Technology Scaling and CMOS Analog Design

Franco Maloberti

Department of Electronics, University of Pavia, Pavia, Italy
franco.maloberti@unipv.it

Abstract

Modern and future ultra-deep-submicron technologies make challenging the analog design especially when power consumption must match digital counterparts. The decrease of the supply voltage reduces the voltage headroom in analog circuits, the gate leakage current increases, the voltage gain decreases in planar bulk transistors, $1/f$ noise deteriorate when using new high-k gate dielectrics. The transistor and passive components mismatches give rise to large inaccuracies. Only some of these problems can be solved at the technology level; others require new circuit topologies and design techniques. The increased digital processing capability not only pushes the analog-digital interface toward more and more digital, thus limiting the analog design to very first interfaces and data converters, but also enables to use digital methods for the correction and digital assisted analog design. Several of the solutions currently proposed for analog pre-processing and data converter design are discussed in this paper.

I. Introduction

In past decades the well known Moore's law has governed the microelectronics evolution [1]. The constant pace of advancements in device and fabrication technology has lead to an exponential progress rates in transistor miniaturization and integration density. As shown in Fig. 1, the number of transistors per chip has increased by 4 order of magnitudes in three decades and very soon the technology progress will allow the integration of 1 billion transistors on a single chip.

The astonishing evolution started in *1958* when J. Kilby at Texas Instruments realized the first *IC* followed few months lather by a similar realization of R. Noyce at Fairchild. The first analog and digital planar products were quite simple like the ones shown in the microphotographs of Fig. 2. Presently, even not taking into consideration memories and pure digital circuits it is quite common to integrate on a single integrated circuit millions of transistors.

The *IC* complexity increases thanks to the technology scaling boosted by the continued improvements in optical projection lithography that enables printing ever finer features, the smallest feature size decreasing by about *30%* every two years. Process technology is now in sub-wavelength mode that with extreme Ultra-Violet (*EUV*) lithography (Fig. 3) obtains the patterning of lines below $50\,nm$ dimensions.

Actually, at the same time, we have a constant reduction of minimum feature size, oxide thickness, supply voltage and

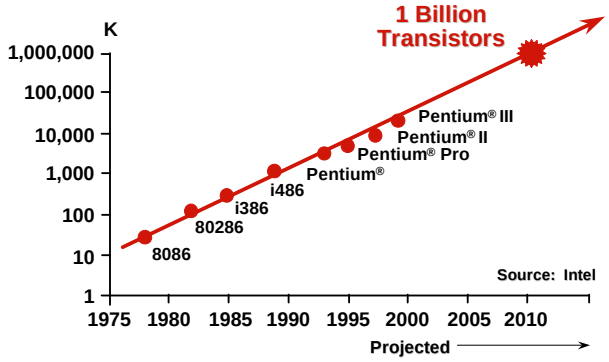

Fig. 1. Number of transistor per chip versus time.

transistor thresholds. The reasons are in the device physics of scaling the bulk *MOSFET* structures. Key challenges are controlling short channel effects like the V_{Th} roll-off and Drain-Induced Barrier Lowering (*DIBL*). Since for minimizing these short channel effects the planar and the lateral aspect ratios of transistors must be preserved, the technology evolution from one generation to the next should scale down all together gate oxide thickness, junction depth and depletion depth.

A major problem that limits the scaling of the oxide thickness is the leakage through the gate oxide by direct band-to-band tunneling. Indeed, if a not negligible current flows from source to drain of an *MOS* transistor highly energetic carries generate electron-hole pairs leading to charges trapped in the oxide. By a positive feedback mechanism the oxide tunneling current steadily increases in time until, eventually, the oxide breakdown occurs. Moreover, before the breakdown occurs the threshold voltage steadily degrades.

Supply voltage reduces each technology generation and contributes, on one hand, to lower the power dissipation but, on the

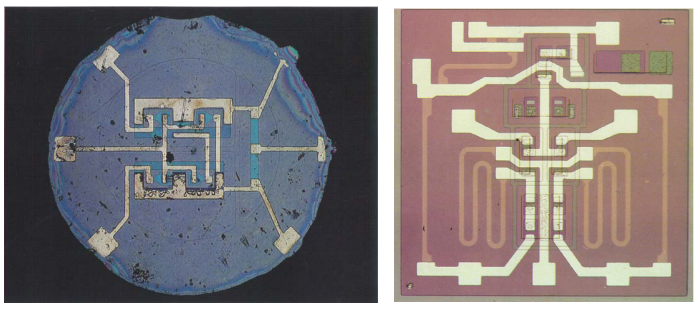

Fig. 2. First planar integrated flip-flop (Fairchaild, 1961) (left) and first analog integrated circuit (μA 702, Faichild -1962) (right).

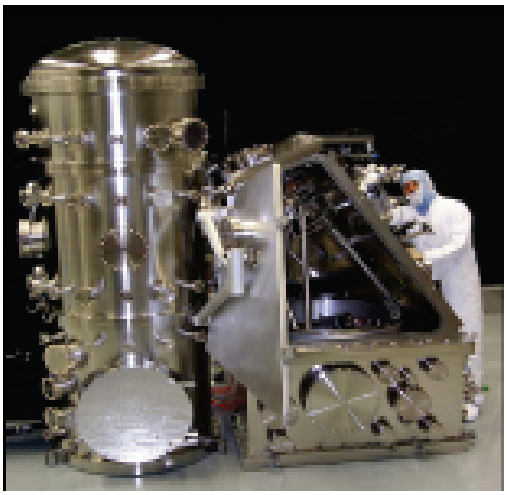

Fig. 3. Equipment for EUV.

other hand, worsens the signal-to-noise ratio (*SNR*) of analog circuits. As the supply voltage reduces the transistor threshold voltage should also reduce at the same rate to maintain enough gate overdrive without reducing the signal swing by more than the shrinking factor. However, lower threshold causes an exponential increase of the sub-threshold leakage that augments the leakage power and, more critical, dissolves the analog signals stored on capacitors. A possible remedy is using a dual threshold technology that reduces the leakage in high threshold devices by orders of magnitude but is much more expensive than the normal technology (one extra mask is required) and complicates the design and the layout routing.

The above points indicate that the push toward deep submicron technology is not so attractive for analog functions that, when implemented alone or with a small digital part, are more conveniently obtained with "analog" technologies whose line-width is *2–3* times larger than the one used for digital. The challenge (technical and cultural) for the analog designer is with large digital systems integrated in deep submicron technologies requiring limited analog functions.

II. Consequences of Scaling on Analog

The future evolution of *CMOS* technologies is predicted by companies and other organizations. According to them it is reasonable to assume that in the next *5–10* years the minimum line-width will go down to *35–25 nm* with equivalent oxide thickness as low as *1–1.5 nm*. If the gate oxide breakdown field will not increases properly, the values for V_{DD}, limited by reliability consideration, will drop well below 1 V [2]. Therefore, since designing high performance analog functions at this voltages is problematic, even for analog is important to find new replacements of silicon dioxide for increasing the maximum field strength and supply voltage.

In addition to the problems caused by a very low supply voltage there are other consequences on the transistor features that, by turns, limit the analog performance.

A. Transconductance

As known the transconductance is the key parameter for analog design whose desirable high value is typically obtained at the cost of an increased bias current. However, for very short channel the carrier velocity quickly reaches the saturation limit at which the transconductance also saturates becoming independent of gate length or bias

$$g_m = \frac{W_{eff} C_{ox} v_{sat}}{2} \tag{1}$$

Since for $T_{ox} = 1.5/, nm$ C_{ox} is about $16fF/\mu^2$, assuming that v_{sat} of electrons at the oxide semiconductor interface is $3.2 \cdot 10^{10} \mu/s$, the transconductance of a transistor with $1\,\mu m$ effective width is at most $0.25\,mA/V$.

A limited transconductance is problematic for analog design: for obtaining high gain it is necessary to use wide transistors at the cost of an increased parasitic capacitances and, consequently, limitations in bandwidth and slew rate.

B. Output Resistance

The output resistance is another important analog parameter. It is influenced by channel length modulation, static feedback and, for deep submicron devices, the weakavalanche effect. The output resistance decreases quickly at gate lengths below $0.1\mu m$ (much faster than L). Fig. 4 shows a possible plot of output resistance versus the effective submicron length.

The output resistance degradation and the saturation of g_m affect the intrinsic gain: using a transconductance limit of $0.25\,mA/V$ when the output resistance becomes $4\,k\Omega$ the intrinsic gain $g_m r_0$ becomes unity and the device cannot be used anymore for amplification purposes. Even using longer lenghts obtainig gain with deep submicron technologies is difficult: it is typically necessary using cascode structures with stack of transistors or circuits with positive feedback.

C. Diffusion and Gate Capacitances

Since with scaling the gate oxide becomes thinner, the specific capacitance C_{ox} increases as the shrinking factor. However, since the gate area decreases as the square of the shrinking factor, the gate-to-source and gate-to-drain parasitics lower as as the process feature size shrinks.

Reducing transistor parasitics is positive but the benefit is contrasted by the increased parasitic capacitance of the interconnections that compensates for the reduction of the

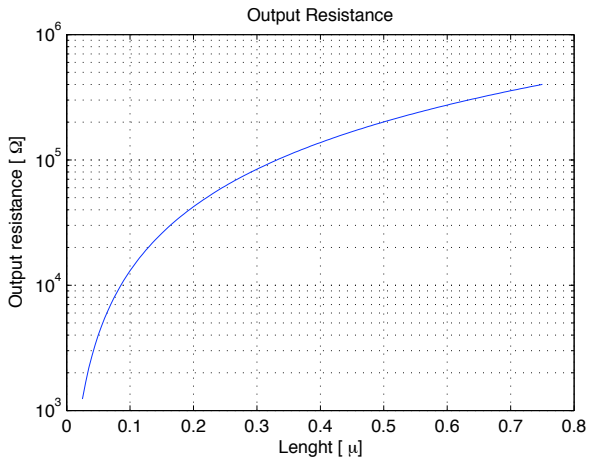

Fig. 4. Output resistance as a function of the effective gate lenght.

gate capacitance or perhaps brings about a larger parasitic contribution. The global effect is that shrinking does not help much in increasing the speed of analog circuit as the position of the non-dominant poles is almost unchanged.

D. Speed

The unity gain frequency of an *OTA* is approximated by $g_m/(2\pi C_L)$, where C_L is the load capacitance. Assume, for simplicity, that the fundamental kT/C limit establishes the value of the load capacitance and that the requires *SNR* is $62\,dB$. With a $0.5\,V$ input signal the noise power integrated over the Nyquist interval must be lower than $144\,\mu V$. Therefore, the value of C_L must be at least $0.2\,pF$.

The required transconductance for securing $f_T = 2\,GHz$ is $2.52\,mA/V$, a larger value than the saturation value of a $1\,\mu m$ wide transistor. Accordingly, the aspect ratio necessary for the input differential pair must be fairly large, in the hundred range.

E. Matching Accuracy

The offset of any analog circuit and the static accuracy of data converters critically depend on the matching between nominally identical devices. The stochastic nature of physical and chemical fabrication steps causes a random error in electrical parameters that gives rise to a time independent difference between equally designed elements. The error typically decreases as the area of the devices. It can be assumed that the stochastic mismatch between the parameters p_1 and p_2 of two elements placed at short distance is given by

$$\sigma(\Delta p)^2 = \frac{A_p^2}{W \cdot L} + S_p^2 d^2 \tag{2}$$

where A_p and S_p are the process-dependent matching parameter at unity area (normally in μm^2) and unity distance between elements (normally in μm). d is the distance between the centroids of the two identically designed elements.

The value of various A_p and S_p depend on the process accuracy and the technology minimum line-width. Namely, the threshold mismatch parameter $A_{V_{th}}$ decreases almost linearly with the oxide thickness reduction, equal to the shrinking factor.

For capacitors, expressing the absolute accuracy by

$$\left(\frac{\Delta C}{C}\right)^2 = \left(\frac{\Delta \epsilon_r}{\epsilon_r}\right)^2 + \left(\frac{\Delta A}{A}\right)^2 + \left(\frac{\Delta t_{ox}}{t_{ox}}\right)^2 \tag{3}$$

it can be assumed that the first term is negligible; the second term, inversely proportional to the area, improves with the accuracy of the perimetrical region whose definition enhances with scaling. The last term is almost constant and can be incorporated in the second term of equation (2). Summing up, A_C^2 decreases almost linearly with scaling.

The inaccuracy of integrated resistor depends on specific resistance and geometrical dimensions

$$\left(\frac{\Delta R}{R}\right)^2 = \left(\frac{\Delta R_\square}{R_\square}\right)^2 + \left(\frac{\Delta L}{L}\right)^2 + \left(\frac{\Delta W}{W}\right)^2 \tag{4}$$

$\Delta R_\square$, that improves with the square root of area, depends on physical and technological parameters whose accuracy linearly improves $A_\square^2$ with shrinking. The other two terms are proportional to $1/\sqrt{A}$ and improve with scaling. Their effect becomes important for deep submicron technologies and give rise to an extra term in equation (2) inversely proportional to the perimeter.

Since at deep submicron the second term of (2) is negligible, for preserving matching the area of transistor or capacitor should diminish as the shrinking factor (and not as its square), while the area of resistors should diminishes less because of the perimetrical term.

F. Substrate and supply noise

In mixed signal applications the substrate noise and the interference between analog and digital supply voltages caused by the switching of digital sections are problematic. The situation becomes more and more critical with scaling as smaller geometries induce higher coupling. Moreover, higher speed and current density augment *EM* issues.

The use of sub-micron technologies with high resistive substrates is advantageous because the coupling from digital sections to regions where the analog circuits are located is partially blocked. However, the designer must face the bounce of the V_{DD} and V_{SS} digital lines which can have values up to hundreds of mV. The use of separate analog and digital supplies is a possible remedy but its effectiveness is limited by the internal coupling between close metal interconnections.

The substrate and the supply noise cause two main limits: the in-band tones produced by non linearities that mix high frequency spurs, and the reduction of the analog dynamic range required for accommodating the common-mode part of spurs. Since the substrate coupling is also a problem for pure digital circuit the submicron technologies is evolving toward *SOI* and trench isolation options.

G. Signal Swing

The continuous shrink of modern process reduces the supply voltage. Since it is necessary to ensure a suitable headroom for keeping transistors in saturation, even if the number of transistors piled up is kept at the minimum the swing of signals cannot be closer to V_{DD} or ground by more than one or two hundred mV. Therefore, the trend of *CMOS* features shown in Fig. 5 predicting supply voltages well below $1\,V$ in the next few

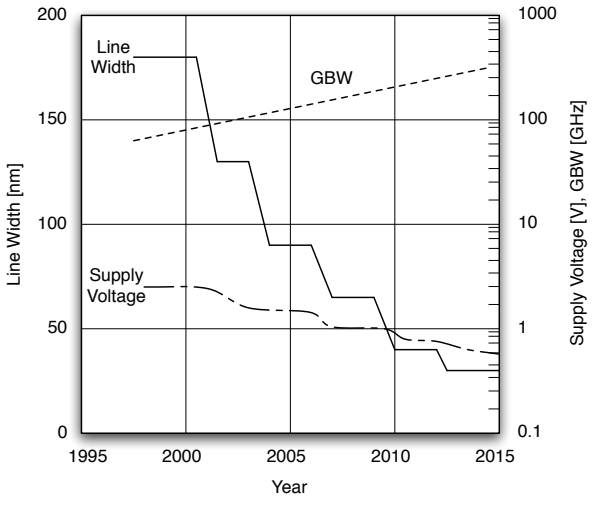

Fig. 5. Trend of analog features in CMOS technologies.

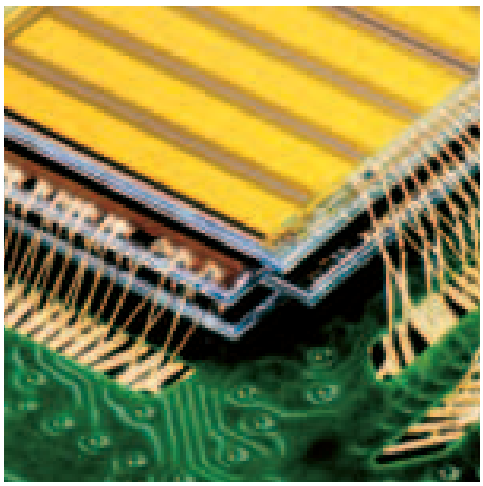

Fig. 6. Stack of chips.

years, also conceives a signal swing as small as few hundred of *mV*, fairly low if it is necessary ensuring a reasonable *SNR*. Low voltage is also problematic for driving *CMOS* switches especially for the ones connected to signal nodes as the on-resistance can become very high or at the limit the switch does not close at all in some interval of the input amplitude.

Different options and scenarios are possible. The most obvious solution is the multi-chip solution, with different circuits possibly assembled on the same package using, in some cases, a stack of chips (Fig. 6). Digital functions and memories are implemented by a single or multiple chips, the analog processing is obtained by a separate chip with suitably high supply voltage and reduced analog digital interference.

The use on the same chip of two supply voltages, one for analog and the other for digital, is another possibility. In some cases the higher voltages is generated on-chip with a charge pump. Obviously the technology or the transistors used in the analog part must be able to sustain the higher analog voltage. Moreover, it is necessary to design suitable analog-digital interfaces for the interactions between the two parts.

The already mentioned multiple threshold technology is another option. The use of a low or, at the limit zero, V_{Th} is good for many analog functions except the switching of voltage that needs medium V_{Th} devices for ensuring the *off* condition.

III. Circuit Design Techniques

Scaling-down the process certainly shifts the boundary digital/analog toward digital. However, analog circuits remain irreplaceable components. Besides the obvious data conversion, it will be always necessary to perform a variety of critical tasks before and after the analog-digital interface, such as amplification, prefiltering, demodulation, signal conditioning for transmission, generation of absolute values (voltages, currents, frequencies). In addition, analog will retain for a long while its advantage over digital when very high frequency and very low power consumption are required [3].

The key limitation of analog circuits is that they operate with electrical variables and not simply with discrete numbers that, in circuit implementations, gives rise to a beneficial noise margin. On the contrary, the accuracy of analog fundamentally relies on matching between components, low noise, offset, and low distortions. For deep submicron technologies the most challenging design issues are:

- Ensure good matching for low-offset and performant *ADC*.
- Obtain high *SNR* with low-voltage supplies.
- Contrast the degradation of analog performances caused by shrinking.

A. Matching of Elements

Since many circuit techniques exploit the equality of two components it is important for a given process obtaining the best matching especially for critical devices. Some of the rules that the designer must follow are common sense or obvious; nevertheless it is important to have them well in mind [4]:

- Devices to be matched should have the same structure and use the same materials.
- The temperature of matched components should be same. For chips consuming significant power the devices to be matched should be located on the same isotherm, which is obtained by a symmetrical placement with respect to the dissipative devices.
- The distance between matched devices should be minimum for having the maximum spatial correlation of fluctuating physical parameters.
- Common-centroid geometries like quad configuration used to implement a pair of transistors should be used to cancel the gradient of parameters at the first order.
- The same orientation of devices on chip should be the same to eliminate dissymmetries due to unisotropic fabrication steps, or to the unisotropy of the silicon itself.
- The surroundings in the layout, possibly improved by dummy structures, should be the same to avoid border mismatches.

Matching must be extended to the case of non-unity *n/m* ratios by separately grouping *m* and *n* matched devices. For ratios that are not a fraction of integer numbers a slight alteration of one of the unity devices for giving 1 plus the fractional part obtains the optimum solution.

B. Low-Voltage Design

A low-voltage supply can be augmented using a charge pump boosting on-chip voltages beyond the available supply; this relaxes the design but may lead to reliability problems in some cases. In addition a number of circuit techniques make possible true low-voltage operation without voltage boosting [5]; these include the use of transistors in the weak inverted region, the use

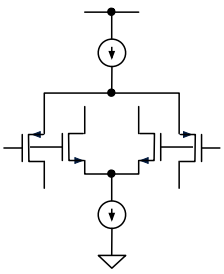

Fig. 7. Complementary differential pair.

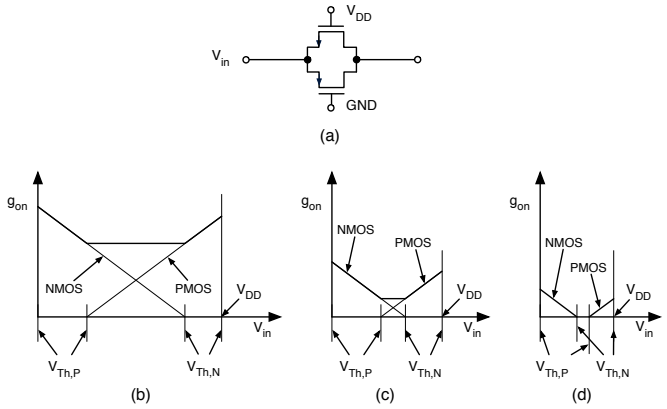

Fig. 8. a) Complementary switch and its on-conductance for scaling supply voltages (b), (c), and (d).

of the body node as a fourth control or signaling terminal, and the use of circuit topologies that eliminate stack of transistors.

For some circuits like the sample and hold the challenge is obtaining an input stage with a rail-to-rail common-mode voltage range. A solution is using complementary p-channel and n-channel parallel pairs (Fig. 7) with a control of the bias currents for extending the common mode input region with constant g_m. Thus, for example, a low-voltage *CMOS* op-amp can be fully-differental with an input stage made by the combination of n-channel and p-channel differential pairs for achieving a wide common-mode input voltage range, a differential output stage made by a low-voltage, class-*AB* rail-to-rail architecture with symmetrical output drive capabilities, and a switched-capacitor common mode feedback.

Sampled-data circuits use switches whose *on* condition is problematic at low voltage supply. Indeed, as shown in Fig. 8, the switch conductance can go to zero if the supply voltage is less than $V_{th,n} + V_{th,p}$. The problem can be resolved by using the switch boosting method and the switched op-amp technique. Fig. 9 shows the conceptual scheme of the switch boost [6]. During the phase at which the switch is off a capacitor C_B is pre-charged to the supply voltage; it is then connected during the complementary phase across source and gate of M_S for giving $V_{GS} \simeq V_{DD}$. The switched voltage should be lower than the supply. However, for special applications is also possible to switch voltages above the supply [7].

The switched op-amp technique is illustrated with the help of Fig. 10 that shows a flip-around $S\&H$. The left terminal of the sampling capacitance C_s is connected to the input during the sampling phase and also to output of the op-amp. During this phase the second push-pull stage of the op-amp is switched-off

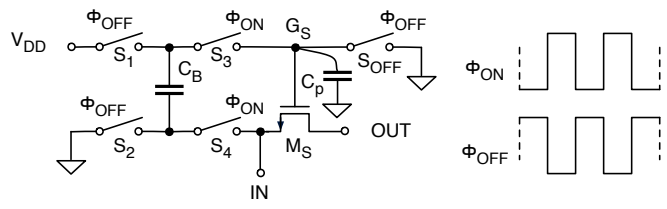

Fig. 9. Conceptual scheme of the switch boosting method.

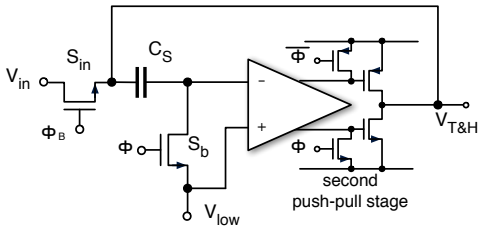

Fig. 10. Flip-aroud S&H with input common-mode voltage and switched OTA.

and becomes active only during the hold period. The switching *off* of the second stage can be obtained without boosting as the used transistors are connected to V_{DD} and ground.

C. Performance Enhancement and Tolerant Architectures

Poor analog performances that are caused by the use of a deep submicron technology are overcame by special analog techniques that, for example, foresee the storing of errors on capacitors afterward used for correcting the limit or by using architectures that are tolerant of analog limitations.

For example, the low gain of op-amps used in sigma-delta modulators is less critical than the first stages of a pipeline scheme. A continuous-time filter is less slew-rate demanding than a switched-capacitor counterpart. Therefore, the analog design trade-offs must include the limits that the deep submicron technology entails.

IV. Digital Assisted Techniques

The rapid advancement of digital technology motivates an increasing use of digital methods that improves the performance of analog circuits and data converters by correcting or calibrating the static and possibly the dynamic limitations.

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

The methods are classified into the following categories:

- Digital correction.
- Foreground calibration.
- Background calibration.
- Dynamic matching.

A. Error Measurement and Correction

The basis of digital corection methods is in the measurement of errors. Accurate error measurement allows for their correction by trimming, or by the storing of errors for a successive digital correction or analog calibration.

Suppose, for example, it is required estimating the difference between capacitors that are expected to be equal [8]. Fig. 11 illustrates a two phase scheme for measuring the difference between C_1 and C_2. During Φ_s the capacitance C_1 samples the voltage V_B, and C_2 and the auxiliary capacitance C_3 are

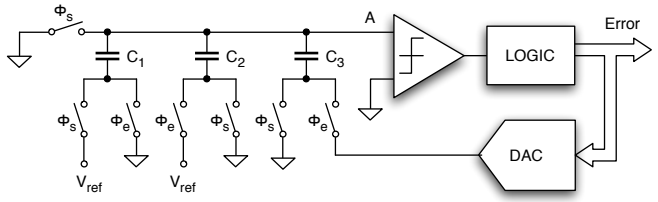

Fig. 11. Circuit for error measurement.

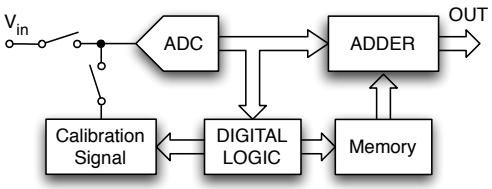

Fig. 12. Conceptual scheme of a foreground calibrated ADC.

connected to ground. During Φ_e the switching of the roles of C_1 and C_2, and a *DAC* output of zero give rise to a voltage at node *A* of

$$V_A = V_{ref}\frac{C_2 - C_1}{C_1 + C_2 + C_3} \tag{5}$$

which is proportional to the mismatch. Observe that the *DAC* and the logic form a successive approximation loop that through the action of C_3 can bring the error to zero. The *DAC* voltage at the end of *N* successive approximation cycles (the number of bits of the *DAC*) is

$$V_A = \frac{V_{ref}(C_2 - C_1) + (V_{DAC} + \epsilon_Q)C_3}{C_1 + C_2 + C_3} \tag{6}$$

where ϵ_Q is less than the *LSB* of the auxiliary *A/D* conversion.

If $\pm V_{ref}$ and $\Delta = V_{ref}/2^{(N-1)}$ are the references and the quantization step of the *DAC*, the range of measurement and its accuracy are

$$|(C_2 - C_1)|_{max} = \frac{V_{ref}}{V_B}C_3; \quad \Delta C_{mism} = \frac{\Delta}{V_B}C_3. \tag{7}$$

Assuming that $V_B \simeq V_{ref}$ the value of C_3 must be larger than the expected mismatch; the number of bits of the *DAC* gives the accuracy of the measure with C_3 as full scale.

The previous method aimed at obtaining a digital measure of an error that is to be accounted for in a digital correction section. Another possibility is using the measure for trimming. The method slightly changes the value of one or more elements to ensure the required analog accuracy. The trimming method was widely used in integrated filters to regulate the time constant of integrators or other parameters for ensuring an accurate frequency response. The use of metal thin-film resistors realized on the top of the passivation layer and a laser trim regulates the resistance values after fabrication, but before packaging. The method is expensive and does not account for temperature and aging effects.

Another form of trimming is to adjust a component value by discrete steps using fuses or anti-fuses that permanently open or close connections. Also, it is possible to use *MOS* switches whose *on* or *off* state is stored in a memory. When trimming with discrete steps it is obviously necessary to use an array of small elements, possibly binary weighted, connected in series or in parallel for connecting or disconnecting them according to a suitable control algorithm.

B. Foreground Calibration

The foreground calibration, typical of *ADC* [9], interrupts the normal operation of the converter for performing the trimming of elements or the mismatch measurement by a dedicated calibration cycle normally performed at power-on or during periods of inactivity of the circuit. Any miscalibration or sudden environmental changes such as power supply or temperature may make the measured errors invalid. Therefore, for devices that operate for long periods it is necessary to have periodic extra calibration cycles.

Fig. 12 shows the conceptual scheme of a foreground calibrated *ADC*. During the measure phase the input connection is disabled and a suitable input is applied to the *ADC*. It can be an analog linear ramp, a sine wave or pseudo-white noise, possibly generated by a slow and accurate on-chip *DAC* driven by an on-chip digital signal. In this case, since the signal is generated by the system itself, the method is called self-calibration.

Various types of inputs can be used to obtain the input-output characteristics and, in turn, the integral non-linearity (*INL*) error. How the differential inputs are generated, and how they are processed to produce the the *INL* are not discussed in detail here being sufficient to know that the calibration provides the *INL* to be stored in a memory. The input switch restores the data-converter to normal operational after the mismatch measurement and every conversion period the logic uses the output of the *ADC* to properly address the memory that contains the correction quantity. In order to optimize the memory size the stored data should be the minimum word-length which, obviously, depends on technology accuracy and expected *A/D* linearity.

The digital measure of errors, that allows for calibration by digital signal processing, can be at the element, block or entire converter level. The calibration parameters are stored in memories but, in contrast with the trimming case, the content of the memories is frequently used, as they are the input of the digital processor.

C. Background Calibration

Methods using background calibration work during the normal operation of the converter by using extra circuitry that functions all the time synchronously with the converter function. This makes the measurement hidden in the background and does not interfere with the normal converter operation.

Often these circuits use hardware redundancy to perform a background calibration on the fraction of the architecture that is not temporarily used. However, since the use of redundant hardware is effective but costs silicon area and power consumption, other methods aim at obtaining the functionality by borrowing a small fraction of the sampled-data circuit operation for performing the self-calibration.

For example, the method of Fig. 13, used for the on-line calibration of an *ADC* obtains the necessary room using a conversion rate higher than the sampling rate [10]. Assume, as

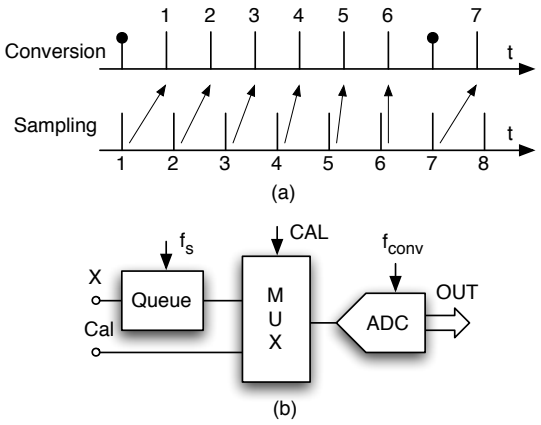

Fig. 13. (a) Timing scheme of a queue based background calibration. (b) Block diagram of a possible implementation.

shown in Fig. 13 (a) that the sampling and conversion frequency are such that $f_{conv} = 7f_s/6$. Therefore, the lower sampling frequency accommodates for 7 conversions in 6 sampling periods, thus making the converter available one extra slot every 7. The shift between sampling and conversion times is managed by a queue as shown in Fig. 13. The queue consists of one or more *S&H*s used for storing the input sample between the sampling instant and the time of its conversion.

A method used in systems that use only a fraction of the available band consists in modulating the error and moving it at an out-of-band frequency [11]. A dedicated circuit performs the band-pass error measure and the possible trimming.

D. Dynamic Matching

The dynamic matching method is another possibility for calibrating unity elements. The goal of the approach is to equal the elements on average instead of performing a static correction of the values. As well as correcting the mismatch, the dynamic matching is also effective for canceling the variations caused by drifts like temperature changes and aging effects.

Consider just two nominally equal inter-changeable elements X_1 and X_2. Assuming $X_2 = X_1(1+\delta)$ and $X_1 + X_2 = 1$, the use of only one element gives rise to

$$Y_1 = \frac{1}{2+\delta} \quad or \quad Y_2 = \frac{1+\delta}{2+\delta} \tag{8}$$

depending whether the element X_1 or X_2 is used. The resulting errors are equal and opposite

$$\epsilon_{1,2} = \mp\frac{\delta}{2+\delta} \tag{9}$$

The dynamic averaging uses one or the other element under the control of a random bit-stream. Accordingly, the error is randomly positive or negative with spectrum given by the bit-stream control. The total power is approximately given by $\delta^2/[(2+\delta)^2 \cdot 12]$ and is almost uniformly spread over the Nyquist interval.

The method is described by using the scheme of Fig. 14 (a) where the current source I_{ref} is splitted into two nominally equal parts by the transistor pair $M_1 - M_2$. The degeneration resistances R_1 and R_2 improve the matching by attenuating the possible limit given by the thresholds mismatch. However, since the current divider is also limited by the resistor mismatch, the accuracy of the division is not enough for high resolution.

The four switches on the top of the current divider obtain the dynamic matching. They use a randomizing signal V_R similar to the one of Fig. 14 (b). To make the average of the output currents I_1 and I_2 equal it is required that the number of 1 equals the number of 0.

The method, illustrated for two elements, can be extended to many elements by using a proper algorithm for randomizing the selection pursuing the goal of transforming the mismatch error into a noise-like term [12]. If only a fraction of the Nyquist band is used a possible shaping of the noise-like spectrum further reduces the mismatch limit.

The control of the dynamic matching of many elements as used in *DAC*s with a thermometric selection of unity elements can be problematic. Typically, equal resistors, capacitors or current sources, are summed up using a random representation of the thermometric input. The selection of the unity element is defined by the randomizer that receives *N* thermometric ones out of *M* input lines and generates a scrambled set of *M* controls, *N* of which are one and the others are zero. However, the number of possible scrambled outputs is *M!* which is a very large number even for a relatively small number of levels: it is 5040 for *M* = 7, 40320 for *M* = 8 and more than 3.6 million for *M* = 10. Such a large number of connections is difficult to code but also are not strictly necessary for obtaining a random result.

What is actually necessary is avoiding frequent repetitions of the same or similar code which would produce tones and not noise-like spectra. Since it is enough to randomize the elements with a subset of possible connections, a very simple solution uses an *M*-port barrel shifter which rotates one increment after each clock. A more effective way is the butterfly randomizer shown in Fig. 15 consisting of a series of butterfly switches coupling inputs to outputs [13]. The use of $log_2 M$ butterfly stages, as done by the scheme of Fig. 15, ensures that any input can be connected to any output; adding more butterfly stages increases the number of possible connections. The control of the butterfly switches can be done using $log_2 M$ bits of a *k-bit RNG* (random number generator) ($k > log_2 M$). More simply the control can be done by successive divisions by 2 of the clock. In this case the method is called *clocked averaging*.

The randomization of elements transforms the mismatch error that produces tones in the output spectrum into a pseudo

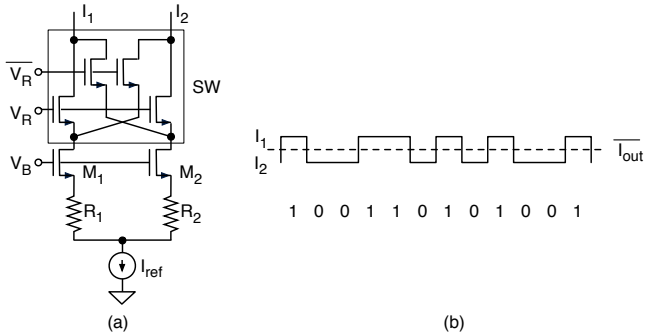

Fig. 14. (a) Current splitting and dynamic matching of the outputs. (b) Bit-stream control and I_1 current.

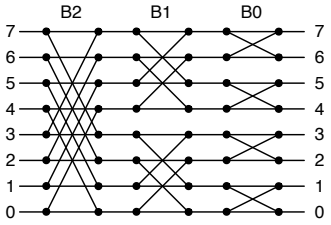

Fig. 15. Three stage butterfly randomizer of 8 elements with three bit control.

noise. Since the result smoothes the tones but does not achieve a reduction in the total error power, then for Nyquist rate data converters the signal plus distortion remains almost unchanged while the linearity improves.

For circuits that use a small fraction of the Nyquist interval it would be more effective having small the noise caused by the mismatch in the used band as the noise out of the band of interest can be removed by filters. This result is obtained by special dynamic matching schemes that achieve a low-pass shaping of the noise. The techniques, used in sigma-delta converters, are the individual level averaging (*ILA*) [14] and the data weighted averaging (*DWA*) [15]. The *ILA* approach aims at exercising each unity element with equal probability for each digital input code. The *DWA* performs a rotation of the elements used by using a starting index that is updated by the addition of the new input code to the content of the index register.

Both methods works well with busy input signals. The *ILA* is more effective for a small number of elements while the *DWA* method works well for 7 or more elements.

V. Conclusions

The design of analog functions using deep submicron technologies is source of new challenges that are faced with different approaches: solving the problem at the circuit level using at the best the analog feature of the technology and staying away from the minimum feature sizes, studying architectures that are tolerant to analog limits, and, more effective, exploiting the increasing digital processing capabilities and use digital circuits to assist the analog performances. Indeed, digitally assisted analog circuits avoid the disadvantages of technology scaling by delegating matching requirements to a digital processor and compensating for non idealities. The relaxed analog circuit constraints translate into higher overall system performance such as higher resolution and better circuit speed.

References

[1] Moore, G. E.; *Cramming more components onto integrated circuits*, Electronics, Vol. 38, n. 8 April 19, 1965.

[2] Matsuzawa, A.; *Analog IC technologies for future wireless systems*, IEICE Trans. Electron, Vol. E89-C, April 2006, pp. 446 – 454.

[3] Vittoz, E.; The design of high performance analog circuits on digital CMOS chips, Journal of Solid-State Circuits, vol. 20, pp. 657 – 666, 1985.

[4] F. Maloberti: *Analog Design for CMOS-VLSI Systems*. Kluwer Academic Press, Boston, Dordrecht, London, 2001.

[5] Maloberti, F.; Francesconi, F.; Malcovati, P.; Nys, O. J. A. P.; *Design Considerations on Low-Voltage Low-Power Data Converters*, IEEE Transactions on Circuits and Systems-I: Fundamental Theory and Applications, vol. 42, no. 11, pp. 853 – 863, 1995.

[6] Abo, A. M.; and Gray, P. R.; *A 1.5-V, 10-bit, 14.3-MS/s CMOS pipeline analog-to-digital converter*, IEEE Journal of Solid State Circuits, vol. 34, pp. 599 – 606, 1999.

[7] Aksin, D.; Al-Shyoukh, M.; Maloberti, F.; *Switch Bootstrapping for Precise Sampling Beyond Supply Voltage*, IEEE J. Solid State Circuits, vol. 41, pp. 1938 – 1943, 2006.

[8] Ohara, H.; Ngo, H. X.; Armstrong, M. J.; Rahim, C. F., and Gray, P. R.; *A CMOS programmable self-calibrating 13-bit eight-channel data acquisition peripheral*, IEEE Journal of Solid-State Circuits, vol. 22, pp. 930 – 938, 1987.

[9] Maloberti, F.; *Data Converters*, Springer, Boston Dordrech, London, in press, 2007.

[10] Blecker, E. B.; McDonald, T. M.;. Erdogan, O. E;. Hurst, P. J; and Lewis, S. H.; *Digital background calibration of an algorithmic analog-to-digital converter using a simplified queue*, IEEE Journal of Solid-State Circuits, vol. 38, pp. 1059 – 1062, 2003.

[11] Fornasari, A.; Borghetti, F.; Malcovati, P.; Maloberti, F.: *On-line calibration and digital correction of multi-bit sigma-delta modulators*, Proceedings of the 2005 International Symposium on VLSI Circuits, IEEE VLSI '05, 2005, pp. 184-187.

[12] van de Plassche, R. J.; *Dynamic element matching for high accuracy monolithic D/A converters*, IEEE Journal of Solid-State Circuits, vol. SC-11, pp. 795 – 800, 1976.

[13] Carley, L. R.; *A noise-shaping coder topology for 15+ bit converters*, IEEE Journal of Solid-State Circuits, vol. 24, pp. 267 - 273, 1989.

[14] Leung, B. H., Sutarja S.; *Multibit Sigma-Delta A/D converter incorporating a novel class of dynamic element matching techniques*, IEEE Trans. Circuit Syst. II, vol. CAS-39, pp. 35 – 51, 1992.

[15] K. Vleugels, S. Rabii, and B. A. Wooley: *A 2.5-V Sigma-Delta modulator for broadband communications applications*, IEEE Journal of Solid-State Circuits, vol. 36, pp. 1887 – 1899, 2001.