Experimental Verification of a ±10% Accurate 1–Dimensional Output Resistance Model for Long Channel Silicon and SiC MOSFETs in Saturation Suitable for Hand Analysis in CMOS IC's

F. S. Shoucair

***Abstract* — Accurate (±5%) quantitative predictions for the output resistance of MOSFETs operating in their saturation region have long been associated with 2-dimensional (2D) modeling of electric fields near the drain junction. The complexities of such analytical and numerical models tend not only to hinder intuitive appreciation of the dominant effects determining the behavior of the output resistance, but to render them unsuitable for hand analysis, and even for efficient computer-aided design, as well. On the other hand, extant simplified 1-dimensional (1D) models approximating channel shortening by the length of the drain's depletion region, (a) typically underestimate the output resistance by 50%, (b) fail to account for non-negligible dynamic range degradation, and (c) predict substrate voltage effects qualitatively opposite to observations. These shortcomings are suggestive of underlying conceptual wants.**

We elaborate a 1D model of the pinchoff region whose extent $\ell'(Q_i, V_{DS}, V_{DSAT})$ is determined by *(i) an "internal" saturation voltage V_{DSAT} corresponding to the limit of laminar flow, the objective criterion which obtains as the surface potential transitions from strong to weak inversion* and, in view of current continuity, *(ii) the inversion charge Q_i injected into the pinchoff region*. This concept, underpinned by physical considerations and guided by observations, yields a continuous drain-source conductance through the triode, transition, and pinchoff regions. We develop a simple expression for the saturation region's incremental conductance:

$$\frac{g_{dpinchoff}}{I_{DSAT}} \approx \left(\frac{1}{V_{DS}}\right)\frac{\ell'(Q_i, V_{DS}, V_{DSAT})}{2L} \approx \frac{constant}{L\sqrt{V_{DS}}}$$

whose dominant underlying effects conforming to the EKV model are explicit, and which is suitable for hand analysis and design of well-tempered contemporary circuits down to the limit of planar CMOS (about 20nm). Its accuracy, which is close to modern PSP-based models', is within experimental and typical process parameter uncertainty (±10%) for silicon and silicon carbide MOSFETs, to whose characteristics we compare it critically.

***Index Terms* — MOSFET, Silicon MOSFET, Laminar flow approximation, Output Resistance, Output Conductance, Quantum–mechanical effects, Drain-source Saturation voltage, Silicon Carbide, Velocity saturation, Circuit design.**

"...The numerical interpretation of the calculation results is nevertheless necessary...as long as it is not obtained, the solutions can be considered incomplete and useless, and the truth that is to be discovered remains just as hidden in the formulas of the analysis as it was in the physical problem itself."

J.B.J. Fourier, 1822 [1]

I. INTRODUCTION

The detailed analysis published by Merckel *et al* in 1972 [2] elaborates the influences of drain diffusion junction depth and curvature, substrate doping concentration, channel carriers' density and mobility, in addition to those of velocity saturation and critical electric field on the output characteristics of long p-channel silicon MOSFETs. Their model, intended for large-signal transistor modeling in computer-aided design, achieved a fit to experimental observations "near 5 percent for a wide range of gate and drain voltages and channel lengths (3.7 to 20.9 μm)". They observed, notably, that the output impedance in saturation would be increased by means of ion implantation to fabricate shallower junctions. Hwang and Yu [3], who subsequently (1987) extended the analysis to Lightly Doped Drain (LDD) n-channel MOSFETs with lengths between 0.68 and 4.68 μm, obtained comparable results, citing "the global error is 5 percent, only 3 percent for L=0.68 μm", as well as "lower channel length modulation effects" among other improvements for down-scaled devices. The analytical models in both [2] and [3] consider the ratio (z/x_j) of the thickness of the inversion layer (z) to drain and/or spacer implant junction depth (x_j) as a parameter influencing the output resistance in saturation. Despite having assumed z ≈ 100Å, thereby likely overestimating the ratios for (z/x_j), other parameters of their models were adjusted plausibly to achieve the overall accuracies which they reported. As

junction depths (x_j) decrease with down-scaling in more modern devices, while the minimum inversion layer thickness does not, — $z_{min} \approx 10$Å near the transition from laminar (triode region) to non-laminar flow (saturation), as dictated by quantum-mechanical considerations — the larger ratios (z/x_j) tend to favor higher output resistances than in earlier process generations, *ceteris paribus*. Unfortunately, extensive data such as provided in [2] is largely absent from the subsequent published literature, ostensibly for being regarded as proprietary by process manufacturers.

Our consideration of the role of energy in this context was suggested by a detailed analysis of the MOSFET's characteristics below saturation (low V_{DS}), whereby we showed that *laminar flow* corresponds to inversion layer carrier mobility μ varying as a power of the transverse field (E_T), $\mu \sim (E_T)^{-1/3}$, in consequence of *energy* and *momentum* conservation under quantum–mechanical confinement, and that such laminar flow is limited to a relatively narrow span of E_T below $E_{Tmax|laminar}(A_o,T)$, which is predictable solely in terms of fundamental constants, temperature (T), and of the semiconductor–insulator interface terrain 'roughness' amplitude A_o [4]. The pertinence of these findings herein concerns the aforementioned inversion layer thickness z_{min}, and the mobility of channel carriers (hence its velocity-field relationships, transverse and longitudinal), whose accurate modeling in the triode region is essential to large signal modeling, as emphasized in [2], and which we discuss in Appendix **2**. These, and other, higher-order effects are critical to the design of modern analog and RF circuits in the range below the 100nm "node", and down to the limit of planar CMOS processes (about 20nm); they are accounted for in surface-potential-based (PSP) models [14] underpinning the "compact modeling" paradigm of contemporary MOSFETs as implemented in most major circuit simulators (but otherwise transparent to the user). Although the PSP models are more accurate than their classical V_{TH}-based forerunners familiar to circuit designers, the latter remain in effect in the global, industry standard, BSIM4 [11], whose latest version has been "fine-tuned" in compliance with compact modeling imperatives.

While our development straddles the boundary between the two methods, as did the well-known EKV model [15], our results pertain to any standard planar/bulk MOSFET structure in principle, silicon or otherwise. Although our treatment is valid for electrons and holes, we refer explicitly to electrons as the "carriers" (n–channel MOSFETs) in the sequel, which is organized as follows:

Section II: Setting the stage: A mostly qualitative discussion of pertinent issues in a 1-D model.

- **A.** Relations between depletion region length, drain current, and incremental output resistance.
- **B.** Why channel current injected into the drain's depletion ("pinchoff") region reduces the length of this region, and estimate of reduction thereof.
- **C.** Velocity saturation's influence on pinchoff region length.

Section III: Output Conductance in saturation.

- **A**. Sources of error in the classical 1D model.
- **B**. Modified 1D model of the pinchoff region.
- **C.** Body effect.
- **D.** Overall characteristics and observations.
- **E.** Circuit design considerations.

Section IV: Summary
Section V: References
Section VI: Appendixes

II. SETTING THE STAGE: A MOSTLY QUALITATIVE DISCUSSION OF PERTINENT ISSUES IN A 1-D MODEL

A. Relations between depletion region length, drain current, and incremental output resistance.

Figure 1 illustrates various features and elements introduced in this section and further referred to throughout the text.

Double integration of Poisson's equation in one dimension in material of dielectric constant ε with a charge density profile $\rho(x)$:

$$\frac{\partial^2 V}{\partial x^2} = -\frac{\rho(x)}{\varepsilon} \tag{1}$$

applied to an abrupt n^+p silicon junction reverse-biased by a potential difference V_R readily yields the length $\ell(V_R)$ of the resulting depletion region extending on the p-side of the junction whose uniform volume doping concentration is N_a [5]:

$$\ell(V_R) = \sqrt{\frac{2\varepsilon_{si}}{qN_a}(V_R + V_{bi})} \tag{2}$$

with q and ε_{si} the electronic charge, and the dielectric constant of silicon, respectively, $V_{bi} \approx 2\phi_F \approx 2\phi_t \ln(N_a/n_i)$ the built-in potential, and $\phi_t = (kT/q)$ the thermal voltage.

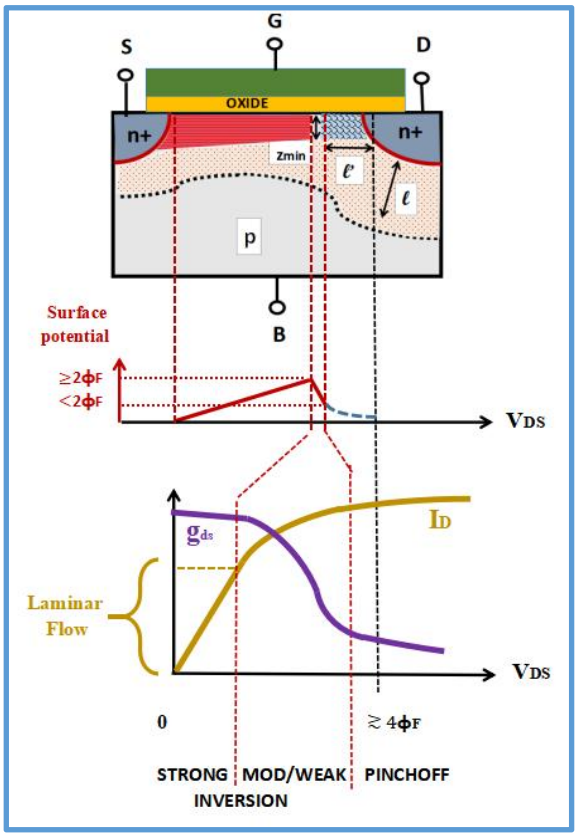


Fig. 1: Cross section of a MOSFET biased in the saturation region, with $V_S = V_B = 0$ and uniform substrate doping concentration. The inversion layer's thickness (not to scale) attains a minimum value $z_{min} \approx 10$Å at 300K, where laminar flow ends (vertical arrow). A "pinchoff" region of extent $\ell' \leq \ell$ begins to form where the surface potential falls below strong inversion ($\lesssim 2\phi_F$), and is "fully formed" when $V_{DS} \gtrsim 4\phi_F$ (see section II.*B*.). ℓ is the depletion region's maximum extent away from the oxide interface, and is typically 10 to 100 times greater than z_{min}. The lower diagram depicting I_D and g_{ds} is representative of typical observations (see Figs. 7a,b).

In a well "tempered" long channel MOSFET[1], the effect of $\ell(V_{DS})$ on the drain current in the saturation region of operation is conventionally expressed as:

$$I_D(V_{DS} \geq V_{DSAT}) \geq \frac{I_{DSAT}}{1 - \frac{\ell(V_{DS})}{L}}$$

$$\approx I_{DSAT}\left[1 + \frac{\ell(V_{DS})}{L}\right] \qquad (3)$$

the latter approximation being valid when $\ell(V_{DS}) << L$, the nominal drawn length between the source and drain diffusions, with the source potential taken as reference. The value of V_{DSAT} in Eq.(3) is a potential near to the value obtained by solving the "3/2 power law" I_D characteristics for the value of V_{DS} at which $dI_D/dV_{DS} = 0$, (also see footnote 3 below, and Appendix **2**). This formulation for the saturation region's drain current models a finite, incremental, drain-source "output" resistance r_{ds} or, equivalently, an incremental conductance $g_{ds} = (1/r_{ds})$ valid in the range of $V_{DS} \geq V_{DSAT}$:

$$g_{ds}(I_D, V_{DS}) \equiv \frac{dI_D}{dV_{DS}} \approx \left(\frac{I_{DSAT}}{L}\right)\frac{d\ell(V_{DS})}{dV_{DS}} \qquad (4)$$

which is smaller than that of the linear (or triode) region of operation, typically by a factor in the range of 10 to 100 for devices of practical dimensions. Since, moreover, the analog voltage gain of MOSFET amplifier stages varies directly as r_{ds}, or products of combinations thereof, the more accurately r_{ds} can be predicted, the better, in turn, is the gain. While accurate values for I_{DSAT} and V_{DSAT} are important to the modeling of the overall (I_D.vs.V_{DS}) characteristics, these large signal quantities are readily estimated by methods set forth in the standard literature (e.g. [6, p.175]). While, as such, we regard their values in Eq.(3) above as known for purposes of the present qualitative discussion, we will further examine the "saturation" current(s) I_{DSAT} in section III.***C.a)*** below, and in Appendix **2**. In the following sub-section (***B***. below), we justify re-locating the aforementioned parameter V_{DSAT} near to the bottom of the knee of the (I_D.vs.V_{DS}) characteristics, that is, *we will adduce evidence for regarding* $V_{DSAT} \approx 2\phi_F$ *as the maximum drain-source potential consistent with laminar flow, or strong inversion,* along the inversion layer up to the (physical) point at which the pinchoff region begins to form. This objective criterion, which is determined by the onset (or lack thereof) of a physical phenomenon, obviates the need for "adjusting" V_{DSAT} so as best to fit experimental observations [6, p.172] absent a better method. Moreover, this criterion has otherwise proven fruitful towards elucidating the physical origins of universal mobility in MOSFET inversion layers [4].

B. Why channel current injected into the drain's depletion ("pinchoff") region reduces the length of this region, and estimate of reduction thereof.

Qualitatively, if the quantity qN_a in Eq.(2) for $\ell(V_R)$ above represents the *fixed*, negative (in p

[1] Free of Drain Induced Barrier Lowering (DIBL), junction and oxide breakdown, hot electron injection, substrate punchthrough, and other detrimental and parasitic effects over its useful range of operation.

type bulk), charge in the pinchoff region of a MOSFET biased in saturation, the additional *mobile* charge ($|Q_i/z|$) emanating from the (electron) conducting channel and flowing towards the drain junction across the pinchoff region, modifies Eq.(2), as follows:

$$\ell'(V_{DS}) = \sqrt{\frac{2\varepsilon_{si}}{\frac{Qi}{z} + qN_a}\,(V_{DS} - V_{DSAT} + V'_{bi})} \qquad (5)$$

wherein $(V_{DS} - V_{DSAT}) \geq 0$ is the reverse-bias across the (n+) drain to p (pinchoff region) junction, with V_{DSAT} here an "*internal*" *saturation voltage* to be defined below, and *different from that described in section A. above,* such that $\ell'(V_{DS}) \leq \ell(V_{DS})$ for any non-zero value of $|Q_i/z|$, z being the finite thickness of the inversion layer (further discussed below), and Q_i the inversion layer's charge per unit area. This effect is similar to the well-known Kirk effect [12] associated with base-spreading of the neutral base layer into a p-type collector region of bipolar transistors, which occurs when the mobile carrier space-charge density in the collector transition region (here Q_i/z) becomes approximately equal to qN_a, the fixed impurity charge density of the high resistivity side of the collector junction, thereby shrinking the extent of the collector depletion region.

Furthermore, however, this is not an "ordinary" pn junction: it comes into existence, physically, i.e. $\ell'(V_{DS}) > 0$, only for $V_{DS} > V_{DSAT}$; its built-in potential V'_{bi} in Eq.(5) above must therefore vanish when the p side of the junction does so, i.e. at $V_{DS}=V_{DSAT}$, yet "appear" with the correct value in the expression for $\ell'(V_{DS})$, as it does for an ordinary reverse-biased junction per Eq.(2) with $Q_i = 0$, for V_{DS} sufficiently greater than "*internal*" saturation potential V_{DSAT}, i.e. for $V_{DS} \gtrsim (V_{DSAT} + V_{bi})$, with $V_{bi} \approx 2\phi_F$ when the uniformly doped p side extension of this junction may be considered fully formed[2]. These three boundary conditions:

$$\begin{cases} \ell'(V_{DS} \approx V_{DSAT}) = 0 \\ V'_{bi}(V_{DSAT}) = 0 \\ V'_{bi}(V_{DSAT} + 2\phi_F) \approx V_{bi} \end{cases}$$

may be met simultaneously by a simple expression of the following form, or by a piece-wise linear approximation thereof for better computational efficiency:

$$V'_{bi}(V_{DS}) = V_{bi}\left[1 - e^{-\beta\left(\frac{V_{DS} - V_{DSAT}}{2\phi_F}\right)}\right]$$

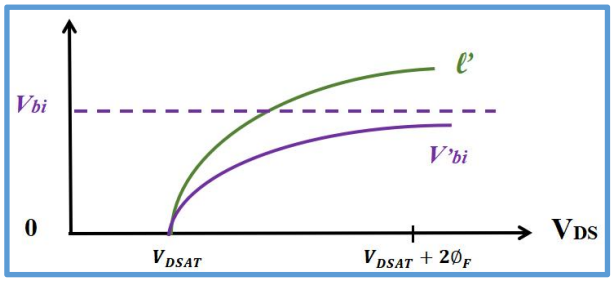


Fig. 2: **Model for the gradual "build-up" of the 1D pinchoff region's depletion region length $\ell'(V_{DSAT}, V'_{bi})$.**

as illustrated in Fig.2 above, which ensures $V'_{bi}(V_{DSAT}) = 0$, and such that

$$V'_{bi}[V_{DS} \geq V_{DSAT} + 2\phi_F] = V_{bi}(1 - e^{-\beta}) \leq V_{bi}$$

attains a value as close as desired to V_{bi} by a suitable choice of a scalar β, e.g. $V'_{bi}(V_{DSAT} + 2\phi_F) \approx 0.9\,V_{bi}$, with $\beta = -\ln(0.1) \approx 2.3$. Then, $\ell'(V_{DS})$ may be written as follows for $V_{DS} \geq V_{DSAT}$:

$$\ell'(V_{DS}) \approx \sqrt{\frac{2\varepsilon_{si}}{\frac{Qi}{z} + qN_a}\left\{(V_{DS} - V_{DSAT}) + V_{bi}\left[1 - e^{-\beta\left(\frac{V_{DS} - V_{DSAT}}{2\phi_F}\right)}\right]\right\}} \qquad (6)$$

This model reasonably neglects a very small potential corresponding to the lateral electric field at the source end of the pinchoff region (of the order of mV or less for long channel devices), which results from a formal integration of Poisson's equation, because the model for the output resistance to be developed sidesteps the problem of an infinite slope for dI_D/dV_{DS} as V_{DS} approaches (external) V_{DSAT} [6, p.172], as will be elaborated in section III below. The complication of including the additional charge (Q_i/z) in the formulation of V'_{bi} above is also neglected because it is an effect of the second order. The validity of this gradual build-up of V'_{bi} and $\ell'(V_{DS})$ is demonstrated in Fig.6 below, wherein it is compared to the classical model with a fixed value for $V'_{bi} = V_{bi}$: the gradual model accounts for the transition in the (g_{ds}/I_D) observations, whereas the fixed-value model not only does not, but also

[2] That is to say: if X_n is the extent of the depletion region on the n+ side of the drain's diffusion junction and N_d the doping concentration thereof: $N_a \times \ell'(V_{DS}) \gtrsim N_d \times X_n$, which is the (fixed) charge neutrality condition for the reverse-biased junction.

overestimates the dynamic range for high incremental output resistance $r_{ds} = 1/g_{ds}$ as well.

Moreover, we have observed experimentally [4, Fig.3], that ***under laminar flow conditions, at low V_{DS}, in the triode region and in strong inversion, the mean potential across the inversion layer ranges between $1.45\phi_t$ and $1.65\phi_t$ as*** $\mathbf{0.4 \leq \left| Q_i/Q_b \right| \leq 1.8}$, hence as the ***surface potential*** ranges correspondingly between approximately $[V_{SB}+2\phi_F+(3x1.45\phi_t)]$ and $[V_{SB}+2\phi_F+(4x1.65\phi_t)]$, as indicated in Fig.3 below.

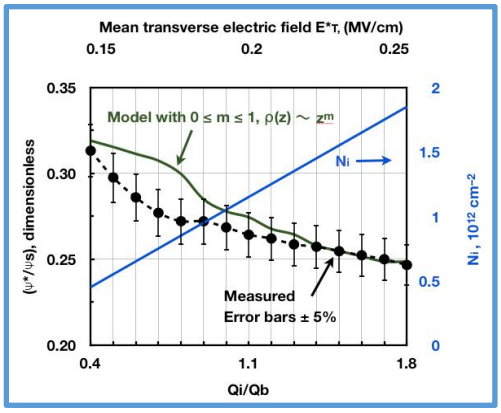


Fig. 3: **Observed ratio of mean to surface potential (ψ^*/ψ_s) under laminar flow conditions, and corresponding inversion layer carrier concentration of commensurate charge sheet model ($N_i = Q_i/q$), both as functions of the ratio (Q_i/Q_b), i.e. with increasing mean transverse electric field E^*_T. The range of observations is from onset of strong inversion through saturation of the surface potential ψ_s (also see Fig.4 in [4] for further details).**

Under these conditions, ***the inversion layer of finite depth z reaches a minimum extent $z_{min} \approx 10$Å at 300K*** (see Appendix **1**) at the drain end of the gradual channel. As the surface potential and the drain current saturate with increasing V_{DS}, the local ratio $\left| Q_i/Q_b \right|$ decreases as the surface potential at the drain end of the channel transitions from strong to weak inversion; that is, as the surface potential falls below $2\phi_F$, as indicated in the middle diagram of Fig.1 above.

The approximate corresponding value of (internal) V_{DSAT} required for this transition from triode — herein understood as corresponding to the laminar flow region — to the onset of saturation is thus:

$$V_{DSAT} \approx 2\phi_F$$

which is nearly independent of V_{SB}, a subtle and key point in this context, because Q_i varies as e^{-V_{SB}/ϕ_t} under weak inversion conditions, as emphasized in [6, p.84], and in agreement with experimental observations [9]. This is the minimum drain voltage which drives the *drain end* of the channel into weak inversion [6, pp.111-113], thereby causing a "pinchoff" region to form, wherein current flow is no longer laminar[3] (see [4], section II.B. for a more extensive discussion). An increase in V_{GS} (increasing Q_i and V_{DSAT} via the ϕ_t term) tends to decrease $\ell'(V_{DS})$, thereby decreasing the output resistance in saturation, as evident in [2, Figs. 7-10], and in [9, Figs. 2,3].

The ratio (ℓ'/ℓ) may thus be estimated using Eqs.(5) and (2) with this value for V_{DSAT}:

$$\frac{\ell'(V_{DS})}{\ell(V_{DS})} \lesssim \sqrt{\frac{\frac{1}{1.4}\left\{(V_{DS}-V_{DSAT}) + V_{bi}\left[1 - e^{-\beta\left(\frac{V_{DS}-2\phi_F}{2\phi_F}\right)}\right]\right\}}{(V_{DS}+V_{SB}+V_{bi})}} \qquad (7)$$

whose value ranges between 0 and 0.79 for $V_{DSAT} \approx 0.7\text{V} \leq V_{DS} \leq 5\text{V}$ with $V_{SB} = 0$ and $V_{bi} \approx 2\phi_F \approx 0.7\text{V}$. This result is made plausible by noting that, whereas the reverse-bias across the drain junction away from the oxide surface is $(V_{DS} + V_{SB})$, the reverse bias across the pinchoff region is the lower value $(V_{DS} - V_{DSAT})$, with $V_{DSAT} \approx 2\phi_F$. The factor (1/1.4) under the radical in Eq.(7) corresponds to $(Q_i/z) \approx 0.4(qN_a)$ in Eq.(6), which is the approximate charge injected into the pinchoff region under laminar flow conditions near to the transition to weak inversion, as shown in Fig.3 above. Since, moreover, $V_{bi}' \leq V_{bi}$ was established for Eq.(6) above, $(\ell'/\ell) \leq 1$ even in the absence of the additional charge term (Q_i/z).

[3] This value for V_{DSAT} is typically lower than the well-known, classical, long channel limiting value, $V_{DSAT} \approx [V_{GS} - V_T(V_{SB})]$, ([5], p.442, [6], p.175), which is obtained by solving the "3/2 power law" I_D characteristics for the value of V_{DS} at which $dI_D/dV_{DS} = 0$, and which corresponds to the (nonphysical) condition $Qi=0$ (further discussed in subsection III.***A***. below). In down-scaled, low voltage, devices, and/or those with channel implants, the values of the surface potential corresponding to various levels of inversion will be different [13], (also see [6], chapter 6).

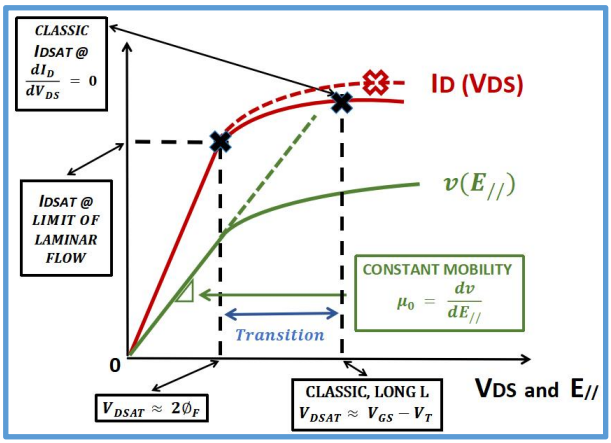


Fig. 4: **Drain current and mobility (velocity vs. longitudinal field, $E_{//}$) characteristics. The internal saturation voltage V_{DSAT} corresponds to the upper limit of the range of laminar flow. The long channel, "classic", limiting value for (external) $V_{DSAT} \approx (V_{GS} - V_T)$ corresponding to the condition $dI_D/dV_{DS} = 0$ neglects the transition region required to fully "build" ℓ, thereby overestimating the dynamic range for high r_{ds} (see Fig.6 for illustration).**

The foregoing qualitative argument as to why $\ell'(V_{DS}) \leq \ell(V_{DS})$ relies only on classical considerations of 1D potential conditions. The role of the gate voltage near the drain end of the channel (i.e. 2D field effects) is implicit, insofar as it corresponds to (i.e. brings about) weak inversion at the source side end of the pinchoff region, as elaborated above. We next consider the role of *energy* on the extent of the pinchoff region $\ell'(V_{DS})$.

C. Velocity saturation's influence on pinchoff region length.

In the previous section, we alluded to the saturation of the mean potential across the inversion layer under laminar flow conditions ($V_{DS} \leq V_{DSAT}$) and $z \approx z_{min}$. As a pinchoff region begins to form (when $V_{DS} \geq V_{DSAT}$), current continuity requires that the average velocity of carriers emanating from the inversion layer gradually increase from v_{drift} towards its upper limiting value v_{sat} in the pinchoff region (here p-type silicon), whether the flow therein is laminar or otherwise in our 1D model. A maximum energy increase $(1/2)m(v_{sat}^2 - v_{drift}^2) \approx (1/2)mv_{sat}^2$, since $v_{sat} >> v_{drift}$, must therefore be expended, hence subtracted from the total available energy $q(V_{DS}–V_{DSAT})$ sustaining the pinchoff region, (m being the electronic mass).

As the potential corresponding to this energy, $(1/2q)mv_{sat}^2 \approx 0.028$V (28mV) at 300K, its relative influence will be greatest (about 4%), near to the onset of saturation, or when $V_{DS} \approx V_{DSAT}$, thereby reducing $\ell'(V_{DS})$ by a maximum of about 2%, so that the estimate rendered by Eq.(7) above, $0 \leq (\ell'/\ell) \leq 0.79$ remains valid given the otherwise greater tolerance range of process, geometrical, and electrical parameters. This range for (ℓ'/ℓ) is, moreover, in good quantitative agreement with that obtained by Merckel *et al* [2, Figs. 11,12] by means of their "simplified 2–dimensional analysis".

While the saturation velocity is a function of temperature, it is relatively constant, $v_{sat} \approx 10^7$ cm/s below 400K (see Appendix **1**.). The observed velocity–longitudinal field ($E_{//}$) relationship in bulk silicon is given in [5, p.46] in graphical form, a useful approximate analytical form of which is[4]:

$$v \approx v_{sat} \frac{(E_{//}/E_C)}{\sqrt{1+(E_{//}/E_C)^2}} \approx \frac{\mu_{eff} E_{//}}{\sqrt{1+(E_{//}/E_C)^2}} \qquad (8)$$

wherein,

$$E_{//} = \frac{(V_{DS} - V_{DSAT})}{\ell'(V_{DS})} \qquad (9)$$

and μ_{eff} is the low field, "effective" mobility. Unfortunately, Ec is an ill-defined "critical" field which is typically chosen, and/or adjusted, to fit observations. Thus Ec was "taken as" $3x10^4$V/cm for n type silicon in [2], whereas its value was estimated at $2.2x10^4$V/cm, then revised to $2x10^4$V/cm "after adjustments" for p type silicon in [3, Table II]. In the industry standard simulation program BSIM4 [11], Ec is evaluated as $(2v_{sat}/\mu_{eff})$, so that the function $v(E_{//})$ used therein (see footnote 4) be continuous at $E_{//} = E_C$. In PSP-based models, on the other hand, a continuous surface potential is evaluated along the channel and along the overlap regions, which includes velocity saturation and quantum-mechanical effects, thereby achieving high accuracy for the output conductance and for higher order conductances and transconductances, all of which are essential to the evaluation of distortion components in RF circuits [14].

III. OUTPUT CONDUCTANCE IN SATURATION

A. Sources of error in the classical 1D model.

The classical value for the saturation voltage in the long channel limit (infinite "Early" voltage):

---

[4] BSIM4 uses the computationally simpler,
$v = \frac{\mu_{eff} E_{//}}{1+(E_{//}/E_C)}$ for $E_{//} < Ec$, and
$v = v_{sat} = 8 \times 10^6\ cm/s$ for $E_{//} \geq Ec$.

$$V_{DSAT} \approx [\, V_{GS} - V_T(V_{SB}) \,]$$

a) Implicitly combines two incorrect, and mutually inconsistent, assumptions:

(i) $Q_i = 0$ at the source side end of the pinchoff region and,
(ii) strong inversion at that point.

This value for V_{DSAT} is typically *larger* than that which corresponds to the upper limit of laminar flow by an amount $\geq \phi_F$, which is the approximate span of the weak and moderate inversion regions for the surface potential [6, p.97], [15]. The corresponding drain current I_{DSAT} is thereby overestimated for typical process parameters and bias conditions, as indicated in sub-section ***D. a)*** below, and in Appendix **2**.

b) Overestimates the ratio ($g_{dpinchoff}/I_D$) by evaluating an expression of the form:

$$\frac{g_{dpinchoff}}{I_D} \approx \frac{1}{2}\left[\frac{\ell(V_{DS})}{L}\right]\frac{\left[1-\frac{\ell(V_{DS})}{L}\right]}{V_{DS}-[\,(V_{GS}-V_T(V_{SB})\,]} \qquad (10)$$

which exceeds observations, typically by 30% or more. Moreover, worse yet, this function is rendered discontinuously by certain simulation programs (though not by current versions of BSIM4 [11]) at the transition from triode to saturation (see Fig.6), because V_{DSAT} is taken to correspond to the drain voltage at which $dI_D/dV_{DS} = 0$; i.e. to the aforementioned (invalid) premise $Q_i = 0$.

c) predicts the wrong qualitative effect (opposite of observations) of V_{SB} on the ratio ($g_{dpinchoff}/I_D$), as further elaborated in subsection ***C.*** below.

B. Modified 1D model of the pinchoff region.

a) $g_{dpinchoff}$

Using Eqs.(4) and (6), and for $V_{DS} > V_{DSAT}$, the ratio[5] ($g_{dpinchoff}/I_D$), a desirable quantity for biasing analog gain stages in the saturation region, is then:

$$\frac{g_{dpinchoff}}{I_D} \approx \frac{1}{2}\left[\frac{\ell'(V_{DS})}{L}\right]\frac{\left[1-\frac{\ell'(V_{DS})}{L}\right]}{(V_{DS}-V_{DSAT})+V_{bi}\left[1-e^{-\beta\left(\frac{V_{DS}-V_{DSAT}}{2\phi_F}\right)}\right]} \qquad (11)$$

which is at a maximum near to $V_{DS} \approx V_{DSAT}$, where the transition from triode to pinchoff begins, hence where the incremental conductance

$$g_{dpinchoff} = (1/r_{dpinchoff}) \approx g_{dtriode}$$

remains relatively high as long as $V_{DSAT} \lesssim V_{DS} \lesssim V_{DSAT} + 2\phi_F$ (see subsection ***c*** below). For $V_{DS} \gtrsim V_{DSAT} + 2\phi_F$, i.e. once $\ell'(V_{DS})$ is fully formed, such that $V'_{bi} \approx V_{bi}$ (see section II.***B.***), and as long as $\ell'(V_{DS})/L \ll 1$, Eq.(11) simplifies to:

$$\frac{g_{dpinchoff}}{I_{DSAT}} \approx \frac{1}{2}\left[\frac{\ell'(V_{DS})}{L}\right]\frac{1}{V_{DS}}$$

$$\approx \left(\frac{1}{2L}\right)\frac{\sqrt{\frac{2\varepsilon_{si}}{1.4(qN_a)}}}{\sqrt{V_{DS}}} \approx \frac{constant}{L\sqrt{V_{DS}}}$$

which is in agreement with observations to within ±10% (see Fig. 6 below), and whose inverse is typically approximated by, and specified as, a fixed "Early" voltage by circuit designers, by analogy with bipolar transistors [6, pp.169-175], [11].

b) **Two drain voltages**, V_{DSX} and V_{DSY} are indicated in Figs.5,6, and 7b below, where the slopes of the "3/2 power law" for the drain current and $g_{dpinchoff}$ are nominally equal (see Appendix **2** for further details, and [15] where the concept is indicated), $g_{dtriode} = g_{dpinchoff}$ at:

$$V_{DSX} \approx V_{DSAT} \approx 2\phi_F$$

and at

$$V_{DSY} \approx \frac{\gamma^2}{4}\left[\sqrt{1+\frac{4\left\{(V_{GS}-V_T)\left[1-\left(\frac{\ell'}{4LR}\right)\right]+V_{DSAT}\right\}}{\gamma^2}}-1\right]^2 - V_{DSAT}$$

$$\lesssim V_{GS} - V_T(V_{SB})$$

[5] Wherein, for simplicity, we have omitted the relatively small potential (about 4%, hence comparable to process parameter spreads) owed to velocity saturation, as evaluated in sub-section II.***C*** above. The reported ±5% accuracy achieved with 2D simulations in [2,3] thus ostensibly implies accounting for this potential, which is straightforward in numerical models. In quasi-1D PSP-based compact models, including BSIM5 and BSIM6, it is built-into evaluations of the surface potential seamlessly, from which large and small signal quantities are derived.

where equality is approached in the long channel (low $E_{//}$) and/or low doping limits (low γ), with V_{DSY} corresponding to a drain current lower than that at which $dI_D/dV_{DS} = 0$. Where velocity saturation is not negligible, for shorter channel devices wherein (Lv_{sat}/μ_{eff}) is comparable to, or lower than, $(V_{GS} - V_T)$, V_{DSY} is reduced, the more so the shorter the channel, and approaches the limit (Lv_{sat}/μ_{eff}), as indicated in Appendix **2d**.

The industry standard BSIM4 [11], uses a similar, though more complicated, function for V_{DSX}, which extends to short channel devices, and in which "V_{gsteff}", defined as a "fitting parameter for moderate inversion condition" [sic], takes the place of V_{DSAT} above, and must be specified accordingly for reasonable results ("V_{gsteff}" otherwise defaulting to 0, which is tantamount to setting $Q_i = 0$ in Eq.(6) for $\ell'(V_{DS})$, a condition which we compare to observations in Fig.6 below). BSIM4 implements a different model for V_{DSY} as well, which requires the specification of a suitable "Early" voltage. As noted in footnote 5 above, BSIM5 and later versions thereof no longer use the classical V_{TH}-based model.

c) **The overall drain-source incremental conductance g_{ds} for positive values of V_{DS}:**

$$g_{ds} = \frac{g_{dtriode} \times g_{dpinchoff}}{g_{dtriode} + g_{dpinchoff}} \qquad (12)$$

is continuous across the transition from triode to saturation, because $g_{dpinchoff}$ in our model matches, and combines with well defined, *finite,* values for $g_{dtriode}$ at $V_{DS} = V_{DSX}$, and for its extension at $V_{DS} = V_{DSY}$. This is in contrast with the classical 1D model described in sub-section III.***A.*** above, where (an incorrect) $g_{dpinchoff}$ is "matched" near to, or at, $dI_D/dV_{DS} = 0$, depending on whether a finite, or an infinite, "Early" voltage is assumed.

The *overlapping* ranges of validity for these conductances being approximately:

$[0 \leq V_{DS} \lesssim V_{DSX}]$, for $g_{dtriode}$, and $V_{DS} \gtrsim V_{DSX}$, for $g_{dpinchoff}$ as indicated in Fig.5 above, Eq.(12) yields:

$$g_{ds} \approx \begin{cases} g_{dtriode} & \text{for } V_{DS} < V_{DSX} \\ (½)\, g_{dtriode} & \text{at } V_{DS} = V_{DSX} \\ g_{ds} & \text{per Eq.(12) for } V_{DSX} \lesssim V_{DS} \lesssim V_{DSY} \\ g_{dpinchoff} & \text{for } V_{DS} \gtrsim V_{DSY} \end{cases}$$

A comparison of Eq.(12) to observations on n-channel MOSFETs in a standard CMOS process for $1/7 \leq (\ell'/L) \leq 1/30$ is shown in Fig.6 below, which indicates agreement within ±10% with our measurements and with those of [2] and [14] under commensurate experimental conditions $(V_{GS} \approx 3V_T, V_{SB} = 0)$.

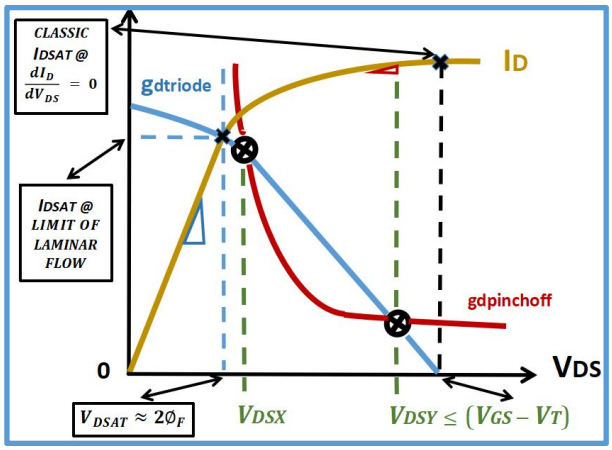


Fig. 5: Schematic representation for the combination of incremental conductances $g_{dtriode}$ (laminar range) and $g_{dpinchoff}$ (pinchoff region). These conductances are equal at V_{DSX} and at V_{DSY}, where circled crosses (⊗) are indicated. The long channel, classical, value for $V_{DSAT} \approx (V_{GS} - V_T)$ fails to account for the region corresponding to the transition of the surface potential at the drain end of the channel from strong to weak inversion and for the potential required to fully form ℓ' as modeled in section II.*B.* above.

We note, moreover, that the ***ratio*** (g_{ds}/I_{ds}) in saturation is largely independent of V_{GS} as it is of V_{SB} . This is readily made plausible by recalling that in saturation, $r_{ds} = (1/g_{ds}) = (r_{triode} + r_{pinchoff}) \approx r_{pinchoff}$ since $r_{pinchoff} >> r_{triode}$, hence variations of V_{GS} and/or V_{SB}, which influence the triode region's r_{triode} most strongly, are largely imperceptible in saturation, while $r_{pinchoff}$ is primarily determined by the ratio $[\ell'(V_{DS})/L]$ per Eq.(11), which is relatively insensitive to V_{GS} and V_{SB}.

C. Body effect.

Measurements of (g_{ds}/I_D) in saturation [9] indicate that this ratio is nearly unaffected by V_{SB}, ***in qualitative contradiction with the predictions of Eq.(10), and of simulation programs modeled thereon:*** Increasing V_{SB} increases V_R in the expression for $\ell(V_R)$ per Eq.(2), which is the dominant term of the numerator of Eq.(10), hence of

the overall ratio ($g_{dpinchoff}/I_D$) rendered thereby. In our modified expression, Eq.(11), by contrast, $\ell'(V_R)$ varies approximately as $\sqrt{V_{DS} - V_{DSAT} + V'_{bi}}$, per Eq.(5), which is relatively insensitive to V_{SB} when $V_{DS} \gtrsim V_{DSY}$.

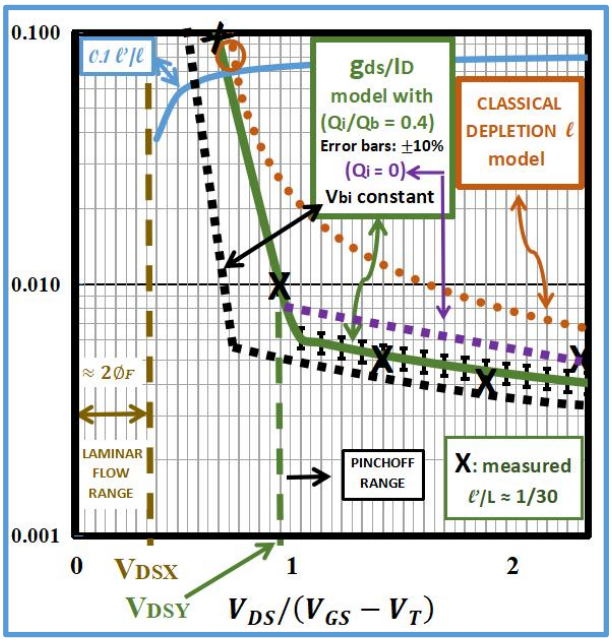


Fig. 6: Comparison of the ratio (g_{ds}/I_D) with g_{ds} per Eq.(12) to measurements on n-channel MOSFETs in a standard silicon CMOS process for $(\ell'/L) \approx 1/30$ and $\gamma \approx 1$; similar agreement is observed in the range of $1/7 \leq (\ell'/L) \leq 1/30$. The upturn at the far right of the data is owed to impact ionization of electrons which is not included in our simple model. The dashed red curve labeled "CLASSICAL DEPLETION ℓ", represents Eq.(10)'s predictions for ($g_{dpinchoff}/I_D$) which (a) is typically ≥ 30% higher than observed values, (b) increases with increasing V_{SB}, contrary to observations, and c) is discontinuous at the transition from triode to saturation (indicated by a circle). The dashed black curve indicated for our model corresponds to a constant, fixed, value for $V'_{bi} = V_{bi}$, in Eq.(6) with $Q_i/Q_b = 0.4$, instead of the gradually growing model thereof (section II.*B*. above), while the dashed purple curve shows the effect of ignoring the charge injected by the inversion layer (i.e. setting $Q_i = 0$ in Eq.(6) with gradual V'_{bi}): Only the full model, combining both effects, fits the data to within ±10%, as indicated by the error bars along the solid green curve.

D. Overall characteristics and observations.

Using the MOSFET's "3/2 power law" for the drain current, various values for, and relations between, bias and incremental quantities of interest may be evaluated (see Appendix **2**), which are indicated in Fig.5 above.

a) The ratio ***R*** of saturation drain current (at $dI_D/dV_{DS} = 0$) to the drain current at the limit of laminar flow, for $(V_{GS} - V_T) \gtrsim \gamma\sqrt{2\phi_F}$:

$$1 \leq \frac{I_{DSAT} @ (dI_D/dV_{DS} = 0)}{I_{DSAT} @ (V_{DSAT} \approx 2\phi_F)} \equiv \boldsymbol{R}$$

$$\lesssim \frac{V_{GS} - V_T}{2\phi_F} \frac{1}{\sqrt{1 + \left[\frac{(V_{GS} - V_T)}{LE_C}\right]^2}} \lesssim \frac{V_{GS} - V_T}{2\phi_F}$$

for typical values of $\gamma \gtrsim 0.55V^{1/2}$, $2\phi_F \approx 0.7V$, $E_C \lesssim 10V/\mu m$ (or $10^5 V/cm$), and mobility characteristics in the range $E_{//} \gtrsim E_C$. These currents are indicated by the smaller crosses, and labeled in Fig.5 above, which is representative of devices with channel lengths in the sub-100 nanometer range [13,14].

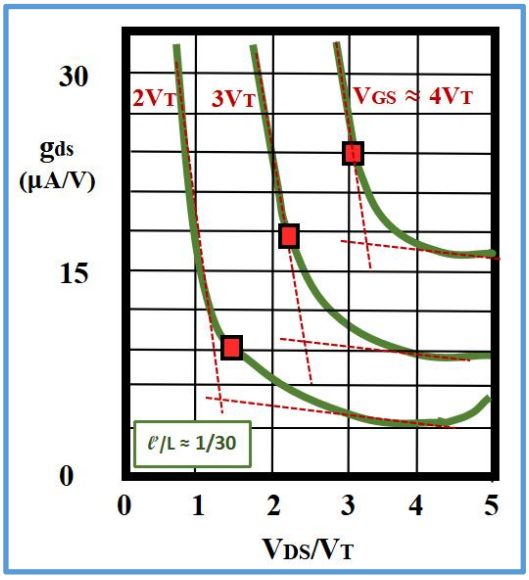


Fig.7a: Observations of g_{ds} for typical long n-channel Si MOSFETs with $(\ell'/L) \approx 1/30$, and $\gamma \approx 1$. The analytical locations of V_{DSY} are indicated by red squares, whereas the graphical estimates thereof are at the intersections of the dashed lines extending the corresponding triode and pinchoff regions' characteristics. The two values differ by less than 10% for all observations. The dynamic range (high r_{ds}) is approximately the span between V_{DSY} and the onset of avalanche multiplication at sufficiently high V_{DS}, which is visible only for the curve labeled ($V_{GS} \approx 2V_T$), near to $V_{DS}/V_T \approx 4$.

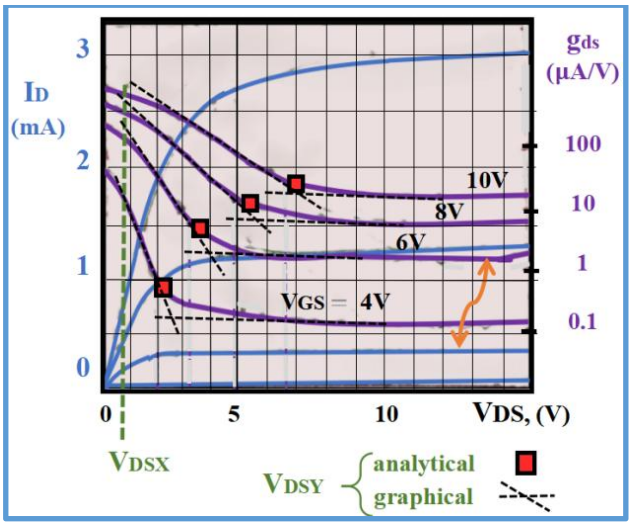


Fig.7b: Observations of I_D and g_{ds} for typical long n-channel SiC MOSFETs with $(\ell'/L) \approx 1/30$, $V_T \approx 2V$, and $\gamma \approx 1$. V_{DSX} and V_{DSY} are defined analytically and illustrated graphically (Fig. 5) in section III.B.*b)* above. To the g_{ds} curve flattening just above 1μA/V and corresponding to $I_{DSAT} \approx 0.25mA$ (orange double-headed arrow), the ratio $(g_{dpinchoff}/I_{DSAT}) \gtrsim 0.004$, and $V_{DSY}/(V_{GS} - V_T) \approx 3.7V/4V \approx 0.92$. This is nearest to the condition of the thick solid (green) curve for our 1D model in Fig. 6 for silicon, where the (g_{ds}/I_D) ratio is nearly identical (≈ 0.004), as expected in view of Eq.(11), whose corresponding "Early" voltage is the inverse ratio $(I_D/g_{ds}) \lesssim 250V$.

b) Analytical expressions for $g_{dtriode}$, $g_{dpinchoff}$, and the approximate potentials V_{DSX} and V_{DSY} at which $g_{dtriode} = g_{dpinchoff}$, as referred to in section III.**B.*b)*** above. Among these, for $(V_{GS} - V_T) \gg 2\phi_F$, and for $V_{DS} \gtrsim V_{DSY} \gtrsim 4\phi_F$, the ratio:

$$\frac{g_{dpinchoff}}{g_{dtriode}} \approx \left(\frac{2\phi_F}{L}\right)\left(\frac{\boldsymbol{R}}{2}\right)\frac{\sqrt{\dfrac{2\varepsilon_{si}}{1.4(qN_a)}}}{\sqrt{V_{DS}}}$$

derived in Appendix **2 c)**, relates the incremental conductances of interest to their respective bias currents as defined in ***a)***, with $\boldsymbol{R} \gtrsim 1$.

E. Circuit design considerations.

The magnitude $\boldsymbol{a_0}$ of the ratio:

$$\boldsymbol{a_0} = \left|\frac{g_m}{g_{ds}}\right| \approx \left|\frac{g_m}{g_{dpinchoff}}\right| \approx$$

$$\left(\frac{2I_D}{V_{GS} - V_T}\right)\left[\frac{2L/I_D}{\ell'(V_{DS})}(V_{DS} - V_{DSAT} + V'_{bi})\right]$$

$$\approx \left[\frac{4L}{\ell'(V_{DS})}\right]\frac{(V_{DS} - V_{DSAT} + V'_{bi})}{(V_{GS} - V_T)}$$

is the approximate low frequency incremental voltage gain of a MOSFET biased in saturation (unloaded), with $g_{dpinchoff}$ per Eq.(11), and $g_m = 2I_D/\big((V_{GS} - V_T)\big)$ its transconductance. The range of validity of this expression is $V_{DS} \gtrsim V_{DSY} \gtrsim V_{DSX} + 2\phi_F \gtrsim 4\phi_F$, as elaborated in section III.***B.b)*** above, which limits the dynamic range (for high r_{ds} or low g_{ds}) of the stage accordingly, V_{DSY} being an increasing function of $(V_{GS} - V_T)$, as shown in Appendix **2 b)**. Both dynamic range and $\boldsymbol{a_0}$ thus bear inverse relationships with gate overdrive.

Moreover, for given values of L and $(V_{GS} - V_T)$, the minimum dc gain:

$$\boldsymbol{a_0} \gtrsim \left[\frac{4L}{\ell'(V_{DS})}\right]\frac{(2\phi_F + V'_{bi})}{(V_{GS} - V_T)}$$

$$\approx \left[\frac{L}{\ell'(V_{DS})}\right]\frac{(16\phi_F)}{(V_{GS} - V_T)}$$

occurs at $(V_{DSY} - V_{DSAT}) \approx (4\phi_F - 2\phi_F)$, with $V'_{bi} \approx 2\phi_F$. Then, with $(L/\ell') \approx 30$, and $(V_{GS} - V_T) \gtrsim \gamma\sqrt{2\phi_F}$, we expect $a_0 \gtrsim 200$ for $V_{DS} \gtrsim 4\phi_F$, consistently with measurements on Si and SiC devices for which we observed $360 \lesssim \boldsymbol{a_0} \lesssim 430$ under the aforementioned conditions, and within the span of the dynamic range as otherwise limited by the onset of avalanche multiplication, which is typically more pronounced at comparable fields for n-channel devices in view of the relative ionization coefficients of electrons and holes in silicon [5, p.47]. While thermal self-heating occurs at high power dissipation, especially so in SiC devices, our simple 1-D model formulated under isothermal conditions neglects these effects, which degrade both mobility and velocity saturation.

IV. SUMMARY

We have elaborated a modified, one dimensional, model of the incremental output ("pinchoff") conductance for long channel MOSFETs biased in saturation. An essential modification entails the choice of an "internal" saturation voltage V_{DSAT} corresponding to the limit of laminar flow along the inversion channel, which (V_{DSAT}) is otherwise traditionally taken along the (I_D.vs.V_{DS}) drain current characteristics near to the condition $dI_D/dV_{DS} = 0$. A second, key, modification resulting from the first, entails accounting for the additional shortening of the length of the pinchoff's depletion region by the physical presence/injection of uninterrupted channel current into this region. These considerations lead to predictions of incremental conductances which are free of discontinuities throughout the (I_D.vs.V_{DS})

characteristics, and in agreement with observations to within the range of variations of process, geometrical, and electrical parameters.

V. REFERENCES

[1] J.B.J. Fourier, "Théorie Analytique de la Chaleur", Chapter I, Article 13, p.11, Paris, Firmin Didot Publishers, 1822. From French Original: ***"Cette interprétation numérique des résultats du calcul est cependant nécessaire … tant qu'elle n'est pas obtenue, les solutions peuvent être considérées comme incomplètes et inutiles, et la vérité qu'il s'agit de découvrir n'est pas moins cachée dans les formules de l'analyse qu'elle ne l'était dans le problème physique lui-même."***

[2] G. Merckel, J. Borel, and N.Z. Cupcea, "An Accurate Large-Signal MOS Transistor Model for Use in Computer-Aided Design", *IEEE Trans. Electron Devices*, vol. ED–19, no.5, pp. 681–690, May 1972.

[3] G-S. Huang and C-Y. Wu, "An Analytic Model for Lightly Doped Drain (LDD) MOSFET Devices", *IEEE Trans. Electron Devices*, vol. ED–34, no.6, pp. 1311–1321, June 1987.

[4] F. S. Shoucair, "Conservation Laws at Physical Origins of Universal Mobility in MOSFET Inversion Layers in Consilience with River Flow in a Gravitational Field", 2021: https://doi.org/10.36227/techrxiv.17088419.v2.

[5] S. M. Sze, *Physics of Semiconductor Devices,* John Wiley & Sons, 1981.

[6] Y.P. Tsividis, *Operation and Modeling of the MOS Transistor*, McGraw-Hill, 1987.

[7] H. D. Young, *University Physics*, Addison–Wesley, pp. 1140-1182, 1992.

[8] R. P. Feynman, R. B. Leyton, M. Sands, *The Feynman Lectures on Physics*, vol. III, *Quantum Mechanics*, Addison–Wesley publishing company, 1966.

[9] F. S. Shoucair , "A Semi-Empirical Model of the MOSFET's Small-Signal Drain Conductance in Saturation for Analog Circuit Design", *IEEE Trans. Electron Devices*, vol. ED–39, no.5, pp. 1246–1248, May 1992.

[10] ———— , "How the Supervention of Carrier Heating Degrades the Mobility Characteristics of MOSFET Inversion Layers", 2023: https://doi.org/10.36227/techrxiv.170244204.41850386/v1.

[11] BSIM 4.8.3 User's Manual, Copyright © University of California, May 2025.

[12] C. T. Kirk, J$_R$., "A Theory of Transistor Cutoff Frequency (f_T) Falloff at High Current Densities", *IRE Trans. Electron Devices*, vol. ED–9, pp. 164–174, March 1962.

[13] G.Curatola, G.Doornbos, Josine Loo, Youri V. Ponomarev, and G.Iannaccone, "Detailed Modeling of Sub-100-nm MOSFETs Based on Schrödinger DD Per Subband and Experiments and Evaluation of the Performance Gap to Ballistic Transport", *IEEE Trans. Electron Device*s, vol. ED-52, no. 8, pp. 1851-1858, August 2005.

[14] G. Gildenblat, X. Li, W. Wu, H. Wang, A. Jha, R. van Langevelde, G.D.J. Smit, A. J. Scholten, and D.B.M. Klaassen," PSP: An Advanced Surface-Potential-Based MOSFET Model for Circuit Simulation", *IEEE Trans. Electron Devices*, vol. ED–53, no.9, pp. 1979–1993, September 2006.

[15] C.C. Enz, F. Krummenacher, and E. A. Vittoz, " An Analytical MOS Transistor Model Valid in All Regions of Operation and dedicated to Low-Voltage and Low-Current Applications", *Analog Integrated Circuits and Signal Processing*, vol.8, pp. 83-114, 1995.

ACKNOWLEDGMENT

The author is indebted to R. Rice, and to Prof. S. Simic, Department of Mathematics and Statistics, San José State University, for valuable technical discussions.

VI. APPENDIXES

Appendix 1. **The saturation velocity** in a periodic silicon lattice is given by [5, p.46]:

$$v_{\text{sat}}(\text{T}) = \frac{2.4\times10^{7}\ \text{cm/s}}{1+0.8\,e^{\text{T}/600\text{K}}} \qquad \text{(A1)}$$

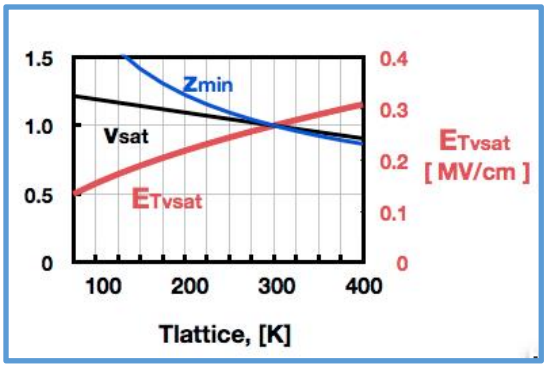


Fig. A1: E_{Tvsat}(T) per Eq.(7) in [10], v_{sat}(T) per Eq.(A1), and z_{min}(T) per Eq.(A2) for silicon, normalized to their respective values at 300K.

which is relatively constant, and near to 10^7cm/s below 400K, while z_{min}(T) = Δz is the minimum extent of the inversion layer allowed by the quantum–mechanical uncertainty relationship (Δp x Δz) ≥ $\hbar$, where $\hbar$ = (h/2π) is Planck's reduced constant, which is associated with the wave nature of electrons [7,8]:

$$z_{min}(\text{T}) \geq \frac{\hbar}{\sqrt{3\text{mkT}}} \qquad \text{(A2)}$$

and z_{min}(300K) ≈ 10Å. The maximum value of the transverse field for which electrons are in thermal equilibrium with the lattice E_{Tvsat}(T) varies approximately as $\sqrt{\text{T}}$ in the range of T ≤ 400K, and E_{Tvsat}(300K) ≈ 0.28 MV/cm.

The variations of v_{sat}, E_{Tvsat}, and z_{min} with temperature are shown in Fig.A1 (reproduced from [10]). The maximum value of E_T for which laminar flow is expected at a given value of T was derived in [4, Eq.(15)]:

$$\left|E_{\text{Tmax|laminar}}(A_o,\text{T})\right| \leq \frac{\hbar^2}{mq\left[A_o+\frac{\hbar}{\sqrt{3mkT}}\right]^3} \qquad \text{(A3)}$$

and is a function only of fundamental constants and silicon-oxide interface "terrain" amplitude A_o. $E_{Tmax|laminar}$(5Å,300K) ≈ 0.23 MV/cm, near to the value of E_{Tvsat}(300K) ≈ 0.28 MV/cm per Eq.(7) in [4], insofar as Eq.(A1) is applicable 'as is', which pertains to longitudinal velocity in a periodic lattice. Since the interface terrain is rougher and non–periodic, however, the value of E_{Tvsat}(300K) is almost certainly an overestimate when E_T > $E_{Tmax|laminar}$(A_o,T), and for A_o < z_{min}(T), when carriers flow along a non–periodic interface.

Appendix 2.

a) Estimate of the ratio *R*, of saturation drain current at $dI_D/dV_{DS} = 0$, to the drain current at the limit of laminar flow (i.e., at $V_{DSAT} \approx 2\phi_F + 1.65\phi_t$) **in the absence of velocity saturation.** These currents are indicated in Figs.4,5,6, and 7b of the main text, and in Fig.A2.1 below.

The "3/2 power law" for the MOSFET's drain current is:

$$I_D = K\left\{(V_{GS} - V_T)V_{DS} - \frac{1}{2}V_{DS}^2 - \frac{2}{3}\gamma\left[(V_{DS} + V_{SB} + 2\phi_F)^{3/2} - (V_{SB} + 2\phi_F)^{3/2}\right]\right\}$$

and valid for $0 \le V_{DS} \le [\, V_{DS}$ at which $dI_D/dV_{DS} = 0\,]$, with $K = \mu(E_T, E_{//})\, C_{ox}\frac{W}{L}$, and mobility μ a function of both transverse (E_T) and longitudinal ($E_{//}$) electric fields. We thus have, upon further simplifying, with $V_{SB} = 0$:

$$\boldsymbol{R} \equiv \frac{I_D\ @\ (V'_{DSAT} = V_{DS}\ corresponding\ to\ dI_D/dV_{DS} = 0)}{I_D\ @\ (V_{DSAT} \approx 2\phi_F + 1.65\phi_t)} \approx$$

$$\frac{\frac{1}{2}(V_{GS} - V_T)^2 - \frac{2}{3}\gamma\left[\left((V_{GS} - V_T) + 2\phi_F\right)^{3/2} - (2\phi_F)^{3/2}\right]}{(V_{GS} - V_T)(2\phi_F + 1.65\phi_t) - \frac{1}{2}(2\phi_F + 1.65\phi_t)^2 - \frac{2}{3}\gamma\left[(2\phi_F + 1.65\phi_t + 2\phi_F)^{3/2} - (2\phi_F)^{3/2}\right]} \times \frac{\mu\,[E_T, (V'_{DSAT})/L]}{\mu\,[E_T, (2\phi_F)/L]}$$

where $\mu\,[E_T, (2\phi_F)/L] \approx \mu_{eff}$ is a constant, "effective", low field mobility, and $V'_{DSAT} \approx (V_{GS} - V_T)$ in the long channel limit. For typical silicon process parameters and mobility variations over the range at hand, the ratio of low to high *transverse* field mobilities above is approximately 2/3, such that the ratio ***R*** ranges between the approximate limits of 1 and (2/3) $[(V_{GS} - V_T)/2]/(2\phi_F) \approx (V_{GS} - V_T)/(6\phi_F) \ge 1$.

b) The two drain voltages at which dI_D/dV_{DS} and the saturation incremental conductances are equal (absent velocity saturation):

$$g_{dtriode}(I_D, V_{DS}) \equiv \frac{dI_D}{dV_{DS}} \approx K\{(V_{GS} - V_T) - V_{DS} - \gamma(V_{DS} + 2\phi_F)^{1/2}\} \tag{A4}$$

$$g_{dpinchoff}(I_D, V_{DS}) \approx \frac{1}{2}\left(\frac{I_{DSAT}}{L}\right)\frac{\ell'(V_{DS})}{(V_{DS} - V_{DSAT}) + V_{bi}\left[1 - e^{-\beta\left(\frac{V_{DS} - 2\phi_F}{2\phi_F}\right)}\right]}$$

$$\le \frac{(K/2)(V_{GS} - V_T)^2}{\boldsymbol{R}}\ \frac{1}{2}\left[\frac{\ell'(V_{DS})}{L}\right]\frac{1}{(V_{DS} - V_{DSAT}) + V_{bi}\left[1 - e^{-\beta\left(\frac{V_{DS} - 2\phi_F}{2\phi_F}\right)}\right]} \tag{A5}$$

Equating these, and solving for the two values desired for V_{DS} yields:

$$\boldsymbol{V_{DSX}} \approx V_{DSAT} + \frac{\ell'(2\phi_F)}{4L\boldsymbol{R}}\left\{\frac{[V_{GS} - V_T(V_{SB})]^2}{2\phi_F + \gamma\sqrt{4\phi_F} + V_{GS} - V_T(V_{SB})}\right\} \approx V_{DSAT} \approx 2\phi_F$$

and

$$\boldsymbol{V_{DSY}} \approx \frac{\gamma^2}{4}\left[\sqrt{1 + \frac{4\left\{(V_{GS} - V_T)\left[1 - \left(\frac{\ell'}{4L\boldsymbol{R}}\right)\right] + V_{DSAT}\right\}}{\gamma^2}} - 1\right]^2 - V_{DSAT}$$

whose limit is $(V_{GS} - V_T)$ in the long channel (low $E_{//}$) and/or low doping limits: $(V_{GS} - V_T) >> 2\phi_F$, and/or $\gamma \to 0$. Moreover, as $(V_{GS} - V_T) \to \gamma\sqrt{2\phi_F}$, its minimum value consistent with setting Eq.(A4) ≈ 0, $(V_{DSY} - V_{DSX})$ decreases as the transition region from strong inversion to pinchoff shrinks (see Fig.7b). This expression is otherwise within ±10% of our observations for Si and SiC devices.

As gate overdrive $(V_{GS} - V_T)$ increases, so does V_{DSY}, which, in turn, ***decreases*** the dynamic range (high r_{ds}/low g_{ds}). There is thus an ***inverse relationship between dynamic range and gate overdrive***.

The locations of these potentials are indicated in Fig.**A2.1** below**,** and in Figs.5,6, and 7b of the main text. Figure **A2.2** shows the variations of $(V_{DSY} - 2\phi_F)/(2\phi_F)$ with gate overdrive $(V_{GS} - V_T)/(2\phi_F)$, both normalized in their range of validity, $(V_{GS} - V_T) \ge \gamma\sqrt{2\phi_F}$, and with γ as a parameter.

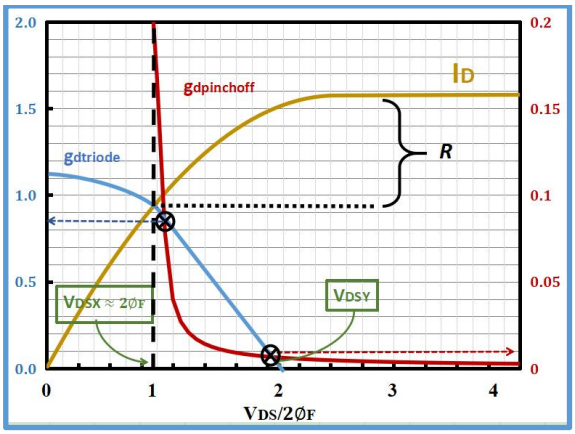


Fig. A2.1: V_{DSX} and V_{DSY} as defined in Appendix 2 b) with body–effect parameter $\gamma \approx 1$. The units on both vertical axes are mA/V; they differ by an order of magnitude, in correspondence to the two intersections of the $g_{dtriode}$ and $g_{dpinchoff}$ curves at V_{DSX} and V_{DSY}. The current I_D is shown for reference only, whose scale is in the mA range, along with the ratio R as defined in Appendix 2 a).

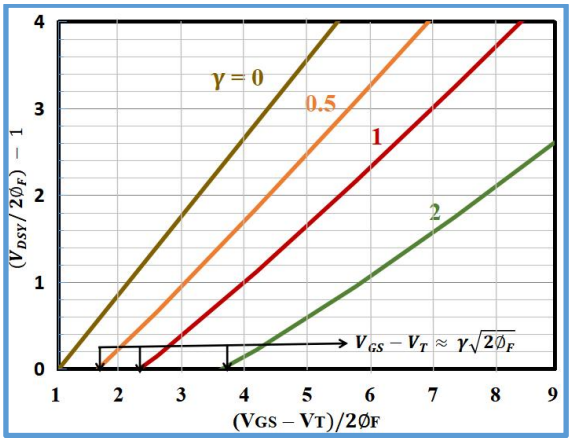


Fig. A2.2: Variations of V_{DSY} as defined in Appendix 2 b), relatively to its minimum value $\approx 2\phi_F$, with γ as a parameter. The range of validity for the various curves is $(V_{GS}-V_T) \geq \gamma\sqrt{2\phi_F}$ since $g_{dtriode} \geq 0$. Note that $V_{DSY} \leq (V_{GS}-V_T)$ for all values of $\gamma \geq 0$. as expected in view of the definition of V_{DSY} in section III.*B*.b), and as illustrated in Fig.5 of the main text and in Fig. A2.1.

c) Combining Eq. (A4) under the condition $(V_{GS}-V_T) \gg 2\phi_F$ and Eq. (A5) with $V_{DS} \geq V_{DSY} \geq 4\phi_F$ we have:

$$\frac{g_{dtriode}}{I_D\ @\ (V_{DSAT} \approx 2\phi_F)} \approx \frac{1}{V_{DSAT}} \approx \frac{1}{2\phi_F} \quad \text{and} \quad \frac{g_{dpinchoff}}{I_{DSAT}\ @\ (V_{DS} \geq V_{DSY})} \approx \frac{1}{2}\left[\frac{\ell'(V_{DS})}{L}\right]\frac{1}{V_{DS}} \approx \left(\frac{1}{2L}\right)\frac{\sqrt{\frac{2\varepsilon_{si}}{1.4(qN_a)}}}{\sqrt{V_{DS}}}$$

hence the ratio: $\frac{g_{dpinchoff}}{g_{dtriode}} \approx \left(\frac{2\phi_F}{L}\right)\left(\frac{R}{2}\right)\frac{\sqrt{\frac{2\varepsilon_{si}}{1.4(qN_a)}}}{\sqrt{V_{DS}}}$, in view of the definition of R in **a)** above.

d) The effect of velocity saturation on V_{DSY} (at constant effective mobility μ_{eff}) is to reduce this potential to a value:

$V'_{DSY} \approx V_{DSY} + \frac{Lv_{sat}}{\mu_{eff}} - \sqrt{{V_{DSY}}^2 + \left(\frac{Lv_{sat}}{\mu_{eff}}\right)^2} \leq V_{DSY}$, with equality approached in the long channel limit where $V_{DSY} \lesssim (V_{GS}-V_T)$. In the short channel limit, on the other hand, i.e. when $\left(\frac{Lv_{sat}}{\mu_{eff}}\right) \ll V_{DSY}$, V'_{DSY} tends to the limit $\left(Lv_{sat}/\mu_{eff}\right)$, such that V'_{DSY} takes on a (positive) value no higher than the lowest of $(V_{GS}-V_T)$ and $\left(Lv_{sat}/\mu_{eff}\right)$, as represented in [2, Fig.13].