УДК 638.235.231

D. Foty[1], G. Gildenblat[2]
[1]Gilgamesh Associates, Fletcher, Vermont, USA, dfoty@sover.net;
Department of Electrical and Computer Engineering,
[2]Pennsylvania State University, USA, gildenblat@psu.edu.

FUNDAMENTAL ASPECTS OF CMOS TECHNOLOGY: BASICS, IMPLICATIONS, AND A ROADMAP TO THE FUTURE

Наведено огляд розвитку напівпровідникової технології та сучасний стан розвитку цього напрямку.

In this paper, the critical importance of scaling theory to the success of CMOS technology will be reviewed and evaluated. The history of CMOS shows that scaling theory has been the dominant theme, but that the evolution of the technology has followed different directions at different times due to constraints imposed by scaling theory. This situation is actually best understood by comparison with the theory of punctuated equilibrium, which is a familiar concept in evolutionary biology. The state of present-day constraints will be considered, followed by discussion of some possible options for continuing the progress of CMOS into the future.

Introduction. Judging by its ubiquity, the MOS transistor is arguably the most successful invention in human history, with more than ten billion microprocessors having been produced and annual U.S. per-capita transistor consumption now in the tens of millions. Given this long-running (and continuing) record of success, it is easy to forget that less than twenty years ago, it was widely believed that CMOS technology, like various predecessors, would enjoy a brief reign as the leading technology and would then be supplanted by something else. Instead, CMOS has become an overwhelmingly dominant technology, with no real end in sight to this success story. At this point, it is worthwhile to carefully revisit the real underlying reasons for this success, both to understand how we reached this point and also to make some judgments about what the future might hold.

In this paper, the critical importance of scaling theory to the success of CMOS technology will be reviewed and evaluated. The history of CMOS shows that scaling theory has been the dominant theme, but that the evolution of the technology has followed different directions at different times due to constraints imposed by scaling theory. This situation is actually best understood by comparison with the theory of punctuated equilibrium, which is a familiar concept in evolutionary biology. The state of present-day constraints will be considered, followed by discussion of some possible options for continuing the progress of CMOS into the future.

History of the "MOS Miracle". The concept of "field-effect" is so obvious that it doubtless must be very old. It is quite likely that more than 10,000 years ago, someone rubbed a dog on a dry day, heard a crackling sound, and noticed that his hand then attracted dust particles. For practical usage, this concept required several millennia of development, including the famous experiments of Franklin, Coulomb, Ampere, and Faraday.

The modern concept of the field-effect device originated with J.E. Lilienfeld [1], who obtained patents for working prototypes in the early 1930s; this work used copper oxide as the substrate material, but was unable to produce practical electron devices. However, the concept involved was radical and well ahead of its time – the development of what was effectively a switch with no moving mechanical parts.

MOS technology became practical in the 1960s with the development of workable silicon dioxide growth on silicon; this ability to produce the insulating layer *in situ* on the substrate was the key advance over earlier work, making the concept of the field-effect electron device a practical reality. In addition, the MOS transistor used trivially-common materials; as Gordon Moore has noted [2], the basic "ingredients" were the sand and aluminum of the earth, combined where necessary with the oxygen of the air.

This allowed for the practical realization of the original vision of the field-effect device. In the MOS transistor, the gate provides a field which causes to come into being a conducting charge layer beneath an insulator; the source and drain diffusions serve to contact this conducting layer, thus allowing its inclusion into a circuit current.

The Slow Development of CMOS. With our hum-drum daily lives now filled with multi-gigahertz microprocessors and wireless LANs, it is easy to forget that CMOS actually got off to a slow and inauspicious start in the semiconductor world. Originally, CMOS was, among its peers, a "slow" technology – slower, for example, than various bipolar technologies (such as ECL, which formed the processor "core" of most mainframe computers) - and even its NMOS cousin.

However, CMOS inherently possessed one very unique attribute; alone among technologies, CMOS logic gates consumed (effectively) no power when in a static condition – power was consumed only during switching operations. This capability kept CMOS in play as a niche technology for some two decades. Over time, it was realized that CMOS had some other unique and pleasant attributes. It was noted that CMOS logic operation was very stable; unlike other technologies, CMOS circuits could be "perturbed" and would settle back to the original logic state without disruption.

It was also realized that, very much unlike bipolar devices, MOS device concepts could be generalized, allowing for the inclusion of the basic device geometries in the design methods themselves. Because of this, device geometry and layout were not fixed (as they were in bipolar devices), but could be varied at will by the designer. This is one of the most important reasons for the astonishing success of CMOS – one which is now usually forgotten. Rather than being an advantage inherent in the *electron device itself*, it was the practical and intellectual flexibility which was possible that proved to be decisive. In design usage, CMOS simply proved to be an easier technology to handle. A useful analogy is to compare the bipolar and CMOS situations with that of wrenches; bipolar structures had to be fixed in geometry and layout, minimizing design flexibility – very much like having to carry around a large number of fixed-size wrenches. In contrast, CMOS geometries and layouts could be generalized and adjusted at will by the designer – much like being able to carry around a single adjustable wrenches.

In contrast to this good story, though, CMOS process technology seemed to be impeded by a number of severe problems which were often regarded as intractable and fatal. For many years, oxide layers were seemingly unable to be grown without including various mobile ions; over time, the gate field would cause slow migration of these ions within the gate oxide, causing the device threshold voltage to change significantly over time – making it very difficult to use the MOS transistor in circuit design. Also, the basic CMOS structure includes a built-in n-p-n-p thyristor; it was widely believed that CMOS could not be operated without switching this intrinsic thyristor into a low-impedance on-state, which would cause a large and catastrophically destructive current to flow. This pre-disposition to latch-up was often cited as an insolvable barrier to the use of CMOS. Throughout the 1970s and into the 1980s, much of the work on CMOS revolved around tackling and attempting to mitigate these kinds of fundamental technology problems. This was the first "stable equilibrium" condition of CMOS technology development.

Origins of CMOS Scaling. However, while these activities were going on, there was a very quiet revolution in the thinking about transistors. As noted above, it was recognized that the MOS transistor was amenable to a simple understanding based on a few technology parameters, which included the basic transistor geometry. During the early 1970s, both Mead [3] and Dennard [4] noted that this endowed the MOS transistor with a new and unique property – that the basic transistor structure could be *scaled* to smaller physical dimensions in a sane and comprehensible way. In essence, one could postulate a "scaling factor" of λ, where λ is the fractional size reduction from one generation to the next (e.g., $\lambda = 0.8$). This reduction factor could then be directly applied to the structure and behavior of the MOS transistor in a straightforward multiplicative fashion. For example, a CMOS technology generation possessed a minimum channel length L_{min}, along with technology parameters such as the oxide thickness t_{ox}, the substrate doping N_A, the junction depth x_j, the power supply voltage V_{dd}, the threshold voltage V_t, etc. The basic "mapping" to the next process, $L_{min} \rightarrow \lambda \cdot L_{min}$, involved the concurrent mappings of $t_{ox} \rightarrow \lambda \cdot t_{ox}$, $N_A \rightarrow N_A / \lambda$, $x_j \rightarrow \lambda \cdot x_j$, $V_{dd} \rightarrow \lambda \cdot V_{dd}$, $V_t \rightarrow \lambda \cdot V_t$, etc. Thus, the structure of the next-generation process could be known

beforehand, easing the process development task, and the behavior of circuits in that next generation could be predicted in a straightforward fashion from the behavior in the present generation. In addition, it was obvious that this "mapping" could be projected successively, and Mead in particular noted that he could foresee no reason that it could not continue over time to the unbelievable (in 1972) dimension of 0.1μm. It was this realization of scaling theory and its usage in practice which has made possible the better-known "Moore's Law" – Moore's Law is a phenomenological observation, but the scaling theory developed by Mead and Dennard is solidly grounded in the basic physics and behavior of the MOS transistor.

Scaling theory amounts to a "photocopy reduction" approach to feature size reduction in CMOS technology. Inherent in this approach is that while the dimensions shrink, scaling theory causes the field strengths (and profiles) in the MOS transistor to remain the same across the different process generations. Thus, the "original" form of scaling theory is *constant field scaling*.

Detour Through Constant Voltage Scaling. Due to the vexing problems inherent in CMOS technology (which were noted earlier), scaling theory received little attention during the 1970s. It wasn't until the 1980s, when these "basic" problems were resolved and CMOS became a practical technology, that scaling theory got appropriate attention. Finally, all of the pieces for the success of CMOS had come together – the process technology was up to the task, and there was a coherent understanding of the tremendous design flexibility that CMOS offered. This provided the incentive for tremendous investments of time and resources into developing process technologies which could take advantage of the tremendous opportunities offered by CMOS scaling. CMOS technology took off in a new direction.

However, constant field scaling carried one inherent problem – it required a reduction of the power supply voltage with each generation. This was considered to be undesirable from a systems-level viewpoint, since it would require completely new boards and systems with each generation – the maintenance of the same voltage, and thus intergenerational "plug compatibility," were considered to be more important. Because of this desire for plug compatibility, CMOS adopted the 5V power supply which had been used by bipolar TTL logic; it was thus decided to maintain this power supply voltage across generations. Thus, there was a deviation from constant field scaling, which was replaced with *constant voltage scaling*. Rather than remaining constant, the fields inside the device increased from generation to generation.

Because of this reality, the simple, parametric, "plain vanilla" MOS transistor had to be modified. In particular, the basic structure of the MOS transistor had to be altered to cope with the increasing fields. This was the situation during the 1980s, when the dominant factor in CMOS was the development of various process technology methods of coping with this reality; in this sense, therefore, the 1980s were the "golden age of process technology" in the CMOS world - particularly for the process technologists, who held the central place of importance in the industry. Many engineers still regard continual tinkering with processes and materials as the central activity of CMOS engineering.

To meet these needs, a number of additions were made to the basic MOS transistor structure. The gate metal was replaced by a polysilicon gate, which could easily be doped, patterned, and etched, and which could serve as a mask for self-aligned formation of the source and drain. Sidewall spacers were added, to allow the source and drain to be set at the proper distance under the gate edge; this also made possible the introduction of the lightly-doped drain structure (LDD), which served to spread out the voltage near the drain and thus reduce the peak field and hot carrier injection. The gate and the source/drain diffusions were silicided, to reduce the sheet resistance. The introduction of ion implantation (rather than simple surface-originated diffusion) allowed for the tailoring of vertical doping profiles – which was critical for preventing latch-up and punchthrough and for setting the threshold voltage. Shallow trench isolation was introduced to permit tighter packing of structures.

This was the "equilibrium state" of CMOS technology during the 1980s; the focus was almost exclusively on increasingly intricate process engineering, in order to maintain the 5V power supply voltage as feature sizes were decreased.

The "Rediscovery" of Constant Field Scaling. However, by the early 1990s, constant voltage scaling was running into considerable difficulty. Due to the continually increasing field, it was becoming more and more difficult to introduce new CMOS process generations; the time between process generations was

approaching three years. There were serious problems with excessive power dissipation and heating, and hot carrier effects were becoming a genuinely serious problem. As a consequence, it was becoming very difficult to reach the targeted L_{min} of the succeeding process generation – and it is this reduction of L_{min} which provides the real "bang" (imrproved performance) of scaling. In addition, the process technologies had continually developed more and more complicated ways of coping with the higher and higher fields; process complexity had reached a disastrous level, requiring a large number of mask steps while processes were becoming rather finicky and too expensive. This created a crisis for the continuing development of CMOS technology.

As a result, the 1990s were a decade in which constant field scaling was rediscovered in CMOS technology; once again, CMOS development changed course and moved in a different direction. As Figure 1 shows, the power supply voltage was scaled along with the minimum channel length; this began with 0.35μm processes, has continued into today's 0.13μm and 90nm processes, and is projected to continue for the foreseeable future – out to at least the 32nm technology node, which is projected to have a power supply voltage of 0.35V.

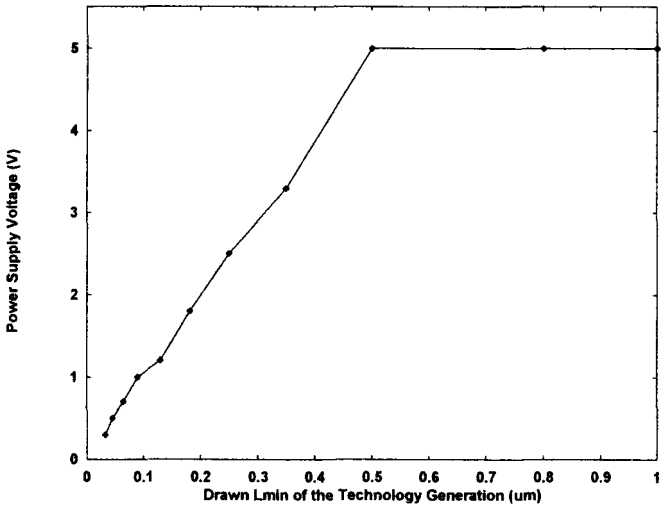

Figure 1. Power supply voltages for various CMOS technology generations

This caused a number of radical changes to the complicated MOS transistor structure which had developed during the 1980s. Not surprisingly, LDD was removed as no longer necessary, and vertical doping profiles became simpler and less abrupt. As a consequence, there was no longer a need for "heroic" process engineering for the successful introduction of the next technology generation. Thus, during the 1990s, in contrast to the 1980s the introduction of new process generations came with breathtaking speed – becoming almost annual rather than tri-annual. This represented the equilibrium state of CMOS development during the 1990s.

However, there is one inherent problem in constant voltage scaling which emerged as intractable during the 1990s. As noted earlier, constant field scaling mandates that the threshold voltage be scaled in proportion to the feature size reduction. However, ultimately threshold voltage scaling is limited by the subthreshold slope of the MOS transistor, which itself is limited by the thermal voltage kT/q. The Boltzmann constant k and the electron charge q are fundamental constants of nature and cannot be changed. This situation is roughly explained in Figure 2; the choice of the threshold voltage in a particular technology is determined by the off-state current goal per transistor (usually in μA per μm of channel width) and the subthreshold slope. With off-current requirements remaining the same (or even tightening) and the subthreshold slope limited by basic physics, the difficulty with scaling the threshold voltage is clear.

Because of this, as Figure 3 shows, during the 1990s the power supply voltage decreased in concert with constant field scaling, but the threshold voltage was unable to scale as aggressively. As Figure 3 also shows, this situation is only going to get worse as feature sizes and power supply voltages continue to scale. This is a fundamental problem for further progress in CMOS. A "solution" introduced widely at the 0.13μm technology node was a "menu" of device "flavors" with different threshold voltages - to allow a designer to selectively use devices for either high performance or low off-current as needed.

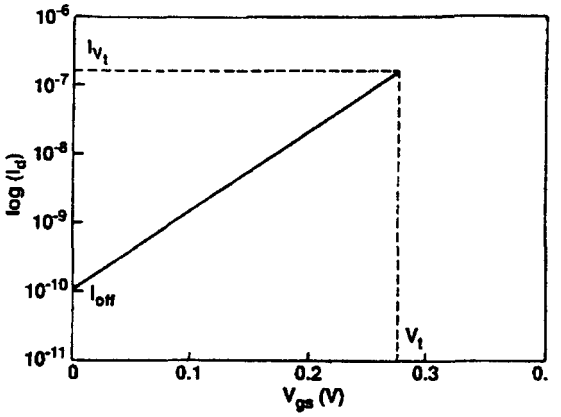

Figure 2. Threshold voltage determination from the off-current goal and the subthreshold slope

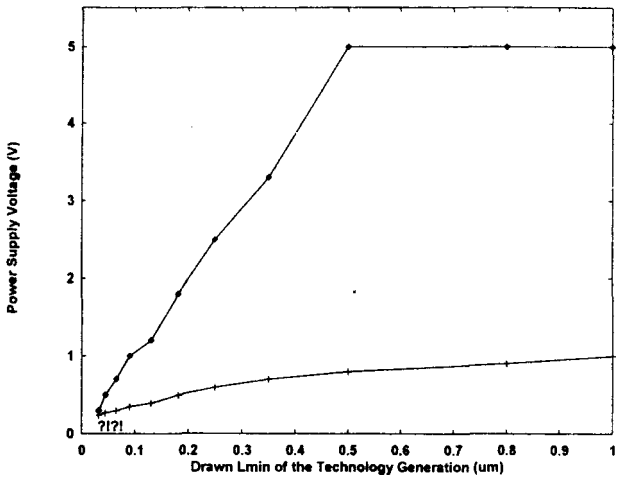

Figure 3. Power supply and threshold voltage for various CMOS technology generations

Implications of Threshold Voltage Non-Scaling. Ten years ago, the emerging difficulties inherent in the combination of constant field scaling and threshold voltage non-scaling were described in several interesting papers [5] – [8]; the contentions published then have proven to be quite accurate, and are worth revisiting.

When the power supply voltage is scaled, the improvements in power dissipation are very large, and do not depend too strongly on the value of the threshold voltage. However, in terms of digital switching speed, the situation is quite unpleasant; the decreasing headroom in ($V_{dd} - V_t$) extracts a severe price. This is illustrated in Figure 4.

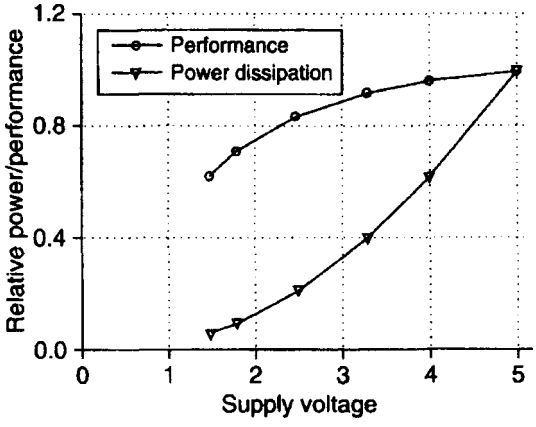

Figure 4. Performance and power dissipation with a fixed threshold voltage

In terms of digital design, the story is not completely dolorous. Under constant field scaling, hot carrier generation decreases exponentially with decreasing field and is vanquished as a problem. Power dissipation decreases much more rapidly than performance is lost, indicating that power dissipation problems are manageable. The penalty in loss of performance is still strong – even when considering the ability to further scale the channel length to recover the "lost" performance.

The analog design issues are more complex, but still revolve around the decreasing range of ($V_{dd} - V_t$); among other consequences, the shrinkage of this range provides less of a range for choosing DC gate biases in analog circuits. When this reality is combined with "traditional" analog design practice, the situation is even worse. Due to long-standing problems with our capabilities with device physics and with the structure of transistor models, analog designers have typically adopted a rule-of-thumb regarding DC gate biases which ensure that the transistors always remain biased in strong inversion: "$V_g \geq V_t + 200mV$." As V_{dd} decreases while V_t does not, it is obvious that this "rule" becomes increasingly problematic. Furthermore, a lower V_{dd} implies a

tighter "voltage budget" between power supply and ground; there is less voltage available to "budget" across the transistors (to keep them in saturation), and larger voltage budgets across the active and cascode transistors forces a reduced voltage budget across any loads, which reduces gain. This budgeting situation becomes very unpleasant when a relatively large V_g causes the transistor to thus have a rather large V_{dsat}; this forces the designer to allocate even more voltage across the transistors to keep them in saturation.

This is clearly a crisis situation for both analog and digital CMOS design. Some approaches to mitigating these difficulties will now be considered.

Re-Inventing Analog CMOS Design Basics. In analog design, these difficulties are catastrophic and are seriously impeding progress. Interestingly, though, the real problems are not technological, but intellectual. The basic methods of analog design were developed some 40 years ago, and have become grossly outdated; in addition, those methods have become increasingly incoherent and non-systematic, making it difficult to move a design along to its goal. Rather than tackling the fundamental problem, though, a tremendous amount of effort has been (and continues to be) expended on trying to prop up the obsolete approaches; a painful example is the persistence of the "$V_g \geq V_t + 200mV$" rule cited above. These sorts of rules are not imposed by physics; instead, they are self-inflicted by antiquated thinking about the MOS transistor and by the methods based on that inadequate thinking.

Back in the 1980s, before this critical situation was of much interest, the idea of interpreting the MOS transistor via the concept of an *inversion coefficient* was suggested by both Tsividis [9] and Vittoz [10]. This amounted to a method of evaluating the MOS transistor via a logarithmic interpretation of the inversion charge Q_i which included the transistor aspect ratio. This beguilingly-simple approach provided a continous and quantifiable method of describing the MOS transistor across weak, moderate, and strong inversion. This kind of approach can permit the discarding of rules which restrict the DC gate biases to strong inversion, allowing the use of lower voltages in analog circuits.

A more modern extension to this interpretation [11] may be found in Figure 5; this interpretation shows the embodiment of the Tsividis/Vittoz concept via the normalized transconductance-to-current ratio (g_m/I_d), and includes the behavior of very short channel devices - and also the reality that even "large" devices do not follow the square law. This provides a template for interpreting the MOS transistor over the entire range of weak, moderate, and strong inversion, in a manner which is a continuous "spectrum" and which is quantified.

Figure 6 shows measured data from a 0.18μm process technology; all of the features described in Figure 5 are evident here, including the separation of the device characteristics in strong inversion for the different channel lengths.

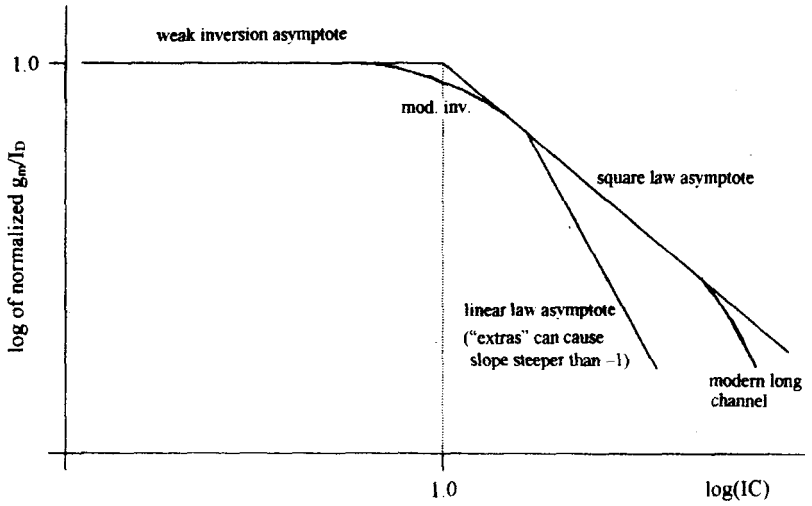

Figure 5. A graphical interpretation of MOS transistor behavior, via plotting the logarithm of normalized g_m/I_d vs. the logarithm of the inversion coefficient IC; all relevant MOS transistor phenomena are embodied in this graph

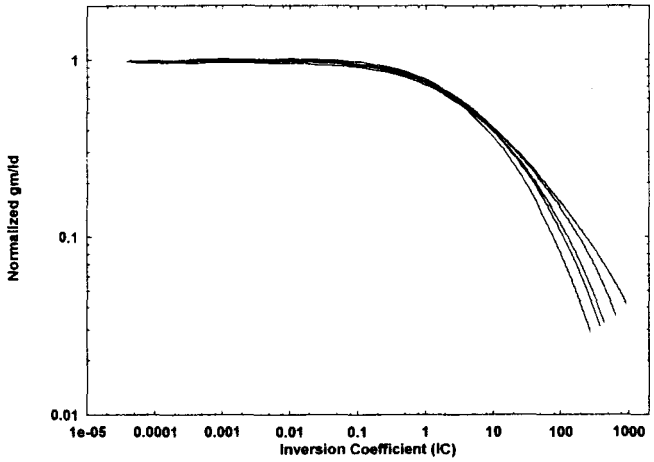

Figure 6. Measured values of normalized g_m/I_d vs. IC from a 0.18μm process; the channel lengths (left to right) in the graph are 0.18μm, 0.27μm, 0.36μm, 0.72μm, and 20.0μm

The description of Figures 5 and 6 is a very powerful method of understanding and employing this fundamental behavior of the MOS transistor – one which supercedes the various four-decade-old methods which grew out of the grossly oversimplified "square law" interpretation of the MOS transistor. It is also solidly based in physics - but in a more complete and modern fashion. It has been shown [12] that this kind of interpretation can provide the basis for new analog design methodologies – methodologies which are deterministic and which allow designers to properly and quantitatively evaluate the various trade-offs inherent in analog design. This interpretation of the MOS transistor also raises the standards for the capabilities of MOS transistor models; such models must fundamentally be able to describe the characteristics of Figure 6 [13], [14] – *otherwise, that model will not be suitable for use in the design methods which grow out of this approach.*

MOS Transistor Modeling for RF-CMOS. As noted above, the exploitation of these more detailed and modern descriptions of the MOS transistor require more fundamental analysis and modeling. In particular, MOS transistor models (and thoughts) must shift from the old "threshold-voltage-based" interpretation to one based on the surface potential. A surface-potential-based description of the MOS transistor is very fundamental, and provides an intrinsically-continuous description from accumulation all the way through strong inversion. Such a description also inherently provides second-derivative continuity of the current through $V_{ds}= 0$; this is required for proper modeling of IM3 distortion in RF-CMOS design.

Surface-potential-based descriptions have historically been mathematically complicated. However, great progress has been made in recent years; this now allows tactical progress against the various RF-CMOS design issues that have been long-term problems. In addition, this also allows strategic progress on the practical exploitation of the modernized description of the MOS transistor which underpins the new stage of "punctuated equilibrium" in CMOS scaling theory.

The leading surface-potential-based MOS transistor model is SP [14]. SP meets all of the conditions described above, using a relatively simple analytical description of the surface potential; as a consequence, the model naturally produces proper asymptotic behavior. Device symmetry also enters naturally, allowing for the proper modeling of current-division circuits.

In addition, a true surface-potential-based description intrinsically describes the MOS device seamlessly over the entire range from accumulation to strong inversion; this natural description of the MOS accumulation region makes SP an ideal model for use in varactor design. This structure also naturally provides a correct description of MOSFET behavior over the entire range of weak, moderate, and strong inversion – as embodied in characteristics such as Figures 5 and 6.

In addition to the RF-CMOS abilities with varactors and IM3 distortion, SP also includes a large-signal NQS description. Taken together, SP is a major advance in MOS transistor modeling for RF-CMOS design; in addition, the nature of the model structure allows for easy extensions to phenomena associated with much smaller MOS transistors than are presently in use.

Low Temperature CMOS Technology. It is also interesting to consider other possible implications that flow from scaling theory. As noted earlier, the central conundrum of modern scaling theory is the

inability to properly scale the threshold voltage; this constraint is basically due to the thermal voltage, kT/q, where k and q are unchangeable. However, the temperature *can* be changed; since the subthreshold slope is proportional to the temperature, a sharper subthreshold slope permits scaling of the threshold voltage. This is illustrated in Figure 7, where it is shown how decreasing the operating temperature allows the off-current goal to be met with a lower threshold voltage.

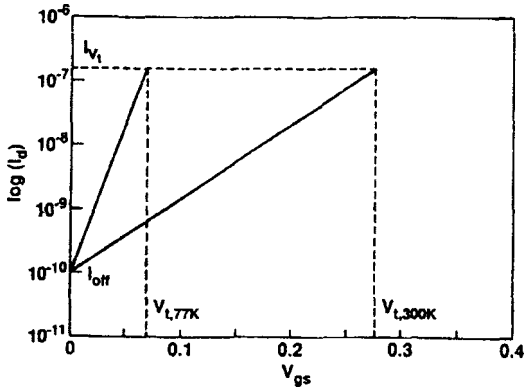

Figure 7. Threshold voltage determination from the off-current goal and the subthreshold slope, showing how 77K can allow a lower threshold voltage

The operation of CMOS at very low temperatures (such as 77K, the temperature of liquid nitrogen) is an idea that has been around nearly as long as CMOS itself. Due to reduced phonon scattering, a lower operating temperature greatly increases the channel mobility, resulting in much higher switching speeds; improvements of four to seven times were routinely reported. However, such operation would require some sort of cooling system, posing practical difficulties for implementation. In addition, as geometries decreased, the performance advantage dwindled – the so-called "saturation velocity" of carriers at 77K is only 30% larger than that at 300K, limiting the performance gains in modern CMOS.

However, from a first-principles viewpoint, scaling theory *should* include temperature. It has been shown [15] that both the power supply voltage and the threshold voltage should scale in proportion to the absolute temperature – thus, both should be scaled to roughly one-quarter of their room temperature values for 77K usage. Thus, the case for "improved performance" in 77K CMOS changes from what it originally was – rather than providing a "free lunch" improvement due to cooling, *the real improvement comes from being able to adjust the transistor so that the threshold voltage can be scaled* [6] – [8]. This is illustrated in Figure 8, where it can be seen that the performance gap between 300K CMOS and 77K CMOS grows rapidly as V_{dd} is decreased. *This is* ***not*** *a consequence of the "physics," but of the way that 77K operation allows for the adjustment of the silicon process for optimal 77K operation.* It should also be noted that carrier freeze-out could possibly be a concern, implying "odd" device behavior at low temperatures; however, it has been shown [6] – [8] that the dynamic behavior of dopants in silicon leads to "normal" device operation at 77K; it is only at much lower temperatures (roughly 30K and below) that thermal energy becomes sufficiently scarce and "unusual" device behavior becomes a factor.

A final factor in low temperature CMOS is hot-carrier reliability. It has long been known that along with a larger drain current, low temperature operation produces a higher level of hot-carrier generation; this is often cited as a fatal weakness of 77K CMOS. However, scaling the power supply voltage will reduce this problem – the only question is if reliability will improve faster than performance is lost. It has been shown [16] that this trade-off is indeed favorable at 77K; in fact, when taken fully into consideration, the improvements in hot-carrier reliability are actually very large – proportionally *larger* than the performance improvements. This large reliability margin can actually be used to push a given technology to *smaller channel lengths* than is possible at room temperature – thus permitting *very large* performance gains vs. the 300K base case. Once again, it should be noted that these benefits do not occur *gratis*, but are available because the technology *can be redesigned for maximally-salutary behavior at 77K.*

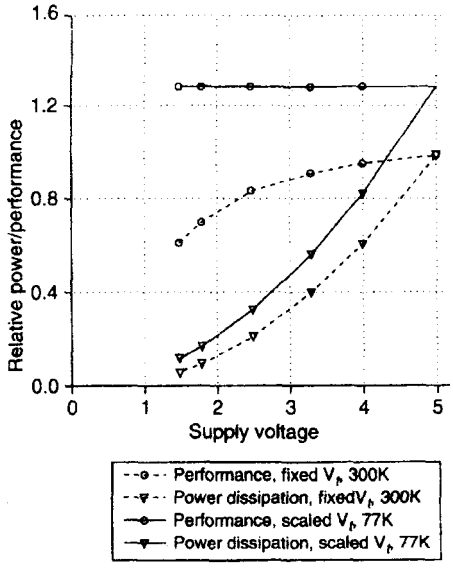

Figure 8. Performance and power dissipation, comparing a 300K (fixed threshold voltage) technology with a 77K (scaled threshold voltage) technology

Until recently, there was one last barrier to 77K CMOS. The use of 77K requires that the pMOS transistor have a p^+ polysilicon gate; however, for simplicity reasons, CMOS used n^+ polysilicon on both nMOS and pMOS devices. However, at the 0.25μm technology node (introduced about 1999), CMOS technology was finally forced to make the change to dual-poly gates. Thus, the last technology barrier to 77K CMOS is now gone; all that is required is a change in the channel tailor implant to one that creates proper threshold voltages for 77K operation.

Conclusions. This paper has examined the evolution of scaling theory as the main driver of CMOS technology. The ability to create scaling theory was a unique aspect of CMOS technology, and has been the main reason for the spectacular success of CMOS. However, along the way, the priorities of the industry have varied drastically as the various challenges inherent in scaling theory have changed over time. During the 1970s and 1980s, process technology issues were dominant, while during the 1990s design and usage issues came to the fore. The story may change again, as there is now a severe need for the development of new gate dielectric materials, and for a new gate material system.

Viewed in this light, scaling theory has remained the dominant force, but has tended to shift its focus about every ten years; there is a decade of stability in the focus of engineering efforts, punctuated by a rapid upheaval in which the "high-priority" issues change drastically. This is a concept familiar in evolutionary biology, that of "punctuated equilibrium" – the form and function of organisms is stable for a long period of time, until the equilibrium is upset by either an innovation or a disaster (usually the latter); the entire system re-arranges itself, and a new equilibrium is established. This seems to be the situation for CMOS technology, and we are entering a new period of upheaval, after which a new equilibrium will be established. While process technology innovations will rise somewhat in importance, it is likely that the largest upheavals will affect circuit design and circuit design methodologies – largely because, to date, these methods have remained virtually unchanged for the past four decades

However, over all those years, the basic theme has remained the same – scaling theory has permitted CMOS to be a powerful yet inexpensive technology. For the most part, challenges from "alternative" devices cannot at this time be based solely on *performance*, but must be based (to a large degree) on *cost*. This should be noted with care; the technological reality is that for analog design and *particularly* for RF design, the bipolar transistor is a much "nicer" device to work with and to use. Although a technically inferior device, the MOS transistor has triumphed due to low cost and great ubiquity. Scenarios for the use of bipolar or BiCMOS technologies in RF-IC design are plausible, but must provide careful evaluation of the *overall cost picture* of the technology choices.

The present situation also indicates that the methods of employing CMOS in circuit design (particularly analog and RF design) are obsolete – and must be replaced with completely new, fully-integrated approaches to both *interpreting* the behavior of the MOS transistor and *directly employing* those methods in circuit design. New thinking is clearly required. Finally, scaling theory strongly favors the use of reduced temperature

operation – despite the obvious practical constraints. While unlikely to be a mainstream technology, the tantalizing opportunities offered by low temperature CMOS operation, combined with recent changes in the international security environment, make this a technology worth considering for certain situations.

Acknowledgements. The authors would like to acknowledge the influence of numerous colleagues over the years in helping to shape his thinking on this subject; here, they would particularly like to thank Peter Bendix, David Binkley, Colin McAndrew, Edward Nowak, and Yannis Tsividis for numerous enlightening discussions over a long and interesting period of time.

1. J. Lilienfeld, U.S. Patent Nos. 1,745,175 (1930), 1,877,140 (1932), and 1,900,018 (1933). 2. Quoted in G. Gilder, "The Soul of Silicon," Forbes ASAP, 1 June 1998. 3. C. Mead, "Fundamental Limitations in Microelectronics – I. MOS Technology," Sol. St. Elec. vol. 15, pp. 819 – 829 (1972). 4. R. Dennard et al., "Design of Ion Implanted MOSFETs with Very Small Physical Dimensions," IEEE J. Sol. St. Circ. vol. SC-9, pp. 256 – 268 (1974). 5. E. Nowak, "Ultimate CMOS ULSI Performance," 1993 IEDM Tech. Dig., pp. 115 – 118. 6. D. Foty and E. Nowak, "MOSFET Technology for Low Voltage/Low Power Applications," IEEE Micro, June 1994, pp. 68 – 77. 7. D. Foty, "The Design of Deep Submicron FETs for Low Temperature Operation," Proceedings of the Symposium on Low Temperature Electronics and High Temperature Superconductors (ed. by S. Raider), pp. 63 – 77 (1993). 8. D. Foty and E. Nowak, "Performance, Reliability, and Supply Voltage Reduction, with the Addition of Temperature as a Design Variable," Proceedings of the 1993 European Solid State Device Research Conference, pp. 943 - 948. 9. Y. Tsividis, "Moderate Inversion in MOS Devices," Sol. St. Elec. Vol. 25, pp. 1099 – 1104 (1982). 10. E. Vittoz, "Micropower Techniques," in Design of MOS VLSI Circuits for Telecommunications (ed. By J. Franca and Y. Tsividis), Prentice Hall, 1994. 11. D. Foty, "Re-Interpreting the MOS Transistor for the 21st Century: Generalized Methods and Their Extension to Nanotechnology," Proceedings of the 21st Nordic VLSI Design Conference (NorChip), pp. 8 – 15 (2003). 12. D. Binkley et al., "A CAD Methodology for Optimizing Transistor Current and Sizing in Analog CMOS Design," IEEE Transactions on Computer-Aided Design of Circuits and Systems vol. CAD-22, pp. 225 – 237 (2003). 13. C. Enz, F. Krummenacher, and E. Vittoz, "An Analytical MOS Transistor Model Valid in All Regions of Operation and Dedicated to Low Voltage and Low Current Applications," Analog Int. Circ. and Signal Proc. vol. 8, pp 83 – 114 (1995). 14. G. Gildenblat et al., "SP: An Advanced Surface-Potential-Based Compact MOSFET Model," IEEE J. Sol. St. Circ. vol. SC-39, pp. 1394 – 1406 (2004). 15. Y. Yi et al., "Temperature-Scaling Theory for Low-Temperature-Operated MOSFET with Deep-Submicron Channel," Jpn. J. Appl. Phys. vol. 27, pp. L1958 – L1961 (1988). 16. D. Foty and E. Nowak, "Performance/Reliability Trade-Off and Optimized nMOSFET Design for 77K CMOS Logic," Proceedings of the Symposium on Low Temperature Electronics and High Temperature Superconductors (ed. by S. Raider), pp. 89 - 95 (1993).

УДК 638.235.231

Райд Фатхи Сахавне
Національний університет "Львівська політехніка",
кафедра САПР

МОДЕЛЮВАННЯ НЕЙРОННИХ МЕРЕЖ ДЛЯ РЕАЛІЗАЦІЇ АЛГОРИТМІВ РОЗПІЗНАВАННЯ

Розглянуті методи опису та зображення нейронних мереж для реалізації алгоритмів розпізнавання. Розроблено послідовні, паралельні та послідовно-паралельні схеми реалізації обчислень у нейронних елементах.

Methods of the description and representation of neural networks of realization of algorithms of recognition are considered. Algorithms of emulation of such networks are developed and realized. Examples of results of work of the program of emulation are resulted.

Завжди існував і існує великий інтерес до проблемно-орієнтованих та спеціалізованих систем обробки і розпізнавання зображень. Це передусім пов'язано з вимогами реального часу, обмежених