text{ENOB}} \times 2 \text{ ERBW})$ [107], is a measure for the required power per achieved resolution per conversion. Today, the state-of-the-art FoM for Nyquist A/D converters is around 1 pJ/conversion.

From Table 2.1 it can be seen that the flash architecture is (barely or) not used at all for accuracies above 6 bits due to the large intrinsic capacitance required. The most prominent drawback of flash A/D converter is the fact that the number of comparators grows exponentially with the number of bits. Increasing the quantity of the comparators also increases the area of the circuit, as well as the power consumption. Other issues limiting the resolution and speed include nonlinear input capacitance, location-dependent reference node time constants, incoherent timing of comparators laid out over a large area, and comparator offsets. To lessen the impact of mismatch in the resistor reference ladder and the unequal input offset voltage of the comparators on the linearity of ADC, several schemes, such as inserting a preamplifier [1] in front of the latch, adding a chopper amplifier [2] and auto-zero scheme to sample an offset in the capacitor in front of the latch or digital background calibration [18] have been developed. Alternatively, the offsets and input capacitance can be reduced by means of distributed pre-amplification combined with averaging [108, 109] and possibly also with interpolation [110]. As thin oxide improves the matching property of transistors, smaller devices can be

Table 2.1 Table of realized flash A/D converters

Reference	N	ENOB	f_s (MS/s)	ERBW (MHz)	P (mW)	FoM (pJ)
[8]	6	5.8	500	160	225	12.6
[9]	6	5.2	500	250	330	17.3
[11]	6	5.2	1300	750	545	11.4
[13]	6	5.2	22	11	0.48	0.6
[14]	6	5.3	1300	600	600	12.7
[15]	6	5.4	2000	1000	310	3.5
[16]	6	5.7	1200	600	135	2.2
[17]	6	5.2	4000	1000	990	13.5
[19]	5	4.7	1000	200	46	1.8
[21]	5	4.0	5000	2500	102	1.3

Table 2.2 Table of realized two-step/multi-step/subranging A/D converters

Reference	N	ENOB	f_s (MS/s)	ERBW (MHz)	P (mW)	FoM (pJ)
[30]	10	9.1	25	12.5	195	14.2
[31]	14	12	100	50	1250	3.1
[32]	12	10.3	50	25	850	13.5
[33]	13	11.1	40	15	800	12.1
[34]	12	10.3	54	4	295	29.2
[35]	10	9.1	160	10	190	17.3
Ch 3 [37]	12	10.5	60	30	100	1.1
[38]	15	10.8	20	2	140	19.7
[50]	12	10.1	40	20	30	0.7

used in newer technology generations to achieve the same matching accuracy; this fact has been exploited by many recent works of flash-type converters to improve the energy efficiency of the conversion.

The realized FoM of the two-step/multi-step or subranging architectures is relatively constant for different values of the ENOB. In Table 2.2 it was shown that the FoM of a two-step/multi-step or subranging converter increases less rapidly than the noise-limited architectures such as multi-stage pipeline. Although the number of comparators is greatly reduced from the flash architecture, path matching is a problem and in some cases the input bandwidth is limited to relatively low frequency compared to the conversion rate [23, 24, 26]. Fine comparators accuracy requirements can be relaxed by including an inter-stage gain amplifier to amplify the signal for the fine comparator bank [22, 25]. While both, D/A converter and input of the residue amplifier require full resolution requirement, the coarse A/D converter section requirement can be relaxed with the digital error correction. If one stage subtracts a smaller reference than it nominally should due to the comparator offset, the subsequent stages compensate for this by subtracting larger references. This widely employed error correction method is referred to as redundant signed digit (RSD) correction, which was firstly developed for algorithmic ADCs in [111, 112] and later utilized in pipelined ADC [59]. Other related methods have also been used [54].

The redundancy allows for quantization errors as far as the residue stays in the input range of the next stage. The errors can be static or dynamic; it is only essential that the bits going to the correction logic circuitry match those which are D/A converted and used in residue formation. The same correction method can easily be expanded to larger resolution stages as well. As a minimum, one extra quantization level is required [113], but for maximum error tolerance the nominal number of comparators has to be doubled. The level of error tolerance on the coarse A/D converter section depends on how much digital error correction range the fine A/D converter section can provide. The correction range varies from ± 3 LSBs in [24] to a much larger value in [22, 25] with an S/H inter-stage amplifier.

If over/under-range protection is used, the offset requirements for the coarse converter can be greatly relaxed; but the fine one shares similar matching concerns as the flash architecture. Interpolation can be applied as well to reduce the number of preamps and their sizes. A balanced design can often achieve energy efficiency per conversion close to that of the pipeline converters. Since the two-step/multi-step architecture is matching limited, calibration can be applied to reduce the intrinsic capacitance. Two approaches can be taken to calibrate out the errors: mixed signal or fully digital. In mixed signal calibration, the erroneous component values are measured from the digital output and adjusted closer to their nominal ones [34]. The correction is applied to the analog signal path and thus requires extra analog circuitry. In the fully digital approach the component values are not adjusted [114]; however, the accuracy of this method depends on the accuracy of the measurement. In the sub-ranging converter the absence of a residue amplifier places stringent offset and noise requirements on the second quantizer, which can be overcome at modest power dissipation through the use of auto-zeroing [48], averaging [50] or background offset calibration [49]. The utilization of time-interleaved second quantizers increases the effective sampling rate [46].

The FoM of the pipeline converter increases as a function of the realized ENOB (Table 2.3). The architecture has evolved by making use of the strengths of the switched capacitor technique, which provides very accurate and linear analog amplification and summation operations in the discrete time domain. When the input is a rapidly changing signal, the relative timing of the first stage S/H circuit and the sub-ADC is critical and often relaxed with a front-end S/H circuit. Consecutive stages operate in opposite clock phases and as a result one sample traverses two stages in one clock cycle. So, the latency in clock cycles is typically half the number of stages plus one, which is required for digital error correction. For feedback purposes, where low latency is essential, a coarse result can be taken after the first couple of stages. The different bits of a sample become ready at different times. Thus, digital delay lines are needed for aligning the bits.

Several techniques for achieving resolutions higher than what is permitted by matching have been developed; the reference feedforward technique [55] and commutated feedback capacitor switching [56] improve the differential non-linearity, but do not affect the integral non-linearity. In 1-bit/stage architecture the

Table 2.3 Table of realized pipelined A/D converters

Reference	N	ENOB	f_s (MS/s)	ERBW (MHz)	P (mW)	FoM (pJ)
[66]	14	12	20	10	720	8.8
[72]	14	12.5	2.5	1.2	145	5.0
[73]	12	10.5	100	50	390	2.2
[74]	10	9.3	80	40	69	1.4
[75]	12	10.4	110	10	97	3.6
[76]	10	9.4	100	50	67	1.0
[78]	10	8.6	200	90	55	0.8
[80]	14	13.1	125	60	1850	1.7
[84]	11	10.5	45	5	81	5.6

capacitive error averaging technique, which has previously been used in algorithmic ADCs [115] can be used [66]. With it, a virtually capacitor ratio-independent gain-of-two stage can be realized. The technique, however, requires two opamps per stage (a modification proposed in [70]) and needs at least one extra clock phase. Pipeline architecture has been found very suitable for calibration [57, 62]. The number of components to be calibrated is sufficiently small, since only the errors in the first few stages are significant as a result of the fact that, when referred to the input, the errors in the latter stages are attenuated by the preceding gain. Furthermore, no extra A/D converter is necessarily required for measuring the calibration coefficients, since the back-end stages can be used for measuring the stages in front of them.

Similar to a subranging converter, over/under-range protection is necessary in a pipeline A/D converter. Since the comparator offset specs are substantially relaxed due to a low stage resolution and over/under-range protection, comparator design in pipeline A/D converters is far simpler than that of the flash ones, and usually does not impose limitation on the overall conversion speed or precision. It is how fast and how accurate the residue signals can be produced and sampled that determines the performance of a pipeline converter, especially for the first stage that demands the highest precision. Negative feedback is conventionally employed to stabilize the voltage gain and to broaden the amplifier bandwidth. It is expected that with technology advancement, the accompanying short-channel effects will pose serious challenges to realizing high open-loop gain, low noise, and low power consumption simultaneously at significantly reduced supply voltages. The tradeoff between speed, dynamic range, and precision will eventually place a fundamental limit on the resolution of pipeline converters attainable in ultra-deep-submicron CMOS technologies.

The realized FoM of the parallel pipeline architectures is severely limited by required power (Table 2.4). Up to a certain resolution, component matching is satisfactory enough and the errors originating from channel mismatch can be kept to a tolerable level with careful design. High-resolution time-interleaved A/D converters, however, without exception, use different techniques to suppress errors. The offset can be rather easily calibrated using a mixed signal [116] or all-digital circuitry [91]. Calibrating the gain mismatch is also possible, but requires more complex circuitry than offset calibration [95, 96]. The timing skew may originate

Table 2.4 Table of realized parallel-pipelined A/D converters

Reference	N	ENOB	f_s (MS/s)	ERBW (MHz)	P (mW)	FoM (pJ)
[94]	8	7.6	75	35	75	5.5
[95]	10	9.4	40	20	650	24.1
[96]	10	9.5	40	10	565	39.0
[98]	10	9.4	120	20	234	8.7
[101]	10	9.7	120	2	75	22.5
[103]	10	9.4	200	60	104	1.3
[104]	8	7.8	150	75	71	2.1
[105]	11	9.4	800	400	350	0.65
[106]	15	12.3	125	60	909	1.5

from the circuit generating the clock signals for different channels or it may be due to different propagation delays to the sampling circuits. Skew can be most easily avoided by using a full-speed front-end sample-and-hold circuit [117]. The A/D converter channels resample the output of the S/H when it is in a steady state, and so the timing of the channels is not critical. However, the S/H circuit has to be very fast, since it operates at full speed.

2.2 Notes on Low Voltage A/D Converter Design

Explosive growth in wireless and wireline communications is the dominant driver for high-resolution, high-speed, low power, and low cost integrated A/D converters. From an integration point of view the analog electronics must be realized on the same die as the digital core and consequently must cope with the CMOS evolution dictated by the digital circuit. Technology scaling offers significantly lowering of the cost of digital logic and memory, and there is a great incentive to implement high-volume baseband signal processing in the most advanced process technology available. Concurrently, there is an increased interest in using transistors with minimum channel length (Fig. 2.5a) and minimum oxide thickness to implement

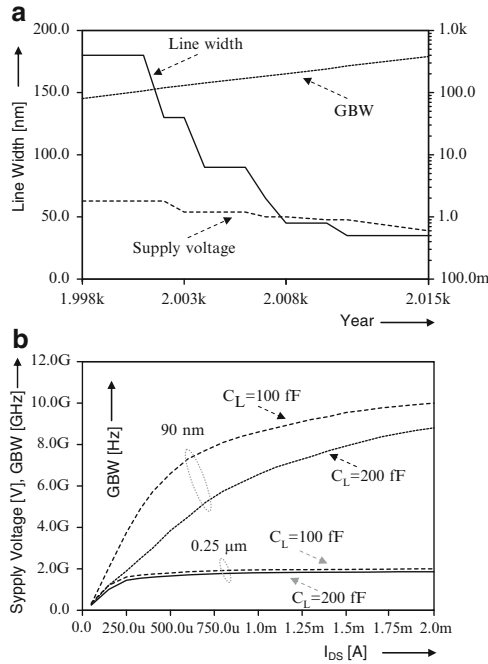


Fig. 2.5 (a) Trend of analog features in CMOS technologies. (b) Gain-bandwidth product versus drain current in two technological nodes

analog functions, because the improved device transition frequency, f_T , allows for faster operation. To ensure sufficient lifetime for digital circuitry and to keep power consumption at an acceptable level, the dimension-reduction is accompanied by lowering of nominal supply voltages. Due to the reduction of supply voltage the available signal swing is lowered, fundamentally limiting the achievable dynamic range at reasonable power consumption levels. Additionally, lower supply voltages require biasing at lower operating voltages which results in worse transistor properties, and hence yield circuits with lower performance. To achieve a high linearity, high sampling speed, high dynamic range, with low supply voltages and low power dissipation in ultra-deepsubmicron CMOS technology is a major challenge.

The key limitation of analog circuits is that they operate with electrical variables and not simply with discrete numbers that, in circuit implementations, gives rise of a beneficial noise margin. On the contrary, the accuracy of analog circuits fundamentally relies on matching between components, low noise, offset and low distortions. In this section, the most challenging design issues for low voltage, high-resolution A/D converters in deep submicron technologies such as contrasting the degradation of analog performances caused by requirement for biasing at lower operating voltages, obtaining high dynamic range with low voltage supplies and ensuring good matching for low-offset are reviewed. Additionally, the subsequent remedies to improve the performance of analog circuits and data converters by correcting or calibrating the static and possibly the dynamic limitations through calibration techniques are briefly discussed as well.

With reduction of the supply voltage to ensure suitable overdrive voltage for keeping transistors in saturation, even if the number of transistors stacked-up is kept at the minimum, the swing of signals is low if high resolution is required. Low voltage is also problematic for driving CMOS switches especially for the ones connected to signal nodes as the on-resistance can become very high or at the limit the switch does not close at all in some interval of the input amplitude. One solution is the multi-chip solution, where digital functions are implemented in a single or multiple chips and the analog processing is obtained by a separate chip with suitably high supply voltage and reduced analog digital interference. The use on the same chip of two supply voltages, one for the digital part with lower and one for the analog part with higher supply voltage is another possibility. The multiple threshold technology is another option.

In general, to achieve a high gain operation, high output impedance is necessary, e.g. drain current should vary only slightly with the applied V_{DS} . With the transistor scaling, the drain assert its influence more strongly due to the growing proximity of gate and drain connections and increase the sensitivity of the drain current to the drain voltage. The rapid degradation of the output resistance at gate lengths below $0.1\ \mu\text{m}$ and the saturation of g_m reduce the device intrinsic gain $g_m r_o$ characteristics. As transistor size is reduced, the fields in the channel increase and the dopant impurity levels increase. Both changes reduce the carrier mobility, and hence the transconductance g_m . Typically, desired high transconductance value is obtained at the cost of an increased bias current. However, for very short channel the carrier velocity quickly reaches the saturation limit at which the transconductance also

saturates becoming independent of gate length or bias $g_m = W_{eff} C_{ox} v_{sat}/2$. As channel lengths are reduced without proportional reduction in drain voltage, raising the electric field in the channel, the result is velocity saturation of the carriers, limiting the current and the transconductance. A limited transconductance is problematic for analog design: for obtaining high gain it is necessary to use wide transistors at the cost of an increased parasitic capacitances and, consequently, limitations in bandwidth and slew rate. Even using longer lengths obtaining gain with deep submicron technologies is not appropriate; it is typically necessary using cascode structures with stack of transistors or circuits with positive feedback. As transistor's dimension reduction continues, the intrinsic gain keeps decreasing due to a lower output resistance as a result of drain-induced barrier lowering (DIBL) and hot carrier impact ionization. To make devices smaller, junction design has become more complex, leading to higher doping levels, shallower junctions, halo doping, etc. all to decrease drain-induced barrier lowering. To keep these complex junctions in place, the annealing steps formerly used to remove damage and electrically active defects must be curtailed, increasing junction leakage. Heavier doping also is associated with thinner depletion layers and more recombination centers that result in increased leakage current, even without lattice damage. In addition, gate leakage currents in very thin-oxide devices will set an upper bound on the attainable effective output resistance via circuit techniques (such as active cascode). Similarly, as scaling continues, the elevated drain-to-source leakage in an off-switch can adversely affect the switch performance. If the switch is driven by an amplifier, the leakage may lower the output resistance of the amplifier, hence limits its low-frequency gain.

Low-distortion at quasi-*dc* frequencies is relevant for many analog circuits. Typically, quasi-*dc* distortion may be due to the variation of the depletion layer width along the channel, mobility reduction, velocity saturation and nonlinearities in the transistors' transconductances and in their output conductances, which is heavily dependent on biasing, size, technology and typically sees large voltage swings. With scaling higher harmonic components may increase in amplitude despite the smaller signal; the distortion increases significantly. At circuit level the degraded quasi-*dc* performance can be compensated by techniques that boost gain, such as (regulated) cascodes. These are, however, harder to fit within decreasing supply voltages. Other solutions include a more aggressive reduction of signal magnitude which requires a higher power consumption to maintain SNR levels.

The theoretically highest gain-bandwidth of an OTA is almost determined by the cutoff frequency of transistor (see Fig. 2.5b for assessment of GBW for two technological nodes). Assuming that the kT/C noise limit establishes the value of the load capacitance, to achieve required SNR large transconductance is required. Accordingly, the aspect ratio necessary for the input differential pair must be fairly large, in the hundred range. Similarly, since with scaling the gate oxide becomes thinner, the specific capacitance C_{ox} increases as the scaling factor. However, since the gate area decreases as the square of the scaling factor, the gate-to-source and gate-to-drain parasitic capacitance lowers as the process is scaled. The coefficients for the parasitic input and output capacitance, C_{gs} and C_{gd} shown in Fig. 2.6a) have been obtained by

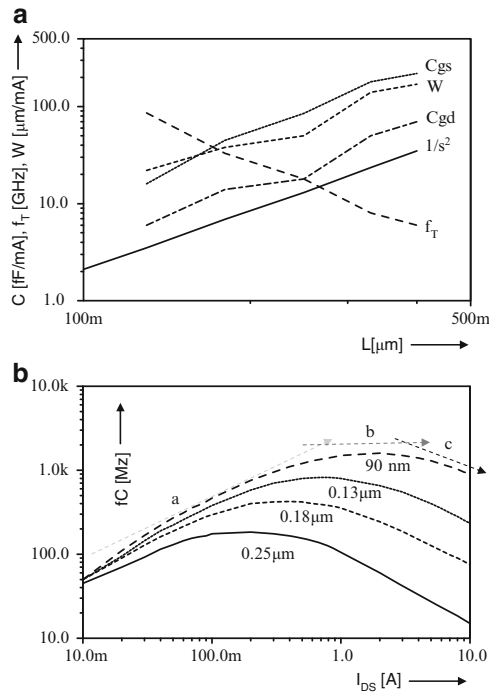


Fig. 2.6 (a) Scaling of gate width and transistor capacitances. (b) Conversion frequency f_c versus drain current for four technological nodes

simulation for conventional foundry processes under the assumption that the overdrive voltage is 0.175 V. Similarly, with technology-scaling the actual junctions become shallower, roughly proportional to the technology feature size. Also, the junction area roughly scales in proportion to the minimum gate-length, while the dope level increase does not significantly increase the capacitance per area. Altogether this leads to a significantly reduced junction capacitance per g_m with newer technologies.

Reducing transistor parasitic capacitance is desired, however, the benefit is contrasted by the increased parasitic capacitance of the interconnection (the capacitance of the wires connecting different parts of the chip). With transistors becoming smaller and more transistors being placed on the chip, interconnect capacitance is becoming a large percentage of total capacitance. The global effect is that scaling does not benefit fully from the scaling in increasing the speed of analog circuit as the position of the non-dominant poles is largely unchanged. Additionally, with the reduced signal swing, to achieve required SNR signal capacitance has to increase proportionally. By examining Fig. 2.6b), it can be seen that the characteristic exhibits convex curve and takes the highest value at the certain sink current (region b). In the region of the current being less than this value (region a), the conversion frequency increases with an increase of the sink current. Similarly, in the

region of the current being higher than this value (region *c*), the conversion frequency decreases with an increase of the sink current.

There are two reasons why this characteristic is exhibited; in the low current region, the g_m is proportional to the sink current, and the parasitic capacitances are smaller than the signal capacitance. At around the peak, at least one of the parasitic capacitances becomes equal to the signal capacitance. In the region of the current being larger than that value, both parasitic capacitances become larger than the signal capacitance and the conversion frequency will decrease with an increase of the sink current.

In mixed signal application the substrate noise and the interference between analog and digital supply voltages caused by the switching of digital sections are problematic. The situation becomes more and more critical as smaller geometries induce higher coupling. Moreover, higher speed and current density augment electro-magnetic issues. The use of submicron technologies with high resistive substrates is advantageous because the coupling from digital sections to regions where the analog circuits are located is partially blocked. However, the issues such as the bounce of the digital supply and ground lines exhibit strong influence on analog circuit behavior. The use of separate analog and digital supplies is a possible remedy but its effectiveness is limited by the internal coupling between close metal interconnections. The substrate and the supply noise cause two main limits: the in-band tones produced by nonlinearities that mix high frequency spurs and the reduction of the analog dynamic range required for accommodating the common-mode part of spurs. Since the substrate coupling is also a problem for pure digital circuit the submicron technologies are evolving toward silicon-on-insulator (SOI) and trench isolation options.

The offset of any analog circuit and the static accuracy of data converters critically depend on the matching between nominally identical devices. With transistors becoming smaller, the number of atoms in the silicon that produce many of the transistor's properties is becoming fewer, with the result that control of dopant numbers and placement is more erratic. During chip manufacturing, random process variations affect all transistor dimensions: length, width, junction depths, oxide thickness etc., and become a greater percentage of overall transistor size as the transistor scales. The stochastic nature of physical and chemical fabrication steps causes a random error in electrical parameters that gives rise to a time independent difference between equally designed elements. The error typically decreases as the area of devices. Transistor matching properties are improved with a thinner oxide [130]. Nevertheless, when the oxide thickness is reduced to a few atomic layers, quantum effects will dominate and matching will degrade. Since many circuit techniques exploit the equality of two components it is important for a given process obtaining the best matching especially for critical devices. Some of the rules that have to be followed to ensure good matching are: firstly, devices to be matched should have the same structure and use the same materials, secondly, the temperature of matched components should be the same, e.g. the devices to be matched should be located on the same isotherm, which is obtained by symmetrical placement with respect to the dissipative devices, thirdly, the distance between matched devices should be minimum for having the maximum spatial

correlation of fluctuating physical parameters, common-centroid geometries should be used to cancel the gradient of parameters at the first order. Similarly, the same orientation of devices on chip should be the same to eliminate dissymmetries due to unisotropic fabrication steps, or to the uniostropy of the silicon itself and lastly, the surroundings in the layout, possibly improved by dummy structures should be the same to avoid border mismatches.

Since the use of digital enhancing techniques reduces the need for expensive technologies with special fabrication steps, a side advantage is that the cost of parts is reduced while maintaining good yield, reliability and long-term stability. Indeed, the extra cost of digital processing is normally affordable as the use of submicron mixed signal technologies allows for efficient usage of silicon area even for relatively complex algorithms. The methods can be classified into foreground and background calibration.

The foreground calibration, typical of A/D converters, interrupts the normal operation of the converter for performing the trimming of elements or the mismatch measurement by a dedicated calibration cycle normally performed at power-on or during periods of inactivity of the circuit. Any miscalibration or sudden environmental changes such as power supply or temperature may make the measured errors invalid. Therefore, for devices that operate for long periods it is necessary to have periodic extra calibration cycles. The input switch restores the data converter to normal operational after the mismatch measurement and every conversion period the logic uses the output of the A/D converter to properly address the memory that contains the correction quantity. In order to optimize the memory size the stored data should be the minimum word-length, which depends on technology accuracy and expected A/D linearity. The digital measure of errors, that allows for calibration by digital signal processing, can be at the element, block or entire converter level. The calibration parameters are stored in memories but, in contrast with the trimming case, the content of the memories is frequently used, as they are input of the digital processor.

Methods using background calibration work during the normal operation of the converter by using extra circuitry that functions all the time synchronously with the converter function. Often these circuits use hardware redundancy to perform a background calibration on the fraction of the architecture that is not temporarily used. However, since the use of redundant hardware is effective but costs silicon area and power consumption, other methods aim at obtaining the functionality by borrowing a small fraction of the sampled-data circuit operation for performing the self-calibration.

2.3 A/D Converter Building Blocks

2.3.1 Sample-and-Hold

Inherent to the A/D conversion process is a sample-and-hold (S/H) circuit that resides in the front-end of a converter (and also between stages in a pipeline

converter). In addition to suffering from additive circuit noise and signal distortion just as the rest of the converter does, the S/H also requires a precision time base to define the exact acquisition time of the input signal. The dynamic performance degradation of an ADC can often be attributed to the deficiency of the S/H circuit (and the associated buffer amplifier).

The main function of an S/H circuit is to take samples of its input signal and hold its value until the A/D converter can process the information. Typically, the samples are taken at uniform time intervals; thus, the sampling rate (or clock rate) of the circuit can be determined. The operation of an S/H circuit can be divided into sample mode (sometimes also referred as acquisition mode) and hold mode, whose durations need not be equal. In sample mode, the output can either track the input, in which case the circuit is often called a track-and-hold (T/H) circuit or it can be reset to some fixed value. In hold mode an S/H circuit remembers the value of the input signal at the sampling moment and thus it can be considered as an analog memory cell. The basic circuit elements that can be employed as memories are capacitors and inductors, of which the capacitors store the signal as a voltage (or charge) and the inductors as a current. Since capacitors and switches with a high off-resistance needed for a voltage memory are far easier to implement in a practical integrated circuit technology than inductors and switches with a very small on-resistance required for a current memory, all sample-and-hold circuits are based on voltage sampling with switched capacitor (SC) technique.

S/H circuit architectures can roughly be divided into open-loop and closed-loop architectures. The main difference between them is that in closed-loop architectures the capacitor, on which the voltage is sampled, is enclosed in a feedback loop, at least in hold mode. Although open-loop S/H architecture provide high speed solution, its accuracy, however, is limited by the harmonic distortion arising from the nonlinear gain of the buffer amplifiers and the signal-dependent charge injection from the switch. These problems are especially emphasized with a CMOS technology as shown in [Section 3.3](#). Enclosing the sampling capacitor in the feedback loop reduces the effects of nonlinear parasitic capacitances and signal-dependent charge injection from the MOS switches. Unfortunately, an inevitable consequence of the use of feedback is reduced speed.

Figure 2.7 illustrates three common configurations for closed-loop switched-capacitor S/H circuits [56, 57, 59, 91]. For simplicity, single-ended configurations

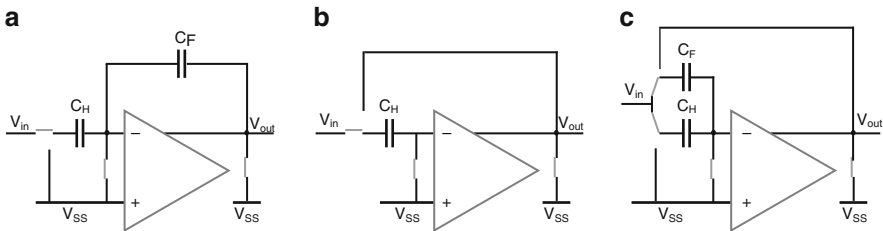


Fig. 2.7 Switched capacitor S/H circuit configurations in sample phase: (a) a circuit with separate CH and CF, (b) a circuit with one capacitor, and (c) a circuit with CF shared as a sampling capacitor

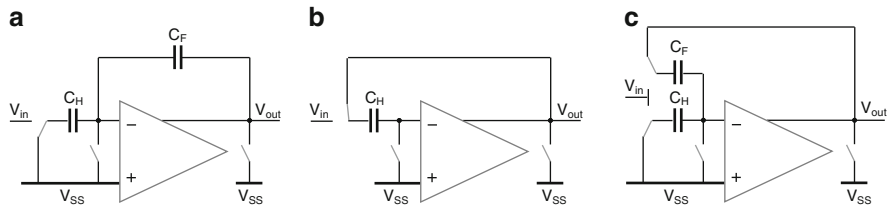


Fig. 2.8 Switched capacitor S/H circuit configurations in hold phase: (a) a circuit with separate C_H and C_F , (b) a circuit with one capacitor, and (c) a circuit with C_F shared as a sampling capacitor

are shown; however in circuit implementation all would be fully differential. In a mixed-signal circuit such as A/D converters, fully differential analog signals are preferred as a means of getting a better power supply rejection and immunity to common mode noise. The operation needs two non-overlapping clock phases – sampling and holding or transferring. Switch configurations shown in Fig. 2.7 are for the sampling phase, while configurations shown in Fig. 2.8 are for hold phase.

In all cases, the basic operations include sampling the signal on the sampling capacitor(s) C_H and transferring the signal charge onto the feedback capacitor C_F by using an op amp in the feedback configuration. In the configuration in Fig. 2.7a, which is often used as an integrator, assuming an ideal op amp and switches, the op amp forces the sampled signal charge on C_H to transfer to C_F . If C_H and C_F are not equal capacitors, the signal charge transferred to C_F will display the voltage at the output of the op amp according to $V_{out} = (C_H/C_F) V_{in}$. In this way, both S/H and gain functions can be implemented within one SC circuit [57,91]. In the configuration shown in Fig. 2.7b, only one capacitor is used as both sampling capacitor and feedback capacitor. This configuration does not implement the gain function, but it can achieve high speed because the feedback factor (the ratio of the feedback capacitor to the total capacitance at the summing node) can be much larger than that of the previous configuration, operating much closer to the unity gain frequency of the amplifier. Furthermore, it does not have the capacitor mismatch limitation as the other two configurations. Here, the sampling is performed passively, i.e. it is done without the op amp, which makes signal acquisition fast. In hold mode the sampling capacitor is disconnected from the input and put in a feedback loop around the op amp. This configuration is often used in the front-end input S/H circuit [56, 59] and will be discussed in more detail in Section 3.3.

Figure 2.7c shows another configuration which is a combined version of the configurations in Fig. 2.7a and Fig. 2.7b. In this configuration, in the sampling phase, the signal is sampled on both C_H and C_F , with the resulting transfer function $V_{out} = (1 + (C_H/C_F)) V_{in}$. In the next phase, the sampled charge in the sampling capacitor is transferred to the feedback capacitor. As a result, the feedback capacitor has the transferred charge from the sampling capacitor as well as the input signal charge. This configuration has a wider bandwidth in comparison to the configuration shown in Fig. 2.7a, although feedback factor is comparable.

Important parameters in determining the bandwidth of the SC circuit are G_m (transconductance of the op amp), feedback factor β , and output load capacitance. In all of these three configurations, the bandwidth is given by $1/\tau = \beta \times G_m/C_L$, where C_L is the total capacitance seen at the op amp output. Since S/H circuit use amplifier as buffer, the acquisition time will be a function of the amplifier own specifications. Similarly, the error tolerance at the output of the S/H is dependent on the amplifier's offset, gain and linearity. Once the hold command is issued, the S/H faces other errors. Pedestal error occurs as a result of charge injection and clock feed-through. Part of the charge built up in the channel of the switch is distributed onto the capacitor, thus slightly changing its voltage. Also, the clock couples onto the capacitor via overlap capacitance between the gate and the source or drain. Another error that occurs during the hold mode is called droop, which is related to the leakage of current from the capacitor due to parasitic impedances and to the leakage through the reverse-biased diode formed by the drain of the switch. This diode leakage can be minimized by making the drain area as small as can be tolerated. Although the input impedance to the amplifier is very large, the switch has a finite off impedance through which leakage can occur. Current can also leak through the substrate.

A prominent drawback of a simple S/H is the on-resistance variation of the input switch that introduces distortion. Technology scales the supply voltage faster than the threshold voltage, which results in a larger on-resistance variation in a switch. As a result, the bandwidth of the switch becomes increasingly signal dependent. Clock bootstrapping was introduced to keep the switch gate-source voltage constant (Section 3.3.3). Care must be exercised to ensure that the reliability of the circuit is not compromised.

While the scaling of CMOS technology offers a potential for improvement on the operating speed of mixed-signal circuits, the accompanying reduction in the supply voltage and various short-channel effects create both fundamental and practical limitations on the achievable gain, signal swing and noise level of these circuits, particularly under a low power constraint. In the sampling circuit, thermal noise is produced due to finite resistance of a MOS transistor switch and is stored in a sampling capacitor. As the sampling circuit cannot differentiate the noise from the signal, part of this signal acquisition corresponds to the instantaneous value of the noise at the moment the sampling takes place. In this context, when the sample is stored as charge on a capacitor, the root-mean-square (rms) total integrated thermal noise voltage is $\overline{v_{ns}^2} = kT/C_H$, where kT is the thermal energy and C_H is the sampling capacitance. This is often referred to as the kT/C noise. No resistance value at the expression is present, as the increase of thermal noise power caused by increasing the resistance value is cancelled in turn by the decreasing bandwidth.

In the sampling process the kT/C noise usually comprises two major contributions – the channel noise of the switches and the amplifier noise. Since no direct current is conducted by the switch right before a sampling takes place (the bandwidth of the S/H circuit is assumed large and the circuit is assumed settled), the $1/f$ noise is not of concern here; only the thermal noise contributes, which is a function of the channel resistance that is weakly affected by the technology scaling [120]. On the other hand, the amplifier output noise is in most cases dominated by the channel

noise of the input transistors, where the thermal noise and the $1/f$ noise both contribute. Because the input transistors of the amplifier are usually biased in saturation region to derive large transconductance (g_m), impact ionization and hot carrier effect tend to enhance their thermal noise level [121, 122]; the $1/f$ noise increases as well due to the reduced gate capacitance resulted from finer lithography and therefore shorter minimum gate length. It follows that, as CMOS technology scaling continues, amplifier increasingly becomes the dominant noise source. Interestingly, the input-referred noise (the total integrated output noise as well) still takes the form of kT/C with some correction factor χ_1 , $\overline{v_{ns}^2} = \chi_1 kT/C_H$. Thus a fundamental technique to reduce the noise level, or to increase the signal-to-noise ratio of an S/H circuit, is to increase the size of the sampling capacitors. The penalty associated with this technique is the increased power consumption as larger capacitors demand larger charging/discharging current to keep up the sampling speed.

2.3.2 Operational Amplifier

In front-end S/H amplifiers or multi-stage A/D converters, precision op amps are almost invariably employed to relay the input signal (or the residue signal) to the trailing conversion circuits. Operating on the edge of the performance envelope, op amps exhibit intense trade-offs amongst the dynamic range, linearity, settling speed, stability, and power consumption. As a result, the conversion accuracy and speed are often dictated by the performance of these amplifiers.

Amplifiers with a single gain stage have high output impedance providing an adequate dc gain, which can be further increased with gain boosting techniques. Single-stage architecture offers large bandwidth and a good phase margin with small power consumption. Furthermore, no frequency compensation is needed, since the architecture is self-compensated (the dominant pole is determined by the load capacitance), which makes the footprint on the silicon small. On the other hand, the high output impedance is obtained by sacrificing the output voltage swing, and the noise is rather high as a result of the number of noise-contributing devices and limited voltage head-room for current source biasing.

The simplest approach for the one-stage high-gain operational amplifier is telescopic cascode amplifier [150] of Fig. 2.9a. With this architecture, a high open loop dc gain can be achieved and it is capable of high speed when closed loop gain is low. The number of current legs being only two, the power consumption is small. The biggest disadvantage of a telescopic cascode amplifier is its low maximum output swing, $V_{DD} - 5V_{DS,SAT}$, where V_{DD} is the supply voltage and $V_{DS,SAT}$ is the saturation voltage of a transistor. With this maximum possible output swing the input common-mode range is zero. In practice, some input common-mode range, which reduces the output swing, always has to be reserved so as to permit inaccuracy and settling transients in the signal common-mode levels. The high-speed capability of the amplifier is the result of the presence of only n -channel transistors in the signal path and of relatively small capacitance at the source of the cascode transistors.

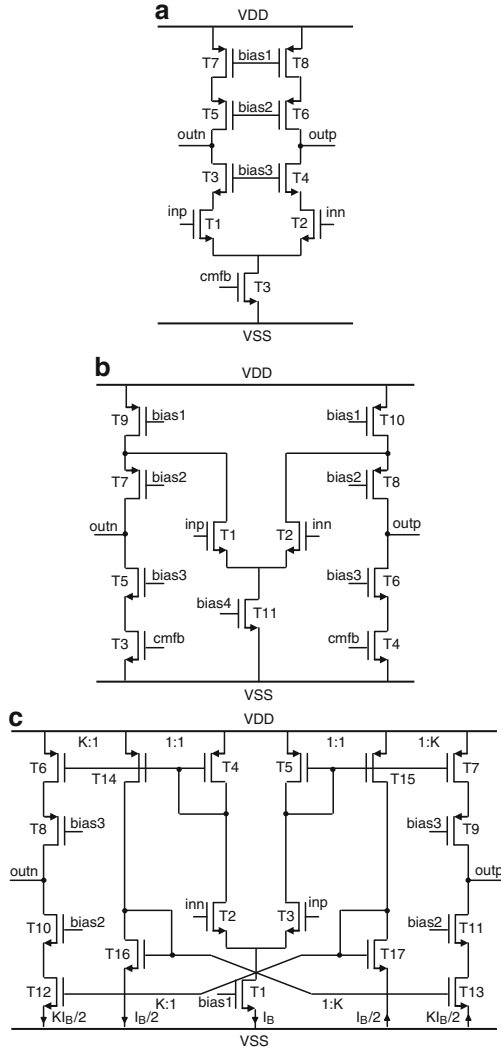


Fig. 2.9 One-stage amplifiers: (a) Telescopic cascode, (b) folded cascode, and (c) push–pull current-mirror amplifier with a cascade output stage

The gain-bandwidth product of the amplifier is given by $GBW = g_{m1}/C_L$, where g_{m1} is the transconductance of transistors T_1 and C_L is the load capacitance. Thus, the GBW is limited by the load capacitance. Due to its the simple topology and dimensioning, the telescopic cascode amplifier is preferred if its output swing is large enough for the specific application. The output signal swing of this architecture has been widened by driving the transistors T_7 – T_8 into the linear region [151]. In order to preserve the good common mode rejection ratio and power supply rejection

ratio properties of the topology, additional feedback circuits for compensation have been added to these variations. The telescopic cascode amplifier has low current consumption, relatively high gain, low noise and very fast operation. However, as it has five stacked transistors, the topology is not suitable for low supply voltages.

The folded cascode amplifier topology [152] is shown in Fig. 2.9b. The swing of this design is constrained by its cascoded output stage. It provides a larger output swing and input common-mode range than the telescopic amplifier with the same dc gain and without major loss of speed. The output swing is $V_{DD} - 4V_{DS,SAT}$ and is not linked to the input common-mode range, which is $V_{DD} - V_T - 2V_{DS,SAT}$. The second pole of this amplifier is located at g_{m7}/C_{par} , where g_{m7} is the transconductance of T_7 and C_{par} is the sum of the parasitic capacitances from transistors T_1 , T_7 and T_9 at the source node of transistor T_7 . The frequency response of this amplifier is deteriorated from that of the telescopic cascode amplifier because of a smaller transconductance of the p -channel device and a larger parasitic capacitance. To assure symmetrical slewing, the output stage current is usually made equal to that of the input stage. The GBW of the folded cascode amplifier is also given by g_{m1}/C_L . The open loop dc gain of amplifiers having cascode transistors can be boosted by regulating the gate voltages of the cascode transistors [140]. The regulation is realized by adding an extra gain stage, which reduces the feedback from the output to the drain of the input transistors. In this way, the dc gain of the amplifier can be increased by several orders of magnitude. The increase in power and chip area can be kept very small with appropriate feedback amplifier architecture [140]. The current consumption of the folded cascode is doubled compared to the telescopic cascode amplifier although the output voltage swing is increased since there are only four stacked transistors. The noise of the folded cascode is slightly higher than in the telescopic cascode as a result of the added noise from the current source transistors T_9 and T_{10} . In addition, the folded cascode has a slightly smaller dc gain due to the parallel combination of the output resistance of transistors T_1 and T_9 .

A push-pull current-mirror amplifier, shown in Fig. 2.9c, has much better slew-rate properties and potentially larger bandwidth and dc gain than the folded cascode amplifier. The slew rate and dc gain depend on the current-mirror ratio K , which is typically between one and three. However, too large current-mirror ratio increases the parasitic capacitance at the gates of the transistors T_{12} and T_{13} , pushing the non-dominant pole to lower frequencies and limiting the achievable GBW. The non-dominant pole of the current mirror amplifier is much lower than that of the folded cascode amplifier and telescopic amplifiers due to the larger parasitic capacitance at the drains of input transistors. The noise and current consumption of the current-mirror amplifier are larger than in the telescopic cascode amplifier or in the folded cascode amplifier. A current-mirror amplifier with dynamic biasing [153] can be used to make the amplifier biasing be based purely on its small signal behavior, as the slew rate is not limited. In dynamic biasing, the biasing current of the operational amplifier is controlled on the basis of the differential input signal. With large differential input signals, the biasing current is increased to speed up the output settling. Hence, no slew rate limiting occurs, and the GBW requirement is relaxed. As the settling proceeds, the input voltage decreases and the biasing current is reduced. The biasing current needs to

be kept only to a level that provides enough GBW for an adequate small-signal performance. In addition to relaxed GBW requirements, the reduced static current consumption makes the design of a high- dc gain amplifier easier. With very low supply voltages, the use of the cascode output stages limits the available output signal swing considerably. Hence, two-stage operational amplifiers are often used, in which the operational amplifier gain is divided into two stages, where the latter stage is typically a common-source output stage. Unfortunately, with the same power dissipation, the speed of the two-stage operational amplifiers is typically lower than that of single-stage operational amplifiers.

Of the several alternative two-stage amplifiers, Fig. 2.10a shows a simple Miller compensated amplifier [154]. With all the transistors in the output stage of this amplifier placed in the saturation region, it has an output swing of $V_{DD} - V_{DS,SAT}$. Since the non-dominant pole, which arises from the output node, is determined dominantly by an explicit load capacitance, the amplifier has a compromised frequency response. The gain bandwidth product of a Miller compensated amplifier is given approximately by $GBW = g_{m1}/C_C$, where g_{m1} is the transconductance of T_1 . In general, the open loop dc gain of the basic configuration is not large enough for high-resolution applications. Gain can be enhanced by using cascoding, which has, however, a negative effect on the signal swing and bandwidth. Another drawback of this architecture is a poor power supply rejection at high frequencies because of the connection of V_{DD} through the gate-source capacitance $C_{GS5,6}$ of T_5 and T_6 and C_C . The noise properties of the two-stage Miller-compensated operational amplifier are comparable to those of the telescopic cascode and better than those of the folded cascode amplifier. The speed of a Miller-compensated amplifier is determined by its pole-splitting capacitor C_C . Usually, the position of this non-dominant pole, which is located at the output of the two-stage amplifier, is lower than that of either a folded-cascode or a telescopic amplifier. Thus, in order to push this pole to higher frequencies, the second stage of the amplifier requires higher currents resulting in increased power dissipation. Since the first stage does not need to have a large output voltage swing, it can be a cascode stage, either a telescopic or a folded cascode. However, the current consumption and transistor count are also increased. The advantages of the folded cascode structure are a larger input common-mode range and the avoidance of level shifting between the stages, while the telescopic stage can offer larger bandwidth and lower thermal noise.

Figure 2.10b illustrates a folded cascode amplifier with a common-source output stage and Miller compensation. The noise properties are comparable with those of the folded cascode amplifier. If a cascode input stage is used, the lead-compensation resistor can be merged with the cascode transistors. An example of this is the folded cascode amplifier with a common-source output stage and Ahuja-style compensation [155] shown in Fig. 2.10c. The operation of the Ahuja-style compensated operational amplifier is suitable for larger capacitive loads than the Miller-compensated one and it has a better power supply rejection, since the substrate noise coupling through the gate-source capacitance of the output stage gain transistors is not coupled directly through the pole-splitting capacitors to the operational amplifier output [155].

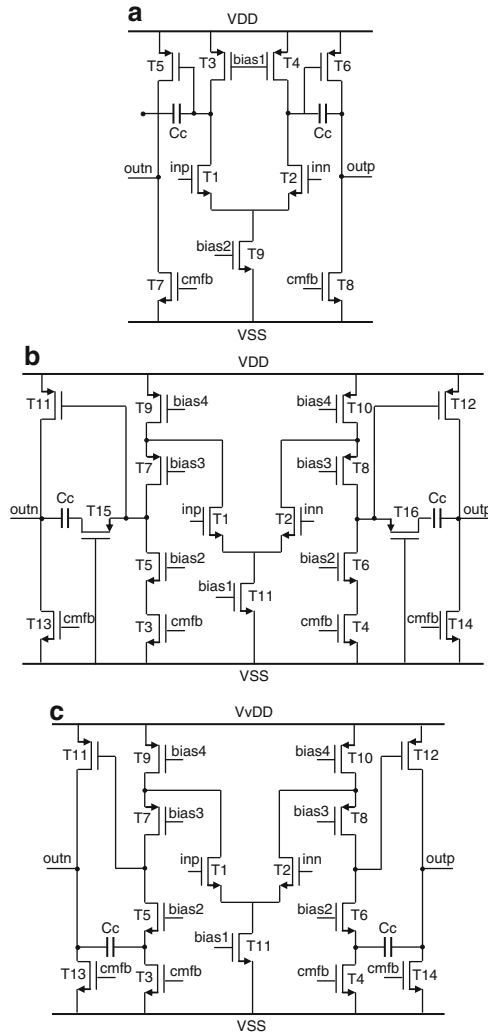


Fig. 2.10 Two-stage amplifiers: (a) Miller compensated, (b) folded cascode amplifier with a common-source output stage and Miller frequency compensation, and (c) folded cascode amplifier with a common-source output stage and Ahuja-style frequency compensation

2.3.3 Latched Comparators

The offset in preamps and comparators constitutes the major source of error in flash-type converters. Simple differential structure with thin oxide devices will keep dominating the preamp architecture in newer technologies. Dynamic performance is crucial at high sample rates with high input frequencies.

Because of its fast response, regenerative latches are used, almost without exception, as comparators for high-speed applications. An ideal latched comparator is composed of a preamplifier with infinite gain and a digital latch circuit. Since the amplifiers used in comparators need not to be either linear or closed-loop, they can incorporate positive feedback to attain virtually infinite gain [171]. Because of its special architecture, working process of a latched comparator could be divided in two stages: tracking and latching stages. In tracking stage the following dynamic latch circuit is disabled, and the input analog differential voltages are amplified by the preamplifier. In the latching stage while the preamplifier is disabled, the latch circuit regenerates the amplified differential signals into a pair of full-scale digital signals with a positive feedback mechanism and latches them at output ends.

Depending on the type of latch employed, the latch comparators can be divided into two groups: static [59, 173, 174], which have a constant current consumption during operation and dynamic [175–177], which does not consume any static power. In general, the type of latch employed is determined by the resolution of the stage. For a low-resolution quantization per stage, a dynamic latch is more customary since it dissipates less power than the static latch (the dynamic latch does not dissipate any power during the resetting period), even though the difference is negligible at high clock rates. While the latch circuits regenerate the difference signals, the large voltage variations on regeneration nodes will introduce the instantaneous large currents. Through parasitic gate-source and gate-drain capacitances of transistors, the instantaneous currents are coupled to the inputs of the comparators, making the disturbances unacceptable. It is so-called kickback noise influence. In flash A/D converters where a large number of comparators are switched on or off at the same time, the summation of variations came from regeneration nodes may become unexpectedly large and directly results in false quantization code output [172]. It is for this reason that the static latch is preferable for higher-resolution implementations.

The static latched comparator from [173] is shown in Fig. 2.11a. When the clock signal is high, T_{10} and T_{11} discharge the latch formed with cross-connected transistors T_{8-9} to the output nodes. When the latch signal goes low, the drain current difference between T_6 and T_7 appears as the output voltage difference. However, some delay in the circuit is present since T_8 and T_{11} have to wait until either side of the output voltage becomes larger than V_T . The other is that there is a static current in the comparators which are close to the threshold after the output is fully developed. Assume the potential of V_{outp} node is higher than that of V_{outn} node. After the short period, T_{11} turns off and the potential of V_{outp} becomes V_{DD} , however, since T_8 is in a linear region, the static current from T_6 will drain during the regeneration period. Since the input transistors are isolated from the regeneration nodes through the current mirror, the kickback noise is reduced. However, the speed of the regeneration circuit is limited by the bias current and is not suitable for low-power high-speed applications.

Figure 2.11b illustrates the schematic of the comparator given in [59]. The circuit consists of a folded-cascode amplifier (T_1 – T_7) where the load has been replaced by a current-triggered latch (T_8 – T_{10}). When the latch signal is high (resetting period),

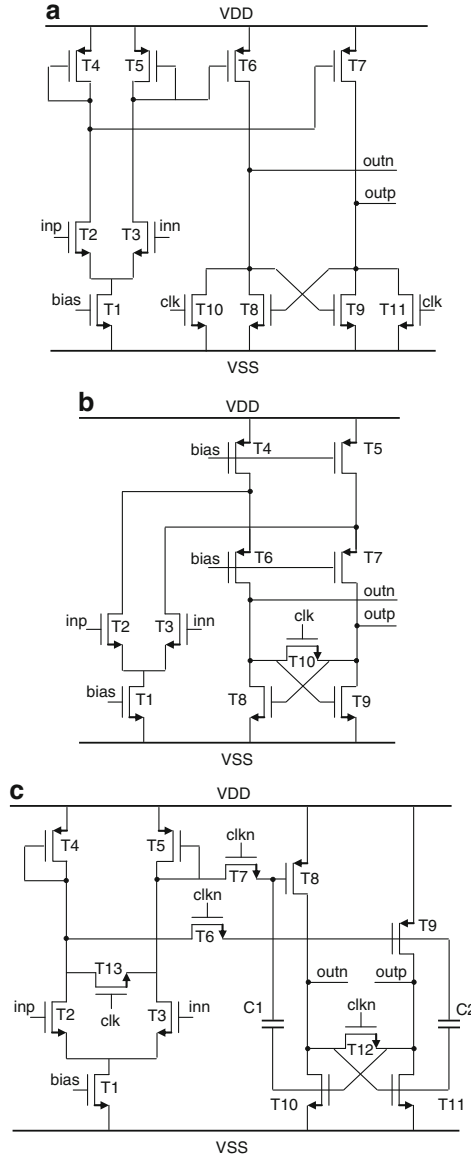


Fig. 2.11 Static latch comparators: (a) [173], (b) [59], and (c) [174]

transistor T_{10} shorts both the latch outputs. In addition, the on-resistance R of T_{10} can give an extra gain A_{reset} at the latch output, $A_{reset} = (g_{m1,2} \times R) / (2 - g_{m8,9} \times R)$, which speed up the regeneration process. However, the on-resistance R , should be chosen such that $g_{m8,9} \times R < 2$ and should be small enough to reset the output at the clock rate. Since all transistors are in active region, the latch can start regenerating

right after the latch signal goes low. The one disadvantage of this scheme is the large kickback noise. The folding nodes (drains of T_4 and T_5) have to jump up to V_{DD} in every clock cycle since the latch output does the full swing. Because of this, there are substantial amounts of kickback noise into the inputs through the gate-drain capacitor of input transistors T_1 and T_2 (C_{GD1} , C_{GD2}). To reduce kickback noise, the clamping diode has been inserted at the output nodes [178].

In Fig. 2.11c design shown in [174] is illustrated. Here, when the latch signal is low (resetting period), the amplified input signal is stored at gate of T_8 and T_9 and T_{12} shorts both V_{outp} and V_{outn} . When the latch signal goes high, the cross-coupled transistors T_{10} and T_{11} make a positive feedback latch. In addition, the positive feedback capacitors, C_1 and C_2 , boost up the regeneration speed by switching T_8 and T_9 from an input dependant current source during resetting period to a cross-coupled latch during the regeneration period. Because of C_1 and C_2 , the $T_8 \sim T_{11}$ work like a cross-coupled inverter so that the latch does not dissipate the static power once it completes the regeneration period. However, there is a large amount of kickback noise through the positive feedback capacitors, C_1 and C_2 . The switches (T_6 , T_7 and T_{13}) have been added to isolate the preamplifier from the latch. Therefore, the relatively large chip area is required due to the positive feedback capacitors (C_1 , C_2), isolation switches (T_6 , T_7 and T_{13}) and complementary latch signals.

The concept of a dynamic comparator exhibits potential for low power and small area implementation and, in this context, is restricted to single-stage topologies without static power dissipation. A widely used dynamic comparator is based on a differential sensing amplifier as shown in Fig. 2.12a was introduced in [175]. Transistors T_{1-4} , biased in linear region, adjust the threshold resistively and above them transistors T_{5-12} form a latch. When the latch control signal is low, the transistors T_9 and T_{12} are conducting and T_7 and T_8 are cut off, which forces both differential outputs to V_{DD} and no current path exists between the supply voltages. Simultaneously T_{10} and T_{11} are cut off and the transistors T_5 and T_6 conduct. This implies that T_7 and T_8 have a voltage of V_{DD} over them. When the comparator is latched T_7 and T_8 are turned on. Immediately after the regeneration moment, the gates of the transistors T_5 and T_6 are still at V_{DD} and they enter saturation, amplifying the voltage difference between their sources. If all transistors T_{5-12} are assumed to be perfectly matched, the imbalance of the conductances of the left and right input branches, formed by T_{1-2} and T_{3-4} , determines which of the outputs goes to V_{DD} and which to 0 V. After a static situation is reached (V_{clk} is high), both branches are cut off and the outputs preserve their values until the comparator is reset again by switching V_{clk} to 0 V. The transistors T_{1-4} connected to the input and reference are in the triode region and act like voltage controlled resistors. The transconductance of the transistors T_{1-4} operating in the linear region, is directly proportional to the drain-source voltage of the corresponding transistor V_{DS1-4} , while for the transistors T_{5-6} the transconductance is proportional to $V_{GS5,6} - V_T$. At the beginning of the latching process, $V_{DS1-4} \approx 0$ while $V_{GS5,6} - V_T \approx V_{DD}$. Thus, $g_{m5,6} \gg g_{m1-4}$, which makes the matching of T_5 and T_6 dominant in determining the latching balance. As small transistors are preferred, offset voltages of a few hundred millivolts are easily resulted. Mismatch in transistors T_{7-12} are attenuated by the gain of T_5 and T_6 , which makes them less critical. To cope with the mismatch problem, the layout of the critical transistors must be drawn as symmetric as possible. In addition to the

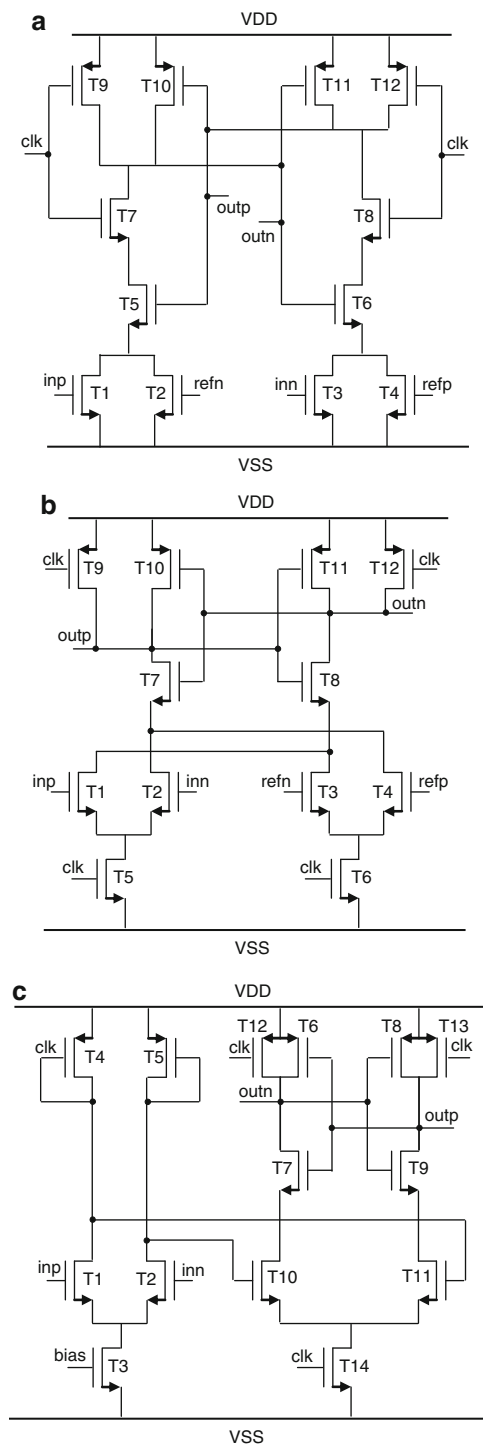


Fig. 2.12 Dynamic latch comparators: (a) [175], (b) [176], and (c) [177]

mismatch sensitivity, the latch is also very sensitive to an asymmetry in the load capacitance. This can be avoided by adding an extra latch or inverters as a buffering stage after the comparator core outputs. The resistive divider dynamic comparator topology has one clear benefit, which is its low kickback noise. This results from the fact that the voltage variation at the drains of the input transistors T_{1-4} is very small. On the other hand, the speed and resolution of the topology are relatively poor because of the small gain of the transistors biased in the linear region.

A fully differential dynamic comparator based on two cross-coupled differential pairs with switched current sources loaded with a CMOS latch is shown in Fig. 2.12b [176]. The trip point of the comparator can be set by introducing imbalance between the source-coupled pairs. Because of the dynamic current sources together with the latch, connected directly between the differential pairs and the supply voltage, the comparator does not dissipate *dc*-power. When the comparator is inactive the latch signal is low, which means that the current source transistors T_5 and T_6 are switched off and no current path between the supply voltages exists. Simultaneously, the *p*-channel switch transistors T_9 and T_{12} reset the outputs by shorting them to V_{DD} . The *n*-channel transistors T_7 and T_8 of the latch conduct and also force the drains of all the input transistors T_{1-4} to V_{DD} , while the drain voltage of T_5 and T_6 are dependent on the comparator input voltages. When clock signal is raised to V_{DD} , the outputs are disconnected from the positive supply, the switching current sources T_5 and T_6 turn on and T_{1-4} compare $V_{inp} - V_{inn}$ with $V_{refp} - V_{refn}$. Since the latch devices T_{7-8} are conducting, the circuit regeneratively amplifies the voltage difference at the drains of the input pairs. The threshold voltage of the comparator is determined by the current division in the differential pairs and between the cross-coupled branches. The threshold level of the comparator can be derived using large signal current equations for the differential pairs. The effect of the mismatches of the other transistors T_{7-12} is in this topology not completely critical, because the input is amplified by T_{1-4} before T_{7-12} latch. The drains of the cross-coupled differential pairs are high impedance nodes and the transconductances of the threshold-voltage-determining transistors T_{1-4} large. A drawback of the differential pair dynamic comparator is its high kickback noise: large transients in the drain nodes of the input transistors are coupled to the input nodes through the parasitic gate-drain capacitances. However, there are techniques to reduce the kickback noise, e.g. by cross-coupling dummy transistors from the differential inputs to the drain nodes [13]. The differential pair topology achieves a high speed and resolution, which results from the built-in dynamic amplification.

Figure 2.12c illustrates the schematic of the dynamic latch given in [177]. The dynamic latch consists of pre-charge transistors T_{12} and T_{13} , cross-coupled inverter T_{6-9} , differential pair T_{10} and T_{11} and switch T_{14} which prevent the static current flow at the resetting period. When the latch signal is low (resetting period), the drain voltages of T_{10-11} are $V_{DD} - V_T$ and their source voltage is V_T below the latch input common mode voltage. Therefore, once the latch signal goes high, the *n*-channel transistors $T_{7, 9-11}$ immediately go into the active region. Because each transistor in one of the cross-coupled inverters turns off, there is no static power dissipation from the latch once the latch outputs are fully developed.

2.4 A/D Converters: Summary

Practical realizations from [Section 2.1.4](#) (Fig. 2.13) show the trend that to a first order, converter power is directly proportional to sampling rate f_S when f_S is much lower than the device technology transition frequency f_T . However, power dissipation required becomes nonlinear as the speed capabilities of a process technology are pushed to the limit. In the case of *constant current-density* designs, there is an optimum power point when total intrinsic capacitance equals total extrinsic capacitance, beyond which power increases yield diminishing returns in speed improvements.

Power dependence on converter resolution is not as straight-forward as its dependence on f_S because converter architecture also varies with resolution. Flash converters dominate at the high-speed low-resolution end of the spectrum while pipelined converters are usually employed at the low-speed high-resolution end typically from 8–12 bits without calibration and up to 15 bits with calibration. The inter-stage gain makes it possible to scale the components along the pipeline,

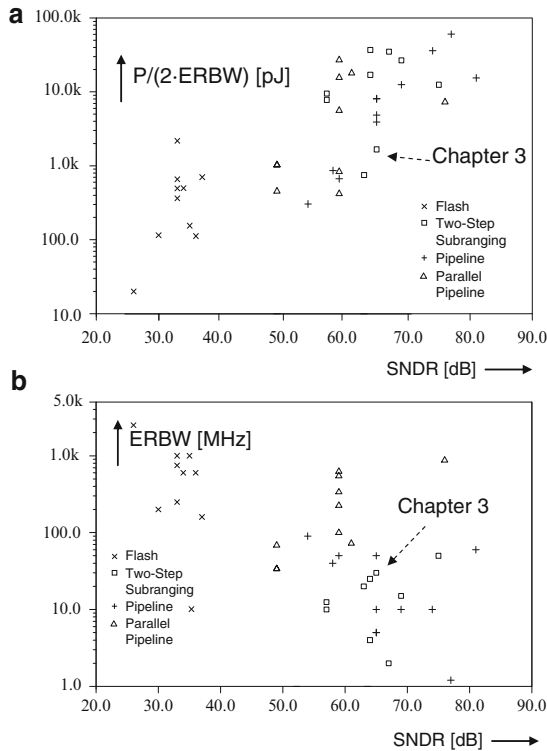


Fig. 2.13 (a) Energy versus SNDR for A/D converters shown in [Tables 2.1, 2.2, 2.3, and 2.4](#). (b) Effective resolution bandwidth versus SNDR for A/D converters shown in [Tables 2.1, 2.2, 2.3, and 2.4](#)

which leads to low power consumption. Resolutions up to 15 bits can be covered with two-step/multi-step/subranging converters or with an architecture which is a combination of these. Pipeline and two-step/multi-step converters tend to be the most efficient at achieving a given resolution and sampling rate specification. Its power proficiency for high resolution will be demonstrated with prototype [\[37\]](#) in the next chapter.

Low-Power High-Resolution Analog to Digital Converters
Design, Test and Calibration

Zjajo, A.; Pineda de Gyvez, J.

2011, XX, 250 p. 100 illus. in color., Hardcover

ISBN: 978-90-481-9724-8