SUMMARY This paper discusses recent trends in the area of low-power, high-performance A/D conversion. We examine survey data collected over the past twelve years to show that the conversion energy of ADCs has halved every two years, while the speed-resolution product has doubled approximately only every four years. A closer inspection on the impact of technology scaling, and developments in ADC design are then presented to explain the observed trends. Finally, we review opportunities in digitally assisted design for the most popular converter architectures. 

key words: A/D conversion, digital calibration, digitally assisted design, technology scaling

2. ADC Performance Trends

2.1 Survey Data and Figure of Merit Considerations

Several surveys on ADC performance are available in literature [5]–[8]. In this section, we will review recent data from designs presented at the IEEE International Solid-State Circuits Conference (ISSCC) and the VLSI Circuit Symposium. Figure 1 shows a scatter plot of results published at these venues over the past twelve years [9]. Figure 1(a) plots the conversion energy per Nyquist sample (\( E_{\text{conv}} \)) against the achieved signal-to-noise-and-distortion ratio (SNDR). This plot purposely avoids dividing the conversion energy by the effective number of quantization steps (\( ENOB \)), as done e.g. in the commonly used figure of merit [5]

\[
FOM = \frac{P}{f_s \cdot 2^{ENOB}}
\]

(1)

where

\[
ENOB = \frac{\text{SNDR (dB)} - 1.76}{6.02}
\]

(2)

Normalizing by the number of quantization steps assumes that the doubling precision of a converter will double its power dissipation, which finds only empirical justification [5]. Fundamentally, if a converter were purely limited by thermal noise, its power would actually quadruple per added bit (see Sect. 3). Nonetheless, since this assumption is somewhat pessimistic for real designs, it is preferable to avoid a fixed normalization between precision and energy altogether when plotting data from a large range of architectures and resolutions. The FOM given by (1) is useful mostly for comparing designs with similar resolution.

Figure 1(a) indicates that the lowest conversion energy is achieved by ADCs with low to moderate resolution, i.e. SNDR < 60 dB. In terms of the FOM given by (1), these designs lie within or near the range of 10–100 fJ/conversion-step, included as dashed lines in Fig. 1(a).

In addition to an ADC’s conversion energy, the available signal bandwidth is an important parameter. Figure 1(b) plots bandwidth against SNDR for the given data
set. In this chart, the bandwidth plotted for Nyquist converters is equal to the input frequency used to obtain the stated SNDR (this frequency is not necessarily $f_s/2$). The first interesting observation from Fig. 1(b) is that across all resolutions, the converters with the highest bandwidth achieve a performance that is approximately equivalent to an aperture uncertainty between 0.1 and 1 ps rms. The dashed lines in Fig. 1(b) represent the performance of ideal samplers with a sinusoidal input and 0.1 and 1 ps rms sampling clock jitter, respectively. Clearly, any of the ADC designs at this performance front rely on a significantly better clock, to allow for additional nonidealities that tend to reduce SNDR as well. Such nonidealities include quantization noise, thermal noise, differential nonlinearity and harmonic distortion. From the data in Fig. 1(b), it is also clear that any new design aiming to push the speed-resolution envelope will require a sampling clock with a jitter of 0.1 ps rms or better.

In order to assess the overall merit of ADCs (conversion energy and bandwidth), it is interesting to compare the locations of particular design points in the plots of Figs. 1(a) and (b). For example, [1] achieves a bandwidth close to the best designs, while showing relatively high conversion energy. The opposite is true for [10]; this converter is the lowest energy ADC published to date, but was designed only for moderate bandwidth. These examples confirm the intuition that pushing a converter toward the speed limits of a given technology will sacrifice efficiency and increase the conversion energy. To show this more generally, Fig. 2 plots the conversion energy of ADCs providing a bandwidth of 100 MHz or more. As we see from comparison with Fig. 1(a), there is now only one design [11] that falls below the 100 fJ/conversion step line. The tradeoff between energy conversion bandwidth is hard to analyze in general terms. There are however, architecture specific closed form results, e.g. as presented in [12] for a pipeline ADC.

Yet another important (but often neglected) aspect in ADC performance comparisons is the converter’s input capacitance (or resistance) and full-scale range. For most ADCs with a sampling-front-end, it is possible to improve the SNDR by increasing the circuit’s input capacitance. Unfortunately, it is difficult to construct a fair single-number figure of merit that includes the power needed to drive the converter input. This is mostly because the actual drive energy will strongly depend on the driver circuit implementation and its full specifications. Therefore, an alternative approach is to calculate a proxy for the input drive energy based on fundamental trade-offs. Such a metric was recently proposed as [13], [14]

$$E_{Q, in} = \frac{C_{in} \cdot ELSB^2}{2^{ENOB}}$$  \hspace{1cm} (3)$$

In this expression, ELSB is the effective size of the quantization step, i.e. the full-scale range (in volts) divided by the number of effective quantization steps ($2^{ENOB}$). For a converter that is limited by matching or thermal noise, the product in the numerator of (3) is independent of resolution, and captures how efficiently the capacitance is put to work. Normalizing by $2^{ENOB}$ then yields a metric similar to (1) in which the energy is distributed across all quantization levels.
The figure of merit defined by (3) is useful as a standalone metric to compare ADCs with otherwise similar performance specs in terms of their input energy efficiency. Consider e.g. the ~9-10-bit, 50-MS/s designs described in [15] (SAR ADC with $C_{in} = 5.12 \text{ pF}$, FOM = 52 fJ/conversion-step) and [16] (pipeline ADC with $C_{in} = 90 \text{ fF}$ and FOM = 119 fJ/conversion-step). For the SAR ADC of [15], we find $E_{G,in} = 1.1 \cdot 10^{-19} \text{ J/step}$, while for the pipeline design of [16], we obtain $E_{G,in} = 2 \cdot 10^{-21} \text{ J/step}$. This result indicates that the drive energy for the pipeline design is approximately two orders of magnitude lower when compared to the SAR design. Whether this is a significant advantage depends on the particular application and system where the converter will be used.

2.2 Trends in Power Efficiency and Speed

Using the data set discussed above, it is interesting to extract trends over time. Figure 3(a) is a 3-D representation of the conversion energy data [Fig. 1(a)] with the year of publication included along the y-axis. The resulting slope in time corresponds to an average reduction in energy by a factor of approximately 2 every 1.9 years.

A similar 3-D fit could be constructed for bandwidth performance. However, such a fit would not convey interesting information, as the majority of designs published in recent years do not attempt to maximize bandwidth. This contrasts the situation with conversion energy, which is subject to optimization in most modern designs. In order to extract a trend on achievable bandwidth, Fig. 3(b) scatter-plots the speed-resolution products of the top three designs in each year. This metric is justified by the speed-resolution boundary observed from Fig. 1(b), in which the straight lines obey a constant product of BW and $\gamma_{\text{ENOB}}$.

A fit to the data in Fig. 3(b) reveals that speed-resolution performance has doubled every 3.6 years; a rate that is significantly lower than the improvement in conversion energy. In addition, as evident from the data points, there is no pronounced trend as far as the top performance point is concerned; designs of the early 2000’s are almost up to par with some of the works published recently. Consequently, the extracted progress rate of speed-resolution performance should be viewed as a relatively weak and error-prone indicator.

3. Impact of Technology Scaling

As discussed above, the power dissipation of A/D converters has halved approximately every 2 years over the past twelve years. Over the same period, CMOS technologies used to implement the surveyed ADCs have scaled from approximately 0.6 $\mu$m down to 45 nm. Today, the choice of technology in which an A/D converter is implemented strongly depends on the application context. For standalone parts, older technologies such as 0.18-$\mu$m CMOS are often preferred (see e.g. [17]). In contrast, most embedded A/D converters usually must be implemented in the latest technologies used to realize large systems-on-chip [18]. Since the number of designs used in embedded SoC applications clearly outweighs the number of standalone parts, we have seen many ADC implementations in aggressively scaled technology over the past several years. Therefore, we will now investigate the role of technology scaling in the context of the trends summarized in the previous section. Broader discussions on the general impact of scaling on analog circuits are presented in [8], [19], [20].

3.1 Supply Voltage and Thermal Noise Considerations

A well-known issue in designing ADCs in modern processes is the low voltage headroom. Since device scaling requires a reduction in supply voltage ($V_{DD}$), the noise in the analog signals must be reduced proportionally to maintain the desired signal-to-noise ratio. Since noise trades with power dissipation, this suggests to first order that power efficiency should worsen, rather than improve, for ADCs in modern technologies.

One way to overcome supply voltage limitations is to utilize thick-oxide I/O devices [21], which are available in most standard CMOS processes. However, using I/O devices usually reduces speed. Indeed, a closer inspection of the survey data considered in this paper reveals that most published state-of-the-art designs do not rely on thick oxide devices, and rather cope with supply voltages around 1 V.
To investigate further, it is worthwhile to examine the underlying equations that capture the trade-off between supply voltage and energy via thermal noise constraints. In most analog sub-circuits used to build ADCs, noise is inversely proportional to capacitance

$$N \propto \frac{kT}{C}$$  \hspace{1cm} (4)

where k is Boltzmann’s constant and T stands for absolute temperature. For the specific case of a transconductance amplifier that operates linearly, we can write

$$f_s \propto \frac{g_m}{C}$$  \hspace{1cm} (5)

Further assuming that the signal power is proportional to \((\alpha \cdot V_{DD})^2\) and that the circuit’s power dissipation is \(V_{DD}\) multiplied by the transistor drain current, \(I_D\), we find

$$\frac{P}{f_s} \propto \frac{1}{\alpha^2 V_{DD}} \cdot \frac{1}{(g_m/ID)} \cdot kT \cdot SNR$$  \hspace{1cm} (6)

The variable \(g_m/ID\) in (6), is related to the gate overdrive \((V_{GS} - V_t)\) of the transistors that implement the transconductance. Assuming MOS square law, \(g_m/ID = 2/(V_{GS} - V_t)\) and in weak inversion \(g_m/ID = 1/(n \cdot kT/q)\), with \(n \approx 1.5\). Considering the fractional swing \(\alpha\) and transistor bias point \((g_m/ID)\) as constant, it is clear from (6) that the energy in noise-limited transconductors should deteriorate at low \(V_{DD}\). In addition, we see that (6) indicates a very steep tradeoff between SNR and energy; increasing the SNR by 6 dB requires a 4x increase in \(P/f_s\).

Since both of these results do not correlate with the observations of Sect. 2, it is instructive to examine the assumptions that lead to (6). The first assumption is that the circuit is purely limited by thermal noise. This assumption holds for ADCs with very high resolution, but typically few, if any, low resolution converters are severely impaired by thermal noise.

To get a feel for typical SNR values at which today’s converters become “purely” limited by noise, it is helpful to plot the data of Fig. 1(a) normalized to a 4x power increase per bit [22]. Figure 4 shows such a plot in which the \(P/f_s\) values have been divided by

$$\left(\frac{P}{f_s}\right)_{min} = 4 \cdot kT \cdot SNR$$  \hspace{1cm} (7)

while assuming \(SNR \approx SNDR\). The pre-factor of 4 in this expression follows from the power dissipated by an ideal class-B amplifier that drives the capacitance \(C\) with a rail-to-rail tone at \(f_s/2\) [23]. Therefore, (7) represents a fundamental bound on the energy required to process a charge sample at a given SNR.

The main observation from Fig. 4 is that the normalized data exhibits a visible “corner” beyond which \((P/f_s)/(P/f_s)_{min}\) approaches a constant value. This corner, approximately located at 75 dB, is an estimate for the SNDR at which a typical state-of-the-art design becomes truly limited by thermal noise. Since ADCs with lower SNDR do not achieve the same noise-limited conversion energy, it can be argued that these designs are at least partially limited by the underlying technology. This implies that over time, technology scaling may have helped improve their energy significantly as opposed to the worsening predicted by (6).

To investigate further, we partitioned the data from Fig. 1(a) into two distinct sets: high resolution (SNDR > 75 dB) and low-to-moderate resolution (SNDR ≤ 75 dB). We then applied a 3-D fit similar to that shown in Fig. 3(a) to each set and extracted the progress rates over time. For the set with SNDR > 75 dB it was found that the conversion energy has halved only every 4.4 years, while for SNDR ≤ 75 dB, energy halves every 1.7 years. The difference in these progress rates confirms the above speculation. For high-resolution designs, (6) applies and scaling technology over time, associated with lower supply voltages, cannot help improve power efficiency. As observed in [8], this has led to a general trend toward lower resolution designs: since it is more difficult to attain high SNDR at low supply voltages, many applications are steered away from using high-resolution ADCs in current fine-line processes. This is qualitatively confirmed in Fig. 5, which highlights the conversion energy data points of ADCs built in CMOS at 90 nm (and \(V_{DD} \approx 1\) V) and below. As we can see from this plot, only a limited number of ADCs built in 90 nm and below target SNDR > 75 dB.

3.2 Example: Impact of Scaling on a Moderate Resolution Pipeline ADC

As we have seen from the above discussion, low-to-moderate resolution converters have benefited strongly from technology scaling. Unfortunately, quantifying the benefits of scaling for these ADCs from first principles is a complex task, primarily because the involved tradeoffs strongly de-
depend on architecture and design specifics. In this subsection, we will therefore only discuss qualitatively the scaling behavior of a moderate resolution pipeline ADC as an example. An analysis that highlights the benefits of scaling in flash and folding ADCs is presented in [24].

Consider the 10-bit, 0.6-μm pipeline ADC described in [25], [26]; this design reflects state-of-the-art in 1996. Close inspection of the design details in [26] reveals that about 30% of the total power in this ADC is dissipated by noise-limited amplifiers. The remaining power is consumed by digital gates, comparators and amplifier stages whose component sizes are set by feature size constraints. To first order, the power dissipation in these latter blocks should scale approximately as $C \cdot V_{DD}^2$ i.e. digital logic gate energy. Since 1997, we have seen a reduction in process $C \cdot V_{DD}^2$ of approximately 325 times (see Sect. 5). Assuming that the power dissipation in the noise limited amplifiers stays constant, this reduction can explain a decrease in overall power dissipation of no more than approximately 3 times. Since we have seen much larger improvements for 10-bit designs in reality, there must be other factors that play a significant role.

For the particular case of pipeline ADCs, a first argument can be constructed based on amplifier self-loading. A circuit portion that is limited by noise typically still carries overhead that reduces with scaling. Especially in high-speed designs, amplifier self-loading and parasitic loading at intermediate circuit nodes plays an important role [27]. Technology scaling helps mitigate these capacitances and therefore improves overall efficiency.

A second factor to consider is the accrual of design experience, improved optimization, the exploitation of process options and refinement of circuit techniques over many generations of technology. For instance, A/D converters in 5-V technologies used to be relatively wasteful in terms of headroom utilization [$\alpha$-term in (6)]. Newer designs are typically optimized to accommodate signal swings as large as $1V_{pp,diff}$ at $V_{DD} = 1$ V. In addition, we are beginning to see designs that efficiently exploit technology options. For instance, the 10-bit pipelined ADC of [28] uses thin-oxide high-performance analog (HPA) devices to achieve high DC gain using simple, power-efficient telescopic transconductance amplifiers.

Lastly, an additional and more general factor to consider is the tradeoff between the transconductor efficiency ($g_m/I_D$) and the transit frequency ($f_T$) of the active devices.

3.3 $g_m/I_D$ and $f_T$ Considerations

Switched capacitor circuits based on class-A operational transconductance amplifiers typically require transistors with $f_T \approx 50 \cdot f_s$ [29]. Even for speeds of several tens of MS/s, it was necessary in older technologies to bias transistors far into strong inversion ($V_{GS} - V_t > 200 \text{ mV}$) to satisfy this requirement. In more recent technologies, very large transit frequencies are available in moderate- and even weak-inversion. This is further illustrated in Fig. 6(a) which compares typical minimum-length NMOS devices in 180-nm and 90-nm CMOS.

For a fixed sampling frequency, and hence fixed $f_T$ requirement, newer technologies deliver higher $g_m/I_D$. This tradeoff is plotted directly, without the intermediate variable $V_{GS} - V_t$, in Fig. 6(b). In order to achieve $f_T = 30 \text{ GHz}$, a 180-nm device must be biased such that $g_m/I_D \approx 9 \text{ S/A}$. In 90-nm technology, $f_T = 30 \text{ GHz}$ is achieved in weak inversion, at $g_m/I_D \approx 18 \text{ S/A}$. From (6), it is clear that this improvement can fully counteract the reduction in $V_{DD}$ when going to a newer process. Note, however, that this advantage can only materialize when the sampling speed is kept constant or at least not scaled proportional to the $f_T$ improvement. This was also one of the observations drawn from Fig. 2(b): a converter that pushes the speed envelope using a new technology typically won’t simultaneously benefit from scaling in terms of conversion energy.
3.4 Architectural Impact

As discussed above, the high transistor speed available in new technologies can be leveraged to improve the energy efficiency of analog blocks (such as amplifiers). For similar reasons, we can also argue that the high transistor speed has had a profound impact on architectural choices and developments. The very high-speed of transistors in 90-nm CMOS and below has led to a renaissance or invention of architectures that were either deemed slow of inefficient in the past. Examples in this category include successive approximation register ADCs [30] and binary search architectures (e.g. [31]). In addition, the high integration density of new processes makes massive time-interleaving (see e.g. [1]) with its associated benefits a possibility (see Sect. 5.2).

4. Minimalistic Design

In addition to technology scaling, the trend toward “minimalistic” and “digitally assisted” designs continues to impact the performance of A/D converters. In this section we will discuss ideas in minimalistic design, followed by a discussion on the importance of digitally assisted architectures in Sect. 5.

Power dissipation in the analog portion of ADCs is strongly coupled to the complexity of the constituent subcircuits. The goal of minimalistic design is to reduce power and potentially increase speed by utilizing simplified analog subcircuits. In architectures that previously relied on precision op-amp based signal processing, there exists a clear trend toward simplified amplifier structures. Examples include inverter-based delta-sigma modulators [32], [33] and various approaches emphasizing op-amp-less implementation of pipelined ADCs [16], [34]–[38]. Especially in switched capacitor circuits, eliminating class-A op-amps can dramatically improve power efficiency. This is for two reasons. First, operational amplifiers typically contribute more noise than simple gain stages, as for example resistively loaded open-loop amplifiers. Secondly, the charge transfer in class-A amplifier circuitry is inherently inefficient; the circuit draws a constant current, while delivering on average only a small fraction of this current to the load. In [39], it was found that the efficiency of a class-A amplifier in a switched capacitor circuit is inversely proportional to the number of settling time constants. For the typical case of settling for approximately 10 or more time constants, the overall efficiency, i.e. charge drawn from the supply versus charge delivered to the load, is only a few percent.

As discussed further in [40], this inherent inefficiency of op-amps contributes to the power overhead relative to fundamental limits. Consider for instance the horizontal asymptote of Fig. 4, located at approximately 300 times the minimum possible $P_{1/f}$. The factor of 300 can be explained for op-amp based circuits as follows. First, the noise is typically given by $\beta \cdot kT/C$, where $\beta$ can range from 5–10, depending on implementation details. Second, charge transfer using class-A circuits, as explained above, brings a penalty of approximately 20x. Third, op-amp circuits usually do not swing rail-to-rail as assumed in (7); this can contribute another factor of two. Finally, adding further power contributors beyond one dominant op-amp easily explains a penalty factor greater than 200...400.

A promising remedy to this problem is to utilize circuits that process charge more efficiently and at the same time contribute less thermal noise. A well-know example of an architecture that achieves very high efficiency is the charge-based successive approximation register (SAR) converter, see e.g. [1], [10], [30]. Such converters have been popular in recent years, primarily because the architecture is well-suited for leveraging the raw transistor speed of new technologies, while being insensitive to certain scaling implications, such as reduced intrinsic gain ($g_m/g_{ds}$). A problem with SAR architectures is that they cannot deliver the best possible performance when considering absolute speed, resolution and input capacitance simultaneously (see Sect. 2.1). This is one reason why relatively inefficient architectures, such as op-amp based pipelined ADCs are still being used and investigated.

In order to make pipelined architectures as power efficient as competing SAR approaches, various ideas are being explored in research. Figure 7 shows an overview of amplification concepts that all pursue the same goal: improve the energy efficiency of the residue amplification.

In scheme (a), the traditional op-amp is replaced by a

![Fig. 7](image-url)

Fig. 7 Energy efficient switched capacitor amplification concepts. (a) Comparator based switched capacitor circuit [34]. (b) Bucket brigade circuit [35], [41]. (c) Dynamic source follower amplifier [16].
comparator [34], which shuts off the capacitor charging current when the final signal value is reached. In scheme (b) a “bucket brigade” pass transistor is used to move a sampled charge packet (q) from a large sampling capacitor (C_S) to a smaller load capacitor (C_L), thereby achieving voltage gain without drawing a significant amount of energy from the supply [35],[41]. Lastly, scheme (c) uses the gate capacitance of a transistor to acquire a charge sample (q_1, q_2). The transistor is then switched into a source-follower configuration, moving all signal charge (q_1 + q_2) to the small capacitance from gate to drain, which also results in voltage amplification [16].

A general concern with most minimalistic design approaches is that they tend to sacrifice robustness, e.g., in terms of power supply rejection, common mode rejection and temperature stability. It remains to be seen if these issues can be handled efficiently in practice. Improving supply rejection, for instance, could be achieved using voltage regulators. This is custom practice in other areas of mixed-signal design, as for example clock generation circuits [42],[43]. Especially when the power of the ADC’s critical core circuitry is lowered significantly, implementing supply regulation should be a manageable task.

A second issue with minimalistic designs is the achievable resolution and linearity. Op-amp circuits with large loop gain help linearize transfer functions; this feature is often removed when migrating to simplified circuits. For instance, the amplifier scheme of Fig. 7(c) is linear only to approximately 9-bit resolution. In cases where simplicity sacrifices precision, it is attractive to consider digital means for recovering conversion accuracy. Digitally assisted architectures are therefore the topic of the next section.

5. Digitally Assisted Architectures

Technology scaling has significantly reduced the energy per operation in CMOS logic circuits. As explained in [44], the typical 0.7x scaling of features along with aggressive reductions in supply voltage have led in the past to a 65% reduction in energy per logic transition for each technology generation.

As illustrated in Fig. 8, a 2-input NAND gate dissipates approximately 1.3 pJ per logic operation in a 0.5-μm CMOS process. The same gate dissipates only 4 fJ in a more recent 65-nm process; this amounts to a ~325 times improvement in only 12 years. The corresponding reduction in ADC conversion energy (based on Sect. 2) amounts to a 64x reduction over the same time. This means that the relative “cost” of digital computation (in terms of energy) has reduced substantially in recent years.

To obtain a feel for how much logic can be used to “assist” a converter for the purpose of calibration and error correction, it is interesting to express the conversion energy of ADCs as a multiple of NAND-gate energy. This is illustrated in Fig. 9 assuming E_{NAND} = 4 fJ and FOM = 100 and 500 fJ/conversion-step, respectively. At low resolutions, e.g. ENOB = 5, a single A/D conversion consumes as much energy as toggling approximately 1000 logic gates. On the other hand, at ENOB = 16, several million logic gates need to switch to consume the energy of a single A/D conversion at this level of precision.

The consequence of this observation is that in a low-resolution converter, it is unlikely that tens of thousands of gates can be used for digital error correction in the high-speed signal path without exceeding reasonable energy or power limits. A large number of gates may be affordable only if the involved gates operate at a low activity factor (e.g. if they are outside the signal path) or if they can be shared within the system. Conversely, in high resolution ADCs, each analog operation is very energy consuming and even a large amount of digital processing may be accommodated in the overall power or energy budget.

The following sub-sections provide a non-exhaustive discussion of opportunities for leveraging digital logic gates in the design of A/D converters.

5.1 Oversampling

The longest standing example of an architecture that ef-
ciently leverages digital signal processing abilities is the oversampling delta-sigma converter. This architecture uses noise shaping to push the quantization error outside the signal band [22]. Subsequent digital filtering creates a high-fidelity output signal, while the constituent analog subcircuits require only moderate precision. Even in fairly old technologies, it was reasonable to justify high gate counts in the converter’s decimation filter, simply because the conversion energy for typical high-SNDR converters is very large.

A new paradigm that might gain significance in the future is the use of significant oversampling in traditional Nyquist converters. An example of such an ADC is described in [45]. As we have noted from Fig. 6(b), migrating a converter with a fixed sampling rate to technologies with higher \( f_T \) no longer improves beyond a certain minimum gate overdrive; it therefore makes no sense to target a transistor \( f_T \) below a certain value. This, in turn, implies that for optimum power efficiency, one should not operate an ADC below a certain clock rate. Consider for example the \( f_T \) versus \( g_m/I_D \) plot for 45-nm technology in Fig. 6(b). For \( g_m/I_D > 20 \text{S}/\text{A} \), \( f_T \) drops sharply without a significant increase in \( g_m/I_D \). At this point, \( f_T \approx 50 \text{GHz} \), implying that it is still possible to build a switched capacitor circuit with \( f_{\text{clock}} \approx 50 \text{GHz} \). To date, there exist only a limited number of applications for moderate- to high speed ADCs that require such high sampling rates, and there will clearly remain a number of systems in the future that demand primarily good power efficiency at only moderate speeds. A solution to this situation could be to oversample the input signal by a large factor and to remove out-of-band noise (thermal noise, quantization noise, and jitter) using a digital filter. Per octave of oversampling, this increases ADC resolution by \( 1/2 \) bit. In a situation where a converter is purely limited by noise, this improvement is in line with the fundamental thermal noise tradeoff expressed in (6).

5.2 Time-Interleaving

Time-interleaved architectures [46] exploit time parallelism and trade hardware complexity for an increased aggregate sampling rate. Time-interleaving can be beneficial in several ways. First, it can help maximize the achievable sampling rate in a given technology. Second, time-interleaving can be used to assemble a very fast converter using sub-ADCs that do not need to operate at the limits of a given architecture or technology. As discussed above, this can help improve energy efficiency.

A well-known issue with time-interleaved architectures, however, is their sensitivity to mismatches between the sub-converters. The most basic issue is to properly match the offsets and gains in all channels. In addition, for the common case of architectures that do not contain a global track-and-hold circuit, bandwidth and clock timing mismatches are often critical [47].

Using digital signal processing techniques to address analog circuit mismatch in time-interleaved converter arrays is an active research area [48]. Basic algorithms that measure and remove gain and offsets in the digital domain have become mainstream (see e.g. [49]), while techniques that address timing and bandwidth mismatches are still evolving.

As with most digital enhancement techniques, the problem of dealing with timing and bandwidth mismatches consists of two parts: (1) an algorithm that estimates the errors, and (2) a mechanism that corrects the errors. For the correction of timing and bandwidth errors, digital methods have been refined substantially over the years [50]. Nonetheless, the complexity and required energy of the proposed digital filters still seems to be beyond practical bounds, even for today’s fine-line technology. As a consequence, timing errors in practical time-interleaved ADC are often adjusted through digitally adjustable delay lines [1], [2], [51], while the impact of bandwidth mismatches is typically minimized by design.

As far as estimation algorithms are concerned, there is a wide and growing variety of practical digital domain algorithms [48]. Particularly interesting are techniques that extract the timing error information “blindly,” without applying any test or training signals [52], [53]. It is clear that algorithms of this kind will improve further and find their applications in practical systems.

5.3 Mismatch Correction

Assuming constant gate area (W-L), transistor matching tends to improve in newer technologies. In matching-limited flash ADC architectures, this trend has been exploited in the past to improve the power efficiency by judiciously downsizing the constituent devices [14]. In order to scale such architectures more aggressively, and at the same time address new sources of mismatch in nano-scale technologies, it is desirable to aid the compensation of matching errors through digital means.

In flash-ADCs, there are several trends in this direction. As illustrated in Fig. 10, the first and most transparent idea is to absorb offset voltages \( V_{OS} \) in each comparator using dedicated “trimm DACs” or similar circuits that allow for a digital threshold adjustment [54]–[56]. The input code for each DAC can be determined at start-up or whenever the converter input is in an idle condition. Alternatively, and for improved robustness, it is also possible to adjust the trimm codes continuously in the background (during normal operation), e.g. using a chopper-based approach [57] or through a two-channel architecture [58].

An important aspect of the arrangement in Fig. 10 is that most of the digital circuitry required to control the trimm DACs is either static during normal operation or can run at very low speeds. This means that there is often no concern about the digital energy overhead for calibration.

In modern technologies, and when near-minimum size
devices are being used, the comparator offsets may necessitate relatively complex trimm-DACs to span the required range with suitable resolution. One idea to mitigate this issue is to include redundant comparators, and to enable only the circuits that fall into the (reduced) trimm-DAC range after fabrication [59]. This scheme can yield good power efficiency as it attacks the mismatch problem along two degrees of freedom. Larger offsets can now be accommodated using smaller trimm DACs, which in turn imposes a smaller overhead in terms of area and parasitic capacitance introduced to signal-path nodes.

The idea of using redundant elements can be pushed further to eliminate the trimm-DACs and the trimm-range issue altogether. In [46], redundancy and comparator reassignment are employed to remove all offset constraints. A similar concept is proposed in [45] (see also [60]), but instead of a static comparator reassignment, a fault tolerant digital encoder is used to average the statistics along the faulty thermometer code. While this approach is conceptually very elegant, it requires a relatively large number of logic operations per sample in the high-speed data path. In light of the data from Fig. 9 (for low ENOB values), such a solution may be efficient only for very aggressively scaled technologies, i.e. 45 nm and below.

This far, the concept of using redundant elements has not yet found widespread use. It remains to be seen if these techniques will become a necessity in technologies for which the variability of minimum-size devices cannot be efficiently managed using trimm-DACs. Extrapolating beyond flash-ADCs, it may one day be reasonable and advantageous to provide redundancy at higher levels of abstraction, as for instance through extra channels in a time-interleaved ADC array [61].

5.4 Digital Linearization of Amplifiers

As pointed out in Sect. 4, power-efficient and minimalistic design approaches are typically unsuitable for high-resolution applications, unless appropriate digital error correction schemes are used to enhance conversion linearity. A generic block diagram of such a scheme is shown in Fig. 11. In [36], it was demonstrated that a simple open-loop differential pair used in a pipeline ADC can be digitally linearized using only a few thousand logic gates. In ADCs, the concept of digital amplifier linearization has so far been applied to the pipelined architecture [36], [62]–[67]. However, it is conceivable to implement similar schemes in other architectures, e.g. delta-sigma modulators [68].

One key issue in most digital linearization schemes is that the correction parameters must track changes in operating conditions relatively quickly; preferably with time constants no larger than 1–10 ms. Unfortunately, most of the basic statistics-based algorithms for coefficient adaptation require much longer time constants at high target resolutions [62], [63]. Additional research is needed to extend the recently proposed “split-ADC” [69], [70] and feedforward noise cancellation techniques [71] for use in nonlinear calibration schemes.

5.5 Digital Correction of Dynamic Errors

Most of the digital correction methods developed in recent years have targeted the compensation of static circuit errors. However, it is conceivable that in the future the correction of dynamic errors will become attractive as well.

One opportunity is the digital correction of errors due to finite slew rate, incomplete settling or incomplete reset of amplifiers [39], [64], [72], [73]. Another opportunity lies in correcting high-frequency nonlinearities introduced at the sampling front-end of high-speed ADCs [74]–[76]. Correcting high-frequency distortion digitally is particularly attractive in applications that already have a complex digital backend that can easily provide additional resources (e.g. inside an FPGA). Such applications include for instance wireless base stations [74] and test and measurement equipment. Most digital high-frequency linearization schemes proposed to date are based on relatively simple, low-order nonlinearity models. However, if digital capabilities in nano-scale technologies continue to improve, dynamic compensation schemes based on relatively complex Volterra series may become feasible [77].

5.6 System-Synergistic Error Correction Approaches

In the discussion so far, ADCs were viewed as “black boxes” that deliver a set performance without any system level interaction. Given the complexity of today’s applications, it is important to realize that there exist opportunities to improve ADC performance by leveraging specific system and signal properties.
For instance, large average power savings are possible in radio receivers when ADC resolution and speed are dynamically adjusted to satisfy the minimum instantaneous performance needs. The design described in [78] demonstrates the efficacy of such an approach.

In the context of digital correction, it is possible to leverage known properties of application-specific signals to “equalize” the A/D converter together with the communication channel [79]–[84]. For instance, the converter described in [79] uses the system’s OFDM pilot tones to extract component mismatch information (see Fig. 12). In such an approach, existing system hardware, as for instance the FFT block, can be re-used to facilitate ADC calibration.

6. Conclusion

This paper has discussed recent trends in the context of low-power, high-performance A/D conversion. Using survey data from the past twelve years, we have observed that power efficiency in ADCs has improved at an astonishing rate of approximately 2x every 2 years. In part, this progress rate is based on cleverly exploiting the strengths of today’s technology. Smaller feature sizes help improve the power dissipation in circuits that are not limited by thermal noise. In circuit elements that are limited by thermal noise, exploiting the high fT of modern transistors can be of help in mitigating a penalty from low supply voltages.

A promising paradigm is the trend toward minimalistic ADC architectures and digital means of correcting analog circuit errors. Digitally assisted ADCs aim to leverage the low computing energy of modern processes to improve the resolution and robustness of simplified circuits. Future work in this area promises to fuel further progress in optimizing the power efficiency of A/D converters.

Overall, future improvements in ADC power dissipation are likely to come from a combination of aspects that involve improved system embedding and reducing analog sub-circuit complexity and raw precision at the expense of “cheap” digital processing resources.

References


Boris Murmann received the Dipl.-Ing. (FH) degree in communications engineering from Fachhochschule Dieburg, Dieburg, Germany, in 1994, the M.S. degree in electrical engineering from Santa Clara University, Santa Clara, CA, in 1999, and the Ph.D. degree in electrical engineering from the University of California, Berkeley, in 2003. From 1994 to 1997, he was with Neuron Mikrolektronik GmbH, Hanau, Germany, where he developed low-power and smart-power ASICs in automotive CMOS technology. Since 2004, he has been an Assistant Professor with the Department of Electrical Engineering, Stanford University, Stanford, CA. His research interests are in the area of mixed-signal integrated-circuit design, with special emphasis on data converters and sensor interfaces. In 2008, Dr. Murmann was a co-recipient of the Best Student Paper Award at the VLSI Circuit Symposium and the recipient of the Best Invited Paper Award at the Custom Integrated Circuits Conference (CICC). In 2009, he received Agilent Early Career Professor Award. He currently serves as a member of the International Solid-State-Circuits Conference (ISSCC) program committee.