1.3 Analog CMOS from 5 Micrometer to 5 Nanometer

Willy Sansen

Professor Emeritus, Katholieke Universiteit Leuven, Leuven, Belgium

1.0 Introduction

The most important application of ICs, today, is most probably the Internet-of-Things. It involves sensor nodes and communications at the lowest-possible power levels. Similar low power levels are required for the next-generation smart-phones as images and video become increasing requirements, in phone, body-worn, and automotive applications. Whatever technology is used, portability requires reduced power consumption.

Analog CMOS in 5 Micrometer technologies dates from the seventies as shown by Moore's Law (Figure 1.3.1). In early chips, Analog circuits were dominant; as time progressed, digital became more prevalent, such that at 5 Nanometer it will overwhelm. Yet, Analog circuits traditionally establish the interface to Digital. Thus, they must be realized in mainstream digital technologies, and correspondingly, are forced to conform to evolving Digital. Note that in Figure 1.3.1, the projections to 10nm and 5nm in 2016 and 2020, respectively, are highly dependent on the success of various competing technologies. Which of FinFETs [1,2], FD-SOI [1,2], SiGe [5], or other alternatives, ascends, is still a matter of current conjecture: A comparison will be given in Figure 1.3.16.

Small channel lengths allow very high f_T values, such as 300GHz at 28nm, and 1.4THz at 10nm as shown in Figure 1.3.2. Velocity saturation has become the most dominant limitation. On the other hand, communication channels now require linear amplifiers and filters of up to 100MHz, and more than 1GHz in the near future. Note that these requirements are still low with respect to the high f_T values potentially available.

At the same time, relatively low baseband frequencies in combination with low supply voltage, will allow the transistors of analog baseband circuitry to operate in moderate or weak inversion. In the new EKV/BSIM6 models, the parameter Inversion Coefficient (IC) is used to measure how deep in weak inversion a MOS transistor operates. Moreover, an optimum operating region can be found, which depends on the actual channel length. This is examined in the next Section 2.0 of this paper. Section 3.0 provides an overview of all design tricks used to reduce power consumption. Well-known examples are the use of negative resistance and negative capacitance. Negative resistances cancel positive elements, thereby leading to lower power consumption. Also noise- and distortion-cancellation are discussed. Section 4.0 discusses digitally-assisted analog, especially for ADCs. Switching amplifiers are added as well. Finally, in Section 5.0, an excursion is taken towards 5nm technologies: FinFET and FD-SOI realizations are compared down to 10nm channel lengths. Vertical nanowire FETs are introduced down to 5nm. While, currently, a first inspection shows that only discrete devices can be used, it is encouraging to note that their I-V characteristics differ little from known ones. Further, it is demonstrated that conventional analog design techniques can still be adopted albeit at much higher frequencies.

2.0 The Inversion Coefficient (IC) as a Design Parameter

A single-transistor amplifier biased at { V_{GS} - V_{T} = 0.2V} yields a Gain-Bandwidth Capacitance over Current product of about 1500MHzpF/mA [6]. This value of { V_{GS} - V_{T} } of 0.2V corresponds to an IC of about 10, as explained next.

The transistor model used is the newly developed BSIM6 model, derived from the previous EKV model [7]. It still includes the well known Shichman-Hodges model [8] in the mid-current region, but extends to weak inversion for low currents and velocity saturation for high currents. The current I_{DS} is normalized to a specific value I_{spec} , which yields IC as indicated in Figure 1.3.3. The overdrive voltage {V_{GS}-V_T} is normalized as well, towards v (as noted on the right side of Figure 1.3.3. Further, the relationship between v and IC (noted on the left side of the figure) is plotted in Figure 1.3.3. It is a very basic relationship, as it does not depend on the channel length, thus maintaining validity for 5nm V-FETs.

The curve in Figure 1.3.3 also shows that negative values of { V_{cs} - V_T } are easily obtained for low values of IC. For example, in Figure 1.3.4, the minimum

supply voltage is shown versus GBW (Gain-Bandwidth Product) for a CMOS inverter amplifier, in which the minimum supply voltage equals $2V_{GS}$. Already in 45nm CMOS, a 1GHz amplifier consisting of a simple CMOS inverter can have a lower supply voltage than a 0.5V operational amplifier [9]. As a result, more and more amplifiers and filters in ADCs and other low-frequency circuits, replace operational amplifiers by CMOS inverters, which offer the advantages of working at lower power levels and at lower supply voltages.

The Inversion Coefficient (IC) is an excellent parameter for design. The curves of the f_Tg_m/I_{DS} product are shown in Figure 1.3.5 [10]. This latter product combines speed, noise, and power consumption in one single Figure-Of-Merit. It is used for comparison of performance of many amplifiers including receiver LNAs. Further, it is the best FOM to guide analog designers in their choice of the biasing point of a transistor in the signal path.

In Figure 1.3.5, it is shown that down to 65nm CMOS, the strong inversion region is present in the middle with weak inversion on the left, and velocity saturation [6] on the right. In velocity saturation, the transconductance reaches its maximum value of $WC_{ox}V_{sat}$ in which v_{sat} is about 100km/s. However, in BSIM6 v_{sat} is represented by $\lambda_c = L_{sat}/L$. This new parameter is the normalized channel length in which L_{sat} is about 20nm. The cross-over value of IC between strong inversion and velocity saturation is reached for IC = $1/(\lambda_c)^2$. For 65nm CMOS, λ_c is about 0.3 and this cross-over value is about 10. When IC = 1, the value of f_T is f_{Tspec} . It is clear from the curves on Figure 1.3.5 that several important observations can be made, as a result of using IC rather than $\{V_{GS}-V_T\}$. Thus, it is clear at 65nm that the optimum biasing values of IC are between 1 and 10. For lower values of IC, f_T decreases considerably, whereas for higher values than 10, the g_m/I_{DS} goes down steeply. For various channel lengths, the optimum values of IC shift to even deeper weak inversion. Thus, for 20nm CMOS, the optimum IC is still unity but for 5nm CMOS the optimum IC decreases to about 0.06. The corresponding values of λ_c and f_{Tspec} are given as well (at the upper right of Figure 1.3.5). Note that for smaller channel lengths, the maximum $f_T g_m / I_{DS}$ product increases and reaches about 22THz/V for 20nm CMOS. Also note that high values of the $f_T g_m / I_{DS}$ product cannot be reached for high IC (or $V_{GS} - V_T$) values. The limit is about 22THz/V at IC = 1. This limit is independent of the particular technology used. It is clear that smaller channel lengths provide much larger values of the $f_T \; g_m/I_{DS}$ product, but that the optimum occurs deeper into weak inversion. Moreover, below 20nm CMOS, the Schichman-Hodges [8] model has now vanished completely.

Finally, the curves in Figure 1.3.5 can also be used to derive the maximum GBW values, which can be reached in deep weak inversion. Thus, for a single-transistor amplifier, they are easily derived directly. For a set of two-stage Miller operational amplifiers [6, 10], GBW values are shown in Figure 1.3.6. Note from the figure that for 65nm CMOS technology, a 1GHz two-stage amplifier can be obtained with IC = 1; but, as technology reduces, operation must proceed deeper and deeper into weak inversion, requiring, for example at 16nm that IC = 0.1, deep indeed!

3.0 Circuit Techniques for Less Power Consumption

The reduction of power consumption at a specific speed of operation, is achieved not only by proper biasing, but also by dedicated circuit techniques. This reduction in power consumption will become even more important as CMOS descends to 5nm, where increasingly complex systems must be integrated on a single chip.

Of the circuit techniques to be examined, the cancellation of parasitic capacitances is the first. It is followed by the use of negative resistors for both higher gain and higher speed. Then, the cancellation of poles by zeros in multistage amplifiers and by use of feedforward is discussed. Finally, cancellation techniques for noise and distortion will follow.

3.1 Cancellation of Capacitances

The cancellation of device and other parasitic capacitances, in order to increase speed for the same power consumption, has been tried long ago as shown in Figure 1.3.7 [11]. The compensation capacitances depend on many transistor parameters; however, as will be shown for MOST (rather than BJT), cancellation is not easily achieved. A gain in bandwidth of a factor of 2 or 3 can be readily obtained; but, beyond these values, mismatching must be kept under control. The most successful applications are in optical receivers [12].

These techniques can be expected to be incorporated within mainstream amplifiers and filters, as well. Thus, they deserve more attention.

3.2 Cancellation of Resistances

Negative resistances have been employed from the days of vacuum-tube oscillators, in which a negative resistor compensated for the loss resistance of an inductor. Present-day CMOS VCOs all use this principle. Negative resistances can also be used to increase the GBW of amplifiers, as demonstrated in Figure 1.3.8 [13]. There, the inclusion of the two transistors MNR with sizes A smaller than unity, creates positive feedback and negative resistors, which partially cancel positive resistances. This increases the GBW by a factor 1/(1-A). For example, in a 65nm design, the 1260MHzpF/mA reached for {V_{GS}-V_T = 0 V}, is increased to 10000MHzpF/mA for A = 0.8. This is six times better than for a single-transistor amplifier! This means that for a given GBW and load capacitance (which indirectly sets the input noise level), the power consumption is only one sixth of that of the single-transistor amplifier. Even higher performance of 40000MHzpF/mA can be reached by nesting the load current mirrors [15], without increasing the equivalent input noise.

Interestingly, in fact, negative resistors have been used in amplifiers for GmC filters, for a long time. They have been included in the source leads of single-transistor configurations such as amplifiers, source followers, and cascodes, as shown in Figure 1.3.9. They partially cancel the $1/g_m$ seen at the source. Such filters achieved the highest performance reported, including power for speed, low noise, and low distortion. It is clear that the inclusion of negative resistors in all analog building blocks has become essential for all categories of analog circuits.

3.3 Cancellation of Poles with Zeros

The cancellation of higher-order poles by zeros is carried out in most higher-order filters. In three-stage amplifiers several higher-order poles can be cancelled, leading to considerable reduction in power. Such an example is shown in Figure 1.3.10 [16]. Two non-dominant poles are cancelled by two zeros, yielding 14000MHzpF/mA. Several such amplifiers have been published recently at ISSCC [17], and many design efforts are on their way. This is clearly an area for further exploration. In all of this, attention must be paid to the additional sources of noise, caused by reduced current levels.

Feedforward is another technique in which poles are cancelled by zeros. Quite often however, pole-zero doublets are generated [18], which degenerate the time response. Caution must be exerted by the designer opting for feedforward to make sure that actual pole-zero cancellation is effectuated.

3.4 Cancellation of Noise

The cancellation of noise is not straightforward, as noise has a dimension of power. Each additional transistor adds power and hence noise. Nevertheless, the noise from the input transistor of an LNA can be cancelled if it proceeds to the output through two paths of opposite phase, as shown in Figure 1.3.11 [19]. Similarly, noise cancellation can be achieved after down-conversion in receivers [20]. Such noise cancellation techniques have not yet been fully developed for integrated amplifiers and filters, but are expected in the near future. They are very attractive as they touch upon the most fundamental limit of any signal processing block, in which power is used to reach a certain SNR at a certain speed.

3.5 Cancellation of Distortion

The most obvious cancellation of distortion is realized in a simple current mirror and differential pair. Both exhibit excellent linearity for small input signals. As well, diode-connected transistors with opposite curvature can always be added [20] to cancel distortion. The opposite curvature within one transistor [6] can also be used to linearize an amplifier. However, it strongly depends on the biasing point. This is why it is difficult to reduce distortion by more than about 10dB or so.

Full distortion cancellation can also be obtained by cross-coupling two differentials pairs [22] provided that the biasing points of the pairs are different according to a specific relationship between transistor sizes and {V_{GS}-V_T} values [6]. Again, the results strongly depend on mismatch.

Finally, a particular interesting case is a filter using an operational amplifier, in which the distortion can be cancelled by means of a negative resistance at the inverting node [23].

4.0 Analog with Digital Assistance

Corrections of mismatch and other errors have been developed primarily for ADCs and for RF applications. As shown in Figure 1.3.12 [24], technology is the primary limitation for SNDR values below 55dB. For higher values of SNDR, noise is the main limitation. Thus, improvements in technology are expected to appear earlier for noise-limited ADCs [25], than for others.

More recently, VCO-based ADCs have been developed in which the input voltage (or current) drives the frequency of a VCO (ICO) [26, 27, 28]. The output frequency or pulse-width is then measured by means of a Time-to-Digital converter, as shown in Figure 1.3.13. By this means, a more highly digital system results, taking less area for smaller channel lengths. In the future, both sigma-delta ADCs and PLLs will benefit from this approach, and are expected to provide better results at lower supply voltages.

Switched-mode amplifiers [29] also provide direct input-voltage to pulse-width conversion, as shown in Figure 1.3.14. Correspondingly, they are expected to provide similar performance improvements at smaller channel lengths, as available in the VCO approach.

5.0 Towards 5nm CMOS

As channels reduce to 10nm, both FinFETs [1,2] and FD-SOI [3,4] are used to reduce the short-channel effects that become increasingly strong for planar bulk devices below 20nm. These devices are sketched in Figure 1.3.15. Note, that while the FinFET has gone three-dimensional, the FD-SOI transistor remains more planar. This allows the gate of a FinFET to surround the channel more closely, yielding better control. This leads to less drain-induced barrier lowering, but the three-dimensional nature of a FinFET generates more extrinsic parasitic capacitance. Correspondingly, its sub-threshold slope is closer to 60mV/decade, providing larger intrinsic gain g_m/g_{ds} , but with larger input-gate capacitance. A comparative table for FinFETs, FD-SOI, and SiGe is shown in Figure 1.3.16. It is apparent that the road for FinFETs towards 5nm is much clearer than for FD-SOI, since the latter technology will eventually suffer from heavy short-channel effects much as does a planar bulk device beyond 20nm.

For FinFETs and FD-SOI, offset and 1/f noise are not much worse than for planar bulk CMOS: the use of intrinsic silicon eliminates the impact of random dopant fluctuations on offset but the effects such as variability of the work-function material in the gate stack that determines the V_T, need to be taken into account. The Pelgrom Factor for bulk CMOS has not decreased below its value for 130nm, remaining constant at about 2.5mVµm [6]. Values for FinFETs and FD-SOI devices are about half of this, as shown in Figure 1.3.17 [30,31]. Moreover, the 1/f noise of a FinFET is comparable to what is found in planar CMOS technology.

Downscaling inevitably leads to higher parasitic capacitances and higher series resistances. This becomes a limitation for very high-frequency applications (such as for mm-wave), where f_T and f_{max} are important. Maintaining a high maximum of f_T , or of g_m/C_{gate} , is indeed challenging for two reasons: the relative increase of the gate capacitance as discussed above comes together with a limitation of the g_m at high inversion levels, due to the strong source degeneration of the series source resistance. In addition, the gate resistance R_g of minimum-length devices increases significantly with downscaling. This is due to the short gate lengths, making R_g narrower, but also due to the fact that gate lengths are smaller than gate contacts, leading to a high-ohmic bottleneck. Below 10nm, R_g must be accounted for not only at mm-wave frequencies, but also for analog circuits such as high-speed data converters.

For low-frequency circuit design in very advanced technologies, operation at low inversion levels is likely. At these inversion levels, the source degeneration is not yet visible and the steep threshold slope in combination with high intrinsic gain can yield excellent analog circuit performance at very-low power, as discussed in Section 2.0 above.

Below 10nm, the downscaling route is more foggy. One possible way to continue the performance improvement is to replace silicon in the channel by

1

a high-mobility material such as a III-V compound or Germanium. However, this does not solve the problem of contacting the gate. One possible way out is to employ a vertical FET where the gate connection is offset from the gate itself. Such a gate is now entirely around the channel as shown in Figure 1.3.18 [32]. An alternative device architecture, which is a less drastic change than a FinFET, is a horizontal channel with a wraparound gate [33], leading to even-better channel control than that of a first-generation FinFET. At the same time, V_{DD} will scale down and operation in weak inversion will become even more important.

For the future, FETs that are based on tunneling rather than on simple charge transfer, come into the picture for generations well below 10nm, offering a subthreshold slope that can be steeper than 60mV/decade. As well, quantum-well devices are in sight. However, for these devices, analog performance is still unclear.

In summary, even if one decides for some applications (such as radio) one decides not to scale analog circuits down further, they will always be needed for complex heavily-downscaled systems-on-chip, for interfacing, clocking, power management, temperature monitoring, and sensing in general.

6.0 Conclusions

In our future, as usual, analog designers will continue to expand their expertise and knowledge in response to changing needs. While devices will change their nature and operate at higher and higher frequencies, their I-V characteristics will remain similar. In the near term, increased speed of MOS circuits, will be reached by operating deeper in weak inversion. Offset and 1/f noise will continue to play a critical role. Thus, in general, it seems that analog expertise is insensitive to technology change.

In conclusion, what can I recommend? Simply put, it is to "learn more, always more". After all, analog design is a discipline where science joins art. Only more learning allows discovery of such art!

Acknowledgment

The support of Dr. Piet Wambacq is acknowledged.

7.0 References

[1] "Group IV channels for 7 nm FinFETs", M. Garcia Bardon, P. Raghavan, G. Eneman, P. Schuddinck, M.Dehan, A. Mercha, A. Thean, D. Verkest, A. Steegen, *Dig. Symp. VLSITechnology*, pp. 88-89, June 2014.

[2] "A 10 nm Platform Technology for Low Power and High Performance Application Featuring FinFET Devices with Multi Workfunction Gate Stack on Bulk and SOI", K.-I. Seo, et al, *Dig. Symp. VLSITechnology*, pp. 12-13, June.
[3] "Performance and Area Scaling Benefits of FD-SOI Technology for 6-T SRAM Cells at the 22-nm Node", C. Shin, M. Cho, Y. Tsukamoto, B.-Y. Nguyen, C. Mazure, B. Nikolic, T.-J. King Liu, *Transactions ED*, pp.1301-1309, June 2010.

[4] "A 3 GHz Dual Core Processor ARM CortexTM-A9 in 28 nm UTBB FD-SOI CMOS with Ultra-Wide Voltage Range and Energy Efficiency Optimization", D. Jacquet, et al, *J. Solid-State Circuits*, pp.812-824, April 2014

[5] "SiGe HBTs in 90 nm BiCMOS technology demonstrating 300KHz/420GHz fT/fmax through reduced Rb and Ccb parasiticis, R. Camillo-Castillo, et al, pp. 227-230, BCTM 2013.

[6] "Analog Design Essentials', W. Sansen, Springer 2006.

[7] "Charge Based MOS Transistor Modeling", C. Enz, E. Vittoz, Wiley, 2006
 [8] "Modeling and Simulation of Insulated-Gate Field-Effect Transistor

Switching Circuits", H. Shichman, D.A. Hodges, *J. Solid-State Circuits*, Vol. 3, pp. 285-289, Sept. 1968.

[9] "0.5 V Analog Circuit Techniques and their Application in OTA and Filter Design", S. Chatterjee, Y. Tsividis, P. Kinget, *J. Solid-State Circuits*, pp.2373-2387, Dec. 2005

[10] "Analog Design Procedures for Channel Lengths Down to 20 nm", W.Sansen, ICECS, pp. 337-340, 2013.

[11] "An Outline of Design Techniques for Linear Integrated Circuits", H. Camenzind, A. Grebene, *J. Solid-State Circuits*, pp.110-122, June 1969.

[12] "A Gigabit Optical Receiver with Monolithically Integrated Photodiode in 0.18 mm CMOS", C. Hermans, P. Tavernier, M. Steyaert, *Proc. ESSCIRC*, pp. 476-479, 2006.

[13] "Integrated PCM Codec", K. Ohri, M. Callahan, *J. Solid-State Circuits*, pp. 38-46, Feb. 1979.

[14] "Power Limits to Amplifiers and Filters", W.Sansen, ISSCC 2012 Short Course.

[15] "A 0.0013mm 2 3.6µW Nested-Current-Mirror Single-Stage Amplifier driving 0.15-to-15nF Capacitive Loads with >620 Phase Margin", Z.Yan, P.-I. Mak, M.-K. La , R. Martins, F. Maloberti, *ISSCC Dig. Tech.Papers*, pp.288-289, Feb. 2014.

[16] "Transconductance with Capacitances Feedback Compensation for Multistage Amplifiers" X .Peng, W.Sansen, *J. Solid-State Circuits*, pp.1514-1520, July 2005.

[17] "A 0.016 mm2 144 μ W Three-Stage Amplifier Capable of Driving 1-to-15 nF Capacitive Load with >0.95 MHz GBW Z.Yan, P.-I. Mak, M.-K. Law, R. Martins, *ISSCC Dig. Tech.Papers*, pp.368-369, Feb. 2012.

[18] "Feedforward Compensation Techniques for High-Frequency CMOS Amplifiers", W. Sansen, Z. Chang, *J. Solid-State Circuits*, pp. 1590-1595, Dec. 1990.

[19] "Wide-Band CMOS Low-Noise Amplifier Exploiting Thermal Noise Canceling", F. Bruccoleri, E. Klumperink, B. Nauta, *J. Solid-State Circuits*, pp. 275-282, Feb. 2004.

[20] "A Blocker-Tolerant Noise-Cancelling Receiver Suitable for Wideband Wireless Applications", D. Murphy, H. Darabi, A. Abidi, A. Hafez, A. Mirzaei, M. Mikhemar, M. Frank Chang, *J. Solid-State Circuits*, pp. 2943-2963, Dec. 2012.
[21] "A Low Power Linearized Ultra-Wideband LNA Design Technique", H.Zhang, X. Fan, E. Sanchez-Sinencio, *J. Solid-State Circuits*, pp. 320-330, Feb. 2009.

[22] "A Precise Four-Quadrant Multiplier with Subnanosecond response" B. Gilbert, *J. Solid-State Circuits*, pp. 365-373, Dec. 1968.

[23] "Cancellation of Opamp Virtual Ground Imperfections by a Negative Conductance Applied to Improve RF Receiver Linearity", D. Mahrof, E. Klumperink, Z. Ru, M. Oude Alink, B. Nauta, *J. Solid-State Circuits*, pp. 1112-1124, May 2014, with "Comments", P.Mohan, E. Klumperink, D. Mahrof, B. Nauta, *J. Solid-State Circuits*, pp. 2083, Sept 2014.

[24] http://www.stanford.edu/~murmann/adcsurvey.html

[25] "A/D Converter Trends: Power Dissipation, Scaling and Digitally Assisted Architectures", B. Murmann CICC, pp. 105-112, 2008.

[26] "A CMOS Time-Digital Converter LSI with Half-Nanosecond Resolution Using a Ring Gate Delay Line" T. Watanabe, Y. Makino, Y. Ohtsuka, S. Akita, T. Hattori, *IEICE Trans. Electr*, pp. 1774-1779, Dec. 1993.

[27] "Delta-Sigma Modulators using Frequency-Modulated Intermediate Values", M. Hovin, T. S. Lande, C. Toumazou, *J. Solid-State Circuits*, pp. 13-22, Jan. 1997.

[28] "A 12-Bit 10 MHz Bandwidth, Continuous-Time SD ADC with a 5-bit, 950 MS/s VCO Based Quantizer", M. Straayer, M. Perrot, *J. Solid-State Circuits*, pp. 805-814, April 2008.

[29] "Switched-Mode Operational Amplifiers and Their Application to Continuous-Time Filters in Nanoscale CMOS", B. Vigraham, J. Kuppambatti, P.Kinget, *J. Solid-State Circuits,* Dec.14.

[30] "Lowest Variability SOI FinFETs Having Multiple V_t by Back-Biasing", T. Matsukawa, K. Fukuda, Y. Liu, K. Endo, J. Tsukada, H. Yamauchi, Y. Ishikawa, S. Ouchi, W.Mizubayashi, S. Migita, Y. Morita, H. Ota, M. Masahara, VLSI Techn. 2014, pp. 114-115

[31] "Low Leakage and Low Variability Ultra-Thin Body and Buried Oxide (UT2B) SOI Technology for 20 nm Low Power CMOS and Beyond", F, Andrieu, et al, *Dig. Symp*, pp. 57-58, June 2014.

[32] "CMOS Nanoelectronics", Pan Stanford Publishing, N. Collaert, September 2012

[33] "Sensor-to-Digital Interface Built Entirely With Carbon Nanotube FETs"M. Shulaker, J, van Rethy, G. Hills, H. Weri, H.-Y. Chen, G. Gielen, P. Wong, S. Mitra. *J. Solid-State Circuits*, pp. 190-201, Jan. 2014.



ISSCC 2015 / SESSION 1 / PLENARY / 1.3



ISSCC 2015 / February 23, 2015 / 10:50 AM

