IEEE 2005 CUSTOM INTEGRATED CIRCUITS CONFERENCE

Scaling of Analog-to-Digital Converters into Ultra-Deep-Submicron CMOS

Y. Chiu [1], B. Nikolić [2], and P. R. Gray [2]

[1] Electrical and Computer Engineering, University of Illinois at Urbana-Champaign
[2] Electrical Engineering and Computer Sciences, University of California at Berkeley

Abstract

This paper presents the opportunities and challenges for scaling A/D converters into ultra-deep-submicron CMOS technologies. With faster transistors and better matching, the trend is to migrate into higher sample rates with lower resolutions. Limited dynamic range at low supply voltages remains the utmost challenge for high-resolution Nyquist converters, and oversampling will become the dominant technique in this arena in the future. Linearity correction with digital calibration is also becoming prevalent as the efficiency of calibration circuitry improves.

Introduction

Explosive growth in wireless and wireline communications is the dominant driver for higher specifications of analog-to-digital converters (ADCs). New applications in wireless communications support multi-mode operation, utilize large portions of bandwidth, such as in the case of ultra-wideband and 60-GHz-band systems, or attempt to re-use the already licensed spectrum, thus requiring a high dynamic range for operation. Similarly, future wireline communication systems commonly extend the signal constellations to increase the data throughput, such as in the case of 10-Gb/s Ethernet or next-generation cable modems. These applications are driving the demand for high-resolution, high-speed, low power, and low cost integrated ADCs.

Technology scaling significantly lowers 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 and minimum oxide thickness to implement analog functions, because the improved device transition frequency, f_T, allows for faster operation. However, scaling adversely affects most other parameters relevant to analog designs. To achieve a high linearity, high sampling speed, high dynamic range, with low supply voltages and low power dissipation in ultra-deep-submicron CMOS technology is a major challenge. In this paper we explore the challenges for ADC design associated with technology scaling. We will examine some circuit, architectural and system design techniques that will allow analog-to-digital converters to utilize transistors available in sub-100-nm technologies.

Technology Divergence with Scaling

Technology scaling doubles the density of digital logic every 2-3 years. Digital circuits have additionally benefited from scaling through increased operating frequencies and lower power consumption. Scaling to 90-nm CMOS technology and beyond is characterized by limited power. To minimize the dissipation, by balancing the switching and leakage power, digital systems choose the appropriate supply voltages and transistor types. Current foundry offerings are characterized by several thin-oxide devices, with different implant-controlled thresholds, and supply voltages that are scaled below the reliability-dictated levels. Furthermore, the same process usually offers thick-oxide I/O devices.

Analog functions can be implemented using either thin- or thick-oxide devices – the thick-oxide devices enjoy the benefit of a larger dynamic range (DR) and the thin-oxide devices harvest a higher operation frequency. This trend will continue in the future as predicted by ITRS (Fig. 1) [87]. Recent requirements to control the digital circuit leakage have slowed down the transistor threshold scaling, with a consequent reduction in supply voltage scaling. As a result, fast analog circuits can use 1V or higher supplies in the next few technology nodes. With continued scaling and the reduction of supply voltages in sub-1V range, the I/O devices will follow to sub-1.8V levels, and analog functions that require a high dynamic range would have to use additional process features or would have to be implemented on a separate chip.

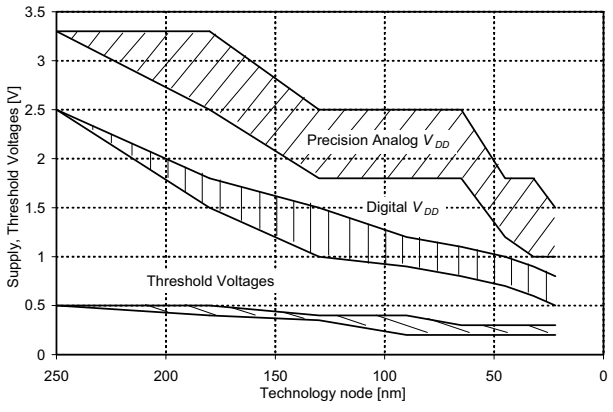


Fig. 1. Scaling of supply and threshold voltages.

Opportunities and Challenges in Scaling Analog Designs

As technology advances, there is an increased incentive for high-speed analog designs to exploit higher f_T of scaled transistors. A number of challenges presented by technology scaling however must be addressed.

Reduced Signal-to-Noise Ratio (SNR). The most prominent challenge for implementing precision analog circuitry in deeply scaled, "digital" processes is the reduction of supply voltages. It lowers the available voltage swings in analog circuits, fundamentally limiting the achievable SNR. To maintain the same dynamic range with a lower supply voltage in a noise-limited design, the circuit noise must also be proportionally reduced. For example, lowering the noise floor in switched-capacitor circuits requires an increase in the capacitor sizes to lower the kT/C noise, hence results in a penalty in power consumption.

Lower intrinsic gain. The intrinsic voltage gain ($g_m r_o$) of an MOS device is one important gauge of device performance for precision analog designs. As scaling 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. 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).

Device leakage. A fundamental advantage of MOS technology is the high quality switch naturally available. 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. Charge storage on capacitive devices will become difficult with leaky transistors attached. In addition, the gate leakage current also violates the high-impedance "summing-node" assumption that underlines the operation of switched-capacitor circuits, especially at lower speeds. In a sense, the gate leakage current resembles the base current in a bipolar junction transistor.

Matching. Transistor matching properties are improved with a thinner oxide [95] [96]. However, devices with small geometries also experience larger mismatch due to higher order terms with either short W or L [97]. When the oxide thickness is reduced to a few atomic layers, quantum effects will dominate and matching will degrade.

Passives. Deeply-scaled CMOS processes target digital applications and frequently lack for high quality passives – inductors and capacitors. Sampled-data systems rely on linear, low-parasitic, high-density capacitors. Either double-poly or metal-insulator-metal (MIM) capacitor requires extra mask layers, hence adds cost. An alternative is the vertical, fringing metal capacitor [101]. These capacitors may benefit from reduced metal pitch and increased number of metal layers with technology scaling.

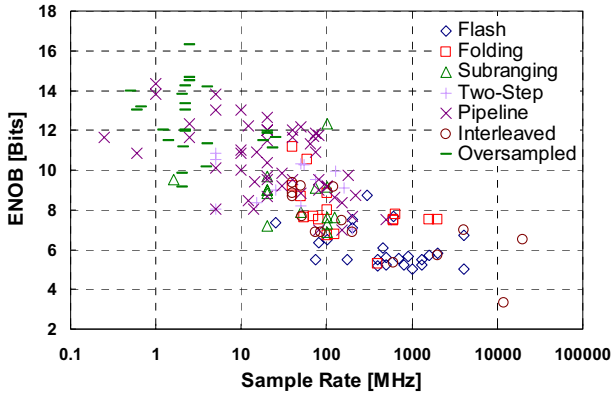


Fig. 2. ADC sample rate vs. ENOB from 1987 to 2005.

Scaling of ADC Architectures

The impact of technology scaling on the performance of various ADC architectures can be dichotomized according to the SNR specifications. While offering steadfast improvement of f_T, the accompanying reduction of supply voltages and the increased channel noise tend to shrink the dynamic range achievable by mixed-signal circuits in deeply scaled CMOS. For high-resolution converters, this inevitably leads to an increase of power consumption to maintain SNR. However, the accuracy of lower-resolution ADCs is limited by component mismatch, which results in a power and area scaling trend similar to that of the digital circuits for fixed conversion speeds (Table 1). The point of watershed at the current technology nodes seems to be between 8 and 10 bits.

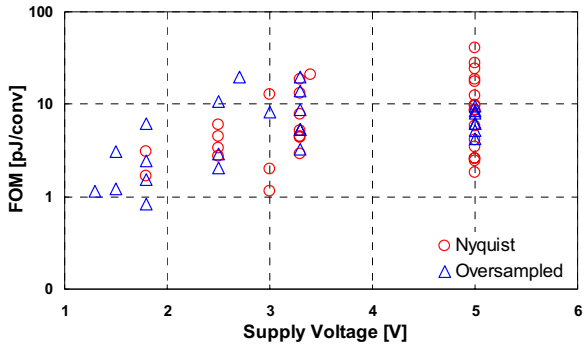
As scaling continues, there is a noticeable trend of a constant migration of the boundaries amongst the conventional A/D architectures (Fig. 2). While oversampled converters are encroaching into the regime used to be dominated by piepline ADCs [73], the later ones are reporting resolutions as low as 5-8 bits, which were only considered suitable for flash-type architectures [37].

A. Flash Converter

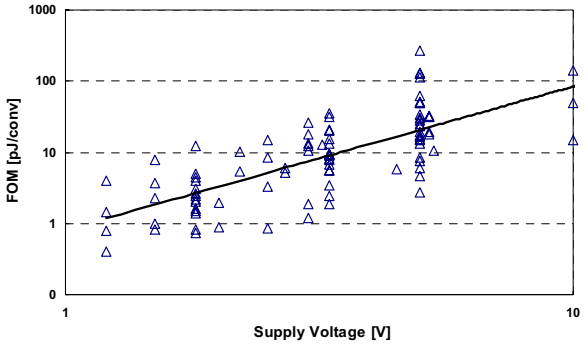
Flash converters are suitable for low-resolution (4-6 bits) applications that require high conversion rates (up to tens of gigahertz) and low latency. Although not a power-efficient architecture, the low latency feature makes it attractive in high-speed communication applications. Flash converters fall into the category of matching-limited scaling scenario.

Table 1. Scaling of Mixed-Signal Circuits

Scaling Parameter	SNR-Limited	Matching-Limited	Digital Circuits
Dynamic Range	$\propto \sqrt{C/kT}$	$\propto \sqrt{WL}$	Word length
Supply Voltage	$1/S$	$1/S$	$1/S$
Speed	1	1	S
Area	S^2	$1/S^2$	$1/S^2$
Power	S	$1/S^2$	$1/S^2$



(a)



(b)

Fig. 3. ADC scaling dichotomy, (a) SNR limited (≥ 12 bits) and (b) matching limited (≤ 10 bits). Data is taken from Fig. 2.

The major source of error in flash-type converters are the static and dynamic offsets sensed by the comparators, including offset errors originating from the comparators, the preamps, the reference ladders, and the relative timing skew of the strobe signals. Out of these, the dominant static offset often derives from the random threshold mismatch in the input devices of the preamps, which are used to suppress the large dynamic errors of small comparators and to mitigate kickback noise. As thin oxide improves the matching property of transistors [92] [96], smaller devices can be 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 figure-of-merit (FOM) or energy efficiency of the conversion

$$FOM = \frac{P}{2^{ENOB} \cdot f}, \qquad (1)$$

Where, f is the sample rate for a Nyquist ADC and twice the effective-resolution bandwidth in oversampled ones. This observation has been confirmed by Fig. 3b – the plot of FOM vs. supply voltage on a log-log scale shows a trend line with a slope of 2, which is predicted by column three of Table 1.

Because of the speed concerns, simple gain stages, such as the resistively-loaded differential pairs, are often used as the preamps in flash-type converters (Fig. 4).

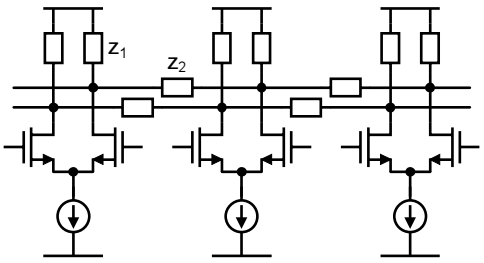
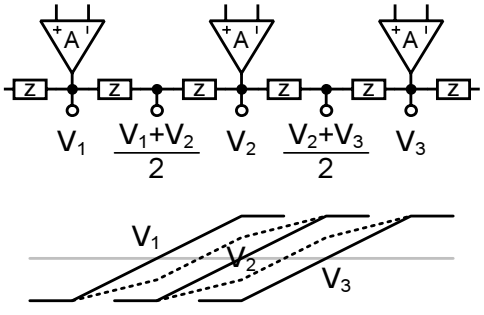


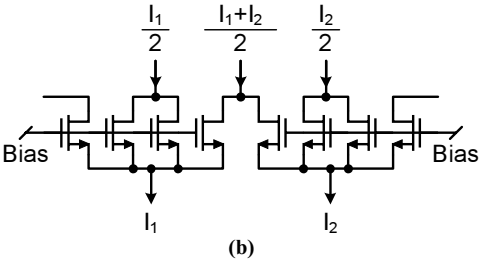
Fig. 4. Preamp with averaging network.

The simplicity of this circuit renders itself amenable to technology scaling and low supply-voltage operation. To further reduce the ADC power and input capacitance, resistive, capacitive [25] [20] [10], or current interpolation [17] can be performed to cut down the total number of preamps required (Fig. 5). The interpolation technique also leads to fewer points tapped off the reference ladder, resulting in a compact layout with less parasitics. Not only does a small layout improve the area and power efficiency, more importantly, it helps to keep the clock signal routing contained, which leads to a better dynamic performance of the converter.

Averaging (Fig. 4) is another common technique to improve the matching performance in preamps [9] [15] [4] [10]. Since the averaging effect improves in general with more amplifiers participating, a wider linear input range of the preamps and closely spaced reference voltages are beneficial. This is also in line with interpolation, where the active region overlap between adjacent differential pairs mitigates the interpolation error due to nonlinearity. Velocity saturation and lower supply voltages are driving flash converter designs converging toward this direction.



(a)



(b)

Fig. 5. Interpolation: (a) resistive/capacitive and (b) current.

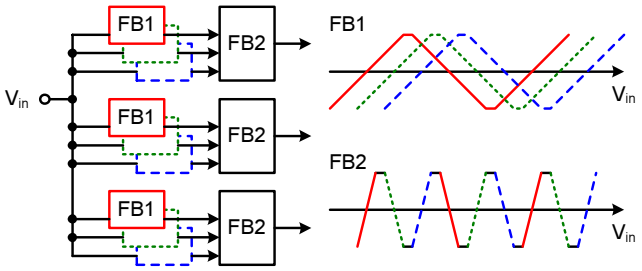


Fig. 6. Cascaded offset parallel folding.

B. Folding Converters

Interpolation reduces the number of preamps, but keeps the number of comparators the same. Signal folding has been developed to further improve the power efficiency of flash converters [16]-[19]. Compared to the flash architecture, the use of signal folding reduces the number of comparators by the folding factor F.[1] The combined hardware efficiency of folding, interpolation, and averaging has lead to CMOS realizations of flash converters with 1.6-GS/s sampling rate [11] or with resolutions of as high as 10-13 bits [13]-[15] at a reasonable power consumption.

The drawback of signal folding is that, if a large folding factor is developed in a single stage of folders, the bandwidth is greatly reduced due to large capacitive loads at the common output of many folding amplifiers; the maximum frequency seen at the folder output is given by

$$f_{\max} = \frac{\pi \cdot f_{in}}{\sin^{-1}(2/F)}, \qquad (2)$$

which can be many times higher than the maximum input frequency [16]. Furthermore, signal "rounding" problem at the folder output prevents further amplitude quantization, which has lead to the architecture of offset parallel folding with zero-crossing detection. This, however, leads to a large number of preamps used in the folders unless a significant amount of interpolation is also performed. Matching concern elevates for folders as, in addition to all other matching requirements, the current sources from different folding amplifiers also need to cancel each other precisely.

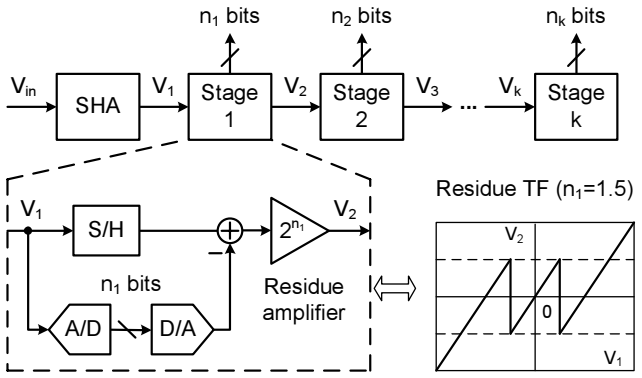


Fig. 7. 1.5-b/s pipeline ADC block diagram.

[1] Note that F coarse comparators are also required to resolve which fold the input signal resides in.

The above observation has lead to the design strategy to develop large folding factors successively by cascading folders with small folding factors [14] [15]. Pre-amplification is also increasingly necessitated by the folders to avoid active region overlap in adjacent folding amplifiers as the supply voltage is reduced, which also provides opportunities for averaging, interpolation, and pipelining.

C. Subranging Converter

The subranging architecture extends the resolution of flash ADCs by arranging the conversion in a coarse-fine two-step fashion. A sample-and-hold amplifier (SHA) and a switch matrix are usually required. In this architecture, latency is traded for low complexity and area/power efficiency. The achievable resolution in CMOS technology with this architecture is usually around 8-10 bits. 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. Averaging and interpolation can be applied as well to reduce the number of preamps and their sizes. A balanced design can often achieve an FOM close to that of the pipeline converters [20] [22] [23].

D. Multi-Stage and Pipeline Converter

Taking the subranging concept to the extreme, a multi-stage ADC resolves the analog input in a cascade of low-resolution stages by passing the conversion residue to the trailing stages. Residue gain is usually provisioned to suppress the noise and nonlinearity contribution of lower ranked stages. The pipeline ADC is a typical multi-stage converter, which inserts SHAs in the residue amplifier to facilitate concurrent operation of all stages to improve throughput at the cost of increased latency. Similar to a subranging converter, over/under-range protection is necessary in a pipeline ADC and is often termed "digital error-correction" (DEC). A commonly used 1.5-bit/stage architecture is illustrated in Fig. 7.

Since the comparator offset specs are substantially relaxed due to a low stage resolution and the DEC, comparator design in pipeline ADCs 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 technology advancement will keep pushing the nondominant poles of these amplifiers to higher frequencies, hence offer the potential of a higher conversion speed.

Nonetheless, the above statement is true only when a close-to-minimum channel length is used. At these dimensions, the accompanying short-channel effects 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 (Fig. 3a).

One such fundamental limit is the dynamic range. The sampling process inherent in switched-capacitor circuits introduces kT/C noise at each pipeline stage when a residue voltage is captured. The sampled noise usually comprises two major contributions – the channel noise of the switches and the amplifier noise. As the switch resistance is only weakly affected by technology scaling [97], it is expected that amplifier will become the dominant noise source in pipeline ADCs in deeply scaled technologies. Based on this observation, efforts were directed to search the optimum stage resolution n and scaling factor γ to minimize the total conversion power [53] [55]. It has been recently pointed out that, for a uniform n and γ, the total ADC power is expressed as

$$P = SNR \cdot kT \cdot f_s \cdot \left(\frac{V_{gs} - V_{Th}}{V_{DD}} \right) \cdot g(n, \gamma, \eta), \qquad (3)$$

where f_s is the sample rate, V_{gs}-V_{Th} is the overdrive voltage of the amplifier input transistors, and $\eta \in [0,1]$ is the speed factor that models the parasitic loading effect depending on the conversion speed [36]. The evaluation of the function $g(.)$ reveals that a choice of 2-3-bit/stage resolution yields the optimum architecture for high-speed pipeline converters.

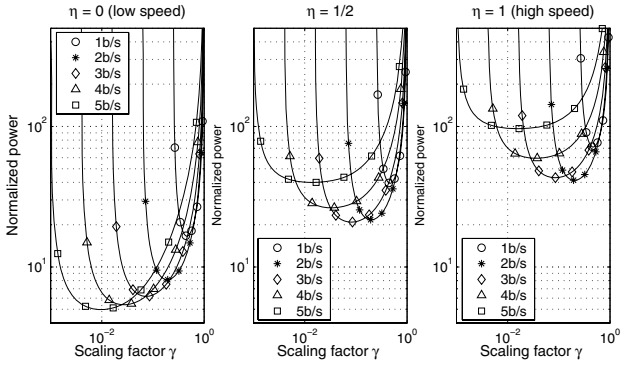


Fig. 8. Optimizing the stage resolution and scaling factor of a pipeline ADC at different sample rate.

E. Oversampled Converter

As mentioned before, technology scaling is commensurate with high sample rate but not high resolution. One technique to trade speed for dynamic range is oversampling – circuit and quantization noises are spread out and later suppressed by a digital decimation filter. Noise shaping is often facilitated to push quantization noise (and part of the circuit noise) further out of the interested signal band. The noise shaper often takes the form of a sigma-delta modulator.

Although oversampling plus noise shaping enables the use of smaller capacitors, other design constraints of the modulator such as linearity share similar concerns as the Nyquist converters.[2] Specifically, the signal swing at the amplifier output is increased due to the lag of the modulator feedback loop in response to a change in the input signal; a high oversampling ratio also stresses the amplifiers for settling speed. However, the decimation filter, being a digital circuit, greatly benefits from technology scaling. It is expected that oversampled ADCs will outperform pipeline converters for high-SNR applications in the near future.

F. Interleaved Converter

Interleaving was originally introduced to improve the power efficiency of a single-path converter as the sample rate is pushed close to the limit set by the fabrication technology [71]. Although the conversion speed of interleaved ADC arrays always surpass their single-path counterparts, the inherent problems of path mismatch and sample clock skew amongst the parallel paths substantially limit the attainable resolution of this converter. A single frontend SHA helps to mitigate the clock skew limitation but often has to be implemented in a different technology due to its high clock rate. Therefore, interleaved ADCs are suitable for 6-8-bit applications that require extremely high speed. Calibration is often engaged to improve the path matching condition. In spite of these difficulties, an interleaved pipeline converter clocked at 20 GS/s has recently been demonstrated [62].

G. Linearity Enhancement with Digital Techniques

It has long been noticed that conversion nonlinearity resulted from component mismatch (such as capacitor ratio errors) can be remedied by recording the error in a memory and successively removed by trimming or digital post processing [59] [91]. It is especially suitable to Nyquist A/D architectures that use an algorithmic approach (non flash-type). A genre of the techniques, a.k.a. digital calibration, were later proliferated to treat amplifier offset and gain errors, switch-induced errors in algorithmic and multi-stage ADCs, and the path mismatch problem in interleaved ADCs [49-52] [54-56] [58] [66] [67] [90]. The recent flurry of research activities in this direction is most likely motivated by Moore's law, which makes the digital calibration circuitry increasingly smaller and more power-efficient [27] [31-33] [35] [38-40] [46] [62] [63] [87] [89].

Just as the oversampling technique trades speed for SNR, the calibration technique can trade digital complexity for precision. Therefore it presents a major opportunity for the design of high-performance ADCs in ultra-deep-submicron CMOS. The combination of oversampling, noise shaping, calibration, and higher device f_T may constitute a viable approach for future high-DR, high-accuracy, and high-speed A/D converters. Hitherto, calibration works on ΣΔ converters have been reported [76] [84]. The concept of digital trimming was also recently reported in a folding ADC [11].

[2] Strictly speaking, only the front stage is subject to these constraints while the input-referred noise and nonlinearity of later stages are much attenuated by the front stage.

Nonetheless, digital treatment on error parameters that are supply voltage and temperature dependent poses serious challenges on the adaptation speed of the calibration algorithm used. This may limit the types of errors that can be reliably treated. The use of simple analog building blocks in the critical paths of high-precision converters also invites common-mode rejection-ratio (CMRR) and power-supply rejection-ratio (PSRR) issues that are typically not included in the calibration loop.

Scaling of Building Blocks

Technology scaling affects the design choices for various ADC building blocks. Amplifiers, sample-and-hold (S/H) circuits and comparators are common building blocks for many A/D architectures.

A. Operational Amplifier

In frontend S/H amplifiers or multi-stage ADCs, precision op amps are almost invariably employed to relay the input signal (or the residue signal) to the trailing conversion circuits. Negative feedback is usually engaged to establish signal transfer fidelity. Operating on the edge of the performance envelope, op amps exhibit intense tradeoffs 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. While technology advancement offers an attractive reward of wider bandwidth, hence a higher sampling rate, with high f_T devices, concerns for matching, low intrinsic gain, hot carrier effect, and sensitivity to process variations often direct a seasoned designer away from using transistors of minimum channel lengths in high-performance op amps.

Nonetheless, resorting to long-channel devices has only a limited capability to recover the needed open-loop gain to achieve a 10-bit or higher conversion accuracy. Multi-stage amplifiers and gain-boosted single-stage amplifiers are becoming popular choices at low supply voltages. Although the necessity for frequency compensation renders a multi-stage amplifier less power-efficient, it may become the only viable op-amp architecture at a supply voltage of less than 1V. CMOS gain boosting (a.k.a. active cascode) exploits the fact that when the load is purely capacitive, the DC gain of an op amp can be increased by enhancing the cascoding effect using negative feedback [92]-[94]. The technique is effective in boosting the output resistance R_o with negligible effect on the effective G_m. Recently, it was demonstrated that further gain is attainable through a recursive gain-boosting technique [36]. The ultimate limit of this approach would be the direct current path to ground at the drains of the cascode devices due to hot carrier induced substrate current and the effective gate resistance due to oxide tunneling current.

In addition to various short-channel effects, the most prominent challenge presented by technology scaling is probably the reduced supply voltage. It has become increasingly difficult to maintain the dynamic range of an op amp while keeping its power consumption low. Recent designs all exploited the differential topology to improve SNR and noise immunity. Stressed by the ever-decreasing supply voltage, pseudo-differential op amps have also been explored more frequently [36] [42] [43]. However, with this architecture, care must be exercised to regulate the common-mode biasing in a multi-stage ADC to avoid unnecessary accumulation of offset voltages. Switched op amp is another approach to achieve high output swing and to save power (it can also function as the sampling switch). While the slow turn-on following a complete op-amp turn-off lead to low conversion speeds in earlier works, a recent attempt demonstrated an 8-bit, 200-MS/s pipeline ADC with partially switched op amps [30].

B. Sample-and-Hold

Inherent to the A/D conversion process is a sample-and-hold (S/H) circuit that resides in the frontend of a converter (and also between stages in a multi-stage 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).

In CMOS technology, switched capacitors are the preferred implementation of the S/H circuits (Fig. 9). The performance of these samplers can be gauged by the small-signal bandwidth when the switches are on and the ratio of the gate capacitance of the switch to that of the sampling capacitor. Technology scaling reduces the associated capacitance while keeping the switch on-resistance nearly constant [97]; not only does this improve the tracking bandwidth of the S/H, it also alleviates the charge injection problem during the turn-off of the sampling switch, which typically results in distortion as the dynamics of the switch-off is quite signal-dependent (even with bottom-plate sampling). In other words, the increase of f_T through technology scaling improves the linearity of the sampling switch. A rule-of-thumb is to use minimum channel length for switches when no critical matching/leakage requirement is concerned.

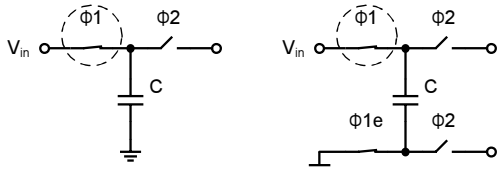


Fig. 9. Top- (left) and bottom- (right) plate sampling.

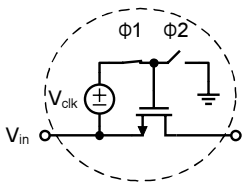


Fig. 10. Concept of clock bootstrapping.

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 [28] [44] [48] [55] [78] [83]. Care must be exercised to ensure that the reliability of the circuit is not compromised.

In most high-performance sampled-data acquisition interfaces, the sampling clock is derived directly or indirectly (e.g., through phase-locking) from an off-chip low-phase noise crystal oscillator. This setup often renders the on-chip clock buffer the dominant source of aperture jitter. The improvement of the rise/fall time of digital gates over each generation of technology desensitizes the clock buffer jitter performance to circuit and supply noises.

C. Preamp and Comparator

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. Circuit techniques addressing these issues [12] [20] will continue to be explored.

Summary

Sustained scaling of high-performance CMOS ADCs in the ultra-deep-submicron regime and the prospect for future performance attainable through technology scaling are analyzed and projected.

Acknowledgments

The authors are grateful to the funding supports provided by Intel Corp., Marco C2S2 and NSF EPDT #0238572.

References

Flash ADC

1. C. Paulus et al., "A 4GS/s 6b flash ADC in 0.13µm CMOS," *VLSI Symposium*, 2004, pp. 420-423.
2. K. Uyttenhove et al., "A 1.8-V 6-bit 1.3-GHz flash ADC in 0.25-µm CMOS," *JSSC*, vol. 38, pp. 1115-1122, July 2003.
3. K. Sushihara et al., "A 7b 450MS/s 50mW CMOS ADC in 0.3mm^2," *ISSCC*, 2002, pp. 170-171.
4. P. C. S. Scholtens et al., "A 6-b 1.6-Gsample/s flash ADC in 0.18-µm CMOS using averaging termination," *JSSC*, vol. 37, pp. 1599-1609, Dec. 2002.
5. M. Choi et al., "A 6-b 1.3-GS/s A/D converter in 0.35-µm CMOS," *JSSC*, vol. 36, pp. 1847-1858, Dec. 2001.
6. K. Sushihara, et al., "A 6b 800MS/s CMOS A/D converter," *ISSCC*, 2000, pp. 428-429.
7. K. Nagaraj, et al., "A dual-mode 700-MS/s 6-bit 200-MS/s 7-bit A/D converter in a 0.25-µm digital CMOS process," *JSSC*, vol. 35, pp. 1760-1768, Dec. 2000.
8. S. Tsukamoto, et al., "A CMOS 6-b, 400-MS/s ADC with error correction," *JSSC*, vol. 33, pp. 1939-1947, Dec. 1998.
9. K. Kattmann et al., "A Technique for Reducing Differential Non-linearity Errors in Flash ADC," *ISSCC*, 1991, pp. 170-171.

Folding ADC

10. W. Zheng-Yu, et al., "A 600-MSPS 8-bit folding ADC in 0.18-µm CMOS," *VLSI Symposium*, 2004, pp. 424-427.
11. R. C. Taft, et al., "A 1.8-V 1.6-GS/s 8-b self-calibrating folding ADC with 7.26 ENOB at Nyquist frequency," *JSSC*, vol. 39, pp. 2107, Dec. 2004.
12. G. Geelen et al., "An 8b 600MS/s 200mW CMOS folding A/D converter using an amplifier preset technique," *ISSCC*, 2004, pp. 254-255.
13. M. Choe, et al., "A 13-b 40-MS/s CMOS pipelined folding ADC w/ background offset trimming," *JSSC*, vol. 35, p. 1781, Dec. 2000.
14. P. Vorenkamp et al., "A 12-b, 60-MS/s cascaded folding and interpolating ADC," *JSSC*, vol. 32, pp. 1876-1886, Dec. 1997.
15. K. Bult and et al., "An embedded 240-mW 10-b 50-MS/s CMOS ADC in 1-mm^2," *JSSC*, vol. 32, pp. 1887-1895, Dec. 1997.
16. G. W. Venes and et al., "An 80-MHz, 80-mW, 8-b CMOS folding A/D converter with distributed track-and-hold preprocessing," *JSSC*, vol. 31, pp. 1846-1853, Dec. 1996.
17. M. P. Flynn and et al., "CMOS folding A/D converters with current-mode interpolation," *JSSC*, vol. 31, pp. 1248-1257, Sept. 1996.
18. B. Nauta and et al., "A 70-MS/s 110-mW 8-b CMOS folding and interpolating A/D converter," *JSSC*, vol. 30, pp. 1302-1308, Dec. 1995.
19. R. J. Van De Plassche et al., "A high-speed 7 bit A/D converter," *JSSC*, vol. 14, pp. 938, Jun. 1979.

Subranging ADC

20. J. Mulder et al., "A 21-mW 8-b 125-MS/s ADC in 0.09-mm^2 0.13-µm CMOS," *JSSC*, vol. 39, pp. 2116-2125, Dec. 2004.
21. R. C. Taft et al., "A 100-MS/s 8-b CMOS subranging ADC with sustained parametric performance from 3.8 V down to 2.2 V," *JSSC*, vol. 36, pp. 331, Mar. 2001.
22. B. P. Brandt et al., "A 75-mW, 10-b, 20-MSPS CMOS subranging ADC with 9.5 effective bits at Nyquist," *JSSC*, vol. 34, pp. 1788-1795, Dec. 1999.
23. M. Yotsuyanagi et al., "Mixed-mode subranging CMOS A/D converter," *JSSC*, vol. 30, pp. 1533-1537, Dec. 1995.
24. C. Mangelsdorf et al., "A two-residue architecture for multistage ADCs," *ISSCC*, 1993, pp. 64-65.
25. K. Kusumoto et al., "A 10-b 20-MHz 30-mW pipelined interpolating CMOS ADC," *JSSC*, vol. 28, pp. 1200-1206, Dec. 1993.

Two-Step ADC

26. M. Clara et al., "A 1.8 V fully embedded 10 b 160 MS/s two-step ADC in 0.18 µm CMOS," *CICC*, 2002, pp. 437-440.
27. H. van der Ploeg et al., "A 2.5-V 12-b 54-Msample/s 0.25-µm CMOS ADC in 1-mm^2 with mixed-signal chopping and calibration," *JSSC*, vol. 36, pp. 1859-1867, Dec. 2001.
28. H. Pan et al., "A 3.3-V 12-b 50-MS/s ADC in 0.6-µm CMOS with over 80-dB SFDR," *JSSC*, vol. 35, pp. 1769-1780, Dec. 2000.
29. R. Jewett et al., "A 12 b 128 MS/s ADC with 0.05 LSB DNL," *ISSCC*, 1997, pp. 138-139.

Pipeline ADC

30. H. Kim, "A 30mW 8b 200MS/s Pipelined CMOS ADC Using a Switched-Opamp Technique," ISSCC, 2005, pp. 284-285.
31. K. Nair et al., "A 96dB SFDR 50MS/s digitally enhanced CMOS pipeline A/D converter," *ISSCC*, 2004, pp. 456-457.
32. L. Hung-Chih et al., "A 15b 20MS/s CMOS pipelined ADC with digital background calibration," *ISSCC*, 2004, pp. 454-455.
33. C. R. Grace et al., "A 12b 80MS/s pipelined ADC with bootstrapped digital calibration," *ISSCC*, 2004, pp. 460-461.
34. B. Hernes et al., "A 1.2V 220MS/s 10b pipeline ADC implemented in 0.13µm digital CMOS," *ISSCC*, 2004, pp. 256-257.
35. E. Siragusa et al., "A digitally enhanced 1.8-V 15-bit 40-MS/s CMOS pipelined ADC," *JSSC*, vol. 39, pp. 2126-2138, Dec. 2004.
36. Y. Chiu et al., "A 14-b 12-MS/s CMOS pipeline ADC with over 100-dB SFDR," *JSSC*, vol. 39, pp. 2139-2151, Dec. 2004.
37. A. Varzaghani et al., "A 600MS/s, 5-bit pipelined analog-to-digital converter for serial-link applications," *VLSI Symposium*, 2004, pp. 276-279.
38. Y. Chiu et al., "Least mean square adaptive digital background calibration of pipelined analog-to-digital converters," *TCAS-I*, vol. 51, pp. 38- 46, Jan. 2004.
39. X. Wang et al., "A 12-bit 20-MS/s pipelined ADC with nested digital background calibration," *CICC*, 2003, pp. 409-412.

40. B. Murmann et al., "A 12-bit 75-MS/s pipelined ADC using open-loop residue amplification," *JSSC*, vol. 38, p. 2040, Dec. 2003.
41. B.-M. Min et al., "A 69-mW 10-bit 80-MS/s pipelined CMOS ADC," *JSSC*, vol. 38, pp. 2031-2039, Dec. 2003.
42. J. Li et al., "A 1.8-V 67mW 10-bit 100MSPS pipelined ADC using time-shifted CDS technique," *CICC*, 2003, pp. 413-416.
43. D. Miyazaki et al., "A 16mW 30MSample/s 10b pipelined ADC using a pseudo-differential architecture," *ISSCC*, 2002, pp. 174-175.
44. W. Yang et al., "A 3-V 340-mW 14-b 75-MS/s CMOS ADC with 85-dB SFDR at Nyquist input," *JSSC*, vol. 36, p. 1931, Dec. 2001.
45. M. Waltari et al., "1-V 9-bit pipelined switched-opamp ADC," *JSSC*, vol. 36, pp. 129-134, Jan. 2001.
46. J. Ming et al., "An 8-bit 80-Msample/s pipelined ADC with background calibration," *JSSC*, vol. 36, pp. 1489-1497, Oct. 2001.
47. L. Singer et al., "A 12b 65MS/s CMOS ADC with 82 dB SFDR at 120 MHz," *ISSCC*, 2000, pp. 38-39.
48. M. Abo et al., "A 1.5-V, 10-bit, 14.3-MS/s CMOS pipeline analog-to-digital converter," *JSSC*, vol. 34, pp. 599-606, May 1999.
49. J. M. Ingino et al., "A continuously calibrated 12-b, 10-MS/s, 3.3-V ADC," *JSSC*, vol. 33, pp. 1920-1931, Dec. 1998.
50. S.-U. Kwak et al., "A 15-b, 5-MS/s low-spurious CMOS ADC," *JSSC*, vol. 32, pp. 1866-1875, Dec. 1997.
51. K. Nagaraj, "Area-efficient self-calibration technique for pipelined algorithmic A/D converters," *TCAS-II*, vol. 43, p. 540, July 1996.
52. M. K. Mayes et al., "A 200-mW, 1-MS/s, 16-b pipelined A/D converter with on-chip 32-b microcontroller," *JSSC*, vol. 31, pp. 1862-1872, Dec. 1996.
53. D. W. Cline et al., "A power optimized 13-b 5-MS/s pipelined ADC in 1.2-µm CMOS," *JSSC*, vol. 31, pp. 294-303, Mar. 1996.
54. T.-H. Shu et al., "A 13-b 10-MS/s ADC digitally calibrated with oversampling delta-sigma converter," *JSSC*, vol. 30, pp. 443-452, Apr. 1995.
55. T. B. Cho et al., "A 10 b, 20 MS/s, 35 mW pipeline A/D converter," *JSSC*, vol. 30, pp. 166-172, Mar. 1995.
56. N. Karanicolas et al., "A 15-b 1-MS/s digitally self-calibrated pipeline ADC," *JSSC*, vol. 28, pp. 1207-1215, Dec. 1993.
57. S. H. Lewis et al., "A 10-b 20-MS/s analog-to-digital converter," *JSSC*, vol. 27, pp. 351-358, Mar. 1992.
58. S.-H. Lee et al., "Digital-domain calibration of multistep analog-to-digital converters," *JSSC*, vol. 27, pp. 1679-1688, Dec. 1992.
59. Y.-M. Lin et al., "A 13-b 2.5-MHz self-calibrated pipelined A/D converter in 3-µm CMOS," *JSSC*, vol. 26, pp. 628-636, Apr. 1991.

Interleaved ADC

60. S. Limotyrakis et al., "A 150MS/s 8b 71mW time-interleaved ADC in 0.18µm CMOS," *ISSCC*, 2004, pp. 258-259.
61. Draxelmayr, "A 6b 600MHz 10mW ADC array in digital 90nm CMOS," *ISSCC*, 2004, pp. 264-265.
62. K. Poulton et al., "A 20GS/s 8b ADC with a 1MB memory in 0.18µm CMOS," *ISSCC*, 2003, pp. 318-319.
63. X. Jiang et al., "A 2GS/s 6b ADC in 0.18µm CMOS," *ISSCC*, 2003, pp. 322-323.
64. S. M. Jamal et al., "A 10-b 120-MS/s time-interleaved analog-to-digital converter with digital background calibration," *JSSC*, vol. 37, pp. 1618-1627, Dec. 2002.
65. W. Ellersick et al., "GAD: A 12-GS/s CMOS 4-bit A/D converter for an equalized multi-level link," *VLSI Symposium*, 1999, pp. 49-52.
66. D. Fu et al., "A digital background calibration technique for time-interleaved analog-to-digital converters," *JSSC*, vol. 33, pp. 1904-1911, Dec. 1998.
67. K. C. Dyer et al., "An analog background calibration technique for time-interleaved analog-to-digital converters," *JSSC*, vol. 33, pp. 1912-1919, Dec. 1998.
68. K. Nagaraj et al., "A 250-mW, 8-b, 52-MS/s parallel-pipelined A/D converter with reduced number of amplifiers," *JSSC*, vol. 32, pp. 312-320, Mar. 1997.
69. K. Nakamura et al., "An 85 mW, 10 b, 40 MS/s CMOS parallel-pipelined ADC," *JSSC*, vol. 30, pp. 173-183, Mar. 1995.
70. CSG. Conroy et al., "An 8-b 85-MS/s parallel pipeline A/D converter in 1-µm CMOS," *JSSC*, vol. 28, pp. 447-454, Apr. 1993.
71. W. C. Black et al., "Time interleaved converter arrays," *JSSC*, vol. 15, pp. 1022-1029, Jun. 1980.

Oversampled ADC

72. R. Brewer, "A 100dB SNR 2.5MS/s Output Data Rate ΔΣ ADC," *ISSCC*, 2005, pp. 172-173.
73. A. Bosi, "An 80MHz 4x Oversampled Cascaded ΔΣ-Pipelined ADC with 75dB DR and 87dB SFDR," *ISSCC*, 2005, pp. 174-175.
74. K. Philips et al., "A continuous-time ΣΔ ADC with increased immunity to interferers," *JSSC*, vol. 39, pp. 2170-2178, Dec. 2004.
75. P. Balmelli et al., "A 25-MS/s 14-b 200-mW ΣΔ Modulator in 0.18-µm CMOS," *JSSC*, vol. 39, pp. 2161-2169, Dec. 2004.
76. J. Silva et al., "Digital techniques for improved ΔΣ data conversion," *CICC*, 2002, pp. 183-190.
77. R. Schreier et al., "A 10-300-MHz IF-digitizing IC with 90-105-dB dynamic range and 15-333-kHz bandwidth," *JSSC*, vol. 37, pp. 1636-1644, Dec. 2002.
78. S. K. Gupta et al., "A 64-MHz clock-rate ΣΔ ADC with 88-dB SNDR and -105-dB IM3 distortion at a 1.5-MHz signal frequency," *JSSC*, vol. 37, pp. 1653-1661, Dec. 2002.
79. Y. Geerts et al., "A high-performance multibit Delta-Sigma CMOS ADC," *JSSC*, vol. 35, pp. 1829-1840, Dec. 2000.
80. Fujimori et al., "A 90-dB SNR 2.5-MHz output-rate ADC using cascaded multibit delta-sigma modulation at 8x oversampling ratio," *JSSC*, vol. 35, pp. 1820-1828, Dec. 2000.
81. Namdar et al., "A 400-MHz, 12-bit, 18-mW, IF digitizer with mixer inside a sigma-delta modulator loop," *JSSC*, vol. 34, pp. 1765-1776, Dec. 1999.
82. V. Peluso et al., "A 900-mV low-power Delta-Sigma A/D converter with 77-dB dynamic range," *JSSC*, vol. 33, pp. 1887-1897, Dec. 1998.
83. T. L. Brooks et al., "A cascaded sigma-delta pipeline ADC with 1.25 MHz signal bandwidth and 89 dB SNR," *JSSC*, vol. 32, pp. 1896-1906, Dec. 1997.
84. J. W. Fattaruso et al., "Self-calibration techniques for a second-order multibit sigma-delta modulator," *JSSC*, vol. 28, pp. 1216-1223, Dec. 1993.
85. P. Brandt et al., "A 50-MHz multibit sigma-delta modulator for 12-b 2-MHz A/D conversion," *JSSC*, vol. 26, pp. 1746-1756, Dec. 1991.
86. B. E. Boser et al., "The design of sigma-delta modulation analog-to-digital converters," *JSSC*, vol. 23, pp. 1298-1308, Jun. 1988.

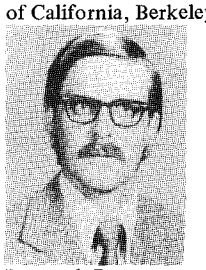
Other references

87. Semiconductor Industry Association. International Technology Roadmap for Semiconductors. http://www.sematech.org/.
88. L. Jipeng et al., "0.9V 12mW 2MSPS algorithmic ADC with 81dB SFDR," *VLSI*, 2004, pp. 436-439.
89. O. E. Erdogan et al., "A 12-b digital-background-calibrated algorithmic ADC with -90-dB THD," *JSSC*, vol. 34, pp. 1812-1820, Dec. 1999.
90. H.-S. Lee, "A 12-b 600 ks/s digitally self-calibrated pipelined algorithmic ADC," *JSSC*, vol. 29, pp. 509-515, Apr., 1994.
91. H.-S. Lee et al., "A self-calibrating 15-bit CMOS ADC," *JSSC*, vol. 19, pp. 813-819, Jun. 1984.
92. B. J. Hosticka, "Improvement of the gain of MOS amplifiers," *JSSC*, vol. sc-14, pp. 1111-1114, Dec. 1979.
93. E. Sackinger et al., "A high-swing, high-impedance MOS cascode Circuit," *JSSC*, vol. 25, pp. 289-298, Feb. 1990.
94. K. Bult et al., "A fast-settling CMOS op amp for SC circuits with 90-dB DC gain," *JSSC*, vol. 25, pp. 1379-1384, Dec. 1990.
95. M. J. M. Pelgrom et al., "Transistor matching in analog CMOS applications," *IEDM*, 1998, pp. 915-918.
96. M. J. M. Pelgrom et al., "Matching properties of MOS transistors," *JSSC*, vol. 24, pp. 1433-1439, May 1989.
97. M. Steyaert et al., "Threshold voltage mismatch in short-channel MOS transistors," Electronics Letters, vol. 30, pp. 1546-1548, 1994.
98. J. M. Rabaey et al., *Digital integrated circuits: a design perspective*, Upper Saddle River, N.J.: Pearson Education, 2003.
99. C. Enz et al., "MOS transistor modeling for RF IC design," *JSSC*, vol. 35, pp. 186-201, Feb. 2000.
100. A. A. Abidi, "High-frequency noise measurements on FETs with small dimensions," *TED*, pp. 1801-1805, Nov. 1986.
101. R. Aparicio, "Capacity limits and matching properties of integrated capacitors," *JSSC*, vol. 37, pp. 384-393, Mar. 2002.

for their prompt, thorough, and competent criticism. The combination of these efforts has resulted in what I believe to be a very interesting issue.

GARY L. BALDWIN
Guest Editor

Gary L. Baldwin (M'66) was born in El Centro, Calif., on October 12, 1943. He received the B.S., M.S., and Ph.D. degrees in electrical engineering in 1966, 1967, and 1970, respectively, all from the University of California, Berkeley.



He was employed by the IBM Corporation, San Jose, Calif., on a work-study program from 1963 to 1966. In the summer of 1968 he was involved in linear integrated circuit design with the Advanced Circuits Research Group, Signetics Corporation, Sunnyvale, Calif. He was an Acting Assistant Professor of Electrical Engineering at the University of California, Berkeley, during 1969 and 1970. He is now a member of the Technical Staff in the Systems Elements Research Department, Bell Laboratories, Holmdel, N.J.

Dr. Baldwin is a member of Eta Kappa Nu and Sigma Xi.

All-MOS Charge Redistribution Analog-to-Digital Conversion Techniques—Part I

JAMES L. McCREARY, STUDENT MEMBER, IEEE, AND PAUL R. GRAY, MEMBER, IEEE

Abstract—**This two-part paper describes two different techniques for performing analog-to-digital (A/D) conversion compatibly with standard single-channel MOS technology. In the first paper, the use of a binary weighted capacitor array to perform a high-speed, successive approximation conversion is discussed. The technique provides an inherent sample/hold function and can accept both polarities of inputs with a single positive reference. The factors limiting the accuracy and conversion rate of the technique are considered analytically. Experimental results from a monolithic prototype are presented; a resolution of 10 bits was achieved with a conversion time of 23 μs. The estimated die size for a completely monolithic version is 8000 mil^2.**

The second paper [3] describes a two-capacitor successive approximation technique, which, in contrast to the first, requires considerably less die area, is inherently monotonic in the presence of capacitor ratio errors, and which operates at a somewhat lower conversion rate. Factors affecting accuracy and conversion rate are considered analytically. Experimental results from a monolithic prototype are presented; a resolution of 8 bits was achieved with an A/D conversion time of 100 μs. Used as a D/A converter, a settling time of 13.5 μs was achieved. The estimated total die size for a completely monolithic version including logic is 5000 mil^2.

I. INTRODUCTION

MOST conventional techniques for analog-to-digital (A/D) conversion require both high-performance analog circuitry, such as operational amplifiers, and digital circuitry for counting, sequencing, and data storage. This has tended to result in hybrid circuits consisting of one or more bipolar analog chips and an MOS chip to economically perform the digital functions [1]. This paper describes a new, all-MOS technique which is realizable in a single chip and performs a 10-bit conversion in 23 μs [6]. It includes an intrinsic sample-and-hold function, accepts both bipolar and unipolar inputs, and is realizable with standard N-channel metal gate technology.

For the realization of a fast, successive-approximation A/D converter in MOS technology, conventional voltage driven R-$2R$ techniques are cumbersome since diffused resistors of proper sheet resistance are not available in the standard single-channel technology. A complex thin-film process must be used. Furthermore, these approaches require careful control of the "ON" resistance ratios in the MOS switches over a wide range of values.

In contrast to its utilization as a current switch, the MOS device, used as a charge switch, has inherently zero offset voltage and as an amplifier has very high input resistance. In addition, capacitors are easily fabricated in metal gate technology. Therefore, one is led to use capacitors rather than resistors as the precision components, and to use charge rather than current as the working medium. This technique, referred to as charge-redistribution, has been used in some discrete component A/D converters for many years [2], [7]. However, these converters have required high-performance operational amplifiers which are difficult to realize in single-channel MOS technology.

This paper describes a new A/D conversion technique using charge redistribution on weighted capacitors, while a companion paper [3] describes another application of the charge redistribution concept using two equal capacitors. In Section II of this paper, the binary-weighted capacitor capacitor array and its operation is described. In Sections III and IV, the limitations on the speed and resolution are analyzed. Ex-

Manuscript received March 28, 1975; revised July 30, 1975. This research was sponsored by the National Science Foundation under Grant GK-40912. This paper was presented at the International Solid-State Circuits Conference, Philadelphia, Pa., February 1975.

The authors are with the Department of Electrical Engineering and Computer Sciences and the Electronics Research Laboratory, University of California, Berkeley, Calif. 94720.

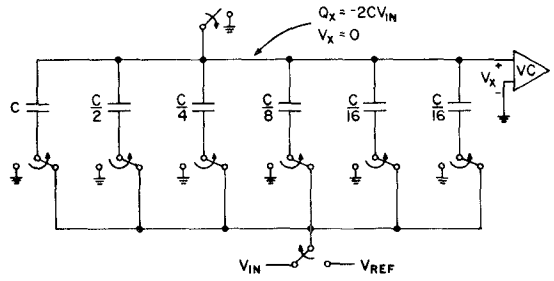


Fig. 1. Conceptual 5-bit A/D converter illustrating the sample mode operation.

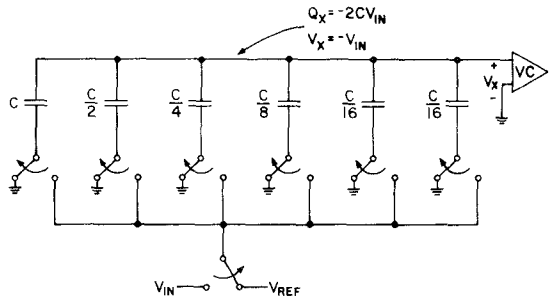


Fig. 2. Pre-redistribution hold mode operation.

perimental results from a prototype system are presented in Section V.

II. Charge Redistribution A/D Conversion Technique Using Binary Weighted Capacitors

The new A/D conversion technique is illustrated with a conceptual 5-bit version of the converter shown in Fig. 1. It consists of a comparator, an array of binary weighted capacitors plus one additional capacitor of weight corresponding to the least significant bit (LSB), and switches which connect the plates to certain voltages. A conversion is accomplished by a sequence of three operations. In the first, the "sample mode" (Fig. 1), the top plate is connected to ground and the bottom plates to the input voltage. This results in a stored charge on the top plate which is proportional to the input voltage V_{in}. In the "hold mode" of Fig. 2, the top grounding switch is then opened, and the bottom plates are connected to ground. Since the charge on the top plate is conserved, its potential goes to $-V_{in}$. The "redistribution mode," shown in Fig. 3, begins by testing the value of the most significant bit (MSB). This is done by raising the bottom plate of the largest capacitor to the reference voltage V_{ref}. The equivalent circuit is now actually a voltage divider between two equal capacitances. The voltage V_x, which was equal to $-V_{in}$ previously, is now increased by $\frac{1}{2}$ the reference as a result of this operation.

$$V_x = -V_{in} + \frac{V_{ref}}{2}.$$

Sensing the sign of V_x, the comparator output is a logic '1' if $V_x < 0$ and is a '0' if $V_x > 0$. This is analogous to the interpretation that

$$\text{if } V_x < 0 \quad \text{then } V_{in} > \frac{V_{ref}}{2};$$

hence the MSB = 1; but

$$\text{if } V_x > 0 \quad \text{then } V_{in} < \frac{V_{ref}}{2};$$

therefore the MSB = 0. The output of the comparator is, therefore, the value of the binary bit being tested. Switch $S1$ is returned to ground only if the MSB b_4 is a zero. In a similar manner, the next MSB is determined by raising the bottom plate of the next largest capacitor to V_{ref} and checking the polarity of the resulting value of V_x. In this case, however,

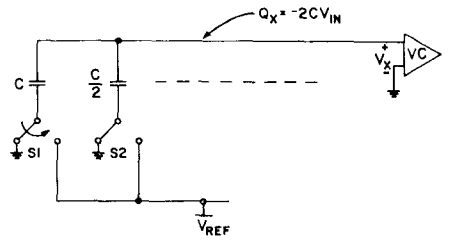


Fig. 3. Redistribution mode operation.

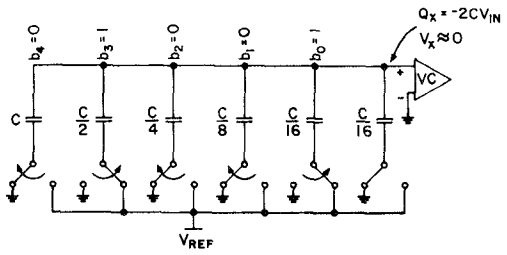


Fig. 4. Example of a final configuration.

the voltage division property of the array causes $V_{ref}/4$ to be added to V_x:

$$V_x = -V_{in} + b_4 \frac{V_{ref}}{2} + \frac{V_{ref}}{4}.$$

Conversion proceeds in this manner until all the bits have been determined. A final configuration is illustrated in Fig. 4 for the digital output 01001. Notice that all capacitors corresponding to a '0' bit are completely discharged. The total original charge on the top plates has been redistributed in a binary fashion and now resides only on the capacitors corresponding to a '1' bit. N redistributions are required for a conversion resolution of N bits. In contrast to earlier charge redistribution techniques, the capacitance of the lower plate switch does not affect the accuracy of the conversion [4]. This fact is evident since the switch capacitance is either discharged to ground or is charged by V_{ref} but never absorbs charge from the top plate. Therefore, the switch devices can be quite large permitting rapid redistributions. On the other hand, the top plate of the array is connected to all the capacitors and to a switch and to the comparator resulting in a large

parasitic capacitance from the top plate to ground. The nature of the conversion process, however, is such that V_x is converged back towards zero—its initial value. Hence, the charge on this parasitic is the same in the final configuration as it was in the sample mode. Therefore, the error charge contributed by this parasitic is very near zero. For the 10-bit converter the parasitic capacitance at the top plate can be 100 times the value of the smallest capacitor and still cause only 0.1 bit offset error. This is equivalent to saying that the smallest capacitor may be much smaller than the parasitic and consequently the largest capacitor may be reduced in value proportionally. Furthermore, the initial value of V_x need not necessarily be zero but can be the threshold voltage of the comparator. This fact allows cancellation of comparator offset by storing the offset in the array during the sample mode. The linearity then is primarily a function of the ratio accuracy of the capacitors in the array.

By only a slight modification of the array switching scheme bipolar voltage inputs can be encoded while still using only the single positive reference. This is achieved by connecting the bottom plate of the largest capacitor to V_{ref} during the sample mode resulting in a stored charge:

$$Q_x = -C_{TOT}\left(\frac{V_{\text{in}}}{2} + \frac{V_{\text{ref}}}{2}\right).$$

Each bit is then tested in sequence just as before except that the largest capacitor is switched from V_{ref} to ground during its test, while all the other capacitors are switched from ground to V_{ref}. Also, as before, a bit value is true if V_x is negative after the test. The expression for V_x again converges back towards zero:

$$V_x = -\frac{V_{\text{in}}}{2} + V_{\text{ref}}\left(-\frac{b_{10}}{2^1} + \frac{b_9}{2^2} + \frac{b_8}{2^3} + \cdots + \frac{b_1}{2^{10}}\right) \approx 0.$$

For a 10-V reference, b_{10} is '0' for $0 \leqslant V_{\text{in}} \leqslant 10$ V, but is '1' for -10 V $\leqslant V_{\text{in}} < 0$. Therefore, b_{10} represents the sign bit and its function is to level shift V_x in order to accommodate negative inputs. Hence, a 10-bit conversion is achieved over the input range ±10 V with negative numbers expressed in 1's complement.

III. Factors Limiting Accuracy

While monolithic circuit technology has had great impact on the cost of many analog circuit functions, such as operational amplifiers, the impact on the cost of A/D and D/A converters has not been as great. This is due to the complexity of a complete converter, and more importantly, to the problem of component matching. Because of the difference between the aspect ratios of diffused resistors of practical value versus those of capacitors, the attainable matching accuracy is higher for capacitors given the same overall die area. The flexibility of capacitor geometry allows them to be made square or even circular so as to optimize matching accuracy, as discussed further in the companion paper [3].

In this technique a mismatch in the binary ratios of capacitors in the array causes nonlinearity. It cannot cause a gain

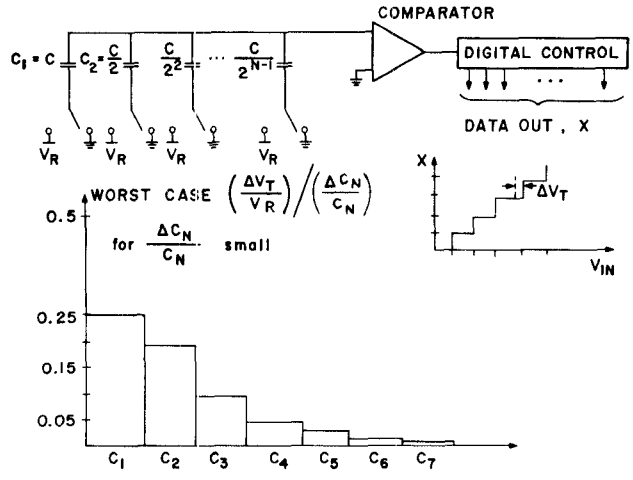


Fig. 5. Sensitivity of A/D conversion linearity to individual capacitor value.

error because the end points of the transfer function curve, V_{out} versus V_{in}, are not dependent on capacitor matching. This results from the fact that no net charge redistribution between capacitors occurs for either zero or full-scale input since all capacitors are either fully discharged or fully charged, respectively. For the same reason no offset error can arise from capacitor mismatch since the mismatch cannot be manifested unless a charge redistribution exists in the final configuration.

First consider the ideal case in which all the capacitors of the general N-bit converter shown in Fig. 5 have the precise binary weight values. For this case the digital output x is a regular staircase when plotted against V_{in}, and every transition occurs at a precise value of V_{in} designated V_T. On the other hand, changing one capacitor from its ideal value by a small amount ΔC_N causes all transition points to shift somewhat, but there will be one worst case transition. The ratio $\Delta V_T/V_R$ is the normalized worst case fraction deviation in transition point from the ideal. This is also a measure of the nonlinearity. The ratio of this deviation to the fractional change in capacitor value $\Delta C_N/C_N$ represents the sensitivity of linearity to individual capacitor value. The plot of sensitivity, also shown in Fig. 5, shows that linearity is very sensitive to a fractional change in the large capacitors, but not very dependent upon similar fractional changes in the smaller capacitors. Therefore, the smaller capacitors have great allowable tolerances. It should be pointed out that actually all capacitors have simultaneous deviations which cause ratio errors and the worst case combination of these must always be considered. Thus, the optimization of the ratio accuracy in the array is a prime consideration.

Capacitor ratio error results from several causes, one of which is the undercutting of the mask which defines the capacitor. Consider two capacitors C_4 and C_2, shown in Fig. 6, which are nominally related by a factor of 2: $C_4 = 2C_2$. During the etching phase of the photomask process a poorly controlled lateral etch occurs called undercut. Let Δx be the undercut length and L_4 be the side length of C_4, also shown in Fig. 6. Then a ratio error is produced between C_4 and C_2 which is proportional to the undercut length:

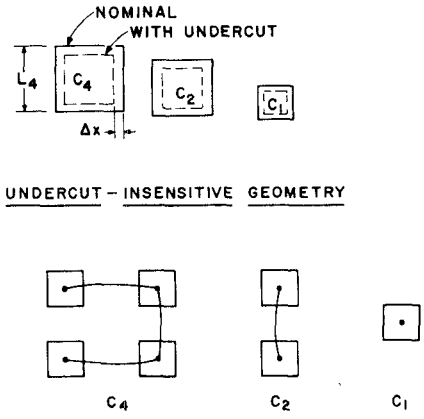


Fig. 6. Capacitor ratio error due to photomask undercut.

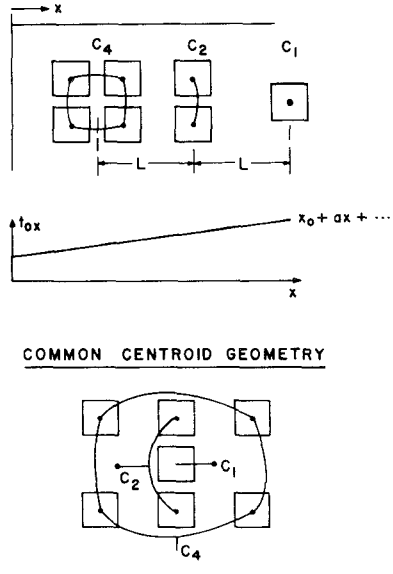


Fig. 7. Capacitor ratio error due to oxide gradients.

$$C_4 = 2C_2(1 + \epsilon); \qquad \epsilon \simeq 4\frac{\Delta x}{L_4}.$$

This problem can be circumvented by a geometry such that the perimeter lengths as well as the areas are ratioed. This can be done, as seen in Fig. 6, by paralleling identical size plates to form the larger capacitors. Thus the capacitor ratios are not affected by undercut.

Long range gradients in the thin capacitor oxide can also cause ratio errors. These gradients arise from nonuniform oxide growth conditions. If this variation in oxide thickness is approximated as first-order gradient, as shown in Fig. 7, then the resulting ratio error is proportional to the fractional variation in oxide thickness:

$$C_2 = 2C_1(1 + \epsilon L)$$

$$C_4 = 4C_1(1 + 2\epsilon L); \qquad \epsilon = \frac{a}{x_0}.$$

Experimentally, values of 10–100 ppm/mil have been observed for the factor ϵ. Error from this source can be minimized by improved oxide growth techniques and by a common centroid geometry. This is done in Fig. 7 by locating the elements of the capacitors in such a way that they are symmetrically spaced about a common center point. In this way the capacitor ratios may be maintained in spite of first-order gradients. Although undercut and oxide gradient errors can be minimized, a random edge variation will still exist and cause small ratio errors.

Any significant variation of small signal capacitance with dc terminal voltage would limit the accuracy because a nonlinearity would result. For MOS capacitors on heavily doped N+ back plates voltage coefficients of less than 24 ppm/V have been observed. This is insignificant at the 10-bit converter level.

Dielectric absorption is a phenomenon in which a residual voltage appears on a capacitor after it has been rapidly discharged. However, MOS capacitors display a relaxation which is unimportant for the 10-bit converter.

The voltage comparison process is fundamental to A/D conversion. The offset voltage of the comparator is usually manifested as an offset error in the digital conversion. Because of the relatively large gate-source voltage mismatch in MOS differential amplifiers, the offset voltage of the all-MOS comparator must be eliminated as a source of error. This can be accomplished either by digital means or by offset cancellation techniques. In this circuit the offset is stored and then subtracted at a later time by the sequence of events illustrated in Fig. 8. During the sample mode, V_{in} remains connected to the bottom plates but switch $S1$ is also ON and precharges the top plate of the array to the threshold voltage of stage $A1$. $S1$ then turns off, but since $A2$ is identical to $A1$ its input is also at the threshold. Since $S2$ is ON in the "up" position the output of $A2$ is saved at a storage node at one input of $A3$. During the subsequent redistribution mode $S1$ always remains OFF but $S2$ turns ON in the "down" position after each redistribution loading the output of $A2$ at the other storage node input of $A3$. This provides a first-order cancellation of switch feedthrough at both storage nodes. Since $A3$ is a difference amplifier, the offset of $A1$ and $A2$ together with the feedthrough of $S1$ have been cancelled. The offset of $A3$ is

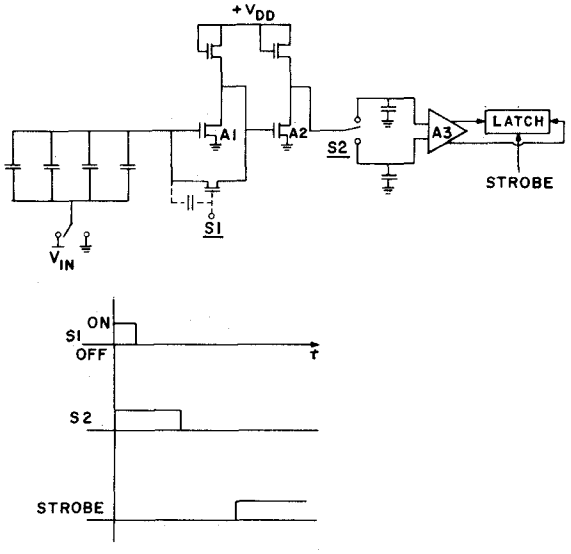


Fig. 8. Comparator offset cancellation.

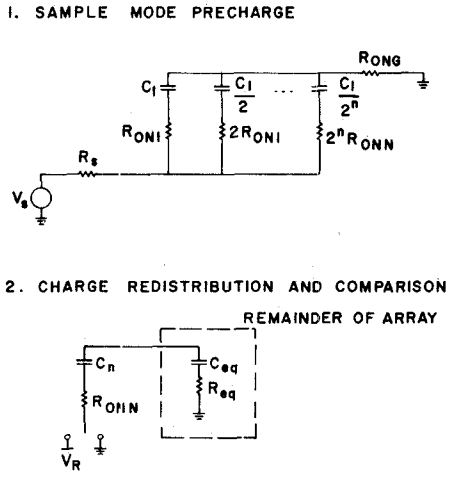


Fig. 9. Factors limiting conversion rate.

therefore the only significant offset in the comparator. This offset is reflected back through the gain stages $A1$ and $A2$ giving an effective offset [5]:

$$V_{OS_{\text{effective}}} = \frac{V_{OS_{A3}}}{G_{A1} \cdot G_{A2}}.$$

The gain product $G_{A1} \cdot G_{A2} \approx 50$ for this circuit.

The conversion technique as described thus far produces a transfer characteristic like that shown in Fig. 5 in which the first transition occurs at one LSB voltage away from the origin. Actually, for minimum quantization error this first transition should occur only $\frac{1}{2}$ LSB voltage away from the origin. The cancellation of this offset may be accomplished by returning the lower plate of the smallest capacitor to $V_{\text{ref}}/2$ rather than ground after the sample mode.

A great advantage of the S_iO_2 dielectric capacitor is its very low temperature coefficient of approximately 20 ppm/°C in contrast with approximately 2000 ppm/°C for diffused resistors and several hundred ppm/°C for thin film and implanted resistors. Thus, capacitor ratios are less sensitive to temperature gradients than resistor ratios. Since component mismatch usually has a greater effect upon linearity than upon gain or offset errors, the temperature coefficient of nonlinearity is lower for this conversion technique in comparison with R-$2R$ approaches.

IV. Factors Limiting Conversion Rate

Compared with many conversion techniques, the successive approximation method used in this circuit is capable of rapid conversion. The conversion requires two distinct operations. The first is the charging of the capacitor array to the input voltage during the sample mode, and the second is the redistribution of the charge on the capacitor array during the successive approximation conversion process. The equivalent circuit during the sample mode is shown in Fig. 9. The total array capacitance C_T is charged through the source resistance R_s, and equivalent ON resistance of the top plate grounding switch R_{ON_G}, and the equivalent ON resistance of the binarily scaled switches. The time constant is thus:

$$\tau_1 = \left[R_S + \frac{R_{ON_1}}{2} + R_{ON_G}\right] C_T.$$

During the redistribution cycle, also illustrated in Fig. 9, the time constant is the series combination of the capacitor being tested and its switch resistance plus the equivalent capacitance and resistance of the remaining elements of the array:

$$\tau_2 = \frac{R_{ON_1}}{2} C_T.$$

The maximum conversion rate is determined by the time required for one sample mode precharge and subsequently for ten redistribution cycles to go to completion, with time allowed for one comparison after each redistribution. In the experimental circuit to be described later, each of these delays contribute to the total conversion time, but it is instructive to consider the fundamental limitations on the conversion rate of this technique.

Assuming that the source resistance R_s can be made small, the acquisition time τ_1 is dependent upon R_{ON_1} and R_{ON_G}. The former can be made small compared to R_{ON_G} since the gate-source and gate-drain capacitance of the bottom plate switches do not result in conversion errors. The gate-drain capacitance of the grounding switch, however, can affect the accuracy of the conversion [8]. In the realization described later this error is cancelled, but assume for the time being that this was not done. The time constant during precharge for this limiting case is

$$\tau_1 = R_{ON_G} C_T = \frac{C_T}{\mu \frac{w}{L_C} C_0 (V_{GS}(\text{ON}) - V_T)}.$$

Assuming an optimum gate drive signal on the grounding

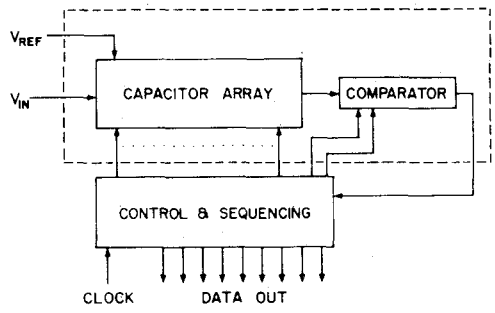


Fig. 10. Complete A/D converter.

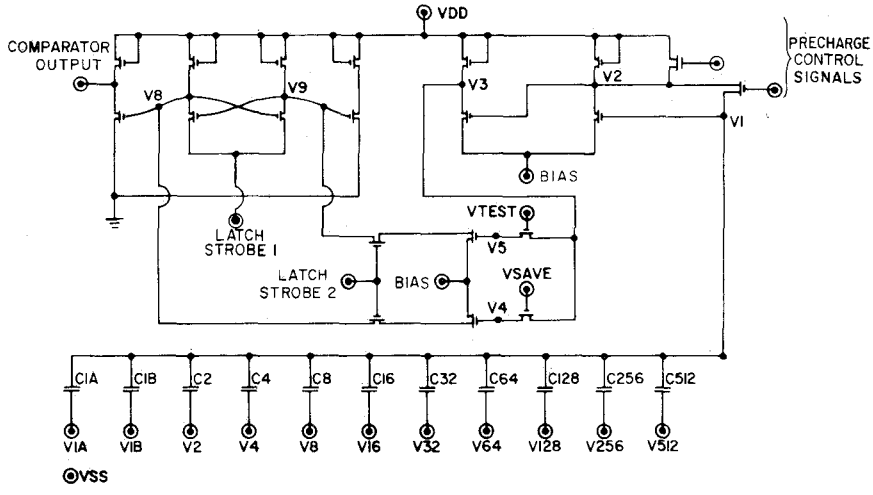


Fig. 11. Schematic of the experimental integrated circuit.

switch, the error voltage is given by

$$V_E = \frac{C_{GS}}{C_T}(V_{GS}(\text{ON}) - V_T)$$

and therefore

$$V_E \tau_1 = \frac{C_{GS}}{\frac{w}{L_C} C_0} \simeq \frac{L_C^2}{\mu}.$$

Here the overlap capacitance has been neglected, and it has been assumed for simplicity that the error charge transferred to C_T is simply the channel inversion layer charge, represented by the quantity $C_{GS}[V_{GS}(\text{ON}) - V_T]$. This result implies for example, that for a 0.1 percent error, 10-μ channel length, and N-channel devices, a minimum acquisition time of about 3 μs is required.

Actually, this source of error is not a fundamental limitation on accuracy since it is simply an offset error and can be cancelled by using either analog or digital techniques. These techniques, however, require considerable time for the sampling and processing of the offset voltage so that this problem does place a practical limit on the conversion rate.

The redistribution time τ_2 is dependent on R_{ON} of the bottom plate switches which can be made arbitrarily small at the cost of the die area. For very large devices, the redistribution time will approach that required for the switches to change their own drain-bulk capacitance.

Following each redistribution, a time interval is required for the voltage comparator to settle to the correct state. This time interval becomes longer as the resolution becomes higher and the minimum overdrive signal becomes smaller. Comparator delay is a fundamental limiting factor in the rate at which conversions can be accomplished.

V. Circuit Description and Experimental Results

Fig. 10 shows a complete A/D converter using the technique described in this paper. In block diagram form the system is composed of a capacitor array, a comparator, and control and sequencing logic. The feasibility of this technique was investigated by fabricating only the critical portions of the circuit, the array and the comparator, in monolithic form using standard N-channel MOS technology. Although the control and sequencing logic was composed of TTL gates, the monolithic realization of this circuit in MOS technology is straightforward.

The experimental integrated circuit schematic is shown in Fig. 11. During the sample mode, offset cancellation is accomplished by the precharge control signals. After precharge is completed, the initial value of V_1 is saved at a storage node V_4 by the action of a transmission gate. After each redistribution, the amplified value of V_1 is transmitted to node V_5 in

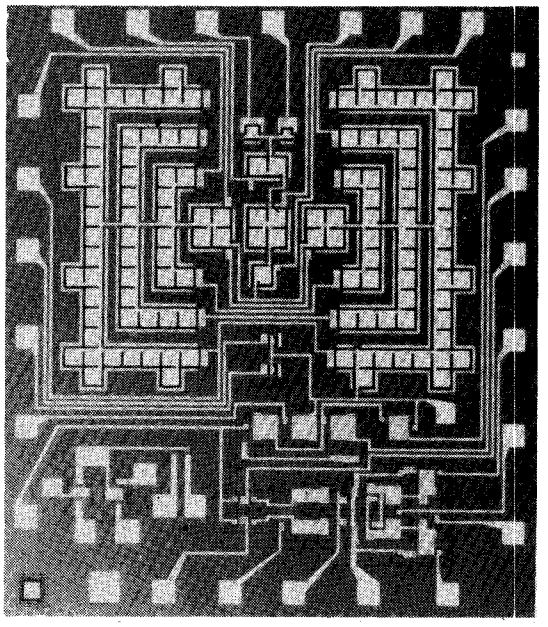


Fig. 12. Die photo.

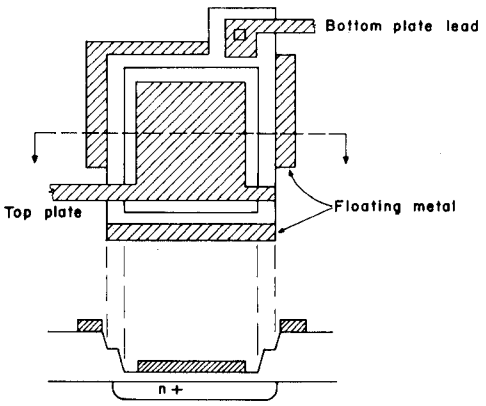


Fig. 13. MOS capacitor structure.

preparation for the subsequent test of the bit value. Since both latch strobe signals are initially high the difference signal at the latch is given by

$$(V_8 - V_9) = (V_4 - V_5) \cdot G_{A3}$$

where G_{A3} is the gain of amplifier $A3$ previously discussed. The total signal gain from the comparator input to the latch input is about 200, hence a 10-mV change in V_1 produces a 2-V differential signal at the latch. When the latch strobe signals go low the cross-coupled pair becomes regenerative and the full gain of the comparator is realized. The final state of the latch is then buffered off chip.

The die photo of Fig. 12 shows the capacitor array totaling 240 pF positioned above the comparator. The extreme right and left segments of the array are connected in parallel to form the largest capacitor. The next two segments on each side compose the next largest capacitor and so on, so that the common centroid is the center of the array. Note that the larger capacitors are composed of small squares 3 X 3 mils to reduce the sensitivity to undercut. The capacitor array has dimensions 75 X 58 mils, and an estimated die size including the digital control is 90 X 90 mils.

The MOS capacitor structure is shown in Fig. 13. Notice that only the metal edge defines the capacitor area, thus taking advantage of the better resolution properties of some positive photoresists. The floating redundant metal strips maintain a relatively constant etchant concentration at every capacitor edge, thereby assuring nearly uniform undercutting effects for all capacitors. The capacitor interconnect spans the thin oxide of each capacitor in a single direction so that capacitor area is insensitive to mask alignment errors. The area of the interconnect over the thin oxide is included in capacitor area calculations, while the area over the thicker oxide is neglected. Interconnect that passes over field oxide causes a parasitic capacitance from the top plate to ground which does not affect circuit performance as previously shown. A first-order cancellation of corner-rounding effects is made by designing each capacitor with an equal number of 90° and 270° corners.

The central question of the feasibility of this approach is the

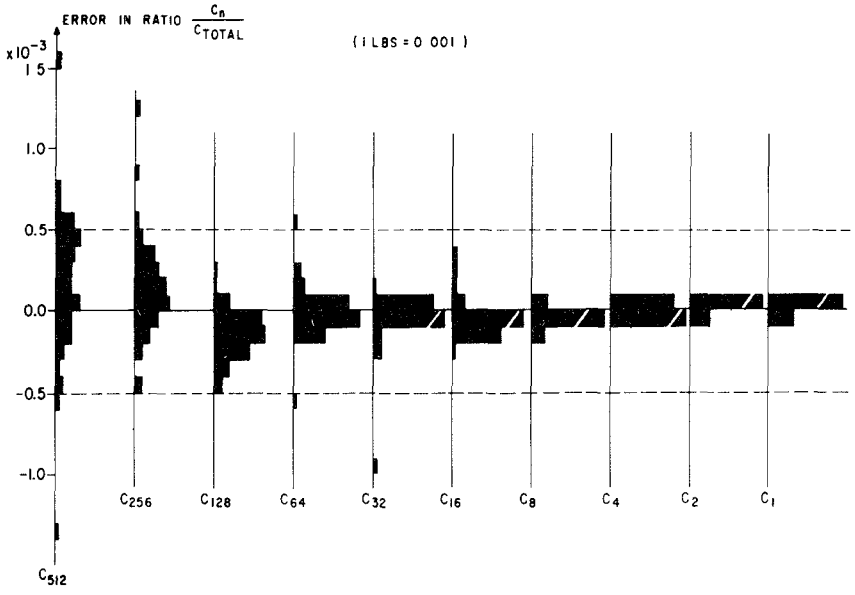


Fig. 14. Distribution of capacitor ratio errors for 47 10-bit converter arrays. Vertical axis is the fractional ratio error while the horizontal axis is the frequency of occurrence for each capacitor.

accuracy with which the capacitor ratios can be maintained using conventional photomasking. Extensive data have been gathered on the ratio accuracy of the array and typical results from three different wafers are shown in Fig. 14. This is the distribution of measured errors in the ratio of each individual capacitor to the total array capacitance. The horizontal axis represents the frequency of occurrence. Assume for a moment that capacitor ratio error were the only factor affecting yield. Then from these data the yield for $\pm\frac{1}{2}$ LSB linearity for 8, 9, and 10 bits resolution would be 98 percent, 94 percent, and 45 percent, respectively, for this sample of 47 functional arrays.

The feasibility of the technique was further studied by operating the experimental chip with external logic as a complete A/D converter. The linearity and offset of a sample of devices were evaluated using the experimental procedure shown in Fig. 15. The output of the A/D was connected to a 12-bit D/A converter. Since V_{in} was a long period ramp, the output of the D/A was a staircase. The difference between these signals is a representation of the total conversion error, and is connected to the y-axis of the plotter. The zero-to-full-scale recording of the output is shown also in Fig. 15. Since the 10-mV marks correspond to the $\frac{1}{2}$ LSB error levels the error is seen to be less than $\frac{1}{2}$ LSB. An expanded recording which permitted the visual resolution of all 1024 states was used for actual verification of the results.

A summary of the measured results is shown in Table I. The sample mode acquisition time is the minimum precharge time required for an accurate conversion of a 5-V step change at the input. The total conversion time corresponds to a 44-kHz sampling frequency.

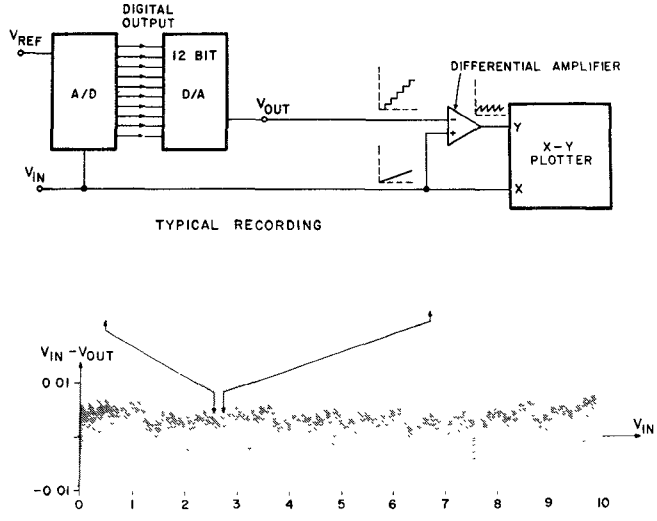


Fig. 15. Experimental measurement of A/D converter nonlinearity.

TABLE I
MEASURED PERFORMANCE DATA

Resolution	10 bits
Linearity	±1/2 LSB
Input voltage range	0–10 V
Input offset voltage	2 mV
Gain error	<0.05 percent (external reference)
Sample mode acquisition time	2.3 μs
Total conversion time	22.8 μs

VI. SUMMARY

A new, all MOS A/D conversion technique has been demonstrated which, with the addition of an external reference, can be used to realize a standard-process A/D converter on a single chip. Experimental data were presented which indicated that 8-bit resolutions can be attained at very high yield and low cost using standard N-channel MOS technology, and that 10-

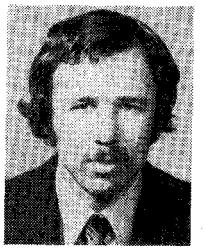
bit resolution can be achieved at somewhat lower yield. It is believed by the authors that more careful control of photolithographic processing would result in very high yield at the 10-bit level and significant yield at even higher resolutions.

ACKNOWLEDGMENT

The authors wish to acknowledge the contributions of D. A. Hodges and R. E. Suárez to the work described in Part I.

REFERENCES

[1] J. A. Schoeff, "A monolithic analog subsystem for high-accuracy A/D conversion," in *ISSCC Dig. Tech. Papers*, Feb. 1973, pp. 18–19.
[2] H. Schmidt, *Analog-Digital Conversion*. New York: Van Nostrand-Reinholt, 1970.
[3] R. E. Suarez, P. R. Gray, and D. A. Hodges, "All-MOS charge redistribution analog-to-digital conversion techniques–part II," this issue, pp. 379–385.
[4] —, "An all-MOS charge-redistribution A/D conversion technique," in *ISSCC Dig. Tech. Papers*, Feb. 1974, pp. 194–195.
[5] R. Poujours et al., "Low level MOS transistor amplifier using storage techniques," in *ISSCC Dig. Tech. Papers*, Feb. 1973, pp. 152–153.
[6] J. McCreary and P. R. Gray, "A high-speed, all-MOS successive approximation weighted capacitor A/D conversion technique," in *ISSCC Dig. Tech. Papers*, Feb. 1975, pp. 38–39.
[7] C. W. Barbour, "Simplified PCM analog to digital converter using capacity charge transfer," in *Proc. 1971 Telemetering Conf.*, pp. 4.1–4.11.
[8] J. McCreary, "Successive approximation analog-to-digital conversion in MOS integrated circuits," Ph.D. dissertation, Univ. California, Berkeley, 1975.



James L. McCreary (S'73) received the B.S. degree in electrical engineering from the University of Texas, Austin, and the M.S. and Ph.D. degrees in electrical engineering from the University of California, Berkeley, in 1968, 1971, and 1975, respectively.

Prior to attending the University of California at Berkeley, where his doctoral research area was analog-to-digital conversion techniques with MOS integrated circuit technology, he spent three years in the U.S. Navy as an Instructor in Nuclear Propulsion Systems. He is currently with the Department of Electrical Engineering and Computer Sciences and the Electronics Research Laboratory, University of California, Berkeley, and will join the Research Staff of Bell Laboratories, Holmdel, N.J., in January 1976.

Dr. McCreary is a member of Sigma Xi.

Paul R. Gray (S'65–M'69) was born in Jonesboro, Ark., on December 8, 1942. He received the B.S., M.S., and Ph.D. degrees from the University of Arizona, Tucson, in 1963, 1965, and 1969, respectively.

In 1969 he joined the Research and Development Laboratory, Fairchild Semiconductor, Palo Alto, Calif., where he was involved in the development and application of new technologies for analog integrated circuits. In 1971 he joined the Department of Electrical Engineering and Computer Science, University of California, Berkeley, where he is now an Associate Professor. Since 1971 he has acted as Consultant to Fairchild Semiconductor and Signetics Corporation.

Dr. Gray is a member of Eta Kappa Nu and Sigma Xi.

All-MOS Charge Redistribution Analog-to-Digital Conversion Techniques—Part II

RICARDO E. SUÁREZ, MEMBER, IEEE, PAUL R. GRAY, MEMBER, IEEE, AND DAVID A. HODGES, SENIOR MEMBER, IEEE

Abstract—**This two-part paper describes two different techniques for performing analog-to-digital (A/D) conversion compatibly with standard single-channel MOS technology. In the first paper, the use of a binary weighted capacitor array to perform a high-speed successive approximation conversion was discussed.**

This second paper describes a two-capacitor successive approximation technique, which, in contrast to the first, requires considerably less die area, is inherently monotonic in the presence of capacitor ratio errors, and which operates at somewhat lower conversion rate. Factors affecting accuracy and conversion rate are considered analytically. Experimental results from a monolithic prototype are presented; a resolution of eight bits was achieved with an A/D conversion time of 100 μs. Used as a digital-to-analog (D/A) converter, a settling time of 13.5 μs was achieved. The estimated total die size for a completely monolithic version including logic is 5000 mil^2.

Manuscript received May 19, 1975; revised July 30, 1975. This research was sponsored in part by the National Science Foundation under Grant GK-40912.

R. E. Suárez was with the Department of Electrical Engineering and Computer Sciences and the Electronics Research Laboratory, University of California, Berkeley, Calif. He is now with the Instituto Venezolano de Investigaciones Cientificas (I.V.I.C.), Caracas, Venezuela.

P. R. Gray and D. A. Hodges are with the Department of Electrical Engineering and Computer Sciences and the Electronics Research Laboratory, University of California, Berkeley, Calif. 97420.

I. INTRODUCTION

AS DISCUSSED in Part I [9] of this paper, widespread application of techniques for digital processing of analog signals has been hindered by the unavailability of inexpensive functional blocks for analog-to-digital (A/D) conversion. Traditional approaches to A/D conversion have required the simultaneous implementation of high-performance analog circuits, such as operational amplifiers, and of digital circuitry

Analog Circuits in Ultra-Deep-Submicron CMOS

Anne-Johan Annema, *Member, IEEE*, Bram Nauta, *Senior Member, IEEE*, Ronald van Langevelde, *Member, IEEE*, and Hans Tuinhout

***Abstract*—Modern and future ultra-deep-submicron (UDSM) technologies introduce several new problems in analog design. Nonlinear output conductance in combination with reduced voltage gain pose limits in linearity of (feedback) circuits. Gate-leakage mismatch exceeds conventional matching tolerances. Increasing area does not improve matching anymore, except if higher power consumption is accepted or if active cancellation techniques are used. Another issue is the drop in supply voltages. Operating critical parts at higher supply voltages by exploiting combinations of thin- and thick-oxide transistors can solve this problem. Composite transistors are presented to solve this problem in a practical way. Practical rules of thumb based on measurements are derived for the above phenomena.**

***Index Terms*—Analog design, breakdown, CMOS, distortion, evolution, future performance, gate leakage, low power, low voltage, mismatch, scaling, technology, UDSM.**

I. Introduction

THE evolution in CMOS technology is motivated by decreasing price-per-performance for digital circuitry; its pace is determined by Moore's Law. To ensure sufficient lifetime for digital circuitry and to keep power consumption at an acceptable level, the dimension-shrink is accompanied by lowering of nominal supply voltages. While this evolution in CMOS technology is by definition very beneficial for digital, this is not so for analog circuits [1]–[3].

Contemporary ICs are mixed-signal systems consisting of a large digital core including amongst others a CPU or DSP and memory, often surrounded by several analog interface blocks such as I/O, D/A, and A/D converters, RF front ends, and more. From an integration point of view all these functions would ideally be integrated on a single die. In this case 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. This paper discusses a number of issues for analog designs in modern and future ultra deep submicron (UDSM) CMOS processes and possible ways to maintain performance [3].

CMOS evolution has come to a point where for analog circuits new phenomena need to be taken into account. A major issue is the decreasing supply voltage. Although the supply voltage has dropped from 5 V in the early nineties down to 1.2 V today, most analog circuits can still be designed. However, a further drop in supply voltages is expected to cause serious roadblocks for analog circuits, because the signal headroom becomes too small to design circuits with sufficient signal integrity at reasonable power consumption levels. Although the analog transistor properties do not really get worse when comparing them at identical bias conditions, lower supply voltages require biasing at lower operating voltages which results in worse transistor properties, and hence yield circuits with lower performance.

A second issue is gate leakage. Gate leakage will increase drastically when migrating to newer technologies. When the gate oxide thickness is reduced with the equivalent of one atomic layer, the gate current increases by approximately one order of magnitude. Despite technological remedies, gate leakage will become part of analog design-especially for long transistors; e.g., in 65 nm technology the current gain of a MOSFET will be as small as unity for a channel length of 30 μm. In this paper, we introduce a bias insensitive frequency $f_{\rm gate}$for quick estimation of the effect of gate leakage. Another issue is gate leakage current mismatch. For large area (long L) transistors mismatch will be dominated by gate leakage mismatch. This effect puts a new upper limit on achievable matching performance. This problem can be coped with by accepting increased power consumption or by using active cancellation techniques.

This paper is organized as follows. Section II reviews the implications of going to lower supply voltages. Section III discusses the trend in a number of bare transistor properties, also illustrating that lower supply voltages degrade circuit performance. In Section IV $f_{\rm gate}$ is introduced while its use and applications are discussed in Section V. Section VI discusses a solution for the reduced nominal supply voltage aspects of newer CMOS generations: operating analog circuits at relatively high voltages, using thick oxide transistors and composite high-voltage transistors.

II. Fundamental Implications of Lower Supply Voltages

From a circuit point of view, plain physics dictates that the power consumption of analog circuits is proportional to the level of signal integrity (e.g., the signal-to-noise ratio, SNR, or the signal-to-noise and distortion ratio, SINAD) and to the signal frequency [4]–[6]. In other words: for analog circuits more performance comes at the cost of higher power consumption. There is a factor between the actual and the fundamental minimum power consumption that takes into account implementation overhead, margins in operating conditions and device spread. A common observation in all power-performance relations is that power consumption rises with decreasing supply voltages. Appendix A presents a short review of a number of

Manuscript received April 11, 2004; revised May 28, 2004.

A. J. Annema and B. Nauta are with the IC Design Group, University of Twente, 7500AE Enschede, The Netherlands (e-mail: a.j.annema@utwente.nl).

R. van Langevelde and H. Tuinhout are with Philips Research Laboratories, 5656AA Eindhoven, The Netherlands.

Digital Object Identifier 10.1109/JSSC.2004.837247

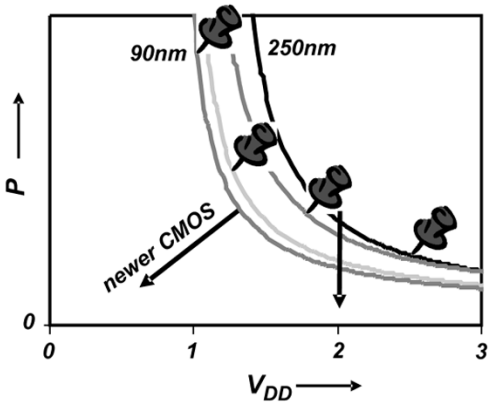


Fig. 1. Minimum power consumption for an (arbitrary) analog circuit (see Appendix A) with fixed topology and performance as a function of the supply voltage, for four technologies. Pushpins correspond to the power consumption in a technology at the nominal supply voltage for each CMOS process.

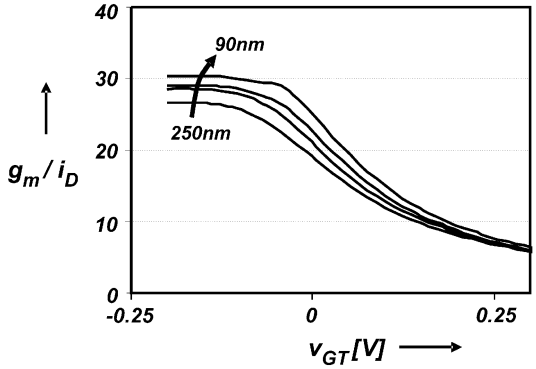


Fig. 2. The transconductance normalized with respect to the gate-overdrive voltage $V_{\rm GT}$, for four technologies.

power-performance relations and discusses exceptions to the "rule."

For a given power budget the performance drops when migrating to newer technologies, simply because of their lower supply voltages. This is probably the single most important effect that (fundamentally and practically) complicates analog designs at low supply voltages. For Fig. 1, a simple unity gain voltage buffer with fixed topology, fixed performance and fixed technology was optimized for minimum power consumption; see also Appendix A. In the optimization process, signal swing, all bias conditions of transistors and device dimensions were optimized [6]. It follows that the minimum power consumption increases with decreasing supply voltages. However, *at constant supply voltage*, porting the circuit to a newer technology *lowers* the required power consumption.

III. Scaling of Conventional Transistor Properties

Apart from issues at circuit level, basic transistor properties also change with CMOS technology evolution. This section reviews a few important properties.

A. DC Properties at Constant Voltage Headroom

This section presents the trend in dc properties of MOS transistors at constant voltage headroom ($V_{\rm DS}$) under typical analog operating conditions (low gate-overdrive voltage $V_{\rm GT}$). Note that these conditions are usually not satisfied when porting designs to newer technologies because of decreasing nominal supply voltages. For reasons of clarity, it is however a fair condition for the comparison of bare transistor properties.

Low-distortion at quasi-dc frequencies is relevant for many analog circuits. Typically, quasi-dc distortion may be due to nonlinearities in the transistors' transconductances and in their output conductances. However, nonlinearities of transconductances are usually not very relevant as they are more or less constant over technologies (under comparable biasing) and because the signal swing $v_{\rm gs}$ is usually low in a feedback or amplifier configuration. Fig. 2 shows the transconductance normalized with respect to the drain current [7] as a function of the gate-overdrive voltage $V_{\rm GT} = V_{\rm GS} - V_T$, derived from measurement on transistors in four different technologies; clearly the normalized g_m hardly changes over technology. For output conductance we find the opposite: it is heavily dependent on biasing, size, technology and typically sees large voltage swings. For this reason the remainder of this section reviews trends in output conductances. As a simple estimate of linearity, let us assume an MOS transistor with a nonlinear output conductance, and an output voltage consisting of a sine superimposed on a dc voltage. The drain current can be approximated by

$$i_D \cong I_D + g_{\rm ds}\cdot\hat{V}\cdot\sin(\omega t) - g_{\rm ds2}\cdot\frac{\hat{V}^2}{4}\cdot\sin(2\omega t) - g_{\rm ds3}\cdot\frac{\hat{V}^3}{24}\cdot\sin(3\omega t) + \cdots \quad (1)$$

The harmonic current components typically add to total harmonic distortion, but can be suppressed with voltage loop gain. Voltage loop gain at quasi-dc frequencies in transistor circuits is the combined effect of a number of $g_m/g_{\rm ds}$ratios and as such also depends on transistor output conductance. Fig. 3 shows graphs of transistor voltage gain and distortion, the latter expressed in output $\mathrm{IP}_3 = \sqrt{24\cdot g_{\rm ds3}/g_{\rm ds}}$, as a function of the effective gate-overdrive voltage $V_{\rm GT}$. The curves are derived from nonlinearly interpolated measurements on devices from four technologies; the length of all transistors was 1 μm for comparison reasons. Note that the curves are independent of the drain-current level.

Fig. 3(a) shows that at constant gate-overdrive voltage $V_{\rm GT}$ and fixed drain source voltage (here 0.3 V) the voltage gain of transistors decreases somewhat with newer technologies: a factor 2 in four generations. Fig. 3(b) shows that with the same scaling the output IP3 improves. The combined effect of these two trends is that at fixed transistor length and at constant voltage headroom, the quasi-dc circuit performance hardly changes over technology. Note that constant voltage headroom implies that no supply voltage downscaling is done.

B. DC Properties at Decreasing Voltage Headroom

With migration to more advanced CMOS processes usually the supply voltage of a circuit realized in that technology is decreased, implying that the voltage headroom and signal swing of individual transistors are also decreased. Fig. 4(a) shows the transistor voltage gain, following from nonlinearly interpolated measurements, now under the assumption that the quiescent

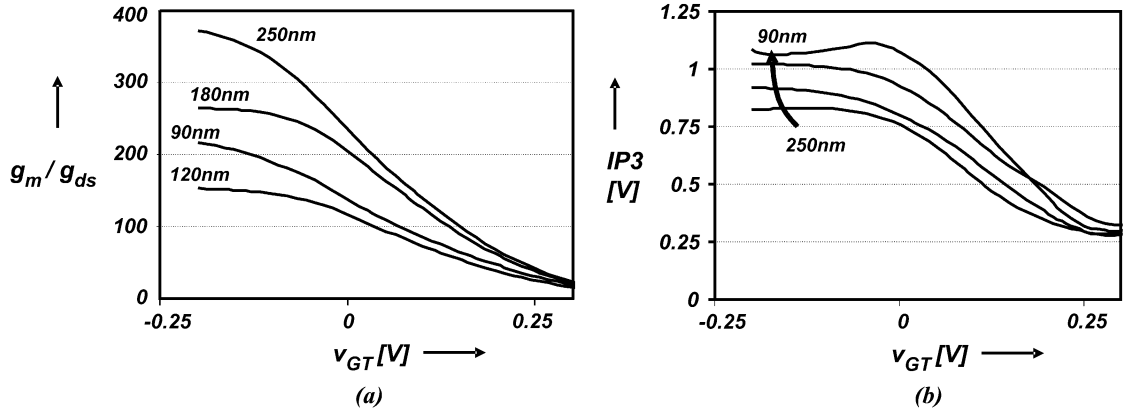


Fig. 3. Various DC-properties of transistors as a function of the gate-overdrive voltage at fixed $V_{\rm DS} = 0.3$ V and $L = 1\ \mu$m for four technologies: (a) the gain and (b) the output IP3.

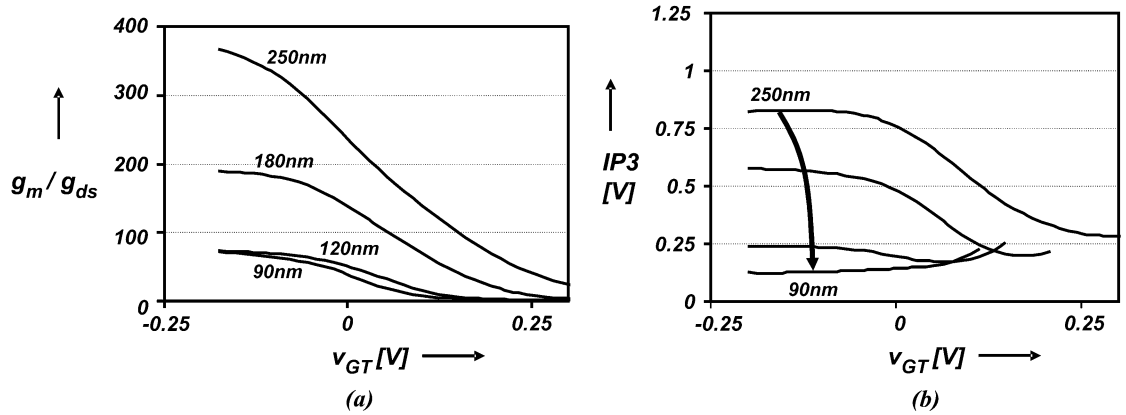


Fig. 4. Various dc properties of transistors as a function of the gate-overdrive voltage with $V_{\rm DS}$ proportional to the nominal supply voltage (0.3 V for 250 nm) and $L = 1\ \mu$m for four technologies: (a) the gain, and (b) the output IP3.

$V_{\rm DS}$ and the signal swing are decreased proportional to the nominal supply voltage, but still at constant transistor length; Fig. 4(b) shows the corresponding output IP3. Note that in order to compare only transistors in saturation, the $V_{\rm GT}$ranges are different for the four technologies.

Under the more realistice assumptions leading to Fig. 4, we see that porting typically results in significantly lowered transistor voltage gain and output IP3. With this scaling scenario, higher harmonic components may increase in amplitude despite the smaller signal; the THD 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.

AC Properties: The ac performance of transistors improves with newer technologies: it is one of the main technology drivers. As a first order estimate, two classes of capacitance are important for speed: the intrinsic capacitances of transistors and the junction capacitances. In this context intrinsic capacitances are all capacitances related to actual MOS operation, including overlap capacitances.

The impact of intrinsic capacitances hardly changes over technology for transistors under analog operating conditions with a fixed length. This can be illustrated using simple scaling theories [8], [9] combined with square-law relations that are satisfactory for illustrating trends. It follows that to first order the unity gain frequency of transistors depends only on effective gate-overdrive voltage and on channel length.

$$f_{UG} \approx \frac{g_m}{2\pi \cdot C_{\rm ox} \cdot WL} \propto \frac{V_{\rm GT}}{L^2} \tag{2}$$

where $C_{\rm ox}$ is the oxide capacitance per unit area. With a fixed transistor length, e.g., for voltage gain or accuracy reasons, the transistor's intrinsic speed hardly changes over technology. Combined with the findings in the previous parts of this section, it follows that there is a clear trade-off between gain and speed via transistor length (see also [10]). Circuits in newer CMOS technologies can hence achieve higher bandwidths but at the cost of degraded quasi-dc performance.

Another aspect of ac performance is the junction capacitance. With technology-scaling both the LDD structures and 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. This allows for better HF and RF performance with technology evolution.

IV. Gate Leakage and F_{GATE}

Besides its impact on conventional properties of circuits and devices, CMOS evolution introduces several new problems in analog design. One of the new phenomena is gate leakage [11]: gate current due to direct tunneling through the thin gate oxide. This leakage depends mainly on gate-source voltage bias and gate area.

One obvious implication of gate leakage is that the gate input impedance includes the conventional input capacitance $C_{\rm in}$ in parallel to a tunnel conductance $g_{\rm tunnel}$. These two have identical area dependence, resulting in an $f_{\rm gate}$ that is area *in*dependent and fairly independent of the drain-source voltage $v_{\rm DS}$. However, the tunnel current density for electrons and holes is different mainly due to the differences in oxide barrier height, resulting in (see Appendix B)

$$\begin{aligned} f_{\rm gate} &= \frac{g_{\rm tunnel}}{2\pi C_{\rm in}} \\ &\cong 1.5\cdot 10^{16}\cdot v_{\rm GS}^2\cdot e^{t_{\rm ox}(v_{\rm GS}-13.6)} \quad \text{(NMOST)} \\ &\cong 0.5\cdot 10^{16}\cdot v_{\rm GS}^2\cdot e^{t_{\rm ox}(v_{\rm GS}-13.6)} \quad \text{(PMOST)} \end{aligned} \tag{3}$$

where $t_{\rm ox}$ is in [nm] and $v_{\rm GS}$ is in [V].

For signal frequencies higher than this $f_{\rm gate}$ the input impedance is mainly capacitive and the MOSFET behaves as a conventional MOSFET. Otherwise, below $f_{\rm gate}$ it is mainly resistive and the gate leakage is dominant. Fig. 5 shows tens of $f_{\rm gate}$ curves based on measurements on transistors with different sizes and bias conditions in a 180-nm technology, and shows the prediction using (3). It follows that in this technology the gate appears to be capacitive for signal frequencies higher than roughly 0.1 Hz, while it appears resistive only at very low signal frequencies.

Analog applications typically apply transistors biased at low and moderate gate-overdrive voltages. The corresponding $f_{\rm gate}$ of a technology is then inside a relatively small frequency band. Fig. 6 shows such $f_{\rm gate}$ bands for four technologies as derived from measurements. This figure clearly illustrates that signal frequencies for which the input impedance appears to be resistive change from roughly 0.1 Hz in 180-nm technologies to about 1 MHz in 65-nm CMOS.

$f_{\rm gate}$ will prove to be useful in the estimation of the impact of gate leakage on other relevant properties of MOS transistors. A number of estimations are given in the next section of this paper.

V. Impact of Gate Leakage

A. Limited Current Gain

Input bias current due to gate leakage is very similar to base current in bipolar technologies and hence known solutions for bipolar circuits can usually be applied in analog CMOS circuits with leaky gates. There are two major differences between the bipolar base current and the CMOS gate current. First, in CMOS the width and length can be selected—and are optimized in analog designs—while in bipolar designs only the emitter area (equivalent to MOSFET width) can be set while the base width (equivalent to MOSFET length) is fixed. The result is that in ultra-deep-submicron processes long CMOS transistors (that are frequently required in conventional analog circuit designs,

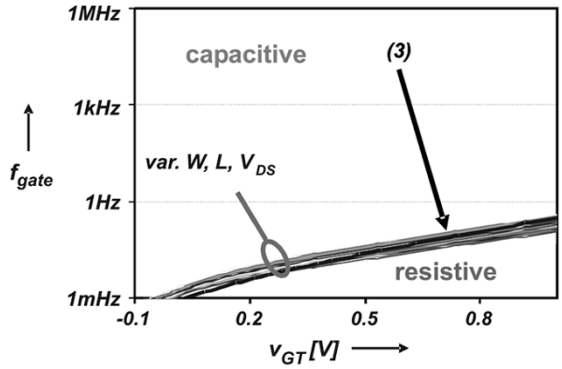


Fig. 5. $f_{\rm gate}$ as a function of the effective gate-overdrive voltage for different NMOS-transistors in 180-nm CMOS, based on measurements and fitted using (3).

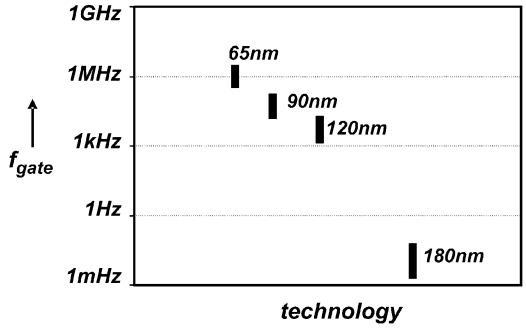


Fig. 6. $f_{\rm gate}$ ranges for typical analog applications, for NMOS transistors in different CMOS technologies. For PMOS transistors, $f_{\rm gate}$ is roughly a factor 3 lower.

for high output resistance or for low mismatch and flicker noise reasons) may have a lower-than-unity current gain. Secondly the bipolar base current is near zero under reverse bias conditions, whereas the gate current is not which should be taken into account, e.g., for switching applications.

The dc current gain can be estimated using $f_{\rm gate}$. In strong inversion and saturation at low gate-overdrive voltages, the drain current is by rough approximation given by the square-law relation. Using the expression for $g_{\rm tunnel}/i_{\rm GS}$ for strong inversion (see Appendix C)

$$\frac{g_{\rm tunnel}}{i_{\rm GS}} \approx \frac{1}{v_{\rm GS}-V_T}+\frac{1}{v_{\rm GS}}+\alpha_{gtun}\cdot t_{\rm ox} \qquad t_{\rm ox}\,\text{in}\,[\text{nm}],\quad \alpha_{gtun}\approx 1/\text{nm}\cdot\text{V}$$

with (3) the gate current at frequencies lower than $f_{\rm gate}$ can be rewritten into

$$\begin{aligned} i_{\rm GS} &= \frac{2\pi C_{\rm ox}\cdot W\cdot L\cdot f_{\rm gate}}{g_{\rm tunnel}/i_{\rm GS}} \\ &= \frac{2\pi C_{\rm ox}\cdot W\cdot L\cdot f_{\rm gate}}{\frac{1}{v_{\rm GS}-V_T}+\frac{1}{v_{\rm GS}}+\alpha_{gtun}\cdot t_{\rm ox}} \quad t_{\rm ox}\,\text{in}\,[\text{nm}] \end{aligned} \tag{4}$$

Substituting the gate-oxide thickness and with typical voltages for analog applications yields the rule-of-thumb estimation of $i_{\rm GS}$ for frequencies below $f_{\rm gate}$

$$i_{\rm GS}\approx \beta_{i\,\rm gs}\cdot C_{\rm ox}\cdot W\cdot L\cdot f_{\rm gate} \quad \text{with}\ \beta_{i\,\rm gs}\approx 1\,\text{V}. \tag{5}$$

Now, using the well-known quadratic approximation for the drain current

$$i_D = \frac{W}{2L} \cdot \mu C_{\rm ox} \cdot (v_{\rm GS} - V_T)^2$$

it follows that at frequencies lower than $f_{\rm gate}$, the current gain of an MOS transistor in strong inversion and saturation is given by (6). A similar relation can be derived for the weak-inversion region:

$$\frac{i_D}{i_{\rm GS}} \approx \frac{1}{L^2} \cdot \frac{\mu \cdot (v_{\rm GT} - V_T)^2}{2 \cdot \beta_{i\,\rm gs} \cdot f_{\rm gate}}. \tag{6}$$

In (6) both the carrier mobility and $f_{\rm gate}$ are about a factor 3 different for NMOS and PMOS transistors; the mobility ratio is specific to silicon, the $f_{\rm gate}$ ratio is due to oxide barrier differences and therefore also material-related. This leads to (7) for both types of transistors:

$$\frac{i_D}{i_{\rm GS}} \cong \vartheta \cdot \frac{v_{\rm GT}^2}{v_{\rm GS}^2 \cdot e^{t_{\rm ox} v_{\rm GS}} \cdot L^2}$$
$$\vartheta = 7 \cdot 10^{-6} \cdot e^{13.6 \cdot t_{\rm ox}} \quad L \text{ in } [\mu\text{m}]. \tag{7}$$

Fig. 7 shows the results of (7) and measurement results for a few transistors in two leaky technologies (gate currents are estimated for the 65-nm generation) at low effective gate-overdrive voltages. Clearly visible from this figure are the strong length dependence and the low current-gain for long transistors in ultra-deep-submicron CMOS technologies. It follows that long transistors cannot be usefully applied anymore in UDSM technologies. The figure also illustrates that the square-law estimation used in the derivation is adequate.

B. Self-Discharge Effects and Droop Rates

A large number of circuit designs apply MOS capacitances for storing charge. Examples include switched-current circuits, PLL loop filters, hold circuits and some switched capacitor circuits. Gate leakage causes a nonzero droop rate of the voltage across MOS capacitances (see Fig. 8) and thereby puts a bound on the maximum usable hold time and the minimum operating frequency. The droop rate of the leaky gate capacitance is

$$\frac{dv_C}{dt} = -\frac{i_G}{C_{\rm in}} = -\frac{i_G}{g_{\rm tunnel}} \cdot \frac{g_{\rm tunnel}}{C_{\rm in}}.$$

With the relation derived in Appendix C it follows that the droop rate in strong inversion is to a good approximation given by (8). A similar relation follows in weak inversion.

$$\frac{dv_C}{dt} \approx -\frac{2\pi \cdot f_{\rm gate}}{\frac{1}{v_{\rm GS}-V_T} + \frac{1}{v_{\rm GS}} + \alpha_{gtun} \cdot t_{\rm ox}}$$
$$t_{\rm ox} \text{ in } [\text{nm}], \quad \alpha_{gtun} \approx 1/\text{nm} \cdot \text{V}. \tag{8}$$

For typical UDSM gate-oxide thicknesses and typical analog operating conditions a rule-of-thumb estimation for the droop rate of MOS capacitances is given in (9):

$$\frac{dv_C}{dt} \approx -\gamma_{dv\,dt} \cdot f_{\rm gate} \left[\frac{V}{s}\right] \quad \text{with } \gamma_{dv\,dt} \approx 1\,\text{V}. \tag{9}$$

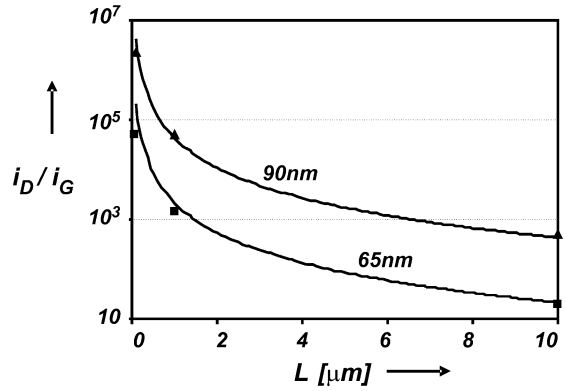


Fig. 7. Low-frequency current gain of MOS transistors in advanced CMOS technologies as a function of gate length, at $V_{\rm GS} = 0.5$ V. The curves follow from (7), 90-nm markers are based on measurements.

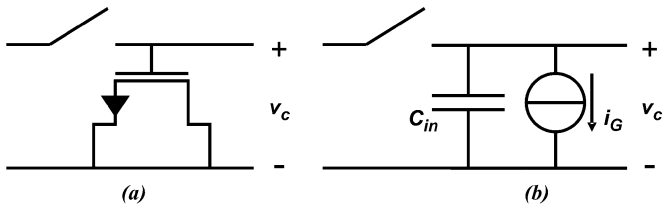


Fig. 8. (a) An MOS capacitance used in an (idealized) track-and-hold circuit, and (b) the equivalent circuit in "hold."

In words, the droop rate of a "stored" voltage on a MOS capacitor, in [V/s], is approximately equal to $f_{\rm gate}$ (in [Hz]). This relation implies that for track-and-hold circuits the maximum hold time for a droop amounting to ΔV is

$$\Delta t \approx \frac{\Delta V}{\gamma_{dv\,dt} \cdot f_{\rm gate}} \quad [\text{s}] \tag{10}$$

Allowing, e.g., 1-mV drop on a sampled-and-held value, the maximum usable hold time is in the millisecond range in 180-nm technologies, which is usually sufficient. However the maximum hold time decreases rapidly with newer technologies, down to a typical value in the low nano second range for 65-nm technologies. Note that this low maximum hold time makes it impossible to apply MOS capacitors in low and medium sample-rate A/D converters. Capacitors must then be realized either using thick-oxide devices, or inter-metal capacitances. If one can only use standard thin-oxide transistors as capacitances, PMOS transistors are half an order better than the NMOS transistors. Similar conclusions hold for PLL loop filters and switched-current circuits.

C. Gate-Leakage Matching and Its Implications

Gate leakage is caused by quantum-mechanical tunneling and depends on the layer thickness and the field strength. As such, it also exhibits spread. Relative spread, or matching, usually limits the achievable level of performance of analog circuits: it sets a lower bound on figures such as offsets in amplifiers and the accuracy in A/D converters.

Because spread and mismatch are dc effects, they do not (from a fundamental point of view) require any additional

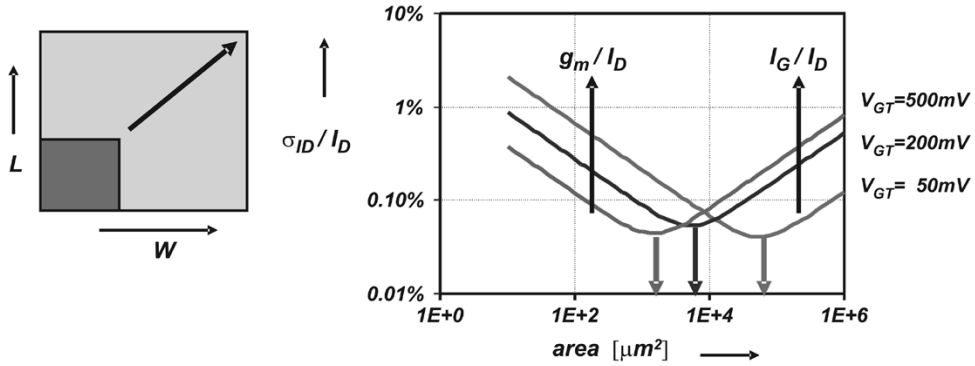


Fig. 9. The spread of an MOS transistor in 65-nm CMOS with linear scaling of W and L as a function of area: there are clear optima in the attainable matching.

power. However, in practice they prove to be a major implementation problem [12]. The usual way to get a sufficient level of matching between MOS transistors is to simply spend area [14], thereby increasing power consumption at a given speed, because larger capacitances have to be charged [15].

Compared to the conventional mismatch sources, gate-leakage mismatch now comes as an extra mismatch source with a *different* area dependency. We found that, excluding defect-like outliers, mismatch of gate leakage current is proportional to the gate current level with a proportionality constant of roughly $0.03/\sqrt{Area}$, where $Area$ is the transistor's gate area in square-microns. Assuming that conventional mismatch and gate current mismatch are uncorrelated, the total relative mismatch of a transistor's drain current is roughly

$$\frac{\sigma_{id}^2}{i_D^2} = \frac{\sigma_{id,\text{conventional}}^2}{i_D^2} + \frac{\sigma_{id,\text{gate leakage}}^2}{i_D^2} = \left(\frac{A_{\text{VT}}}{\sqrt{WL}} \cdot \frac{g_m}{i_D}\right)^2 + \left(\frac{X_{\text{IGS}}}{\sqrt{WL}} \cdot \frac{i_G}{i_D}\right)^2 \quad (11)$$

where $X_{\text{IGS}} \approx 0.03$.

The first term is the conventional mismatch due to mismatch in threshold voltage, with A_{VT} the matching coefficient [12]–[14]. This A_{VT} is a technology-related factor that is roughly proportional to the gate-oxide thickness, saturating in UDSM technologies around 2–3 mVμm [13]. The second term is the mismatch in gate current as introduced here. Note that the mismatch in the current factor β of the transistors is neglected, which is allowed for practical values of V_{GT} [14].

1) Improving Matching: The Classical Way: The classical way—without incorporating gate current mismatch [14]—to decrease mismatch is to spend area and to set an optimum g_m/I_D ratio. For a number of applications the input referred mismatch of transistors $\sigma_{vgs} = \sigma_{id}/g_m$ is relevant; minimization of input-referred mismatch typically comes down to maximizing g_m/I_D or increasing the gate-area in some way. Decreasing output referred mismatch σ_{id} can be achieved by lowering g_m/I_D or by increasing the gate-area of the transistor. By linearly scaling W *and* L of a transistor, its bias settings are unchanged while the matching improves proportionally to the scale factor. When scaling only device widths, and keeping L constant the current level increases proportionally and matching improves as the square root of the width scale factor [16].

2) Improving Matching: Including Gate-Leakage Effects: If gate leakage mismatch is accounted for, like in (11), the conventional mismatch decreasing rules cannot be applied any more. The impact of gate-leakage mismatch on the overall mismatch of MOS transistors is again best illustrated using the square-law relation; this relation is sufficient for rough estimation purposes. With expression (7) for the current gain of a MOS transistor it follows that

$$\frac{\sigma_{ID}^2}{i_D^2} = \left(\frac{\varsigma}{\sqrt{WL}}\right)^2 + \left(\frac{\xi \cdot L^2}{\sqrt{WL}}\right)^2 \quad (12)$$

where ς is related to conventional mismatch and ξ is related to gate current mismatch. In the classical way the matching could be improved by spending more area in any way. In UDSM technologies, (12) shows that by spending more area the conventional mismatch contribution always decreases but that at the same time the gate-mismatch contribution may increase. This latter contribution increases if the transistor length increases, as is the case with linear scaling of W and L of the transistors. Hence, with linear scaling by increasing W and L, the total relative mismatch in drain current may increase for large gate areas, which effectively limits the maximum usable transistor area. For 180-nm and 120-nm CMOS technologies this maximum usable area is very large and hence the attainable levels of matching are very good. However, for the 90-nm (measured matching) and 65-nm (estimated matching) generation this yields maximum usable areas in the order of respectively 10^4 μm^2 and 10^3 μm^2: then gate-leakage mismatch is a significant effect that limits the attainable matching. This is illustrated in Fig. 9. In this case, active mismatch cancellation techniques or matching-insensitive designs are required.

On the other hand, scaling only the transistor width and scaling the current levels in proportion, both the gate-leakage contribution and the conventional matching term improve with the square root of the scale factor. This is illustrated in Fig. 10. In this case essentially any level of matching can be obtained at the expense of area and power consumption. Active matching cancellation techniques are not required here, but can be used to break the area–power-matching relation.

D. Noise

Just as any current across a junction, gate leakage exhibits shot noise with current density $S_{IG} = 2q \cdot I_G$. As such, it is

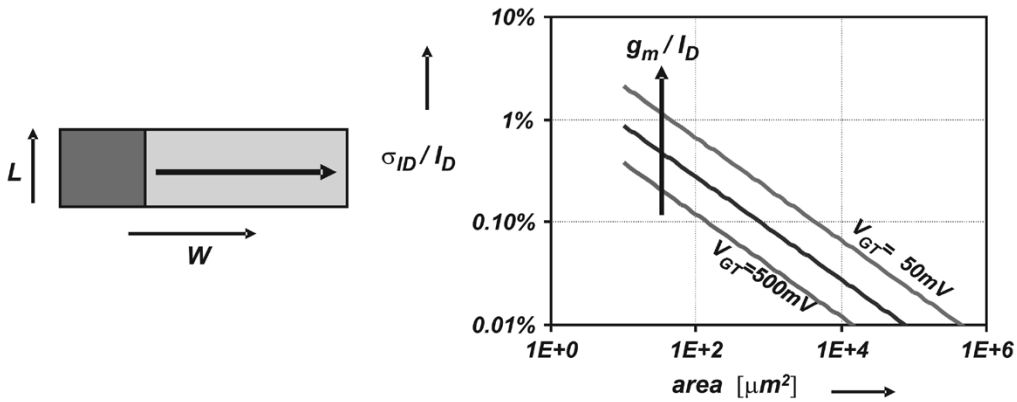


Fig. 10. The spread of an MOS transistor in 65-nm CMOS with width scaling, and constant L as a function of area: no minimum in the attainable matching at the cost of increased power consumption.

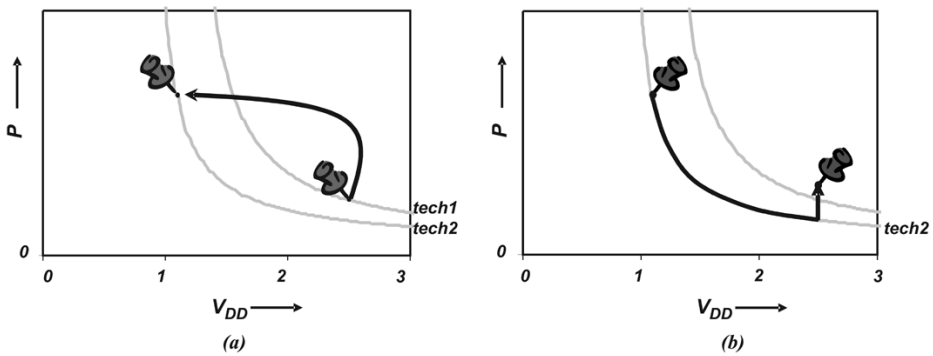


Fig. 11. Porting older designs to newer CMOS technology: (a) with supply voltage down scaling with a significant power consumption increase, and (b) a high-voltage version of the original circuit can operate at higher-than-nominal supply voltages with its associated lower power consumption. Implementation overhead to enable high-voltage operation increases power consumption.

equivalent to base currents in bipolar transistors. Note that at RF frequencies effects like the induced gate noise also contribute to the total gate noise [17], [18]. Noise in the gate current will therefore limit noise performance in analog circuits in UDSM CMOS.

VI. A Solution: Living Outside Rails

The two major problems associated with analog circuit design in UDSM technologies are the low supply voltage and the gate-leakage related effects. One strategy to deal with the low-supply drawback is to operate critical parts of analog circuits at a supply voltage significantly higher than the nominal supply voltage for the CMOS process used. This typically reduces the power consumption significantly at a given level of performance [19]–[21], but requires a focus on lifetime issues such as oxide breakdown [22], hot carriers [23], [24], NBTI [25] and junction breakdown [26]. Generally junction breakdown is not a major issue while hot carrier degradation does not play a significant role at supply voltages lower than 1 V. The other two effects need to be limited by a suitable limitation of terminal-pair voltages. Techniques known from high-voltage I/O circuits can be readily used for this. A brief review is presented at the end of this section.

Analog Circuits at High Supply Voltages: The effect of operation of analog circuits in UDSM CMOS at a high supply voltage, up to a few times as high as the nominal supply voltage $V_{\mathrm{DD,nom}}$ for the process, is illustrated in Fig. 11. Both curves in Fig. 11(a) and (b) show the minimum power consumption as a function of the supply voltage, for a circuit in a CMOS technology at constant performance, with optimized bias settings and device dimensions for each technology and supply voltage [6]. In these figures, the upper curves (tech1) correspond to an older technology while the lower curve (tech2) is a newer technology. For analog, the migration to a newer CMOS technology with its associated lower nominal supply voltage results in increased power consumption at fixed performance, indicated by the a) arrow in Fig. 11(a). Note that the jump to another curve corresponds to going to another technology with the same circuit topology. However, implementing the circuit in such a way that it runs at a high supply voltage can significantly decrease power consumption; this is indicated by arrow b) in Fig. 11(b). For reliability reasons typically circuit overhead is required, resulting in the upward part of the b) arrow.

An obvious advantage of migration to more advanced CMOS technologies is that it enables selective application of low-voltage transistors with their specific advantages and disadvantages. Especially interesting is the digital computational power that can solve many deterministic analog inaccuracies (e.g., mismatch and distortion).

Running Analog Circuits at High Supply Voltages: For circuits operating at high supply voltages, a number of robust high-voltage-tolerant transistors can be used to replace the standard transistors that can only reliably operate up to nominal supply

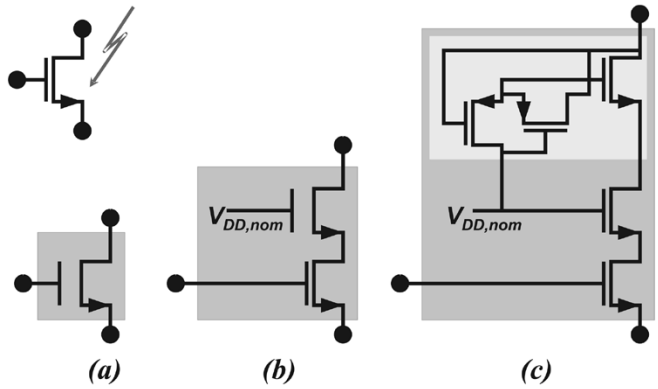


Fig. 12. Ways to implement high-voltage tolerant transistors in standard CMOS: (a) thick-oxide transistor; (b) (thick-oxide) cascode; (c) retractable cascode composite transistor.

voltages. Fig. 12 presents three examples known from high-voltage (HV) I/O circuits. A number of circuits and blocks with analog behavior implementing the retractable cascode HV transistors shown in Fig. 12(c) were discussed in [3] and [27]. These HV transistors enable direct reuse of most older circuit architectures, running at supply voltages corresponding to the original design: high supply voltages when related to the nominal supply voltage of the CMOS process used. Extended-drain transistors that can be realized in standard CMOS could also be used; such structures are described in [28].

The easiest way is to use the (commonly available) thick-oxide transistor [Fig. 12(a)] that is comparable to a two-generations-old standard transistor. However, in order to benefit from technology scaling compound structures using thin-oxide transistors, as in Fig. 12(b) or (c),[1] typically outperform the thick-oxide transistor in Fig. 12(a) in the fields of matching, $1/f$ noise and output impedance. These compound structures have some disadvantages: they are asymmetric, do not solve gate-leakage issues, and require suitable cascode voltages at power-up.

Careful selection of the analog sections to run at high supply voltages, and careful selection of the best type of transistor (thin oxide, thick oxide, or compound) will to a great extent circumvent one of the main roadblocks in UDSM CMOS technolgies: the low nominal supply voltage. Note that thick-oxide transistors also solve gate-leakage issues as their gate leakage is usually negligible.

VII. Conclusion

Modern and future UDSM CMOS introduce several new problems for analog circuit design. From a fundamental point of view, lowering the analog supply voltage leads to an increase in power dissipation at constant performance. This increase in power dissipation becomes drastic as the supply voltage approaches the threshold voltage plus a few hundred millivolts, as illustrated in Fig. 1.

[1]The retractable cascode structure typically uses one cascode with a fixed gate voltage, and one or more cascodes with variable gate voltages. In Fig. 12(c), the variable gate voltage is soft-switched by two PMOS transistors between the drain voltage of the switched transistor and the nominal supply voltage, whichever is highest.

When migrating to modern technologies the quasi-dc analog transistor properties hardly change, as long as constant transistor lengths and terminal voltages are used. However, if the supply voltage is reduced according to the technology roadmap, the analog performance is lowered because of the lower bias voltages. Nonlinear output conductance in combination with reduced voltage gain pose limits with respect to linearity of (feedback) circuits.

Gate leakage becomes a serious problem in upcoming technologies, especially if long transistors are used. A parameter $f_{\rm gate}$ is introduced that enables quick estimations of gate-leakage related effects. Besides gate leakage itself, mismatch in gate leakage introduces new limitations. Mismatch cannot be tackled anymore by simply spending more area for transistors: when the transistor length L is increased an upper limit to matching is encountered. Here, increasing area does not automatically improve overall matching anymore, except if higher power consumption is accepted or active cancellation techniques are applied.

Operating critical parts at higher supply voltages, by exploiting combinations of thin- and thick-oxide transistors can solve the low voltage as well as the gate leakage problems. Composite transistors are presented in Fig. 12 to solve this problem in a practical way. In summary: unlike digital designs, analog circuits benefit from technology scaling if the supply voltages are *not* scaled down.

Appendix A

In this Appendix, known performance–power relations for active circuits are briefly reviewed and their impact is discussed.

A. Performance in SNR: Total Integrated Noise

A number of papers on the relation between analog performance and power consumption specify the performance in only its signal-to-noise-ratio (SNR) and the signal bandwidth [4], [5]. Typically, only the total integrated thermal noise is taken into account. In these papers, a system as depicted in Fig. 13(a) is used: an unspecified analog circuit represented by a resistance or conductance.

Including only thermal noise integrated over the total noise bandwidth of the circuit, the integrated noise voltage and the required (class-A) bias current are

$$\bar{v}_n^2 = \frac{kT}{C} \quad \text{and} \quad I_{\rm bias} = 2\pi f_{\rm sig} C \hat{V}.$$

These equations lead to a minimum power consumption of an analog circuit given by

$$P = \frac{8\pi kT \cdot \mathrm{SNR} \cdot f_{\rm sig}}{\eta_{\rm vol} \cdot \eta_{\rm cur}}$$

where $\eta_{\rm vol}$ is the ratio between the peak-peak signal swing and the supply voltage [15], $\eta_{\rm cur}$ is the efficiency of using supply current [15], kT are Boltzmann's constant and the temperature, respectively, SNR is the circuit's signal-to-noise ratio as a power ratio, $f_{\rm sig}$ is the signal frequency, $\hat{V}$ is the signal amplitude.

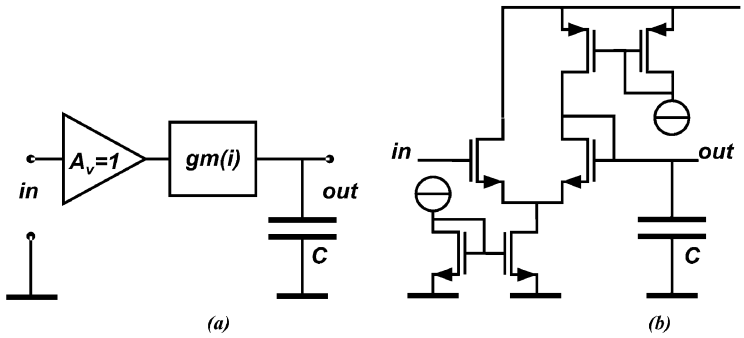


Fig. 13. (a) Circuit representation assumed for the power-performance relations: a circuit with an output resistance and an output load, and (b) actual circuit topology used in [6].

Taking into account only thermal noise, the power consumption required for some SNR is technology independent if both the parameters $\eta_{\rm vol}$ and $\eta_{\rm cur}$ are invariant over technology. In general, $\eta_{\rm vol}$ and $\eta_{\rm cur}$ increase with newer technologies because of voltage overhead (e.g., to accommodate gate-source overdrive voltage and saturation [15]) and due to the use of folded structures.

B. Performance in SNR: Bandwidth Limited Noise

For many circuits the total integrated thermal noise at the output is irrelevant: only the noise within some frequency band is of interest. In this case, neither the thermal noise nor the required bias current is related to the output capacitance: this capacitance is now purely parasitic. For this type of circuit, the integrated thermal noise voltage is[2]

$$\bar{v}_n^2 = \frac{4kT}{gm} BW$$

where BW is the relevant frequency band.

For ordinary active electronic components, the transconductance g_m is a function of the bias current and of some voltage. For bipolar transistors and MOS transistors

$$g_m = \frac{qI_{\rm BIAS}}{kT} \quad \text{and} \quad g_m \cong \frac{2I_{\rm BIAS}}{v_{GT}}$$

where a lower bound to the effective gate-source overdrive voltage v_{GT} for MOS transistors is weak-inversion (bipolar-like) operation. These two expressions can be captured in one: $g_m = I_{\rm BIAS}/v_{\rm OD}$ where $v_{\rm OD}$ is something like equivalent effective overdrive voltage. The minimum power consumption to reach a certain SNR is then

$$P = \frac{16kT \cdot {\rm SNR} \cdot BW}{\eta_{\rm vol} \cdot \eta_{\rm cur}} \cdot \frac{v_{\rm OD}}{\widehat{V}}.$$

It also follows that the power consumption is proportional to the targeted SNR. However, lowering the supply voltage (and signal swing) without decreasing the $v_{\rm OD}$ increases the required power consumption. Note that this situation always occurs with bipolar transistors and MOS transistors in moderate or weak inversion.

[2]Actually the noise should be corrected with the circuit noise excess factor. For simplicity reasons this is neglected here.

C. Performance in SINAD: Total Integrated Noise

In many analog circuits, both noise and distortion are relevant to the performance. For those circuits the signal-to-noise-and-distortion ratio (SINAD or SNDR) may be the right way to express the performance. Examples of circuits for which this type of performance is relevant include switched-capacitor circuits and track-and-hold circuits. The circuit corresponds to that in Fig. 13(b). In general, it is an unspecified analog circuit driving an explicit load capacitance as shown in Fig. 13(a). The actual analog circuit can be anything ranging from a nonlinear resistance to an analog amplifier with any amount of global or local feedback. In [6], it was shown that, taking the total integrated thermal noise into account, for weakly nonlinear analog circuits with a dominant load capacitor the absolute minimum power consumption is

$$P_{\rm min} = 2 \cdot \hat{V} \cdot gmi\left(2kT \cdot \pi \cdot f_{\rm sig} \sqrt[n]{\frac{2\alpha_n}{\hat{V}}} \cdot {\rm SINAD}^{\frac{2n+1}{2n}} \cdot \frac{(2n+1)^{\frac{2n+1}{2n}}}{n\hat{V}}\right)$$

where n is the ordinal number of the dominant harmonic, α_n links higher harmonics to the first harmonic[3] as $h_n = \alpha_n (h_1)^n$, and $gmi(g)$ is the inverse of the function between the conductance and bias current of a circuit.

In the derivation of this expression, distortion, and noise are traded against each other in such a way that a maximum performance–power consumption ratio is reached. Note that this general expression is much more complex than its SNR-based counterpart, presented in Section A. For ordinary weakly nonlinear analog circuits, substituting the $gmi(x)$ function and $\alpha_n$results in a circuit-specific SINAD-P relation with many similarities to the SNR-P relation mentioned above; see [6] for two examples. The biggest difference between the SNR expression and the SINAD expression is that the latter contains a multiplicative term that increases with decreasing signal swing, and hence with the lower supply voltage that comes with newer CMOS generations.

[3]These variables describe the nonlinearity of the analog circuit and follow from a Taylor series expansion of the dc-transfer curve of the circuit.

With the assumptions leading to the SNR limit (marginally preventing both clipping and slewing) straightforward mathematics shows that the complex SINAD-P expression collapses to the SNR-P one[4] in Section A:

$$\lim_{n\to\infty} P_{\min} = 8\pi kT f_{\text{sig}}\,\text{SINAD}|_{D\to 0} \equiv 8\pi kT f_{\text{sig}}\,\text{SNR}.$$

D. Performance in SINAD: Bandwidth Limited Noise

For circuits dealing with a SINAD performance specification, either the total integrated noise or bandwidth limited noise may be relevant. In the case of bandwidth-limited noise, there is no need for an explicit load capacitance. As a direct consequence of the absence of any required bandwidth limitation, from a mathematical point of view harmonic distortion is zero. The relation between power consumption and bandwidth-limited SINAD therefore equals the relation for power and bandwidth-limited noise under B.

E. Summary

From a fundamental point of view it can be concluded that lowering the supply voltage increases the power-performance ratio, with exception of the simplest case described under A. Moreover, any voltage overhead ΔV worsens the power-performance ratio with decreasing supply voltage V_{DD}, typically introducing a multiplicative term $V_{\text{DD}}/V_{\text{DD}} - \Delta V$ in the power relation.

F. Exceptions to the Rule

Most analog circuits comply with the power-performance relations discussed here. Obvious exceptions to this rule are circuits that are overly robust in some aspect, for example most flash A/D converters that minimize mismatch issues by spending area, or that cannot satisfy some scaling issues, e.g., low-noise amplifiers.

Flash A/D Converters: Practical findings indicate that the power consumption in flash A/D converters is not determined by thermal noise issues, see, e.g., [15]. Typically, low-resolution flash converters aim to reach a certain level of matching, by spending area, at some operating frequency. Under these conditions matching and speed requirements determine the power consumption, and consequently the power consumption of flash A/D converters decreases with newer CMOS generations because of better matching properties. However, when using active matching techniques [15], [29], [30], the need to spend area for matching is absent and the power–performance relation is determined by SNR issues again [15]. Note that it is a fundamental property of dc-type disturbances that no power is needed for their minimization. The apparent independence of SNR and power consumption in flash A/D converters is therefore due to the practical way mismatch is dealt with.

Low-Noise Amplifiers: Other circuits that do not comply with the discussed power-performance relations include RF low-noise amplifiers (LNAs). In the derivation of these relations, the signal swing is optimized for a given supply voltage. However, in LNA-type circuits the signal swing is fixed and lower supply voltages result in somewhat degraded bias circuitry and in a lower implementation overhead [15]. The overall result is that for circuits with a fixed very low signal swing, the power–performance ratio can improve with newer CMOS generations [31], [32].

[4]This is true for at least most of the $g_m(I)$ functions that can be realized in standard electronics.

Appendix B

Using a simplified relation for gate current based on the gate current model in MOS Model 11 [33], we can write the following relation for a MOSFET in saturation. In this relation, the effects of overlap regions are neglected:

$$i_{\text{GS}} = A \cdot v_{\text{INV}} \cdot v_{\text{GS}} \cdot \exp(B \cdot v_{\text{GS}})$$

where v_{INV} is the effective gate bias, given by

$$v_{\text{INV}} = m \cdot \varphi_T \cdot \ln\left(1 + \exp\left[\frac{v_{\text{GS}} - V_T}{m \cdot \varphi_T}\right]\right)$$

and A and B are constants given by

$$A = \frac{I_{\text{GINV}}}{2} \cdot \exp\left[-\frac{3}{2} \cdot \frac{B_{\text{INV}}}{\chi_B}\right] \quad \text{and} \quad B = \frac{3}{8} \cdot \frac{B_{\text{INV}}}{\chi_B^2}.$$

In the above, m determines the subthreshold slope ($\text{m} = 1.3$), χ_B is the oxide potential barrier ($\chi_B = 3.1$ V for electrons, $\chi_B = 4.5$ V for holes), and I_{GINV} and B_{INV} are physical parameters dependent on oxide thickness t_{ox}, channel length L and channel width W. For electrons, we can write

$$I_{\text{GINV}} = 1.6 \cdot 10^{-4} \cdot \frac{WL}{t_{\text{ox}}^2} \quad t_{\text{ox}} \text{ in [m]}$$
$$B_{\text{INV}} = 2.9 \cdot 10^{10} \cdot t_{\text{ox}} \quad t_{\text{ox}} \text{ in [m]}.$$

As a simple approximation, we get the following expressions for the factors A and B, now with the oxide thickness in [nm] and assuming NMOS transistors. Note that gate current is proportional to the total gate area.

$$A = WL \cdot \frac{1.6 \cdot 10^{14}}{t_{\text{ox}}^2} \cdot \exp(-14 \cdot t_{\text{ox}}) \quad t_{\text{ox}} \text{ in [nm]}$$
$$B = 4.5 \cdot t_{\text{ox}}.$$

For the input capacitance C_{GG}, we find (also neglecting the overlap regions) the following relation, where C_{OX} is the total oxide capacitance. Note that this term is also proportional to the gate area.

$$C_{GG} = \frac{2}{3} \cdot C_{\text{OX}} \cdot \frac{\partial v_{\text{INV}}}{\partial v_{\text{GS}}} = \frac{2}{3} \cdot \frac{WL \cdot \varepsilon_{\text{ox}}}{t_{\text{ox}}} \cdot \frac{\partial v_{\text{INV}}}{\partial v_{\text{GS}}}$$

with

$$\frac{\partial v_{\text{INV}}}{\partial v_{\text{GS}}} = \left(1 + \exp\left[-\frac{v_{\text{GS}} - V_T}{m \cdot \varphi_T}\right]\right)^{-1}.$$

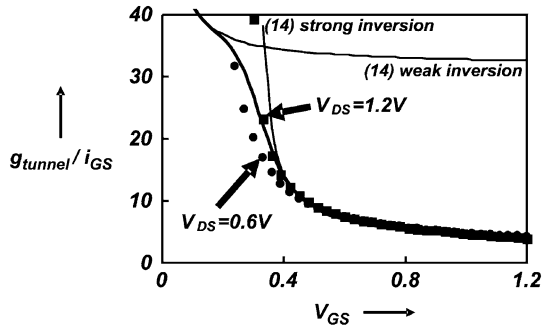


Fig. 14. Ratio $g_{\rm tunnel}/i_{\rm GS}$ as a function of gate bias $v_{\rm GS}$. Markers are measurement results and curves are predicted results using (13) and both expressions of (14).

The frequency where the imaginary part and the real part of the input impedance are equal is

$$f_{\rm gate} = \frac{1}{2 \cdot \pi \cdot C_{GG}} \cdot \frac{\partial i_{\rm GS}}{\partial v_{\rm GS}}$$

with

$$\frac{\partial i_{\rm GS}}{\partial v_{\rm GS}} = A \cdot \exp(B \cdot v_{\rm GS}) \cdot \left[v_{\rm INV} \cdot (1 + B \cdot v_{\rm GS}) + v_{\rm GS} \cdot \frac{\partial v_{\rm INV}}{\partial v_{\rm GS}} \right].$$

it follows that in strong inversion and saturation, where $v_{\rm INV} \approx v_{\rm GS} - V_T$ and $\partial v_{\rm INV}/\partial v_{\rm GS} = 1$ and $C_{GG} = (2/3) \cdot C_{\rm OX}$, $f_{\rm gate}$ for NMOS transistors can be approximated by

$$f_{\rm gate} \cong \frac{\vartheta_{fgate} \cdot v_{\rm GS}^2}{\exp(\upsilon_{fgate} \cdot t_{\rm ox})} \exp(\zeta_{fgate} \cdot t_{\rm ox} v_{\rm GS}) \quad t_{\rm ox} \text{ in } [\rm nm]$$
$$\vartheta_{fgate} \approx 1.5 \cdot 10^{16}\ \rm Hz/V^2$$
$$\upsilon_{fgate} \approx 13.6/\rm nm$$
$$\zeta_{fgate} \approx 1/\rm nm \cdot V.$$

Similar relations can be derived for the moderate and weak inversion regions, and the linear region. For PMOS transistors $f_{\rm gate}$ is roughly a factor 3 lower due to the higher oxide potential barrier $\chi_{\rm B}$.

APPENDIX C

The relation between gate conductance and gate current can be used in a number of expressions that link some gate-leakage related property to the size-independent $f_{\rm gate}$. In this paper the properties discussed are the dc-current gain and the self-discharge droop-rate of MOS-capacitances. Using Appendix B, the tunnel conductance $g_{\rm tunnel} = \partial i_{\rm GS}/\partial V_{\rm GS}$ can be readily calculated. The normalized gate conductance is then

$$\frac{g_{\rm tunnel}}{i_{\rm GS}} = \frac{\partial v_{\rm INV}/\partial v_{\rm GS}}{v_{\rm INV}} + \frac{1}{v_{\rm GS}} + \alpha_{gtun} \cdot t_{\rm ox} \quad t_{\rm ox} \text{ in } [\rm nm], \quad \alpha_{gtun} \approx 1.13/\rm nm. \quad (13)$$

In strong inversion and weak inversion, respectively, this relation can be approximated by the following ones. Note that the dimensions are correct because of implicit multiplication by appropriate scale factors.

$$\frac{g_{\rm tunnel}}{i_{\rm GS}} \approx \frac{1}{v_{\rm GS} - V_T} + \frac{1}{v_{\rm GS}} + \alpha_{gtun} \cdot t_{\rm ox} \quad t_{\rm ox} \text{ in } [\rm nm], \quad \alpha_{gtun} \approx 1.13/\rm nm$$
$$\frac{g_{\rm tunnel}}{i_{\rm GS}} \approx \frac{1}{m \cdot \varphi_T} + \frac{1}{v_{\rm GS}} + \alpha_{gtun} \cdot t_{\rm ox} \quad t_{\rm ox} \text{ in } [\rm nm], \quad \alpha_{gtun} \approx 1.13/\rm nm. \quad (14)$$

As an example, Fig. 14 shows measured data for a 10 μm $*$ 10 μm transistor in a 120-nm technology, with the more exact expression and the two approximations for the weak inversion and strong inversion region. Clearly the simple relations comply very well to measurements.

ACKNOWLEDGMENT

The authors would like to thank J. Schmitz for fruitful discussions on lifetime issues, A. Kumar and A. Heringa from Philips Research for providing the 65-nm CMOS parameters used in this paper, and the anonymous reviewers for their valuable suggestions.

REFERENCES

[1] Y. Taur, "CMOS design near the limit of scaling," *IBM J. Res. & Dev.*, vol. 46, no. 2/3, pp. 213–222, 2002.
[2] D. D. Buss, "Technology in the internet age," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, 2002, pp. 18–21.
[3] A. J. Annema, B. Nauta, R. van Langevelde, and H. Tuinhout, "Designing outside rails constraints," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, 2004, pp. 134–135.
[4] J. O. Voorman and D. Blom, "Noise in gyrator capacitor filters," *Philips' Res. Rep.*, vol. 26, no. 4, pp. 114–133, 1971.
[5] E. A. Vittoz, "Low-power design: Ways to approach the limits," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, 1994, pp. 14–18.
[6] A. J. Annema, "Analog circuit performance and process scaling," *IEEE Trans. Circuits Syst. II*, vol. 46, no. 6, pp. 711–725, 1999.
[7] D. Foty, D. Binkley, and M. Bucher, "Starting over: $\rm g_m/I_d$-based MOSFET modeling as a basis for modernized analog design methodologies," in *Proc. Nanotech*, vol. 1, 2002, pp. 682–685.
[8] R. H. Dennard, F. H. Gaensslen, H. N. Yu, V. L. Rideout, E. Bassous, and A. R. LeBlanc, "Design of ion-implanted MOSFET's with very small physical dimensions," *IEEE J. Solid-State Circuits*, vol. 9, pp. 256–268, 1974.
[9] H. Iwai, "CMOS technology—Year 2010 and beyond," *IEEE J. Solid-State Circuits*, vol. 34, no. 3, pp. 357–366, Mar. 1999.
[10] M. Garg, S. S. Suryagandh, and J. S. Woo, "Scaling impact on analog performance of sub-100 nm MOSFET's for mixed mode applications," in *Proc. ESSDERC*, 2003, pp. 371–374.
[11] R. van Langevelde, A. J. Scholten, R. Duffy, F. N. Cubaynes, M. J. Knitel, and D. B. M. Klaassen, "Gate current: Modeling, ΔL extraction and impact on RF performance," in *IEDM Tech. Dig.*, 2001, pp. 289–292.
[12] M. J. M. Pelgrom, H. P. Tuinhout, and M. Vertregt, "Transistor matching in analog CMOS applications," in *IEDM Tech. Dig.*, 1998, pp. 915–918.
[13] J. Dubois, J. Knol, M. Bolt, H. Tuinhout, J. Schmitz, and P. Stolk, "Impact of source/drain implants on threshold voltage matching in deep-submicron CMOS technologies," in *Proc. ESSDERC*, 2002, pp. 115–118.
[14] M. J. M. Pelgrom, A. C. J. Duinmaijer, and A. P. G. Welbers, "Matching properties of MOS transistors," *IEEE J. Solid-State Circuits*, vol. 24, no. 10, pp. 1433–1440, Oct. 1989.
[15] K. Bult, "Analog design in deep-submicron CMOS," in *Proc. ESSCIRC*, 2000, pp. 11–17.
[16] E. Klumperink and B. Nauta, "Systematic comparison of HF CMOS transconductors," *IEEE Trans. Circuits Syst. II*, vol. 50, no. 10, pp. 728–741, Oct. 2003.

[17] A. J. Scholten, L. F. Tiemeijer, R. van Langevelde, R. J. Havens, A. T. A. Zegers-van Duijnhoven, and V. C. Venezia, "Noise modeling for RF CMOS circuit simulation," *IEEE Trans. Electron Devices*, vol. 50, no. 3, pp. 618–632, Mar. 2003.

[18] R. van Langevelde, J. C. J. Paasschens, A. J. Scholten, R. J. Havens, L. F. Tiemeijer, and D. B. M. Klaassen, "New compact model for induced gate current noise," in *IEDM Tech. Dig.*, 2003, pp. 867–870.

[19] T. A. F. Duisters and E. C. Dijkmans, "A -90-dB THD rail-to-rail input opamp using a new local charge pump in CMOS," *IEEE. J. Solid-State Circuits*, vol. 33, no. 7, pp. 947–955, Jul. 1998.

[20] E. van Tuijl, J. van den Homberg, D. Reefman, C. Bastiaansen, and L. van de Dussen, "A 128 fs multi-bit $\Sigma\Delta$ CMOS audio DAC with real-time DEM and 115 dB SFDR," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, 2004, pp. 368–369.

[21] M. Keskin, U. K. Moon, and G. C. Temes, "A 1-V 10-MHz clock-rate 13-bit CMOS $\Sigma\Delta$ modulator using unity-gain-reset opamps," *IEEE J. Solid-State Circuits*, vol. 37, no. 7, pp. 817–824, Jul. 2002.

[22] "Special issue on scaling limits of gate oxides," *Semicond. Sci. Technol.*, vol. 15, no. 5, 2000.

[23] P. Heremans, R. Bellens, G. Groeseneken, and H. Maes, "Consistent model for the hot-carrier degradation in n-channel and p-channel MOSFET's," *IEEE Trans. Electron Devices*, vol. 35, no. 12, pp. 2194–2209, Dec. 1988.

[24] R. Woltjer, G. Paulzen, H. Pomp, H. Lifka, and P. Woerlee, "Three hot-carrier degradation mechanisms in deep-submicron PMOSFET's," *IEEE Trans. Electron Devices*, vol. 42, no. 1, pp. 109–115, Jan. 1995.

[25] T. Yamamoto, K. Uwasawa, and T. Mogami, "Bias temperature instability in scaled p+ polysilicon gate p-MOSFET's," *IEEE Trans. Electron Devices*, vol. 46, no. 5, pp. 921–926, May 1999.

[26] G. A. M. Hurkx, H. C. de Graaff, W. J. Kloosterman, and M. P. G. Knuvers, "A new analytical diode model including tunneling and avalanche breakdown," *IEEE Trans. Electron Devices*, vol. 39, pp. 2090–2098, 1992.

[27] A. J. Annema, G. G. Geelen, and P. C. de Jong, "5.5 V tolerant I/O in a 2.5 V 0.25 μm CMOS technology," *IEEE J. Solid-State Circuits*, vol. 36, no. 3, pp. 528–538, Mar. 2001.

[28] H. Ballan and M. Declerq, *High Voltage Devices and Structures in Standard CMOS Technologies*. Boston, MA: Kluwer, 1999.

[29] K. C. Dyer, D. Fu, S. H. Lewis, and P. J. Hurst, "An analog background calibration technique for time-interleaved analog-to-digital converters," *IEEE J. Solid-State Circuits*, vol. 33, no. 12, pp. 1912–1919, Dec. 1998.

[30] H. V. D. Ploeg, G. Hoogzaad, H. A. H. Termeer, M. Vertregt, and R. L. J. Roovers, "A 2.5-V 12-b 54-Msample/s 0.25-μm CMOS ADC in 1-mm^2 with mixed-signal chopping and calibration," *IEEE J. Solid-State Circuits*, vol. 36, no. 12, pp. 1859–1867, Dec. 2001.

[31] Q. Huang, F. Piazza, P. Orsatti, and T. Ohguro, "The impact of scaling down to deep submicron on CMOS RF circuits," *IEEE J. Solid-State Circuits*, vol. 33, no. 7, pp. 1023–1035, Jul. 1998.

[32] P. H. Woerlee, M. J. Knittel, R. van Langevelde, D. B. M. Klaassen, L. F. Tiemeijer, A. J. Scholten, and A. T. A. Zegers-van Duijnhoven, "RF-CMOS performance trends," *IEEE Trans. Electron Devices*, vol. 48, no. 8, pp. 1776–1782, Aug. 2001.

[33] R. van Langevelde, A. J. Scholten, and D. B. M. Klaassen, "Physical background of MOS model 11, level 1101," Philips Electronics N.V., 2003. NL-TN 2003/00239, http://www.semiconductors.philips.com/Philips_Models/mos_models.

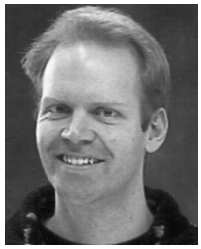
Anne-Johan Annema (M'00) received the M.Sc. degree in electrical engineering and the Ph.D. degree from the University of Twente, Enschede, The Netherlands, in 1990 and 1994, respectively.

In 1995, he joined Philips Research, Eindhoven, The Netherlands, where he worked on a number of electronics-physics-related projects ranging from low-power low-voltage circuits, fundamental limits on analog circuits in conjunction with process technologies, high-voltage in baseline CMOS to feasibility research of future CMOS processes for analog circuits. Since 2000, he has been with the IC Design Group, Department of Electrical Engineering, University of Twente, Enschede, The Netherlands. He is also a part-time consultant in industry. In 2001, he co-founded ChipDesignWorks. His current research interest is in physics, analog and mixed-signal electronics, and deep-submicrometer technologies and their joint feasibility aspects.

Bram Nauta (S'89–M'91) was born in Hengelo, The Netherlands, in 1964. In 1987, he received the M.Sc. degree (*cum laude*) in electrical engineering from the University of Twente, Enschede, The Netherlands. In 1991, he received the Ph.D. degree from the same university on the subject of analog CMOS filters for very high frequencies.

In 1991, he joined the Mixed-Signal Circuits and Systems Department, Philips Research, Eindhoven, the Netherlands, where he worked on high-speed A/D converters. From 1994, he led a research group in the same department, working on analog key modules. In 1998, he returned to the University of Twente, as full Professor heading the IC Design Group in the MESA+ Research Institute and Department of Electrical Engineering. His current research interest is analog CMOS circuits for transceivers. He is also a part-time consultant in industry. In 2001, he co-founded Chip Design Works. His Ph.D. thesis was published as a book: *Analog CMOS Filters for Very High Frequencies* (Boston, MA: Kluwer, 1993). He holds 11 patents in circuit design.

Dr. Nauta received the Shell Study Tour Award for his Ph.D. work. From 1997 to 1999, he served as an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—PART II: ANALOG AND DIGITAL SIGNAL PROCESSING, and in 1998 he served as Guest Editor for the IEEE JOURNAL OF SOLID-STATE CIRCUITS. In 2001, he became an Associate Editor for the IEEE JOURNAL OF SOLID-STATE CIRCUITS and he is also a Member of the Technical Program committee of ESSCIRC and ISSCC. He is a corecipient of the IEEE ISSCC 2002 Van Vessem Outstanding Paper Award.



Ronald van Langevelde (S'94–M'95) received the M.Sc. degree in electrical engineering in 1994 and the Ph.D. degree in 1998, both from the Eindhoven University of Technology, The Netherlands.

He is currently with Philips Research Laboratories, Eindhoven, The Netherlands, where he has developed the surface-potential-based MOS Model 11. His research interests include MOSFET device physics, circuit-level MOSFET modeling, and distortion analysis in circuit design.

Hans P. Tuinhout received the M.Sc. degree in electrical engineering from Delft University of Technology, Delft, The Netherlands, in 1980.

Since then, he has worked at Philips Research, Eindhoven, The Netherlands, on process and device characterization for silicon CMOS and BiCMOS integrated circuit technologies. His current research activities are focused on high precision dc I-V measurements for characterization of device mismatch.