ABSTRACT:

This report advances the need to develop creatively new design strategies for low voltage, low power, high performance, integrated circuits. Attention is focused on analog cells, which are particularly vulnerable to process vagaries and the accentuated performance sensitivities to critical modeling parameters spawned by low voltage operation. Device operation in triode and weak inversion regimes is proposed for specialized applications, as are new modeling initiatives and the need to implement adaptive hardware as a means of mitigating the deleterious effects of model, circuit, and layout uncertainties.
1.0. INTRODUCTION

The continuing shrinkage of transistor channel lengths in deep submicron CMOS technology is accompanied by an absolute requirement for proportionately smaller power bus voltages. Complementing the ongoing penchant for device downsizing is the omnipresent directive to package ever-increasing heights of high performance functionality into the monolithic circuits that comprise the foundation of portable and handheld electronic systems. Implicit to this portability mission is the requirement of harmonizing the low voltage biasing of extremely short channel devices with overall circuit operation at acceptably low power levels. The combined goals of low voltage and low power operation are especially vexing in analog circuit cells. In addition to achieving the traditional design goals of input-to-output (I/O) gain and I/O impedance levels, these cells must also meet operating specifications for such metrics as stability margins, dynamic range, settling times, performance sensitivity to temperature, and equivalent input noise.

The vast majority of high performance analog circuit cells realized in MOSFET technologies traditionally exploits transistors operating in saturation. This design tack derives largely from the fact that the channel inversion layer of saturated devices does not extend throughout the geometric channel extending from the source region to the drain region. In deep submicron devices, the inversion layer extends over only a small fraction of the geometric channel length and, in fact, is concentrated proximate to the source region. As a result, the electrical channel length is smaller than its geometrical value, which gives rise to improved frequency response capabilities, as monitored by the short circuit gain-bandwidth product of the device. Moreover, the charge depletion zone resultantly established in saturation between the drain region and the shortened charge inversion layer reduces the effective gate-drain capacitance to its generally small overlap value, thereby mitigating potential bandwidth deterioration in high gain common source stages.

The price paid for transistor operation in saturation is the need for a relatively large magnitude of drain-source voltage. In N-channel devices, for example, the drain-source voltage, \( V_{ds} \), must be at least as large as the drain saturation voltage, \( V_{dsat} \), which to first order is proportional to the difference between the gate-source voltage, \( V_{gs} \), and the threshold potential, \( V_h \). In order to ensure reasonably linear signal processing performance, the requirement of \( V_{ds} \geq V_{dsat} \) must be sustained over all anticipated input signal levels, which may translate into the necessity of establishing drain-source biases that significantly exceed the drain saturation level. Such requisite biasing severely restricts topological design flexibility in light of the fact that the allowable power bus voltages for deep submicron technologies are necessarily small to preclude voltage breakdown of integrated transistors.

2.0. FUNDAMENTAL PROPOSAL

The voltage dilemma addressed in the preceding section, together with the everlasting need for mixed signal systems capable of high speed data and information processing at seemingly unrealistically low power dissipation levels, encourages a serious focus on innovative new strategies and analytical techniques focused on low voltage CMOS integrated circuit design. The suggested low voltage design scenario invariably embraces MOS devices operated in triode volt-ampere characteristic regions where for NMOS units, \( V_{ds} < V_{dsat} \). Depending on intended circuit function in suitable applications, it also allows for device operation in subthreshold regimes.
where $V_{gs} < V_{th}$. In contrast to relevant philosophies and beliefs proffered a scant decade or so ago, a present day suggestion that high performance analog circuit cells can be realized with MOS technology devices operating in triode regions or even subthreshold regimes is far from ludicrous. The subject suggestion is worthy of definitive study principally because of dramatic advances witnessed in device processing technology. In particular, reliable and reproducible MOS transistors boasting drawn channel lengths from 65 nm to 130 nm are commonplace in the extant state of the art. The short channel lengths indigenous to these deep submicron transistors give rise to relatively large lateral channel fields for even modest levels of drain-source biasing. As a result, measurable, fast drift components of channel currents are supported easily. In concert with the small internal device capacitances implicit to deep submicron transistors, these lateral fields render plausible broadband frequency responses and quickly settling time domain responses to abrupt signal excitations.

Several critical issues must be addressed to promote analog circuit realization with MOS technology devices operated in triode and weak inversion regimes. These issues minimally embrace physical device modeling inclusive of the reliable parameterization of active devices for computer-aided circuit analysis and design, innovative new circuit topologies expressly developed for low voltage operating environments, and circuit compensation techniques that promote low voltage and low power operation while mitigating the circuit level effects of processing vagaries and environmental uncertainties.

2.1. DEVICE MODELING AND PARAMETERIZATION

As noted previously, state of the art, high performance, CMOS analog circuit cells invoke the constraint of saturation domain operation on most of their utilized active devices. It is therefore understandable that the SPICE model parameters typically supplied by process foundries are optimized largely to ensure simulation accuracy only when MOS technology devices operate exclusively in saturated regimes. Although foundry models, and particularly the ubiquitous Level 49 SPICE model, are commonly applied to transistors that are constrained to operate in triode domains, the resultant simulation accuracy is questionable. Even if simulation propriety is sustained over frequency to within acceptable errors, the fact that most Level 49 parameters are non-physical curve fitting numbers renders impossible a physically meaningful consideration of device scaling effects, thermal sensitivities, signal-induced memory episodes, nonlinear responses to signal overdrive conditions, and a host of other phenomena. While simulation accuracy in triode regimes is at best questionable with models supplied by process foundries, recent circuit design work for an ultra-low power biomedical system application leaves little doubt that the subject models are imprecise for MOS devices operating in weak inversion.$^{[1]}$

In view of the foregoing shortfalls, any effort to forge a realistic design methodology for low voltage, low power CMOS integrated circuits must commence with an appropriate device modeling effort. The developed models must deliver accurate computer-aided simulations of circuits for all applicable device operating modes, just as Level 49 ostensibly does for the saturation domain. Additionally, the models must promote an insightful understanding of the inherently complex relationships between circuit level parameters and physical phenomenology at both device and layout levels. This insightful understanding is an absolute necessity in that it underpins innovative new circuit topologies that perform optimally in the face of low voltage and low standby power operating mandates. Most importantly, it is essential that the modeling effort evolve models that are synergistic with those embodied in the Cadence design suite, for it is unrealistic to expect that a creative foray into low voltage and low power CMOS design
methodologies can change immediately the engineering culture pervasive of current integrated
circuit design and layout practices. To this end, a suitable macromodel wrapped around Level 49
is envisaged for the initial phase of the program, wherein the developed wrapper enhances design
insights and reliably extends the range of applicability of Level 49 to all device domains of
operation over all signal frequencies of practical interest.

Fortunately, a substantial body of relevant modeling literature serves as a sturdy
foundation upon which a viable low voltage design strategy for CMOS technology can be built.
For example, the EKV Compact MOSFET Model embraces all MOS device operating regions,
inclusive of weak inversion, moderate inversion, and strong inversion\(^2\). Although the model is
analytically rigorous and sensitized to physical phenomenology, it nonetheless exploits several
approximations that may indeed prove inappropriate for devices featuring channel lengths
smaller than 130 nM. These approximations and the model empiricisms to which they give rise
must be definitively tested, evaluated, and modified in accordance with the mandates of analyses.
Moreover, a parameterization scenario in terms of appropriate static and dynamic data must be
contrived before the model can be applied confidently to low voltage circuit design problems.
To this end, a modest level of follow-on research is required to confirm both the accuracy and
the physical propriety of the EKV structure\[^3\]-\[^9\]. Although the EKV model has already been
embedded in numerous circuit simulators, parameters allowing for meaningful model usage are
rarely, if ever, supplied by commonly used process foundries. Thus, the aforementioned re-
search must necessarily embody the development of pertinent computational algorithms that pro-
duce reliable and realistic numerical model parameters in terms of physical principles and
appropriate device characterization data\[^10\]-\[^12\].

2.2. CIRCUIT DESIGN

Once appropriate device models for low voltage and low power system applications are
forged, extraction algorithms for model parameters are written, and the propriety of both said
models and the associated parameterization methods is confirmed, the development of circuit de-
sign methodologies for particular system applications can be initiated. The systems currently
envisaged include baseband modules for communication systems, high-bit resolution data
converters, biomedical health care monitoring and health maintenance systems, a broad variety
of sensor systems used in military electronics, and intelligent navigation systems. Many of the
circuit topologies currently exploited in CMOS circuits are likely to remain pertinent, with but
minor biasing and other variations to accommodate low voltage circuits. Others, such as the to-
tem pole arrangement consisting of the drain-source circuit series interconnection of common
source driver, common gate cascode, and active current source load, are likely to be abrogated
because of the limited voltage headroom afforded by necessarily low power bus voltages.

The core requirement of the low voltage repertoire is a transconductor amplifier that
linearly transforms an input signal of specified dynamic range to a signal current flowing into a
high impedance output port\[^13\]. In addition to a linear voltage -to- current I/O characteristic, a
suitable general purpose transconductor amplifier ideally delivers a transconductance, or short
circuit output current -to- input voltage gain, that varies linearly with an applied static control
voltage. At a minimum, the transconductance versus control voltage relationship must be a well-
behaved monotonic function. Such a transconductor boasts utility in applications that include
voltage amplifiers, power amplifiers, mixers, oscillators, impedance converters, peak and abso-
lute detectors, gyrators (used to transform a capacitance load to an inductive input impedance),
and a host of other fundamental circuit cells.
The composite field effect transistor (COMFET), which is abstracted in Figure (1a) and shown in conjunction with an embedded offset voltage, $V_k$, in Figure (1b), may be a first order foundation of an optimal transconductor[14]. This contention stems from archival information suggesting that with suitable topological modifications, the COMFET can be made to behave as a single NMOS transistor having very low and even zero or negative threshold voltage[15]-[20]. A low threshold voltage obviously supports the design objective of low voltage operation.

![Fig. (1).](image)

(a). Schematic diagram of the composite field effect transistor, or COMFET. (b). COMFET with offset biasing applied. In both circuits, the drain current, $I_d$, is controlled by the effective gate voltage, $V_{gse}$.

An appreciation of COMFET attributes begins by observing that in either of the two diagrams in Figure (1), the indicated drain current, $I_d$, is given to first order by

$$I_d = \frac{K_{nn}}{2}(V_{gsn} - V_{hn})^2 = \frac{K_{pp}}{2}(V_{sgp} - V_{hp})^2,$$

(1)

assuming that both transistors operate in saturation. In this expression, $V_{gsn}$ and $V_{hn}$ respectively symbolize the applied gate-to-source voltage and the threshold voltage of the N-channel transistor, and $V_{sgp}$ and $V_{hp}$ respectively designate the applied source-to-gate voltage and threshold voltage of the P-channel device. Moreover, the N-channel device has a transconductance coefficient, $K_{nn}$, of

$$K_{nn} = \mu_n C_{oxn} \eta_n,$$

(2)

where $\mu_n$ is the mobility of electrons in the inverted channel, $C_{oxn}$ is the density of oxide capacitance, and $\eta_n$ represents the gate aspect ratio. Similarly, the P-channel unit has

$$K_{pp} = \mu_p C_{oxp} \eta_p,$$

(3)

with $\mu_p$ representing the mobility of holes in the inverted channel, $C_{oxp}$ the density of oxide capacitance, and $\eta_p$ the gate aspect ratio. If (1) is solved for $V_{gsn}$ and $V_{sgp}$ in terms of the drain current, $I_d$,

$$V_{gsn} = V_{hn} + \sqrt{\frac{2I_d}{K_{nn}}}$$

$$V_{sgp} = V_{hp} + \sqrt{\frac{2I_d}{K_{pp}}}.$$  

(4)

Since the voltage, $V_{gse}$, indicated in Figure (1b) is the voltage sum, $(V_{gsn} + V_{sgp} - V_k)$, (4) yields
If an effective threshold voltage, $V_{he}$, is defined, such that

$$V_{he} \triangleq V_{hn} + V_{hp} - V_k,$$

and an effective transconductance coefficient, $K_{ne}$, is introduced in accordance with the stipulation,

$$\frac{1}{\sqrt{K_{ne}}} \triangleq \frac{1}{\sqrt{K_{nn}}} + \frac{1}{\sqrt{K_{pp}}} = \frac{\sqrt{K_{pp}} + \sqrt{K_{nn}}}{\sqrt{K_{nn}K_{pp}}},$$

it is clear that (5) implies

$$I_d = \frac{K_{ne}}{2} (V_{gse} - V_{he})^2.$$  \hspace{1cm} (8)

Equation (8) suggests that the composite structure can be viewed electrically as a single MOSFET having a transconductance coefficient of $K_{ne}$ and an effective threshold voltage of $V_{he}$. But observe in (6) that $V_{he}$ can be made small, zero, or even negative by an appropriate choice of the offset voltage, $V_k$.

Figure (2) offers a simple way of realizing the embedded voltage, $V_k$, as the gate-source voltage of an N-channel transistor, $M2$, which is biased to operate in saturation. The COMFET in this diagram consists of transistors $M1a$ and $M1b$. Transistors $M3$ and $M4$ are identical devices that are interconnected as a current mirror to ensure that the current, $I_d$, conducted by transistor $M1a$ is identical to the current flowing in the second half of the COMFET pair, transistor $M1b$. The embedded voltage, $V_k$, derives from

$$I_k - I_d = \frac{K_{nk}}{2} (V_k - V_{hk})^2.$$  \hspace{1cm} (9)
where $K_{nk}$ is the transconductance coefficient of $M2$ and $V_{hk}$ is the threshold voltage of the device. Assuming that the current, $I_k$, conducted by the constant current source is significantly larger than the COMFET drain current, $I_d$, (9) gives

$$V_k \approx V_{hk} + \sqrt{2I_k/K_{nk}}, \quad (10)$$

which emulates the desired constant offset voltage abstracted in Figure (1b). The circuit at hand is rendered utilitarian through the addition of transistor $M5$. If the gate aspect ratio of $M5$ is $\alpha$-times larger than that of transistor $M4$, a current of $\alpha I_d$ is provided to an arbitrary load. This load can be a simple resistance or impedance, as in the case of an amplifier application, or it can be a suitable N-channel current source if a transconductor is the design goal.

The circuit just discussed and appearing schematically in Figure (2) is but one example of a low voltage circuit topology that exploits the COMFET interconnection of an N-channel and a P-channel transistor. Numerous other COMFET configurations suitable for transconductor and voltage amplifier applications are possible \cite{21}. Many of these alternative structures are amenable to MOS devices operating in triode and weak inversion regimes, as well as in saturation domains. In addition, it is to be understood that the COMFET approach itself is but one example of a plausible low bias voltage topology. It is not presented herewith to suggest that it is the only innovative circuit topology appropriate for low bias voltage applications. Instead, it is offered merely as a means of demonstrating that innovative topologies for low voltage networks are likely not to mirror the canonic cells stereotypically exploited in analog MOS design problems for which device operation in saturation mode is an a priori assumption.

### 2.3. CIRCUIT COMPENSATION

The continuing down scaling of device feature sizes is necessarily accompanied by increased uncertainties in device processing, particularly with respect to geometrical dimensions, the profiles of device dopant concentrations, and many of the physical parameters (mobility, permeability, diffusion lengths, Debye lengths, etc.) that impact attainable circuit performance. These uncertainty issues are exacerbated by low voltage biasing, which effectively forces circuit designers to implement monolithic circuits whose performance sensitivity to the key circuit level parameters that are influenced by processing uncertainties is higher than levels deemed acceptable in earlier generation designs. An additional design-oriented complication derives from the fact that numerous parameters indigenous to the models supplied by processing houses are rarely applicable to circuits in which the function of some of the utilized active devices spans all possible domains (saturation, triode, weak inversion) of operation. This modeling unreliability is largely the result of empiricisms invoked to represent multidimensional, incompletely understood, or otherwise complicated physical phenomena.

Device and even layout-related parametric uncertainties notwithstanding, the low voltage circuits required of new generation communication and data processing systems must nonetheless deliver reliable and predictable I/O performance. Appropriate circuit compensation techniques, some of which may lead to tunable, standardized cell, “drop-in” architectures for monolithic analog and mixed signal chips, must therefore be viewed as an essential component of low voltage design methodologies.

It is possible that the aforementioned compensation amounts to little more than the implementation of appropriate negative feedback used to control such metrics as gain, I/O impedance levels, bandwidth, settling times, and the like. Although prudently applied feedback
enjoys laudable parametric desensitization advantages, its utility in high performance low voltage circuits is likely to be limited. A notable limitation of feedback is its propensity toward poor closed loop transient responses or even outright closed loop instability unless care is exercised to ensure an open loop characterized by a dominant pole frequency response$^{[22]}$. Unfortunately, open loop pole dominance often limits the attainable closed loop bandwidth, and it may require circuit measures that manifest increases in standby power dissipation. The potential instability implications of feedback are rendered especially daunting at high signal frequencies when extreme parametric uncertainties and/or excessive input signal strengths force the active network undergoing compensation to operate nonlinearly$^{[23]}$.

In the subsections that follow, abridged discussions of arguably atypical circuit compensation measures are presented. The topics on which attention is focused are not intended to be an exhaustive treatment of the type of compensation, or tuning, measures foreseen as essential to the proper operation of low voltage circuits. Instead, these topics serve as a plausible foundation for the development of more advanced tuning circuits, and they offer a glimpse as to the kinds of compensation that can be realized currently.

2.3.1. Predictable and Constant Transconductance

As noted earlier, the transconductor amplifier is a fundamental building block of amplifiers, active filters, and other types of networks. Since the forward transconductance of a transconductor sets an open gain of the amplifier, pole frequencies, zero frequencies, and the quality factor of a filter, and other metrics in a wide variety of different types of circuits, the ability to set accurately the transconductance value is a potentially critical design issue. The fundamental features of a relatively simple compensation network yielding constant and predictable transconductance are depicted in Figure (3). The transconductance undergoing tuning is presumed to have a transconductance, $G_m$, which increases monotonically with the voltage, $V_c$, applied to its control port. The operational amplifier (op-amp) must have an open loop voltage gain that is sufficiently large to enable the presumption of a virtual ground at its inverting input port. Moreover, this amplifier must have an input impedance that is sufficiently large to allow for the tacit neglect of input current. Accordingly, the integrating capacitor, $C$, develops a voltage, $V_c$, with the polarity indicated in the figure, and the resultant sum of the capacitor current and transconductor output current, flows through the resistance, $R$, in the direction delineated in the figure. If linearity prevails in both the transconductor and the op-amp,

\[
\frac{dV_c}{dt} = \frac{V_{ref}}{RC} \left( 1 - G_m R \right).
\]
Since the input to the subject subcircuit is a constant reference voltage, \( V_{\text{ref}} \), the output control voltage, \( V_c \), must be a constant in the steady state. It follows from (11) that \( G_m = 1/R \) under steady state circumstances. To the extent that resistance \( R \) can be realized accurately, \( G_m \) is resultant prescribed. Observe that if \( G_m \) decreases from \( 1/R \) because of biasing or any other uncertainty or spurious perturbation, the voltage derivative, \( dV_c/dt \), is positive, thereby indicating that \( V_c \), and thus \( G_m \), increases with time until steady state operation is reestablished at the control voltage yielding \( G_m = 1/R \). On the other hand, if \( G_m \) increases above \( 1/R \), \( dV_c/dt \) in (11) is negative, which promotes a continually decreasing control voltage and transconductance until the steady state corresponding to \( G_m = 1/R \) is achieved. Other constant transconductance networks, inclusive of switched capacitor structures, appear in the literature\(^{[24]} \).

A simple example demonstrating the utility of the compensation network in Figure (3) is offered in Figure (4). In principle, the indicated transconductor network can be an interconnection of any number of transconductor amplifiers as long as all transconductors for which transconductance control is desired are matched to the transconductor in the reference cell. To be sure, the effectiveness of this compensation approach is limited by non-ideal integrators and mismatches among the active cells. But numerous alternative compensation structures worthy of further investigation and possible modification for low voltage electronics have been developed\(^{[25]-[28]} \). In addition to effecting predictable amplifier transconductances, these more complicated networks are capable of controlling the quality factor of active filters and even the frequencies of the deterministic poles and zeros deemed critical for acceptable I/O performance.

![Transconductor Network](image)

![Reference Cell](image)

**Fig. (4).** Simple example of the use of the compensator in Figure (3) to control the transconductances of the active elements embedded in a general active network.

### 2.3.2. Common Mode Bias Compensation

Any presumably balanced differential amplifier is vulnerable to mismatched electrical characteristics between transistor and passive component pairs. The immediate effect of the operational imbalance deriving from such mismatches is deteriorated common mode rejection
ratio, which in turn spawns potentially significant bandwidth compression, reduced dynamic range, and several other performance shortfalls. Because device mismatches are inevitable in deep submicron technologies, common mode compensation aimed toward preserving balanced differential operation is an inherent necessity in high performance integrated circuits\(^{[29]-[32]}\).

Voltage biasing issues are especially acute in differential amplifiers that exploit active current source loads to realize high gain. The relatively simple differential pair in Figure (5) dramatizes the biasing issue at hand. In this circuit, the small signal differential voltage gain, \(V_{do}/V_{di}\), is large because the effective load resistance imposed on each N-channel device is the parallel combination of the relatively large drain-source channel resistances of the N-channel and P-channel units. But since each drain node is the junction of a P-channel current source placed in series with an N-channel transistor that behaves as a current sink under quiescent operating conditions, the accurate stipulation of the quiescent value, \(V_Q\), of the voltages at the two output ports is virtually impossible. Indeed, if all transistors have infinitely large channel resistances, voltage \(V_Q\) is analytically indeterminate. Even if the P-channel transistors are supplanted by passive resistances, \(V_Q\) remains problematic because of active device and resistance mismatches.

\[
\begin{align*}
V_{bias} & \quad M3 & \quad M4 & \quad V_{do} \\
V_Q & \quad V_{di} & \quad M1 & \quad M2 & \quad V_{bias} \quad R_i \\
V_Q + \frac{V_{do}}{2} & \quad V_Q - \frac{V_{do}}{2} \\
V_Q + \frac{V_{di}}{2} & \quad V_Q - \frac{V_{di}}{2}
\end{align*}
\]

\textbf{Fig. (5).} Differential amplifier using P-channel current sources in the drain load circuits of the N-channel transistors to achieve high voltage gain.

In order to mitigate the foregoing biasing issue, the circuit in Figure (6) is proposed\(^{[33]}\). This compensation network appends the matched transistor pairs, \(M5-M6, M7-M8, \) and \(M9-M10\), to the differential amplifier shown in Figure (5). In addition, two large, but matched, resistances of value \(R\) are connected between the drain nodes of the original N-channel transistors. Because attention is directed herewith to the biasing of the basic differential cell, only quiescent voltages are delineated in the subject diagram. Of particular interest is the static common mode voltage, \(V_{cm}\), established at the junction of the two resistances, \(R\). If the static output port voltages are \(V_{Q1}\) and \(V_{Q2}\), as indicated in the schematic diagram, this voltage is

\[
V_{cm} = \frac{V_{Q1} + V_{Q2}}{2}.
\]

Since the desired output port voltages are \(V_{Q1} = V_{Q2} \triangleq V_{Q}\), a voltage equal to \(V_{Q}\) is applied as a reference bias to the gate of transistor \(M5\). In actual practice, this reference bias may differ slightly from \(V_{Q}\) because of component mismatches.
When $V_{cm} = V_Q$, transistors $M5$ and $M6$ function as a balanced pair, and the currents conducted by $M5$, $M6$, $M7$, and $M8$ are identical and equal to $I_k/2$, where it is understood that $I_k$ is realized as a high output impedance, N-channel current sink. The gate aspect ratios of these four devices are chosen carefully to ensure that all transistors are conductive over all reasonable perturbations of the common mode voltage, $V_{cm}$. When operating circumstances induce an increase in $V_{cm}$ over its desired value, $V_Q$, the drain current conducted by transistors $M6$ and $M8$ becomes larger than that of $M5$ and $M7$, which gives rise to an increase in the voltage developed at the drain of transistor $M5$. Because the drain of $M5$ is incident with the gate terminals of both transistors $M9$ and $M10$, the currents flowing in $M9$ and $M10$ decrease, and the voltages at the drain terminals of these two transistors also diminish. Since these drain terminals are connected to the output ports of the original differential pair, the common mode voltage, $V_{cm}$, correspondingly decreases, thereby effecting compensation of the original increase in $V_{cm}$. Precisely the opposite scenario to the one just described occurs when $V_{cm}$ falls below $V_Q$.

### 2.3.3. Adaptive Biasing

MOS technology amplifiers and transconductors using transistors in the I/O signal path that are biased for operation in weak inversion or low current triode regimes suffer from an inherent operating limitation. In particular, while these configurations boast micropower consumption under quiescent circuit conditions, they are incapable of delivering the large output currents mandated by robust input signals. A plausible circumvention of this dilemma in differential amplifiers is the implementation of so called adaptive biasing, wherein the biasing currents flowing through the transistors embedded in the differential input stage are automatically adjusted in response to applied differential signal swing\[^{[34]}-^{[36]}\]. In particular, adaptive biasing boosts the input stage current when large differential input signals are applied, while remanding these input stage currents to their quiescent design levels when small input signals prevail. Of course, simple current mirroring can translate any boost of input stage currents to the output stage of the considered network, thereby allowing for the possibility of sustaining an output response that is a nominally linear function of the input signal level.
An examination of the attributes of adaptive biasing begins with a consideration of the simple transconductor in Figure (7). The amplifier is shown driving a capacitive load, $C_l$. In view of the fact that this load capacitor faces a relatively high output impedance, the circuit at hand is suitable for use as an integrator, which is a fundamental building block of transconductor-based biquadratic filters\(^{37-38}\). The two input voltages, $V_{i1}$ and $V_{i2}$, applied to the balanced pair comprised of transistors $M1$ and $M2$ are conveniently expressed as

\[
V_{i1} = V_{ci} + \frac{V_{di}}{2}, \\
V_{i2} = V_{ci} - \frac{V_{di}}{2},
\]

where $V_{di}$ represents the differential input signal applied to the gate terminal of transistor $M1$ to the gate terminal of $M2$. On the other hand, $V_{ci}$ designates the common mode input voltage, which includes requisite gate biasing for both of the input stage transistors. Thus, $V_{i1} = V_{i2} = V_{ci}$ under quiescent, or zero signal, conditions. The $M1$-$M2$ pair drives P-channel diode-connected loads consisting of the matched transistors, $M3$ and $M4$. Observe that transistor $M6$ mirrors the current flowing through $M4$. If the gate aspect ratio of transistor $M6$ is $K$-times that of $M4$, and the gate aspect ratio of $M5$ is $K$-times that of transistor $M3$. In the quiescent state, for which $I_1 = I_2 = I_k/2$, the static current conducted by both transistors $M6$ and $M8$ is $K_I, K$ is the tail current conducted by the current sinking transistor, $M9$. The static voltage, $V_k$, is the $M9$ gate-source biasing that supports current $I_k$. Transistors $M5$ and $M7$ are respectively matched to $M6$ and $M8$ and are inserted to maintain balanced quiescent operation. Note that like the relationship between transistors $M6$ and $M4$, the gate aspect ratio of $M5$ is $K$-times larger than the gate aspect ratio of transistor $M3$.

As suggested earlier, the principle objective of adaptive biasing applied to balanced differential amplifiers is to increase the tail current, $I_k$ in Figure (7), as a function of the applied differential input voltage, $V_{di}$. The foundation of such an adaptive subcircuit is the current differencing scheme—originally developed for bipolar technology—abstracted in Figure (8)\(^{39}\). If
the current flowing into terminal 1 is $I_1$, which might be a mirrored version of the drain current flowing into transistor $M1$ in the differential amplifier of Figure (7), transistor $Ma$ conducts a current of $I_1$. Since the $Ma-Mb$ pair is a current mirror with identical gate aspect ratios, the drain current conducted by transistor $Mb$ is also $I_1$. Now, if current $I_2$, which is possibly a mirrored version of the $M2$ drain current in Figure (7), flows into terminal 2, the drain of transistor $Mc$ necessarily conducts a current, $(I_2 - I_1)$. It follows that the current flowing into terminal 3 is $P(I_2 - I_1)$, since the $Mc-Md$ pair is a current mirror whose ratio of gate aspect ratios is $P$. Note that if $I_1 = I_2$, the gate-source voltage of $Mc$, and hence of $Md$, falls to its threshold level, thereby yielding no current flowing into terminal 3. Moreover, if $I_1 > I_2$, transistor $Mc$, which always operates in saturation, is forced into cutoff since a saturated N-channel transistor cannot support a negative drain current. Because of the current mirror topology formed by transistors $Mc$ and $Md$, zero current is resultantly conducted by transistor $Md$ when $I_1 > I_2$. Letting $I_3$ designate the current flowing into terminal 3, the foregoing observations can be summarized as

$$I_3 = \begin{cases} P(I_2 - I_1), & I_2 > I_1 \\ 0, & I_2 \leq I_1 \end{cases}$$

(14)

Figure (9) depicts the amplifier in Figure (7) with adaptive biasing incorporated. Two adaptive kernels are used to allow tail current boosting for both positive and negative differential input voltages, $V_{ds}$, which respectively correspond to $I_1 > I_2$ and $I_1 < I_2$, where $I_1$ and $I_2$ are the signal dependent drain currents of transistors $M1$ and $M2$. Thus, when $I_1 > I_2$, transistor $M14$ conducts a current of $P(I_1 - I_2)$, while transistors $M15$ and $M17$ are cutoff. On the other hand, $I_2 > I_1$ forces $M15$ to conduct $P(I_2 - I_1)$, while constraining transistors $M14$ and $M16$ to cutoff. Note that the current, $I_1$, which activates the adaptive cell comprised of transistors $M7$, $M13$, $M15$, and $M17$ and is the current conducted by $M1$ in the differential input stage, derives from the current mirror formed of transistors $M3$ and $M11$. The same current, $I_1$, applied to the $M8-M12-M14-M16$ adaptive cell is forged by the $M3-M5$ mirror. The current, $I_2$, which is the current conducted by $M2$ in the differential pair, is applied to the $M8-M12-M14-M16$ cell as the current response to the $M4-M10$ mirror. The same current activating the other adaptive cell derives from the $M4-M6$ mirror. Transistor $M18$, whose gate aspect ratio is $K$-times larger than that of $M4$, $M6$, and $M10$, is inserted to source a current of $KI_2$ to the output port. Transistor $M19$ is identical to $M18$ to preserve balance while conducting a current of $KI_1$. Observe that the drain of transistor $M19$ drives a diode-connected transistor, $M21$. In turn, the gate of $M21$ is incident
with the gate of transistor $M20$, which is identical to $M21$. Accordingly, the drain of $M20$ conducts $KI_1$. In conjunction with the drain current, $KI_2$, of transistor $M18$, this current renders a current of $K(I_2 - I_1)$ available to the load, which in this case is a simple capacitance, $C_i$.

The transconductance amplifier of Figure (7) with adaptive biasing incorporated to allow for signal dependent increases in the tail current, $I_k$, of the differential input stage.

The amount of current boost, say $B(V_{di})$, which is defined as the factor by which the original differential tail current, $I_k$, is increased in response to differential input excitation, can be evaluated for either positive or negative input differential voltage, $V_{di}$. Of course, either differential input voltage condition gives the same boost factor, but it is necessary to stipulate the considered condition to identify which of the two adaptive subcircuits in Figure (9) is active.

Arbitrarily taking $V_{di} > 0$, which implies $I_1 > I_2$ and thus a negative output current, say $I_{out}$, of $K(I_2 - I_1)$, the effective tail current, $I_t$, in the adaptive network under consideration is

$$I_t = I_k + P(I_1 - I_2) = I_k - P\left(\frac{I_{out}}{K}\right).$$

If transistors $M1$ and $M2$ operate in the weak inversion domain, their volt-ampere characteristics can be approximated as $^{[40]}-^{[41]}$

$$I_1 \approx \left(\frac{W}{L}\right)I_{do} e^{V_{gs1}/nV_T},$$

$$I_2 \approx \left(\frac{W}{L}\right)I_{do} e^{V_{gs2}/nV_T},$$

where $(W/L)$ is the gate aspect ratio of $M1$ and $M2$, and $I_{do}$ is a thermally sensitive current parameter that is dependent on processing variables and a host of physical semiconductor
parameters. The voltage, $V_{gs1}$, is the voltage developed across the gate-source terminals of $M1$, and $V_{gs2}$ is its counterpart voltage for transistor $M2$. Additionally, $V_T$ is the familiar thermal voltage of the devices, while $n$ is a semi-empirical current-versus-voltage slope parameter that is related to the ratio of Fermi-to-thermal potentials and is typically in the range of one-to-three. Since the differential input voltage, $V_{di}$, is simply the gate-source voltage difference, $(V_{gs1} – V_{gs2})$, (16) leads to

$$\frac{I_1}{I_2} = e^{V_{di}/nV_T}, \quad (17)$$

which is independent of the somewhat nebulous parameter, $I_{do}$, provided $M1$ and $M2$ are matched transistors. And since the effective tail current, $I_t$, is merely the sum of currents $I_1$ and $I_2$, the last result delivers

$$I_1 = I_t \left(\frac{e^{V_{di}/nV_T}}{1 + e^{V_{di}/nV_T}}\right),$$

$$I_2 \approx I_t \left(\frac{1}{1 + e^{V_{di}/nV_T}}\right). \quad (18)$$

Now, the current difference, $(I_2 - I_1)$, can be written as

$$I_2 - I_1 = \left(\frac{1 - e^{V_{di}/nV_T}}{1 + e^{V_{di}/nV_T}}\right)I_t = \frac{I_{out}}{K}. \quad (19)$$

If (19) is solved for $I_t$ and the solution is substituted into (15), the output current, $I_{out}$, conducted by the load imposed on the adaptive amplifier is expressible as

$$I_{out} = K(I_2 - I_1) = KI_k \left[\frac{1 - e^{V_{di}/nV_T}}{P + 1 - (P - 1)e^{V_{di}/nV_T}}\right]. \quad (20)$$

An inspection of (20) reveals that the original tail current, $I_k$, is multiplied by a bracketed factor that is functionally related to the differential input voltage and the ratio, $P$, of gate aspect ratios invoked in the adaptive compensation cell. In the present case, this bracketed factor is a negative number, but only because of the a priori assumption of $V_{di} > 0$ and hence, $I_1 > I_2$. Had $V_{di} < 0$ been presumed initially, precisely the negative of the subject factor would have appeared in the resultant expression for the output, or load, current. It therefore makes sense to define the current boost factor, $B(V_{di})$, as the absolute value function,

$$B(V_{di}) = \left|\frac{1 - e^{V_{di}/nV_T}}{P + 1 - (P - 1)e^{V_{di}/nV_T}}\right|. \quad (21)$$

An interesting aspect of the boost factor in (21) is that it clearly suggests that the adaptive network in Figure (9) extols positive feedback. In particular, $B(V_{di})$ becomes infinitely large at a value, $V_{es}$, of differential voltage, known as the escape voltage, that is given by
While positive feedback indeed prevails, the circuit is not necessarily unstable and the load current certainly does not ramp toward infinitely large value. For $V_{di} > V_{es}$, transistors $M1$ and $M2$ leave (or “escape”) their weak inversion domains and enter into strong inversion in either triode or saturated regimes. In other words, (21) is meaningless for $V_{di} > V_{es}$.

![Diagram of tail current boost factor of the adaptive transconductance amplifier.](image)

Fig. (10). Tail current boost factor of the adaptive transconductance amplifier of Figure (9). The voltage scale is normalized to the quantity, $nV_T$, where $V_T$ is the thermal voltage and $n$ represents the current slope factor in weak inversion. Only normalized input voltages corresponding to weak inversion of transistors $M1$ and $M2$ are considered.

Figure (10) plots $B(V_{di})$ as a function of the normalized differential input excitation, $V_{di}/nV_T$, for various values of the factor, $P$. The effects of only values of $V_{di}$ smaller than $V_{es}$ are displayed. It should be understood that $P = 0$ corresponds to no adaptive current enhancement. An inspection of (21) reveals that $B(V_{di}) < 1$ for $P = 0$, thereby indicating that the differential current is always less than the original tail current, $I_k$. Note in Figure (10), however, that appreciable current boosting can be achieved when $P$ exceeds unity. Moreover, the amount of boost increases monotonically with the differential input voltage for all values of $P$.

### 3.0. CONCLUSIONS

In view of the rampant desire to package state of the art communication, data processing, and information transmission systems in the form of portable electronics, there is an
omnipresent need to develop creatively new and possibly unified design strategies for low voltage, low power, integrated mixed signal electronic circuits. The recent dramatic advances with respect to the processing of deep submicron MOS technology transistors render transparent the feasibility of abandoning the traditional wisdom of operating MOS transistors in analog circuits in exclusively saturated domains. Instead, saturation might be supplanted in all but extremely high frequency signal processing applications with operation in triode and even weak inversion regimes.

In order to identify the general types of applications for which acceptable levels of performance can be attained in low power operating regimes, and in order to contrive a practical design strategy for relevant circuit development and implementation, several research and development tasks must be undertaken. These tasks are codified in this report as modeling, circuit design theory and practice, and circuit compensation. It is anticipated that the requisite modeling can build on the EKV Compact MOSFET Model and its companion disclosures, subject to the caveat that any new modeling and device parameterization scenarios be synergistic with extant design trends and practices. The circuit design effort fundamentally entails a definitive investigation of novel circuit topologies capable of delivering high performance levels, while remaining amenable to low voltage operation. It is expected that several of these new topologies can derive from standard cell networks that can be modified for optimal performance at voltage levels that are lower than those recommended in the archival literature. Entirely new circuit topologies can also be reasonably projected. In the compensation phase of the project, adaptive techniques for adjusting power levels in nominal proportion to applied signal stresses are underscored. The text herewith presents a viable technique for weakly inverted transistors, but modified versions of this adaptation approach are envisaged for both the triode and saturation domains.

The traditional integrated circuit design project begins with system definition and optimization, followed by the development of circuit topologies that deliver the performance highlighted by the individual system blocks. Once the basic circuit cells are designed, a process foundry is identified and the challenging task of effectively fitting the process to the circuit requirements commences. Since processes can rarely be modified, this task reduces to circuit modification and inevitable design tradeoffs. The ultimately envisaged circuit design task is expected to embrace semiconductor device design and optimization as an implicit component of the topological development task; that is, circuit design might ultimately merge with device design and definition. While the strategy for effecting this new generation design philosophy and technique remains a work in progress, it may very well discard the body of modeling work currently used in favor of models that implicitly incorporate device physical considerations into the network design procedure. In view of the continuing shrinkage of device feature sizes, these physical considerations are likely to entail a view of semiconductor properties from the perspectives of quantum physics\textsuperscript{[42]}. The work overviewed by this report is expected to serve as a catalyst for this genuinely challenging, but ultimately essential and pragmatic, endeavor.

4.0. REFERENCES


