Conservation Laws at Physical Origins of Universal Mobility in MOSFET Inversion Layers in Consilience with River Flow in a Gravitational Field

F. S. Shoucair

Abstract—The salient properties of charge flow (or current) along the MOSFET’s inversion layer are shown to be analogous to a river’s flow in a gravitational potential field, insofar as both are fundamentally governed by energy conservation principles, and their laminar and turbulent conditions determined by friction losses at shallow depths. We formulate a simple, yet accurate, model for a non-uniform mobile charge density $\rho(z)$ giving rise to a mean potential $\psi^*$ across an inversion layer of finite extent, which we measure by means of a novel, sensitive, experimental method involving nulls of harmonic distortion components ($D_2 \approx D_3 \approx 0$) of the drain current under sinusoidal excitation below saturation. We thus establish analytically that the low-field, “universal” effective, mobility $\mu_{\text{eff}}$, long reported to vary as $\sim (E^* T)^{-1/3}$ for transversal fields

$$E^* T = -\frac{1}{\varepsilon_{\text{si}}} \left[ \eta Q_i + Q_b \right] \leq 0.5 \ \text{MV/cm},$$

is manifestation and consequence of both energy and momentum conservation under laminar flow conditions and quantum mechanical effects, in which case the inversion layer’s mean thickness $<z>$ also varies as $\sim (E^* T)^{-1/3}$ up to a maximum value $E^* T \approx 0.35 \ \text{MV/cm}$ at 300K, determined only by interface “terrain” amplitude and fundamental constants, while $\eta$ varies continuously between 1/2 and 1/3. We reconcile our model and observations with those of Stern–Howard, Schwarz-Russek, and Takagi et al, and establish, analytically and experimentally, that the higher order, derivative, parameter $\theta_T$ quantifying $\mu_{\text{eff}}$’s modulation by $E^* T$ varies as $\sim (E^* T)^{-5/3}$ under these conditions, thereby further corroborating the foregoing effects and interpretations thereof.

Index Terms—Distortion nulls, Effective mobility, Energy conservation, Inversion layer, Laminar flow, MOSFET, Quantum mechanics, River flow.

Physics is mathematical, not because we know so much about the physical world, but because we know so little: it is only its mathematical properties that we can discover. [1]

I. INTRODUCTION

The MOSFET$^1$ has been a critical component of modern CMOS$^2$ integrated circuits since their inception in the 1960’s. The latter’s evolution towards miniaturization continues unabated, whose main purpose is to increase the density and speed of ubiquitous digital and analog circuits. The physical mechanisms underlying the operation of MOSFETs properly continue to receive much attention in the technical literature because they are key to ongoing developments of accurate and efficient models and simulations for devices and circuits, especially as the latter are downscaled. Indeed the technological and commercial import of these mechanisms was recognized from the outset, as attested by Stern and Howard’s landmark analysis of the inversion layer in 1967 [5], by Schwarz and Russek’s empirical model of the same in 1983 [3], and by the indispensable experimental data of Takagi et al, in 1994 [4], whose scattering models have not been revised, ostensibly in the interim, in any substantial manner. While the pertinent literature is otherwise considerable, these prominent forerunners circumscribe the essential features of the subject at hand, thereby providing both

---

$^1$ Metal Oxide Semiconductor Field Effect Transistor.

$^2$ Complementary Metal Oxide Semiconductor.
context and reference for our analyses and observations.

The characteristics of the inversion layer, or current–carrying channel, which underlie the MOSFET’s current conduction, include its “effective” mobility $\mu_{\text{eff}}$, conductance, transconductance, various associated capacitances, and transit/switching speeds, all of which are key to the performance of digital and analog circuit functions. In modern, downscaled, devices, whose gate oxides are near to, and below, $t_{\text{ox}} \approx 20\text{Å}$, $\mu_{\text{eff}}$ is observed to fall more steeply with increasing $E_{*T}$ than expected from its “universal” behavior, whereby $\mu_{\text{eff}}$ varies as $\sim (E_{*T})^{-1/3}$, which devices with thicker oxides and inversion layers have displayed over many prior technological generations. This mobility degradation is widely attributed to the growing, and increasingly dominant, influence of irregularities and/or “scattering” effects occurring at the Si/Gate–insulator interface, as $E_{*T}$ exceeds 0.5 MV/cm approximately. These trends occur in tandem with higher bulk doping concentrations, higher gate insulating dielectric constants, and thinner gate oxides, among other evolving physical parameters, as entailed by downscaling. Consequently, accurate models of inversion layers of extents comparable to a single silicon inter–atomic distance ($\approx \pm 5\text{Å}$) are required towards further developments, not only because their deleterious effects increasingly encroach upon critical MOSFET performance, but because the latter’s complex interactions are likely further to complicate the task of disentangling their respective influences as well.

In section II, we establish the similar, general, relationships pertaining to a river’s flow under gravitation, and to the MOSFET’s current flow under low longitudinal field conditions, which relate the depths $z$ of the respective “fluids” to the relevant parameters of these systems. These considerations readily lead to the origins of the behavior of “universal mobility”$^3$, $\mu_{\text{eff}} \sim (E_{*T})^{-1/3}$, as a consequence of laminar flow, and energy and momentum conservation under quantum-mechanical uncertainty relationships, as elaborated in II.C, along with a maximum value for $E_{*T}$ beyond which laminar flow ceases.

Typical models implemented in standard simulation programs achieve computational efficiencies by modeling $\mu_{\text{eff}}$ and its modulation coefficient $\theta_T$ by means of “curve fitting” parameters (such as “UA”, “UB”, “UC”, etc., in the industry standard BSIM4), which must be optimized for a given process, hence repeated (and/or verified) for each process generation, so as to meet prescribed levels of accuracy. Moreover, a key premise underlying MOSFET conduction models in archetypal circuit simulation programs is the assumption of a “negligibly thin” inversion layer, or a charge sheet. While this assumption simplifies drain current computations, it forsakes consequential information about the charge distribution $\rho(z)$ within, and mean potential $\psi^*$ across, the inversion layer, both of which are critical to the conceptual modeling of channel conductivity, and of the modulation thereof by transverse electric fields $E_{*T}$. The limitations and drawbacks entailed by these considerations are discussed in section III, along with those resulting from constant mobility models.

Absent efficient models derived from first principles, this quasi–Procrustean approach has entailed ever more delicate refinements and compromises as novel effects have gained prominence with the advent of downscaling. While some of these effects are inconsequential to certain circuit functions, they are detrimental to those, analog especially, wherein such performance parameters as distortion components are important, among other high order effects (e.g. noise), yet rendered inaccurately by charge sheet models [11]. Our results indicate that (i) the inversion layer’s detailed characteristics remain consequential even as

---

$^3$ The Greek letter $\mu$ and the Roman letter M of our modern alphabet originate from the Egyptian hieroglyphic symbol ˄˄˄˄ for water.
this layer shrinks with process downscaling and, thus, (ii) the conventional assumption of a “negligibly” thin inversion layer may not warrant its outright neglect, absent contrary and/or contextual justification.

In section IV, we apply an experimental method originally conceived to suppress harmonic distortion components in analog integrated filters [10,11], to measure the mean potential \( \psi^* \) across the inversion layer, a novel element in this context, which yields the corresponding mean transverse \( E^*_T \), (including \( \eta \)), as well as the ratio of mean to surface potentials (\( \psi^*/\psi_s \)) from which we deduce \( \rho(z) \), and subsequently model analytically in section V. The accuracy of this model is confirmed, in section VI, by further, sensitive, measurements of the modulation of \( \mu_{\text{eff}} \) by \( E^*_T \), namely the parameter \( \theta_T = (1/\mu) (d\mu/d(E^*_T) \), which are in agreement with our model’s analytical predictions that \( \theta_T \) should vary as \( (E^*_T)^{-5/3} \).

Our overall results are summarized in section VII, while the pertinent features of the Stern-Howard [5] and Schwarz-Russek [3] models are summarized in appendix A2, to whose models and data we compare and contrast ours throughout the present manuscript.

Fig. 1. Top: View of a wide (\( W \)) river of depth \( z(x) \ll W \) above its bed, flowing in the positive \( x \)

The analyses to follow aim to establish such relationships between the aforementioned phenomena as derivable from first principles. We thus bring conservation of energy and momentum laws to bear upon these from the outset, which are seldom, if at all, invoked explicitly in this context. This approach, thus underpinned and framed by a scaffolding of fundamental principles\(^4\), provides not only a variant pedagogical perspective on the complex effects attending the characteristics of the MOSFET’s inversion layer, but such foundational safeguards as well, as are vital to assessing the validity of prospective physical constructs and interpretations thereof, especially in light of a considerable body of (often disparate) experimental observations extant in the literature — our reference section listing a mere representative sample thereof.

II. PARTIAL ANALOGY BETWEEN THE FLOW OF ELECTRONS IN A VOLUME ENCLOSED AN INVERSION LAYER AND THE FLOW OF A BROAD RIVER IN A GRAVITY FIELD, BOTH CONSIDERED FROM THE PERSPECTIVE OF ENERGY (\( \mathcal{E} \)) CONSERVATION.

The table in appendix A1 summarizes the correspondence between parameters of the gravitational and electrical fields pertaining, respectively, to the river and to the MOSFET’s inversion layer. The variable \( z \) represents depth of river and depth (thickness) of inversion layer, as indicated in Fig. 1.

\(^4\) Symmetry, in nature, with respect to displacements in time implies the conservation of energy. Symmetry with respect to position in \( x,y,z \), implies the conservation of that component of momentum. Many, though not all, laws of physics are symmetrical with respect to spatial dimensions: the former include gravity, electrodynamics, and strong nuclear interactions, the latter weak nuclear interactions (\( \beta \) decay) [13].
direction at an angle $\alpha$, the slope of the bed to the horizontal.

**Bottom:** Cross sectional view of an inversion layer’s volume: electrons flow in the positive $x$ direction, from source towards drain. The “height” of an electron above the “floor” level (the gate oxide plane) is the thickness of the inversion layer $z(x) < W$; the direction of $W$ is not shown. The ratio $(E_f/E^*_f)$ corresponds to the slope of the “riverbed” above, with $E_f$ the longitudinal electric field. The wide arrows represent various energies entering or leaving the volume, only some of which are conserved.

A. River [6]

The gravitational potential energy difference (loss) between depths $z(x)$ and $z(x + \Delta x)$ above the bed is $mg\Delta z$, where $(\Delta z/\Delta x) = \tan \alpha$, and $(1/2)mv^2_{drift}$ is the kinetic energy gain between $x$ and $(x + \Delta x)$. Energy loss to friction, or viscosity, along the riverbed is proportional to kinetic energy, and inversely to depth $z(x)$. The coefficient of hydraulic friction $K_F \geq 0$, is a proportionality constant between the latter quantities. While potential and kinetic energies are conserved, energy lost to friction is not, and we have:

$$\frac{\Delta}{\Delta x} \left[ \mathcal{E}_{\text{potential}} + \mathcal{E}_{\text{kinetic}} \right] + \mathcal{E}_{\text{friction}} = 0 \quad (1)$$

hence,

$$\frac{\Delta}{\Delta x} \left[ mg \left[ z(x) - \Delta x \tan \alpha \right] + \frac{1}{2} mv^2_{drift} \right] + \frac{K_F}{z(x)} \frac{1}{2} mv^2_{drift} = 0 \quad (2)$$

dividing through by $mg$ and taking the limit of the differences readily yields:

$$\frac{d}{dx} \left( z(x) + \frac{1}{2g} v^2_{drift} \right) - \tan \alpha + \frac{K_F}{z(x)} \frac{1}{2g} v^2_{drift} = 0 \quad (3)$$

If the flow of the river is steady at a mean velocity $v_{drift}$, such that $Q$ is the volume, assumed constant, flowing across a given cross-section per unit of time, the following continuity equation holds:

$$z(x) \cdot v_{drift} = Q \quad (4)$$

Substituting $v_{drift}$ from (4) into (3) yields the slope $(dz/dx)$ as a function of $z(x)$:

$$\frac{dz}{dx} = \tan \alpha \left[ 1 - \frac{K_F Q^2}{2g \tan \alpha z(x)^3} \right] \quad (5)$$

The relative magnitudes of $K_F$ and $\alpha$ determine the magnitude and sign of $(dz/dx)$, and the integral curves of (5) yield the profile $z(x)$ of the free surface.

**Isoclines** are straight lines parallel to the bed, which correspond to constant values of $(dz/dx)$, a stream of uniform depth, laminar flow, and a river flowing into a reservoir. When $z(x)$ is large enough so that the ratio inside the square brackets is near to unity, the flow is parallel to the slope of the riverbed, since $(dz/dx) \approx \tan \alpha$.

If the ratio inside the square brackets changes sign, at shallow depths where $z(x)$ is small for example, the flow may incur discontinuities in height and velocity, or hydraulic “jumps”, thereby becoming turbulent or chaotic, such that the river becomes a torrent. Similar phenomena readily occur when dispensing a liquid from a bottle whose neck is narrower than its body, such as a typical (750 ml) wine bottle: as the neck is tipped downward gradually, the fluid first trickles out in a continuous stream until such an angle is reached as causes the fluid to fill the neck’s cross section approximately. Upon exceeding this angle, part of the fluid “backs up” inside the bottle, thereby emerging irregularly, in “fits and starts”, as its flow becomes turbulent or chaotic. Such effects are especially conspicuous when observed in slow motion, with a dense fluid such as warm honey.

Since laminar flow is the condition analogous to the standard, presumed, flow of electrons in MOSFETs’ inversion layers under conditions of low longitudinal fields $(E_{\parallel} << E^*_f)$, (5) may serve as reference for the forthcoming comparison. The case $(dz/dx) = 0$, corresponding to a level surface, or to a fluid in equilibrium at the bottom of a bucket, was previously invoked by C. Séquin [7], to
promote visualization of the behavior of certain charge-controlled devices, including charge-coupled devices (CCD’s) and MOSFETs, whereas non-laminar flow conditions were indicated as pertinent to MOSFETs operating in saturation, where they are typically biased in analog circuits [8]. Although the latter conditions entail more complicated dynamics [12], beyond the purview of the present analysis, an extension thereof in this respect may prove instructive in view of the ubiquity and practical importance of such circuits.

B. Inversion Layer

The potential energy difference for an electron traveling an incremental distance $\Delta x$ along the channel in the positive $x$ direction, i.e., from source towards drain and “falling” a distance $\Delta z$ in the positive $z$ direction inside the inversion layer (i.e., from the bulk towards the gate’s “attractive” potential) is:

$$ q \left[ V(x) - V(x + \Delta x) \right] = q\Delta V = -q(E_T^z dz + E_{II} dx) \quad (6) $$

For an inversion layer charge density varying nearly linearly with $z$, i.e., $\rho(z) \sim z$, which will be established from observations of the ratio of mean to surface potentials ($\psi*/\psi_s$) in section IV below:

$$ \rho(z) \approx \rho_o \left( \frac{z}{z_{max}} \right)^m \quad \text{with} \quad m \approx 1 \quad (7) $$

we have, equivalently to the continuity (4) for the river:

$$ I_D = \frac{dQ_t}{dt} = \frac{d}{dt} \left[ \frac{1}{2} q \left( \frac{\rho_o}{z_{max}} \right) W z(x) \ dx \right] $$

$$ = \frac{1}{2} q \left( \frac{\rho_o}{z_{max}} \right) W z(x) v_{drift} \quad (8) $$

wherein $Q_t$ is the total mobile charge in a cross-sectional slice $dx$ of the inversion layer, and the constant drain current $I_D$, flowing at a mean velocity $v_{drift}$, is the counterpart of the river’s constant volume $Q$ flowing per unit of time and width. In this case:

$$ v_{drift} \approx \left( \frac{2l_p z_{max}}{q \rho_o W z(x)} \right) = \mu_{eff} E_{II} \quad (9) $$

is the “classical” counterpart of (4), whose value is approximately $250 \text{m/s}$ (when $\mu_{eff} \approx 500 \text{cm}^2/\text{V} \cdot \text{s}$, $E_{II} \approx 50 \text{V/cm}$), near to the speed of sound in dry air at standard temperature and pressure, for $(I_D/W) \approx 0.1 \mu \text{A/\mu m}$, $z(x) \approx 20 \text{Å}$, and $(\rho_o/z_{max}) \approx 10^{24} \text{ m}^3$, the latter volume density corresponding to a surface density of $\approx 10^{12} \text{cm}^{-2}$ for a commensurate charge-sheet model (see Fig.4).

If we account for an energy loss term proportional to $(\frac{1}{2})m v_{drift}^2$ as in the case of the river above then, proceeding similarly, taking differences to their differential limits and dividing through by $q$, we have the following statement of conservation of energy:

$$ - E_T^z \frac{dz}{dx} - E_{II} + m \frac{2l_p z_{max}}{q \rho_o W} \frac{d}{dx} \left[ \frac{1}{z(x)} \right]^2 = 0 \quad (10) $$

which may readily be put in the form:

$$ \frac{dz}{dx} \approx - \left( \frac{E_{II}}{E_T^z} \right) \left[ 1 + K_F \left( \frac{m}{2q E_T^z} \right) \left( \frac{2l_p z_{max}}{q \rho_o W} \right)^2 \frac{1}{z(x)^3} \right] $$

$$ \quad \left( \frac{m}{2q E_T^z} \right) $$

$$ = - \left( \frac{E_{II}}{E_T^z} \right) \left[ 1 + \left( \frac{m}{2q E_T^z} \right) \left( \frac{2l_p z_{max}}{q \rho_o W} \right)^2 \frac{1}{z(x)^3} \right] $$

$$ \quad (11) $$

and is the counterpart of (5) in the foregoing river case. Equation (11) thus lends itself to the following interpretations:

a) Large $z$ (thick inversion layer ~ deep river): $(dz/dx) \approx -(E_{II}/E_T^z)$, which is the familiar “gradual channel” approximation when the ratio $(E_{II}/E_T^z) \ll 1$, as typically the case in standard MOSFET circuit operation, and is in the range of $(1/2000)$ to $(1/6000)$ for our data (Figs. 3–5, and 7, below).

b) Small $z \to 0$ (thin inversion layer ~ shallow river): $(dz/dx)$ tends to the limit $(-K_F)\mu_{eff}$, which indicates that electron flow is
predominantly determined by surface scattering conditions, with $K_F$ corresponding to “friction” along the insulating interface (the “riverbed”). Hence the magnitude of $(dz/dx)$ may vary locally, along $x$ if $K_F$ does so, such that a “gradual” channel no longer obtains along the entire length of the conducting channel. If, furthermore, $|K_F|$ should be sufficiently large, or (worse), a function increasing as a power of $(1/z)$, then $(dz/dx) \approx -K_F(x,z)$ would behave in like manner, thereby yet again invalidating a “gradual” channel assumption and, importantly, result in mobility decreasing at a rate other than universal curves’ power of $(E^* T)^{-1/3}$. Such phenomena are observable in [4, I. Figs.1–5] for electrons and holes under both low temperature and high $E^* T$ conditions.

c) The “pinch-off” condition corresponds to the latter situation whereby, in view of current continuity, the velocity of carriers along the channel (between source and pinch-off point) may vary as dictated by $K_F(x,z)$, while that between the pinch-off point and the drain increases owing to the local, accelerating, longitudinal field $E_\parallel$. Indeed, velocity “saturation” is observed under these circumstances, whose value is near to $10^5$ cm/s at 300K for electrons and holes on $<100>$ silicon [9],[12]. While the transverse electric field $E^*_T$ no longer dominates the characteristics of the channel’s conductance, the latter’ modeling is complicated by the resulting two-dimensional effects.

d) Laminar flow, with $z(x)$ not restricted to large values, occurs along isoclines, or curves along which $(dz/dx)$ is constant: (5) and (11) may be solved for constant values of $(dz/dx)$. In the case of special interest whereby $E_\parallel << E^*_T$, and the effects of $K_F$ negligibly small, the result is proportionality between $z$ and $(E^*_T)^{-1/3}$, which is therefore consequence of our starting premise $\rho(z) \sim z$, per (7). While the expression for the quantum–mechanical, average, extension $z_{QM}$ derived by Stern and Howard [5] resulted from an ostensibly different function, $\rho(z) \sim z^2 e^{\beta z}$ (see appendix A2), yet they obtained very nearly the same functional relationship $<z_{QM}> \sim (E^*_T)^{-1/3}$. In section II.C.2 below, we show that conservation of energy and momentum readily lead to $\mu_{eff} \sim (E^*_T)^{-1/3}$ when $z \sim (E^*_T)^{-1/3}$; that is to say, laminar flow is the physical condition underlying universal mobility behavior.

C. Similarities and Distinctions between River and Inversion Layer Cases

While (5) and (11) are similar in general form, the foregoing analogy between the river and inversion layer cases is apposite though limited, because the density of the “incompressible fluid” in the former is considerably higher than that of electrons, whose volume density is more akin to that of a diffuse “gas”. Since, moreover, the motion of electrons in the inversion layer is governed by Schrödinger’s equation as constrained by the boundary conditions of the system (see appendix A2), this motion may be construed as the diffusion of a probability amplitude from one point to the next along the x axis. Whereas ordinary diffusion of a gas along a thin tube gives rise to real exponential functions, Schrödinger’s equation yields complex waves for an “oscillating gas” [13]. While their respective $K_F$ parameters, constant or otherwise, can thus be expected to be different qualitatively and quantitatively, further study of the effects contributing to $K_F$, as formulated above, is relevant to the ongoing technological downscaling of integrated devices, circuits, and systems.

1) Dependence of inversion layer thickness on $E^*_T$ and $T$

To the extent that the inversion layer may be regarded, conceptually, as the combination of a classical ($\zeta_{cl}$) and a quantum–mechanical ($\zeta_{QM}$) component [3], the following relationships obtain:
\[ \frac{3}{2} kT = qE_T^* z_{cl} = \frac{1}{2} m v_{th}^2 \quad (12a) \]

where \((3/2)kT\) is the average translational kinetic energy of a molecule of an ideal gas at a given temperature \(T\), with \(k\) Boltzmann’s constant, hence

\[ Z_{cl} = \frac{3}{2} \frac{kT}{q} E_T^* = \frac{1}{2} \frac{m}{q} v_{th}^2 \quad (12b) \]

and

\[ v_{th} = \sqrt{\frac{2 q E_T^* z_{cl}}{m}} = \sqrt{\frac{3 kT}{m}} \quad (12c) \]

is the (rms) value of the thermal velocity \(v_{th} \approx 1.15 \times 10^5\) m/s at 300K, whereas the quantum–mechanical component \(z_{QM}\) resulting from the uncertainty relationship \((\Delta p \Delta z) \geq \hbar\), where \(\hbar = h/2\pi\), \(h\) being Planck’s constant, hence

\[ \frac{p^2}{2m} = \frac{1}{2m} \left( \frac{h}{z_{QM}} \right)^2 = qE_T^* z_{QM} = \frac{1}{2} m v_{QM}^2 \quad (13a) \]

which may be construed as an energy term when put in a form similar to (12a), such that

\[ Z_{QM} = \left[ \frac{h^2}{2mqE_T^*} \right]^{1/3} \quad (13b) \]

is the counterpart of (12b), and a velocity

\[ v_{QM} = \sqrt{\frac{2qE_T^* z_{QM}}{m}} = \frac{h}{m z_{QM}} \quad (13c) \]

the counterpart of (12c).

Since \(v_{th} \gg v_{drift} \approx 250\) m/s for our experimental conditions (see section VI), and since \(z_{cl}\) and \(z_{QM}\), per (12b) and (13b) are related uniquely for a given value of \(E_T^*\), we may put \(S = (z_{cl} z_{QM})\) and solve the conservation of energy constraint at fixed \(x\) and \(T\):

\[ \frac{d}{dz} \left[ qE_T^* S z + \frac{1}{2m} \left( \frac{h^2}{z} \right)^2 \right] = 0 \]

yielding the value of \(z\) corresponding to the total, conserved, energy \(E\):

\[ z \bigg|_{(\Delta \varepsilon = 0)} = \left[ \frac{S^2 h^2}{mqE_T^*} \right]^{1/3} \quad (14a) \]

which varies as \((E_T^*)^{-1/3}\), in correspondence with laminar conditions per II.B.d) above and, unlike the river, is subject to the minimum value imposed by the uncertainty relationship implicit in (13a), as accounted for in the next subsection, and without which (14a) would allow arbitrarily small \(z\) with increasing \(E_T^*\).

2) Maximum value of \(E_T^*\) for which laminar flow may be expected

The same uncertainty relationship \(\Delta p \Delta z \geq \hbar\) associated with momentum

\[ p = mv_{th} = \sqrt{3mkT} \]

in the \(z\) direction in turn yielding:

\[ \Delta z \geq \frac{\hbar}{\sqrt{3mkT}} \approx 10\ \text{Å} \quad (14b) \]

at 300K, the inversion layer’s thickness may be expected to “clear” the Si/oxide interface terrain (see Fig.2a) when:

\[ z \bigg|_{(\Delta \varepsilon = 0)} - \Delta z \approx \left[ \frac{S^2 h^2}{mqE_T^*} \right]^{1/3} \frac{h}{\sqrt{3mkT}} \geq A_o = \text{amplitude of interface terrain} \]

hence when

\[ |E_T^*| \leq E_{TMAX}(A_o,T) = \frac{S^2 h^2}{mq \left[ A_o + \frac{h}{\sqrt{3mkT}} \right]^{1/3}} \]

(15)

which may serve as an upper limit estimate for laminar flow under ideal conditions of negligibly small \(K_F\), and is a function only of terrain roughness amplitude \(A_o\) and fundamental constants.

Figure 2(a) displays the variations of \(z_{cl}\) per (12b), \(z_{QM}\) per (13b), \(z \bigg|_{(\Delta \varepsilon = 0)}\) per (14a), and \(\Delta z\) per (14b); the vertical arrows located at
(\(E^*\gamma\)\(\text{MAX}\)(300K) \(\approx 0.34\) MV/cm occur where \(z|_{(\Delta \varepsilon = 0)} - \Delta z \approx 15\text{Å} - 10\text{Å} = 5\text{Å}\) just “clears” a terrain of amplitude \(A_o \approx 5\text{Å}\).

The function (15) shown in Fig 2(b) with terrain “roughness” amplitude \(A_o\) as a parameter, takes on the approximate value 0.34 MV/cm (where \(S \approx 1.15\)) at 300K, with \(A_o = 5\text{Å}\) commensurate with silicon’s lattice dimensions and with the \(\pm 5\text{Å}\) interface amplitude observed for thermally oxidized <100> surfaces. The maximum electric field’s value expected from (15), for \(A_o = 0\) at 300K is near to 1.16 MV/cm, as indicated by the dashed arrows.

As far as may be discerned from their published figures [4, I, Figs.4,5], (15) appears to be in quantitative agreement with Takagi et al’s data with respect to the temperature dependence of this maximum value for \(E^*\gamma\), namely \(\sim T^{3/2}\) in the limit \(A_o = 0\), and qualitatively as to its relative insensitivity to the different masses of electrons and holes, and to surface orientations [4, II].

Equation (15)’s close agreement with our observations is readily apprehended in
Figs. 3,4,5,7, and in section VI wherein our experimental results are discussed.

3) Establishing that the variations of \( \mu_{\text{eff}} \) as \( \sim (E^* T)^{-1/3} \), per universal curves, result from conservation laws, laminar flow conditions, and quantum–mechanical effects.

To the total energy \( [qE^* (z_{cl} + z_{QM})] \), which is the sum of (12a) and (13a), we may ascribe the equivalent velocity \( v_{eq} \sim v_{drift} \sim v_{QM} \gg v_{drift} : \)

\[
v_{eq} = \sqrt{\frac{2qE^* (z_{cl} + z_{QM})}{m}} \tag{16}
\]

Momentum (strictly only the \( x \) component thereof), and energy conservation within an incremental inversion layer volume then yield:

\[
m(v_{drift} + v_{eq}) = \text{constant} \tag{17}
\]

and

\[
m(v_{drift}^2 + v_{eq}^2) = \]

\[
m [ (v_{drift} + v_{eq})^2 - 2v_{drift} \cdot v_{eq} ] = \text{constant} \tag{18}
\]

wherein the first term inside the square brackets is constant, per (17). Since \( v_{drift} = (\mu_{\text{eff}} E_{//}) \), it follows that the product:

\[
(v_{drift} \cdot v_{eq}) = \mu_{\text{eff}} E_{//} \sqrt{\frac{2qE^* (z_{cl} + z_{QM})}{m}} \tag{19}
\]

must be constant. Since \( v_{eq} \) corresponds nearly to the total energy in the inversion layer’s volume, for which \( z \sim (E^* T)^{-1/3} \) established by (14a) is the condition for laminar flow per II.B.d) above — and consistently with the solution of Schrödinger’s equation as constrained by the boundary conditions of the system, per (A2.2) — the product \( [E^* (z_{cl} + z_{QM})] \), under the radical in (19) varies as \( \sim E^* (E^* T)^{-1/3} \sim (E^* T)^{2/3} \) which, in turn, requires:

\[
\mu_{\text{eff}} (E^* T)^{1/3} = \text{constant} \tag{20}
\]

so that \( \mu_{\text{eff}} \) varies as \( \sim (E^* T)^{-1/3} \), as is indeed the widely reported behavior of “universal mobility” curves for \( E^* T \leq 0.5 \text{ MV/cm} \), for both electrons and holes [2],[4]. Hence \( \mu_{\text{eff}} \sim (E^* T)^{-1/3} \) is consequence, and manifestation, of the simultaneous concurrence of (i) laminar flow conditions, (ii) energy and momentum conservation, and (iii) quantum–mechanical quantization effects along the \( z \) direction resulting in \( z \sim (E^* T)^{-1/3} \).

III. CONDUCTIVITY, MOBILITY, AND THE CHARGE SHEET MODEL

Under laminar flow conditions, the conductivity of a thin sheet of charge at depth \( z \) inside the inversion layer in a \( (W-x) \) plane, parallel to that of the gate insulator, is \( \sigma(x,z) = [q \rho(x,z) \mu(z)] \), where \( q \) is the electronic charge, \( \rho(x,z) \) the charge density at \( x \) along the channel, and \( \mu(z) \) the average mobility of charge carriers along the \( x \) direction, which is related to the drift velocity of (9):

\[
v_{\text{drift}} = \mu(z) \frac{dv}{dx} = \mu(z) E_{//} \tag{21}
\]

where \( E_{//} \) is the longitudinal electric field, along the \( x \) direction.

The channel’s conductance is proportional to the integral of \( \sigma(x,z) \) over the depth of the inversion layer, namely to the sum of the conductivities of incremental sheets therein. The resulting drain current, in turn, being proportional to \( W \) to conductance, and inversely to \( L \), we have the following general functional relationship between the parameters of interest:

\[
I_D = \left( \frac{w}{T} \right) \int_0^L \int_0^{z_{\text{max}}} \sigma(x,z) dz \, dx
\]

\[
= \left( \frac{w}{T} \right) \int_0^L \int_0^{z_{\text{max}}} q[\sigma(x,z) \mu(z)] dz \, dx \tag{22}
\]

wherein the integral over \( dz \) is taken between the limits of the inversion layer’s thickness \([0, z_{\text{max}}]\), and that over \( dx \) between those of the channel length \([0, L]\). This double integral is customarily simplified by assuming (a) constant mobility along the \( z \) axis, \( \mu(z) \approx \mu_{\text{eff}} \) thereby neglecting the effects of transverse fields \( E^* T \) on mobility and, (b) an inversion layer of zero thickness — the
“charge–sheet” approximation — such that \( \rho(x,z) = \rho(x) \) is a function of \( x \) only, or of the potential \( V(x) \), since \( (dV/dx) = E \). Under these simplifying assumptions, the drain current may be expressed in closed form, by integration along the channel length:

\[
I_D = \left( \frac{W}{L} \right) \mu_{\text{eff}} \int_0^L q \rho(V) dV = \mu_{\text{eff}} C'_{\text{ox}} \left( \frac{W}{L} \right) [a_1 V_D + a_2 V_D^2 + a_3 V_D^3] + \ldots \text{higher order terms} \tag{23}
\]

wherein \( \rho(V) \) is a known, explicit, function of the terminal voltages applied to the transistor, \( V_S = 0 \) and \( V_D \) are the potentials at its source \( (x = 0) \) and drain \( (x = L) \) terminals respectively, \( C'_{\text{ox}} \) the gate oxide capacitance per unit area, \( a_1, a_2, \ldots \) \( a_n \) are constants determined by terminal bias and signal potentials, and the higher order terms are negligibly small in this context. For a given value of \( I_D \), \( \mu_{\text{eff}} \) is then the only unknown quantity in (23), which is thereby deduced readily from standard measurements. We note that, as simplification \( (a) \) above sets \( \mu_{\text{eff}}(E^*_{\text{T}}) = \mu_{\text{eff}} \) in (23), the value of \( E^*_{\text{T}} \) corresponding to a given \( I_D \), which is essential to the determination of the function \( \mu_{\text{eff}}(E^*_{\text{T}}) \), is variously assumed to be an “average” [2], or an “effective” [2],[4],[14], value \( E^*_{\text{T}} = -(1/\varepsilon_{\text{Si}})[\eta Q_i + Q_b] \), with \( \eta = 1/2 \) for electrons [2]–[4],[14], \( \eta = 1/3 \) for holes [4], with \( Q_i \) (corrected [17]), and \( Q_b \), the total inversion and bulk charges, respectively.

A. Limitations of the charge–sheet and constant mobility models

In view of Poisson’s potential field equation associated with a given charge distribution (see appendix A), the assumption of an inversion layer of zero thickness — entailing the tacit neglect of its spatial density profile — results in the loss of information about both the electric field \( E(z) = -(q/\varepsilon_{\text{Si}}) \int \rho(x,z) \, dz \), and the potential \( \psi(z) = -\int E(z) \, dz \) within the inversion layer, with \( \varepsilon_{\text{Si}} \) the permittivity of silicon, since these integrals are taken between identical limits \([0,z_{\text{max}}(x)]\) when \( z_{\text{max}} = 0 \), as indeed the potential difference across any spatial region of zero extent is zero. The surface potential \( \psi_s \) at the Si/insulator interface, may nonetheless be calculated accurately from knowledge of the total inversion layer charge \( Q_i \), per (25) below, which is represented by \( q\rho(V) \) in (23) above [8],[9].

If \( \mu(z) \) is not assumed constant in (22), and an inversion layer of finite thickness \( z_{\text{max}} \neq 0 \) is assumed, then \( E(z) \) and \( \psi(z) \) are, in principle, known functions given \( \rho(x,z) \), but the drain current in (22) may no longer be expressed in analytical closed form as in (23), because \( \mu(z) \) may not properly be removed from the integrand of (22). Consequently, and crucially for purposes of quantifying the dependence of mobility on transverse fields \( (E^*_{\text{T}}) \) from observations of \( I_D \), the function \( \mu_{\text{eff}}(E^*_{\text{T}}) \) may no longer be “disentangled” from the result of the integration of the product \( \sigma(x,z) = [q\rho(x,z)\mu(z)] \) in (22), nor, thereby, extracted unambiguously from standard measurements of \( I_D \), unless the function \( \rho(x,z) \) is known to a sufficient degree of accuracy.

The following section is devoted to the description of a method for achieving the latter objective under laminar flow conditions of foregoing section II.B.d).

IV. AN ACCURATE METHOD FOR MEASURING THE MEAN POTENTIAL \( (\psi^*) \) AND TRANSVERSE ELECTRIC FIELD \( (E^*_{\text{T}}) \) ACROSS THE INVERSION LAYER

The function \( \rho(z) \) in (22) may be deduced from such bias conditions as result in low/null harmonic distortion components \( (D_2 \approx D_3 \approx 0) \) for \( I_D \), when sinusoidal signals of suitable amplitudes are applied to the transistor’s terminals simultaneously, thereby modulating \( E^*_{\text{T}} \), as set forth in [10],[11]: under these experimental conditions, nulls of \( D_3 \) occur uniquely because substrate effects counteract those of mobility, such that \( a_3 = D_3 = 0 \) in (23). When \( a_2 \approx D_2 \approx 0 \) simultaneously,
\( I_D \) is then (nearly) proportional to \( V_D \), hence the channel conductance constant, and the conducting channel obeys Ohm’s Law, which corresponds to laminar flow since constant \( (dz/dx) = (dz/dV)(dV/dx) = (dz/dV)E_0, \) per II.B.(d), yields proportionality between \( z(x) \) and \( V(x) \), hence between \( I_D \) and \( V_D \) per (8).

By contrast, traditional methods [2],[4] entail an extrapolation of the drain current characteristics at low drain–source potential, to a “threshold” voltage whose definition is neither well defined nor unique [8], thereby resulting in various inconsistencies. These, and other, derivative methods [14]–[17], have thus left unresolved the physical origins of the “universality” of mobility curves, among various related observations.

The method implemented below side–steps such inaccuracies and inconsistencies as incurred by the latter methods, by setting forth the objective criterion \( (D_2 \approx D_3 \approx 0) \) for determining the mean electric field \( E^*_T \) across the inversion layer from measurements of \( \psi^* \). This method furthermore yields \( \rho(z) \), \( \mu_{gs}(E^*_T) \), and the latter derivative function \( \theta_T \) accurately, thereby rendering these quantities in a manner that is mutually consistent, and verifiable by direct observation, as will be shown in the sequel.

A. Measuring the mean potential \( \psi^* \) across the inversion layer

The object of the following analysis is to determine the mean potential, \( \psi^* = b^*\phi_t \), across an inversion layer of extent \( z_{max} \), (from which \( E^*_T \), \( \eta \), and \( \theta_T \) follow), in terms of the known surface potential \( \psi_s = b\phi_b \) of the charge sheet model, as given by (25) below. Further details about experimental conditions, methods, and \( \theta_T \) are elaborated in section VI below. Confirmation that our observations correspond to laminar flow will be at hand after the necessary analytical and experimental evidence has been adduced (see section VI.B).

In the ensuing, we use standard MOSFET terminology for an n channel device, and denote \( V_B = (V_{SB} + 2\phi_F + b^*\phi_b) \), wherein \( \phi_F \) and \( \phi_t = kT/q \) are the Fermi and thermal potentials respectively.

The mean potential \( \psi^* \) across the inversion layer is:

\[
\psi^* = b^*\phi_t = \ln \left[ \left( \frac{2}{\gamma} \right) \sqrt{V_B} \left( \frac{1}{r} - 1 \right) \right]
\]

(24)

where \( r = \Delta V_B/\Delta V_{GS} \) is the ratio of increments of applied \( V_B \) and \( V_{GS} \) bias voltages which result in observed nulls of \( D_3 \), per (3) in [10], and \( \gamma \) is the body effect coefficient [8]. The mean, observed, value of \( r \) over the range of our data is 0.39. The corresponding, observed, values of \( \psi^* = (b^*\phi_t) \) range between \((1.45\phi_t)\) and \((1.65\phi_t)\), (see Fig.3), and the analytical values of the surface potential \( \psi_s = (b\phi_b) \) range between \((4.5\phi_t)\) and \((6.7\phi_t)\), as obtained by solving the transcendental [8, p.79]:

\[
\psi_s = V_{GS} - V_{T0} + \gamma \sqrt{2\phi_F - \gamma \sqrt{\psi_s + \phi_t \left[ e^{(\psi_s - V_B + \psi_t)/\phi_t} - 1 \right]}}
\]

(25)

for \((\psi_s - V_B) = (b\phi_b)\) to 0.1% accuracy for the observed bias conditions, with \( V = 0 \). The ratio of derivatives,

\[
\left( \frac{\partial\psi_s}{\partial V_{GS}} \right) / \left( \frac{\partial\psi_s}{\partial V_B} \right) = r
\]

yields the right-hand side of (24), which relates \( r \) to \( b^* \), thereby determining \( b^* \) and \( \psi^* \). The observed variations of the mean potential \( \psi^* \) across the inversion layer, and our analytical predictions for these (see section V below), are shown in Fig.3 for ready comparison with the empirical Schwarz–Russek model [3] (see appendix A2): the latter may be seen to be a reasonable approximation over the range of our data, namely the span of \( E^*_T \) ranging from onset of strong inversion through saturation of \( \psi_s \).

The observed ratio \((\psi^*/\psi_s)\) thus ranges between \((1.45)/(4.5) \approx (1/3)\) near to the onset of strong inversion, and \((1.65)/(6.7) \approx (1/4)\) near to the saturation of \( \psi_s \) (see Fig.4). A ratio of \((1/3)\) is expected from Poisson’s equation.
for a spatially uniform charge distribution, whereas a ratio of \((1/4)\) corresponds to a spatially linear distribution \(\rho(z) \sim z\). These results are suggestive of the analytical model to follow, which must account for the foregoing observations.

**B. Determining the mean transverse electric field \(E^*_T\) across the inversion layer**

The mean electric field \(E^*_T(V)\), which is modulated by an applied drain–source potential \(V = V_{ds}\), is obtained from \(Q_i(\psi_s)\) and \(Q_b(\psi_s)\) \cite{8, p.79}:

\[
Q_i(\psi_s) \approx -\gamma C'_{ox} \left[ \psi_s + \Phi_e(\psi_s - V_{th} + V)/\psi_t - \sqrt{\psi_s} \right]
\]

\[
Q_b(\psi_s) \approx -\gamma C'_{ox} \sqrt{\psi_s} + V
\]

wherein \(C'_{ox}\) is the gate oxide capacitance per unit area. The mean transverse electric field is then:

\[
E^*_T \approx - \frac{1}{\epsilon_{si}} \left[ \frac{Q_i(\psi_s)}{m(E^*_T)^{1/2}} + Q_b(\psi_s) \right]
\]

wherein \(0 \leq m(E^*_T) \leq 1\) is the power of \(z\) in \(\rho(z) \sim z^m\) per \((7)\), and to which corresponds \((1/2) \leq \eta(E^*_T) \leq (1/2)\), per \((28)\). The correspondence between the values of \(m\) and \(\eta\) extracted from our measurements of \((\psi^*/\psi_s)\), and the mean values of \(z\) resulting for the inversion layer are shown in Fig.5 below, wherein:

(a) \(m(E^*_T)\) varies approximately linearly with \((Q/Q_b)\) near to \((Q/Q_b) \approx 1\); i.e., for \(E^*_T \approx 0.2\) MV/cm.

(b) \(\eta(E^*_T) = 1/[m(E^*_T) + 2]\) is a related function of inversion level \((Q/Q_b)\).

(c) the ratio \((\psi^*/\psi_s)\) varies between \((1/3)\) and \((1/4)\) (as per Fig. 4 above).

(d) the mean inversion layer thickness \(<z> = (\psi^*/E^*_T)\) resulting from our observations of \(\psi^*\) and \(E^*_T\) (per Fig.3) is commensurate with the predictions of Stern and Howard \cite{5}, and Schwarz and Russek \cite{3}.

---

**Fig. 3** Observed mean potential \(\psi^*\) across inversion layer and predictions of \((33)\). Comparison with Schwarz–Russek’s model \cite{3}, (see appendix A2), over a span of \(E^*_T\) ranging from onset of strong inversion \([\psi_s \approx 4.5\phi_i\text{ above } (V_{SB} + 2\phi_T)]\) through saturation of the surface potential \([\psi_s \approx 6.7\phi_i\text{ above } (V_{SB} + 2\phi_T)]\). All measurements pertain to long channel devices \((L \approx 100\mu m)\) with \(E^*_T\) near to \(0.2\) MV/cm, where \(\mu_{eff}(E^*_T) \sim (E^*_T)^{1/5}\) per “universal” mobility curves.

**Fig. 4** Observed ratio of mean to surface potential \((\psi^*/\psi_s)\), and corresponding inversion layer carrier concentration of commensurate charge sheet model \((N_i = Q/q)\), both as functions of the ratio \((Q/Q_b)\), i.e. with increasing \(E^*_T\); the model’s values are obtained from \((33)\). The range of observation is from onset of strong inversion through saturation of \(\psi_s\).
V. ANALYTICAL MODEL OF THE INVERSION LAYER’S CHARGE DENSITY PROFILE DEDUCED FROM EXPERIMENTAL OBSERVATIONS

In view of the observations described in section IV.A above we posit, for an inversion layer of extent \( z_{\text{max}} \) supporting a given, total charge \( Q_c = Q_t + Q_b \):

\[
b(z) \approx b^*(\frac{z}{z_{\text{max}}})^m; \quad 0 \leq z \leq z_{\text{max}}
\]

(29)

such that \( \psi^* = b^*\phi_t \) represents the mean potential across \( z_{\text{max}} \), since \( b(z_{\text{max}}) = b^* \) when \( z = z_{\text{max}} \) and, ignoring minority holes,

\[
\rho(z) \approx N_a \left[ e^{b^*(\frac{z}{z_{\text{max}}})^m} - 1 \right]
\]

(30)

for the corresponding mobile charge density within the inversion layer, pending experimental corroboration, which is equivalent to a linear approximation \( \rho(z) \sim z \) when \( m \approx 1 \) and \( (z/z_{\text{max}}) \ll 1 \) (see Figs. 6(b) and (c)).

Equating the inversion charge of the sheet model, \( Q_s(\psi_s) \) per (26), to that posited by (30) for an inversion layer of finite extent \( z_{\text{max}} \),

\[
Q_t = \int_0^{z_{\text{max}}} q\rho(z) \, dz \\n\approx \int_0^{z_{\text{max}}} qN_a \left[ e^{\psi(z/z_{\text{max}})^m} - 1 \right] \, dz
\]

(31)

establishes an \textit{implicit} analytical relationship between the surface potential \( \psi_s = b^*\phi_t \) of the sheet model, per (25), and the desired mean potential \( \psi^* = b^*\phi_t \), derived from (31). Integrating Poisson’s equation with \( \rho(z) \) thus chosen, twice to obtain the surface potential \( \psi_s = (b^*\phi_t) \), once more to obtain the mean potential \( \psi^* = (b^*\phi_t) \) over \([0, z_{\text{max}}] \), thus yields:

\[
\frac{\psi^*}{\psi_s} = \frac{b^*}{b_s} = \frac{1}{m+3}
\]

(32)

whereby \( \eta = 1/(m+2) \) for the mean electric field \( E^*\) in (28), which is thus obtained as an intermediate result.

For \( 0 \leq m(E^*) \leq 1 \), the parameter \( \eta \) is thus expected to vary continuously between \((1/2)\) and \((1/3)\), per (28), as observed for standard devices with <100> surfaces, and predictably as a function of the relative magnitudes of inversion \((Q_t)\) and bulk \((Q_b)\) charges, hence as a function of the level of inversion (see Fig. 5). Because this transition is especially prominent in the range of bias conditions whereby neither \( Q_t \) nor \( Q_b \) is overwhelmingly dominant over the other, our measurements are concentrated on this particular range, namely \( 0.4 \leq (Q/Q_b) \leq 1.8 \) (see Figs. 3–5,7), which corresponds, approximately, to \( 0.15 \text{ MV/cm} \leq E^* \leq 0.25 \text{ MV/cm} \), where laminar flow obtains, and over which we observe \( D\beta \) nulls.

Proceeding thus, we derive in closed form:
which yields $b^*$ readily, given $b_\text{s} = (\psi_s/\phi_t)$ as obtained from (25). The ratio (33) thus reduces to the expected value $(3!)/(4!) = (1/4)$ in the charge–sheet limit $b^* \to 0$, and matches observations of $(b^*/b_\text{s}) = (\psi^*/\psi_s)$ nearly perfectly for $(Q/Q_b) \geq 1$, where $m \approx 1$ as observed in Fig.4, and otherwise to within ±5%, on an average, over the full observation span. The left–hand side of (33) indicates that the ratio $(b^*/b_\text{s}) = (\psi^*/\psi_s)$ decreases as $\psi_s$ increases, i.e., as inversion level increases: physically, this means that as $E^*_T$ increases, the increasing electron charge contributes mainly to an inversion layer’s mobile $Q_i$ concentrated nearer to the oxide, while the fixed depletion charge $Q_b$ remains relatively unaltered. These trends are indicated by the arrows attached to the various quantities depicted in Fig.6(c).

Fig. 6(a) Uniform inversion layer charge density; $|Q_i/Q_b| \approx 0.4$, $\eta = (1/2)$, $(\psi^*/\psi_s) = (1/3)$, $\psi^* = b^*\phi_t \approx 1.45\phi_t \approx 37mV$, $\psi_s = 111mV$ (see Figs.3 and 4); this condition is observed at low values of $|Q/Q_b| < 1$ where the “flat” depletion charge profile for $Q_b$ is dominant. The standard depiction of a “flat” inversion charge profile between 0 and $z_{\text{max}}$ becomes physically unrealistic as $|Q/Q_b|$ increases, because there must be a finite transition region near to the oxide interface wherein $\rho(z)$ drops towards zero, which is a critical boundary condition of the system. The assumption of a fixed $\eta = (1/2)$ thus becomes increasingly erroneous as $|Q/Q_b|$ increases past unity (see Fig.7), which is consequential towards the accurate evaluation of $\mu_{\text{eff}}$ (per section III).

(b) $|Q/Q_b| \approx 1.8$, $\eta = (1/3)$, $(\psi^*/\psi_s) = (1/4)$, $\psi^* = b^*\phi_t \approx 1.65\phi_t \approx 43mV$, $\psi_s \approx 172mV$; unlike in (a), the boundary condition $\rho(z=0) = 0$ holds for our model, per (29) and (30), wherein $\rho(z) \sim z$ is deduced from our measurements of $(\psi^*/\psi_s)$, and is here a valid approximation of:

(c) the Stern–Howard model [5] (see appendix A2) for an inversion layer charge density $\rho(z) \sim (z^2/2)e^{-\beta z}$, $\beta \approx (1/15)$, whose mean value for $<z> \approx (3/2\beta) \approx 22.5\text{Å}$, and $z_{\text{max}} \approx 45\text{Å}$. The comparison with our model — the right triangle whose hypotenuse’s slope is $(\rho_{z_{\text{max}}})$ per (7) — corresponds to an equivalent charge–sheet density near to $10^{12} \text{cm}^{-2}$ (see Fig.4). The areas, between 0 and $z_{\text{max}}$, under the Stern–Howard curve, and our model’s, are nominally equal, and correspond to
the total inversion charge \( Q_i \). The potential values indicated correspond to condition (b). The vertical arrows attached to \( \beta \) (\( \rho \neq \rho_{ \max } \)), and \( \rho_{ \max } \) indicate the trend in these quantities with increasing inversion level, or increasing \( E_{*T}^\ast \).

The ratio of mean electric field \( (E_{*\rho}^\ast) \) across the inversion layer corresponding to \( \rho(z) \) per \( (30) \), to that \( (E_{*u}^\ast) \) of a uniform charge distribution supporting the same total charge \( Q_i \), and obtained as an intermediate result is:

\[
\frac{E_{*\rho}^\ast}{E_{*u}^\ast} = \left[ \frac{2}{b^\ast} \left( \frac{b^\ast - 1}{b^\ast} - \left( \frac{b^\ast}{2!} \right)^2 \right) \right]^{\frac{3}{2}}
\]

whose charge-sheet limit is indeed \((2/3)\), as expected when \( b^\ast \) approaches 0:

\[
\lim_{b^\ast \to 0} \frac{E_{*\rho}^\ast}{E_{*u}^\ast} = \left[ \frac{2}{b^\ast} \left( \frac{b^\ast - 1}{b^\ast} - \left( \frac{b^\ast}{2!} \right)^2 \right) \right]^{\frac{3}{2}} = \frac{2}{3}
\]

in which case the mean transverse electric field \( E_{*T}^\ast \), per \( (28) \), is:

\[
- \varepsilon_{sf} E_{*T}^\ast = \frac{Q_i}{2} \left( \frac{E_{*u}^\ast}{E_{*u}^\ast} \right) + Q_b
\]

wherein the ratio \( (E_{*\rho}^\ast / 2E_{*u}^\ast) = \eta \) yields the expected result \((1/2)(2/3) = (1/3)\) for an inversion layer varying linearly with \( z \), or \( m = 1 \) in \((29)-(32)\), as indicated above.

The foregoing results thus indicate how \( \psi^\ast, \eta^\ast, E_{*T}^\ast \), and \( \eta \) are related, which are determined by the level of inversion (i.e., the relative magnitudes of \( Q_i \) and \( Q_b \)), and by the charge density profile within the inversion layer \( \rho(z) \). The traditional assumption of a value \( \eta = (1/2) \) fixed \( \text{a priori} \) — as illustrated in Fig.6(a), and corresponding to constant \( \rho(z) \) — which does not account for these effects, may thereby be expected to give rise to observable discrepancies left wanting of clarification and reconciliation, as indeed noted in [2]–[4].

VI. TRANSVERSE–FIELD MODULATION PARAMETER \( (\theta_T) \)

Conceptually, \( \theta_T \) represents \( \text{a first order Taylor series expansion of} \) the expression for the mobility \( \mu_{\text{eff}}(E_{*T}^\ast) \) associated with the MOSFET’s drain current: the constant \( \mu_{\text{eff}} \) in \((23)\) is replaced by \( \mu_{\text{eff}}[1 + \theta_T \cdot V(x)] \), where \( V(x) \) is the potential along the channel’s direction \( (x) \), and \( [\theta_T \cdot V(x)] \ll 1 \). The drain current \( I_D \) is then obtained by integrating the function \([1 + \theta_T \cdot V(x)] \cdot \rho[V(x)]\) from 0 to \( L \), in the conventional manner \([10],[11]\). As such, the single parameter \( \theta_T \) effectively subsumes the “averaging”, per \((22)\), of the effect of \( E_{*T}^\ast \) on the conductivities of the infinitesimal charge sheets in the \( z \) direction (between \( 0 \) and \( z_{\max} \)).

A. Experimental conditions

All measurements were made on long \( n \)-channel devices \((L \approx 100 \mu m) \) fabricated in a standard, industrial, CMOS process. The range of \( E_{*T}^\ast \) for our data is centered near to \( 0.2 \) MV/cm, where universal mobility \( \mu_{\text{eff}}(E_{*T}^\ast) \sim (E_{*T}^\ast)^{-1/3} \) behavior prevails, as reported in the literature \([2],[4]\).

While our over–arching objective is to assess the validity and limits of the foregoing analytical model, we summarize the substantial quantitative agreement of our observations with this model in the next subsection.

B. Experimental results

\( \theta_T \) was extracted from nulls of third harmonic current distortion components \((D_2 \approx 0, D_3 = 0)^5 \) in the triode region of operation, with applied \( V_{DS} = V_{\text{peak}} \sin(2\pi ft) \), where \( V_{\text{peak}} < V_{DS\text{sat}}, f = 1 \text{kHz}, \) and \( (E_{*T}^\ast / E_{\rho}^\ast) \geq 2000, \) \([10]\):

\[
\text{signal amplitude conditions (see II.B in [10])}, \text{which is an advantage for the purpose at hand.}
\]

---

5 Although \( \theta_T \) may likewise be obtained from nulls of \( D_3 \), those of \( D_2 \) are much more sensitive to bias and
\[ \theta_T(E^*_T) = \left( \frac{y}{8} \right) \left[ \frac{V_B}{1 - \frac{\sqrt{V_B}}{\sqrt{V_B}}} \right]^3 \] (37)

Fig. 7 below compares these observations to the predictions of the foregoing model, as set forth in section IV.B, whereby \( E^*_T \) is as per (28):

\[ \theta_T(V) |_{V=0} = \left( \frac{1}{\mu} \right) \frac{d\mu}{dV} |_{V=0} = \left( \frac{1}{\mu} \right) \frac{\partial E^*_T}{\partial \mu} \frac{dE^*_T}{dV} |_{V=0} \]

\[ \approx \frac{1}{3} \frac{1}{V_{GS} - V_{T0} + \sqrt{V_B} + 2\Phi_F} \] (38)

While the agreement of (38) is within \( \approx \pm 5\% \) throughout the observation range, the qualitative effect of varying values of \( m \) (hence of \( \eta \)) is most conspicuous, which illustrates the inadequacy of a uniform inversion layer model (the curve labeled \( m = 0 \), or fixed \( \eta = 1/2 \)): the data deviate from the latter, increasingly as \( m \) increases between 0.2 and 1.05 (i.e., \( \eta \) decreasing between \( 1/2 \) and \( 1/3 \)) approximately, as also illustrated in Fig.5, whereas (38) accounts for these effects while rendering \( \theta_T \) explicitly in terms of terminal potentials and device parameters.

\[ \theta_T(V) = \left( \frac{1}{\mu} \right) \frac{d\mu}{dV} = \left( \frac{1}{\mu} \frac{\partial E^*_T}{\partial \psi^*} \frac{\partial \psi^*}{\partial \nu} \frac{d\nu}{dx} \right) dV \]

\[ \approx \frac{-1}{3E^*_T} \left[ \frac{1}{<z>} \left( -\frac{E_{II}}{E_T^*} \right) \frac{1}{E_{II}} \right] \sim \frac{(E_T^*)^2}{(E_T^*)^2} \sim \frac{(E_T^*)^5}{(E_T^*)^7} \] (39)

which is indeed observed in Fig. 7. In view of our analogy with a river’s flow, the same result may be deduced from (4) and/or (9), wherein \( v = \mu \mu_{eff} \) from (21), such that:

\[ \frac{1}{\mu} \frac{d\nu}{dE^*_T} = \frac{1}{\nu} \frac{dE^*_T}{dE_T^*} d\nu = \frac{dz}{dE_T^*} \]

\[ \sim (E_T^*)^{-\frac{2}{7}} \left[ E_T^* \right]^{-\frac{1}{3}} \sim - (E_T^*)^{-\frac{4}{5}} \]

In view of (11), moreover, we may determine independently of section II.C.2, the absence of \( D^3 = 0 \) observations for \( E^*_T \geq 0.25 \) MV/cm for our devices, that is, why laminar flow conditions no longer prevail beyond this limit: at the upper end of our data range, where \( E^*_T \approx 0.25 \) MV/cm, we have \( z \approx 15 \mu \text{m} \), while \( z \) is decreasing at the nearly constant rate \( dE_T^*/dz \approx -10 \mu \text{m}/(0.1 \text{MV/cm}) \), (see Fig.5). For this value of \( z \), the non-unity term in the denominator of (11) is approximately \( 1/30 \), hence negligible. The same term multiplied by a factor of \( K_T(E^*_T/E_{II}) \) appears in the numerator of (11), and increases as \( 1/z \) with decreasing \( z \). Since, moreover, \( (E^*_T/E_{II}) \) is of the order of \( 4 \times 10^3 \), and the viscosity of
(uncompressed) air is $1.81 \times 10^{-4}$ poise at 20°C [18], the non-unity term is of the order of 0.7, hence the magnitude of the rate $dz/dE_{*r}$ increasing as $(1/z)$ is near to 1.7 times its previous value. Hence $z$ may be expected to drop by a further 10Å (i.e. from 15Å to 5Å) when $E_{*r} \leq 0.35$ MV/cm, in good agreement with the estimate provided by (15), and as illustrated in Figs.2(a),(b). While the foregoing argument is tentative, the requisite conditions for non–laminar flow may plausibly ensue, even neglecting ($K_F$)’s scattering components — to the extent that the electron “gas” is compressed into a layer $<z>$ of the order of silicon’s lattice constant ($\approx 5.4Å$), and of the $\approx \pm 5Å$ amplitude of typical Si/oxide “terrain” irregularities, as observed by electron microscopy in good–quality devices.

VII. SUMMARY

We have elucidated physical origins of the dominant phenomena governing current conduction in the MOSFET’s inversion layer, below saturation, by comparing the latter to a broad river’s flow in a gravitational potential field, whereby energy conservation considerations are fundamental.

We have drawn upon such familiar concepts of fluid mechanics as readily apprehended by common sense and familiar experience, to derive analytical expressions promoting physical insight into subtle, intertwined, effects subsumed by the charge sheet, simplifying, assumptions of standard textbooks and circuit simulation models of the MOSFET’s inversion layer. Such interrelationships have been made manifest as obtain between the inversion layer’s charge density profile $\rho(z)$, its mean $<z>$ and maximum ($z_{\text{max}}$) extents, the low–field effective mobility ($\mu_{\text{eff}}$), and the influence of scattering ($K_F$), and quantum mechanical, effects.

A sensitive method for measuring the mean potential ($\psi^*$) across the inversion layer was applied under conditions of harmonic distortion component nulls ($D_2 \approx D_3 \approx 0$) of the drain current under sinusoidal excitation, below saturation. We showed, thereby, how $\rho(z)$ may be deduced accurately from the ratio ($\psi^*/\psi_s$), as the surface potential ($\psi_s$) increases from onset of strong inversion towards saturation, and how such information underpins the physical origins of disparate observations of long standing. We found and established:

1. Non–uniform, or non–constant, $\rho(z)$ causes variations of $1/3 \leq \eta \leq 1/2$, which determine the mean transverse electric field $E_{*r} = -(1/e_{\text{Si}})[\eta Q_i + Q_b]$ in terms of the inversion $Q_i$, and bulk $Q_b$, charges.

2. Laminar current flow and maximum drift velocity, which prevail when $D_2 \approx D_3 \approx 0$, result from energy conservation, concurrently with the gradual channel approximation entailing $<z>$ varying as $\sim (E_{*r})^{-1/3}$. The latter relationship is nearly identical to that obtained by Stern and Howard [5] from solutions of Schrödinger’s equation — which embodies conservation principles — as constrained by the boundary conditions pertinent to the system. Moreover, whereas the latter solutions resulted in a fixed value $\eta = (11/32)$, our measurements indicate that $\eta$ varies with the level of inversion as per (1) above. The value $(11/32) \approx (1/3)$ nearly corresponds to the result of our model $\rho(z) \sim z$, which is a close approximation of the Stern–Howard model, and otherwise substantially in accord therewith.

3. The behavior of “universal mobility” curves, whereby $\mu_{\text{eff}}$ also varies as $\sim (E_{*r})^{-1/3}$, is consequence and manifestation of laminar flow conditions. We established this result, analytically, by invoking energy and momentum conservation, as well as the prevalence of quantum–mechanical effects.

4. An analytical expression for an upper limit to the transverse electric field for which laminar flow may be expected, $(E_{*r})_{\text{MAX}}(300K) \approx 0.35$ MV/cm, which is a function of Si/oxide interface “terrain roughness” amplitude and fundamental constants only, in quasi-quantitative
agreement with extant observations with respect to temperature and surface orientation for both electrons and holes, as available in the literature.

(5) The higher order, derivative, parameter \( \theta \) quantifying the modulation of \( \mu' \) by \( (E^*)^{\frac{5}{3}} \) varies as \( \sim (E^* T)^{-\frac{5}{3}} \) under laminar flow conditions. The close agreement between our analytical predictions and experimental observations provide corroborating evidence for the validity of our interpretations of the physical effects underlying the foregoing models, and for the consistency of the measurements thereto related.

(6) The widely reported degradation of \( \mu' \) for \( E^* T \geq 0.5 \) MV/cm approximately, and/or where \( <z> \) is commensurate with irregular interface “terrains” (\( \approx 5 \) Å), is symptomatic of non-laminar current flow along some part of the inversion layer. While the parameters characterizing the presumed models for phonon, surface roughness, and Coulomb scattering have, heretofore, been interpreted under the paradigm of a fixed value for \( \eta \) (set a priori), and thus extracted from deviations from universal mobility behavior \( \mu' \sim (E^* T)^{-\frac{5}{3}} \), data similar to Takagi et al. may warrant reinterpretation under the foregoing analyses and results, especially so with respect to the range where \( E^* T \geq (E^* T)_{\text{MAX}} \) 300K) \( \approx 0.34 \) MV/cm where laminar flow ceases, and the influence of terrain roughness becomes increasingly dominant: the task of characterizing the function \( K_F(x, z, E_0, E^* T, \text{etc.}) \) pertaining to such interface characteristics and/or to turbulent flow may thereby be simplified since it need not account for the \( (E^* T)^{-\frac{5}{3}} \) behavior of mobility under laminar conditions.

APPENDIXES

A1. Functional correspondence between parameters of the gravitational and electrical fields pertaining, respectively, to the river and to the MOSFET’s inversion layer. Note that \( z \) represents depth of river and depth (thickness) of inversion layer, as indicated in Fig. 1.

| Gravitational                  | Electrical                     |
|-------------------------------|--------------------------------|
| Particle mass \( m \)          | Particle (electron) mass \( m \) |
| Acceleration \( g \)           | Charge density \( \rho(z) \)   |
| Time \( t \)                   | Position \( z \)                |
| Position \( z(t) = \int g dt^2 + z(t_0) \) | Potential \( \psi(z) \) |
| Velocity \( v = dz/dt \)       | Electric field \( E = -\psi/dz \) |
| Newton (2nd Law) \( d^2z/dt^2 = g \) | Poisson \( E = \psi_{\text{total}}/\varepsilon_0 \) (electrostatics) |
| Water volume in \( \Delta x \) \( Q = \int \rho(x, z) dz \) | Total charge in \( \Delta x \) \( Q = q \int \rho(x, z) dz \) |
| Flow rate across \( \Delta x \) \( R = d(Q)/dt \) | Current across \( \Delta x \) \( I = d(Q)/dt \) |

A2. The Stern–Howard [5], and Schwarz–Russek models [3]:

\[
[T - q \phi(z) - E] \xi = 0 \quad (A2.1)
\]

where \( T \) is a kinetic energy operator and \( \phi(z) \) the electrostatic potential obeying Poisson’s equation (see appendix A1 above), with \( \rho(z) \) the sum of the densities of inversion and fixed charges in the depletion layer. The boundary condition \( \xi(0) = 0 \) was used for the oxide–silicon interface, where a
small envelope amplitude $\mathcal{E}(0)$ is expected, while those on $\phi(z)$ were: $\phi(z) \to 0$ as $z \to \infty$, and

$$
\frac{\epsilon_{ox}}{\epsilon_{sl}} \frac{d \phi}{dz}|_{z=0^-} = \epsilon_{sl} \frac{d \phi}{dz}|_{z=0^+}
$$

The density of states for electrons $\rho(E)$ results from these considerations which, in turn, leads to a desired, self-consistent, approximate solution $\rho(z) \sim (1/2)\beta^2 z^2 e^{-\beta z}$, as shown in Fig.6(c), wherein the constant $\beta$ is:

$$
\beta = \left[ \frac{48\pi m^2}{\epsilon_{sl} h^2} \left( Q_b + \frac{11}{32} Q_i \right) \right]^{1/3} \tag{A2.2}
$$

and the factor $(11/32)$ is a fixed value for the factor $\eta$ in our expression for $E^*_T$ per (28). The average value of the inversion layer thickness $z$ weighted by the charge distribution $\rho(z)$ is then $<z> \approx 3(2\beta)^{-1} \approx 22.5 \text{Å}$, which is a maximum when $Q_i \to 0$, and considerably thinner than the depletion layer, which they evaluated at 1.2µm for their process, even in this limiting condition. The general agreement between the results of the Stern–Howard model and ours — except for the latter’ variable $\eta$ per our experimental observations — is simply owed to the fact that (A.2.1) and (10) are equivalent starting premises embodying conservation laws for the system at hand.

Schwarz and Russek formulated the following empirical expression for a uniform inversion layer thickness $z_{\text{max}}$, which they postulated to correspond to $(z_c + z_{\text{out}} \approx 2z_c)$ as per (12b) and (13b) above, and by matching $z_{\text{max}} \approx 45 \text{Å}$ per Stern and Howard (Fig.6):

$$
z_{\text{max}}(E^*_T) = \frac{1.56 l_t}{E^*_T} + \frac{1.24 \times 10^{-5}}{(E^*_T)^2} \tag{A2.3}
$$

where $\phi = kT/q \approx 26 \text{mV}$ at room $T$, and $E^*_T$, evaluated with fixed $\eta = \frac{1}{2}$ is in units of V/cm. The mean potential across the inversion layer predicted by (A2.3), or $(1/2)z_{\text{max}} E^*_T$, is compared to our observations thereof, and to that of our analytical model in Fig.3, while our model’s predictions for $<z>$ are compared to (A2.3) in Fig.5. The second term on the right hand side of (A2.3), which they attributed to “quantum–mechanical broadening”, and based on [5], served two purposes: (a) it introduced the $(E^*_T)^{-1/3}$ dependence of $z$ on $E^*_T$ which they found necessary to match the observed variations of $\mu_{eff}(E^*_T)$ as $(E^*_T)^{-1/3}$ when this term dominates the first (i.e., at higher $E^*_T$ values) and, (b) it was “calibrated” by the factor $(1.24 \times 10^{-5})$ such that the first and second terms on the right hand side of (A2.3) are approximately equal. The overall result was that (A2.3) achieved, approximately, the desired doubling of the value of $z_c$, and a potential across the inversion layer near to $1.5\phi_c \approx 39 \text{mV}$ throughout the data range, as illustrated by Fig.3. The arguments presented in section II.C.3 above, which invoke considerations of energy and momentum conservation (as “built–into” Schrödinger’s equation), namely $z \sim (E^*_T)^{-1/3}$, are thus in quantitative and qualitative agreement with Stern and Howard’s (A2.2), while only in quantitative agreement with Schwarz and Russek’s empirical (A2.3), the latter’ lack of theoretical justification notwithstanding (which they acknowledged).

REFERENCES

[1] B. Russell, An outline of Philosophy, Meridian Books, New York, 1960, p.163.
[2] S. C. Sun and J.D. Plummer, “Electron mobility in inversion and accumulation layers on thermally oxidized silicon surfaces”, IEEE Trans. Electron Devices, vol. ED–27, no.8, pp. 1497–1508, August 1980.
[3] S. A. Schwarz and S. E. Russek, “Semi–empirical equations for electron velocity in Silicon: Part II — MOS inversion layers “, IEEE Trans. Electron Devices, vol. ED–30, no.12, pp. 1634–1639, December 1983.
[4] S. Takagi, A. Toriumi, M. Iwase, and H. Tango, “On the universality of inversion layer mobility in Si MOSFET’s: Part – I Effects of substrate impurity concentration, Part – II, Effects of surface orientation”, IEEE Trans. Electron Devices, vol. ED–41, no.12, pp. 2357–2368, December 1994.
[5] F. Stern and W.E. Howard, “Properties of semiconductor surface inversion layers in the electric quantum limit”, Physical Review, vol. 163, no.3, pp. 816–835, November 1967.
[6] T. Von Karman and M. A. Biot, Mathematical Methods in Engineering, McGraw–Hill, New York and London, 1940.
[7] C. Séquin, in C. Mead and L. Conway, Introduction to VLSI systems, Addison Wesley, 1980.
[8] Y. P. Tsividis, Operation and Modeling of the MOS Transistor, McGraw–Hill, 1987.
[9] S. M. Sze, Physics of Semiconductor Devices, John Wiley & Sons, 1981.
[10] F. S. Shoucair and W. R. Patterson, “Analysis and Modeling of nonlinearities in VLSI MOSFETs including substrate effects”, IEEE Trans. Electron Devices, vol. ED–40, no.10, pp. 1760–1767, October 1993.
[11] F. S. Shoucair, “Modeling, decoupling and suppression of MOSFET distortion components”,

MOSFET Inversion layer, IEEE Trans. ED.  
F. S. Shoucair, 25 November 2021, p.19/20
MOSFET Inversion layer, IEEE Trans. ED.

F. S. Shoucair, 25 November 2021, p.20/20

F. S. Shoucair received a Ph.D. in electrical engineering from Columbia University, NY, USA. etc....