Substrate Current at Cryogenic Temperatures: Measurements and a Two-Dimensional Model for CMOS Technology

ALBERT K. HENNING, MEMBER, IEEE, NELSON N. CHAN, MEMBER, IEEE, JEFFREY T. WATT, AND JAMES D. PLUMMER, FELLOW, IEEE

Abstract—**This work characterizes the temperature, channel length, and voltage dependences of substrate current, and presents a local model describing this behavior using Shockley's lucky electron (LE) model as a basis. For n-channel (p-channel) devices, the model is extended using a Maxwell–Boltzmann (MB) distribution of hot-electron (hole) energies above (below) the conduction (valence) band minimum (maximum). The model has been implemented in CADDET, a 2-D device simulator, and is able to explain all of the important features of substrate current which have been reported to date. The model is discussed in the context of works which look at both the local and physical nature of the impact ionization phenomenon. Based on this discussion, the model's parameters are shown to have a solid physical basis, requiring no reliance on curve fitting. The agreement between data and simulations thus enhances physical understanding of substrate current in MOSFET's, and warrants confident design of CMOS technologies for cryogenic operation.**

I. Introduction

SILICON is the material of choice for the fabrication of high circuit density, low defect density, and high-speed integrated devices. CMOS technology provides the additional advantage of low power dissipation. Device performance enhancement can be obtained through liquid-nitrogen-temperature operation [1], [2]. CMOS device enhancements have been shown in particular [3]–[6], and specific system enhancement factors have also recently been reported [7].

However, low-temperature operation may exacerbate the generation of substrate current by impact ionization. This in turn correlates to possible increased device degradation [8], [9], under either dc or dynamic stress. Reliable device design thus requires minimization of substrate current I_B, making an accurate 2-D model of I_B extremely attractive. In particular, I_B is far easier to monitor than gate current (I_G), or the amount and location of trapped oxide charge—yet it provides a measure of device degradation caused by I_g, due in part to the correlation between I_B and I_G [10]. Once I_B is well-predicted, one can move toward the more complicated tasks of measuring and modeling I_G and insulator trapping and trap generation, toward a full 2-D dc device simulator.

In initiating this task, then, we characterized I_B for a CMOS technology. Modeling efforts began with Shockley's LE model [11]. However, one of our goals was to characterize the technology, to determine if reliability problems related to I_B would occur at liquid-nitrogen temperature. That is, we wished to determine if cryogenic I_B considerations would constrain power supply voltages to too low a level. During these temperature measurements, it became clear the existing physical models were inadequate. We then used temperature as a tool to differentiate between appropriate physical models. Finally, we used the model to address the issue of power supply voltage for a submicrometer liquid-nitrogen-temperature CMOS technology based on these I_B considerations.

II. Experimental Details

CMOS transistors were fabricated in the Stanford IC Lab using a conventional n-well process [12] with t_{ox} = 385 Å and 0.85 μm < L_e < 25 μm. n^+ poly and boron channel implants were used for both device types; the effect of the compensating implant in the p-channel can be important [13] at cryogenic temperatures. Measured threshold voltages were V_{TN} = 0.35 V and V_{TP} = −1.2 V for T = 293 K. Low V_D transconductance and V_T were monitored before and after I–V measurements at each temperature to ensure that the measurements did not degrade the device. Substrate current was measured for 0 < V_G < 5 V, 1.5 < V_D < 5 V and 77 < T < 300 K. The n-channel devices use a standard As source–drain, with no LDD structure: as shown in previous work [14], certain n-channel LDD structures can lead to serious problems for liquid-nitrogen-temperature operation, as trapped charge above the LDD region, combined with freezeout effects, turns off the device channel. The p-channel devices used conventional boron source–drain regions, with $x_j \simeq 0.35$ μm.

III. Measurement Results

The impact ionization process leading to the measurement of substrate current is shown in Fig. 1 for an n-channel device. Channel electrons enter the high-field region

Manuscript received June 10, 1986; revised August 20, 1986. This work was supported in part by the JSEP under Contract DAAG29-84-K-0047.

A. K. Henning, J. T. Watt, and J. D. Plummer are with the Integrated Circuits Laboratory, Stanford University, Stanford, CA 94305.

N. N. Chan is with Intel Corporation, Santa Clara, CA 95051.

IEEE Log Number 8611167.

0018-9383/87/0100-0064$01.00 © 1987 IEEE

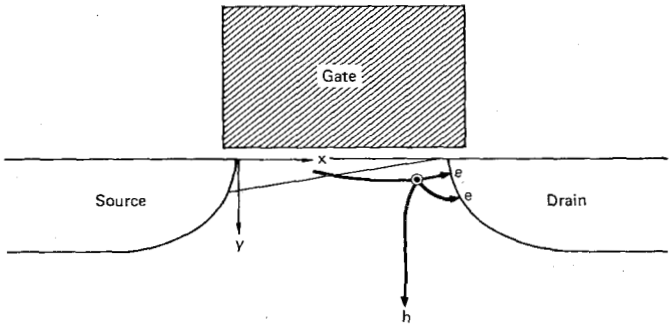


Fig. 1. Schematic of impact ionization current, leading to measurement of substrate current, in an n-channel MOSFET.

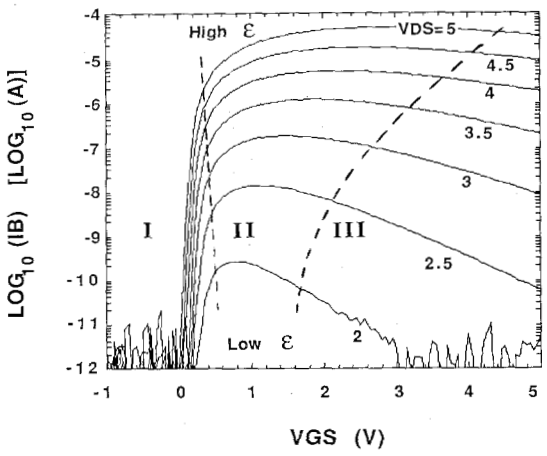


Fig. 2. Substrate current versus gate voltage for an n-channel MOSFET, at various drain voltages. $L_e = 0.85$ μm; $T = 77$ K. "High" and "Low" refer to V_D regions with different temperature coefficients for I_{BMAX} versus T (see text for discussion).

of the device, acquiring enough energy to break a Si–Si bond (nominally E_g, the silicon bandgap). A number of these carriers will in fact break a bond, generating free hole–electron pairs. The holes are swept into the body of the device, where they are measured as substrate current. A mirrored situation holds for p-channel devices.

Fig. 2 shows I_B data for an n-channel FET with $L_e = 0.85$ μm at 77 K. Region I ($V_G < V_T$) indicates the expected exponential behavior of I_B with increasing V_G. In this subthreshold region, I_D is an exponential function of V_G, and I_B is linearly related to I_D (see discussions of the model below). Region II ($V_D > V_G - V_T$) is the peak substrate current regime, where the field in the pinchoff region (and, thus, the number of channel carriers with energy exceeding E_g) has reached a maximum for a particular value of V_D. Region III ($V_D < V_G - V_T$) shows the expected decrease in I_B as the device goes out of saturation and the field in the pinchoff region diminishes. A good model should predict I_B accurately in all three regions.

Any model must also explain the effect first noted by Eitan *et al.* [15], which is demonstrated in Fig. 3 for our n-channel devices. At large drain voltages, we see that the peak I_B (normalized to I_D as in [10] to remove the temperature dependence of the channel current itself) for a device with $L_e = 1.15$ μm, increases with decreasing

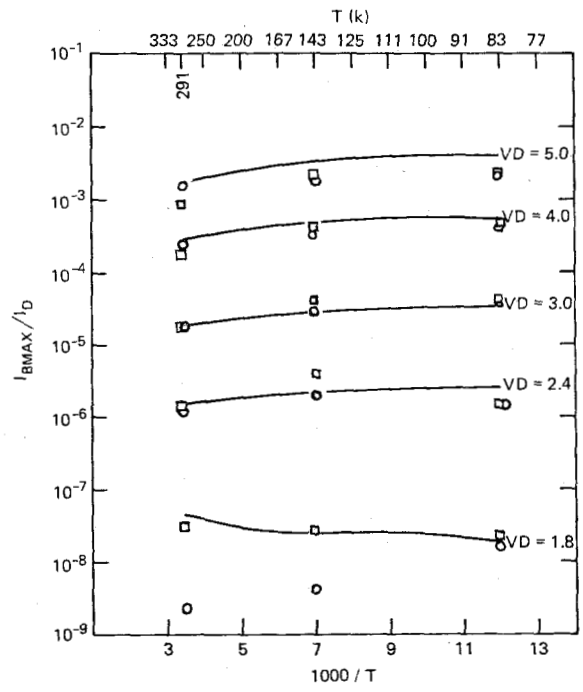


Fig. 3. Peak substrate current versus temperature for an n-channel MOSFET. $L_e = 1.15$ μm. ○ = LE model alone (after [26]); □ = MB plus LE model.

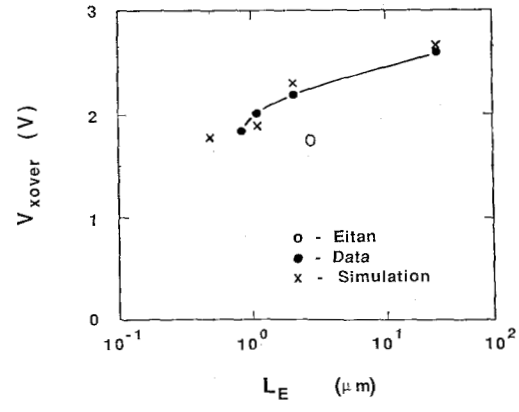


Fig. 4. Crossover voltage versus channel length for n-channel devices. (•, ×): $t_{ox} = 385$ Å. ○, $t_{ox} \approx 250$ Å.

temperature. However, at lower drain biases, the opposite is true. This is precisely the effect one would want in a CMOS technology designed for cryogenic operation to avoid hot-carrier degradation effects. (Operation of a cryogenic CMOS technology above this critical supply voltage might also be possible, provided carrier transport across the Si–insulator interface, and trapping in the insulator itself, are reduced.) We call that V_D for which I_{BMAX} is roughly independent of temperature V_{xover}. V_{xover} equals approximately 2 V in Fig. 3.

We have extended Eitan's work, and plotted V_{xover} versus L_e for our n-channel devices in Fig. 4, from which one may infer a power supply voltage of slightly less than 2 V is required for a cryogenic CMOS technology designed so that hot-carrier effects (as manifested by substrate current I_B) are no worse at low T than at room T.

This inference is accompanied by a caution, however, since other criteria may be important when determining the reliability of a technology allowing scaled temperature and power supply. In particular, if one looks strictly at I_G and requires it to be constant as a reliability constraint, then the power supply may be considerably higher than 2 V for a cryogenic technology [6].

We note that we were unable to observe V_{xover} directly in our p-channel devices, due to a minimum measurable I_B in our system of roughly 100 fA; however, extrapolation of our p-channel I_{BMAX} versus T versus V_D data is consistent with this 2-V design requirement.

We have shown the (V_{xover}, L_e) data point from [15] alongside our own in Fig. 4. Eitan's t_{ox} (250 Å) is roughly 60 percent of ours, and his peak channel doping (4.6×10^{16}) is roughly five times greater than that in our n-channel devices. Both insulator thinning and increased doping would be expected to cause field increases in a scaled device; one may infer, then, that V_{xover} is weakly dependent on scaling to smaller geometries. A stronger dependence might be expected, except that the channel current travels deeper in a scaled device, and so does not experience the same magnitude of peak field [16]. Thus, if reliability is a concern, and if V_{xover} is an indication of susceptibility to—or protection from—hot-carrier effects, a performance enhancement trade-off will need to be made between temperature scaling and geometry scaling.

It should be noted that Lau *et al.* [17] demonstrate a phenomenon similar to V_{xover} in their description of I_B/I_D versus inverse pinchoff field: at lower pinchoff field, I_B/I_D decreased as T decreased. However, it was not possible to determine the peak I_B from their data; nor was any explanation of the phenomenon made.

IV. Simulation Details

Wolff [18] was the first to explore the phenomenon of impact ionization in silicon. Concentrating on the high-field regime, he arrived at a form for the ionization rate

$$\alpha \sim \exp\left(-\frac{\text{const}}{\mathcal{E}^2}\right) \tag{1}$$

Wolff's treatment assumed zero temperature and equilibrium (constant) field. In addition, it was a global model for the bulk semiconductor, and so did not address issues pertinent to MOSFET operation, where knowledge of local fields, mobility, and ionization are essential for effective simulation. Finally, the range of fields explored occurs only for submicrometer devices following constant-voltage scaling (e.g., $L_e = 0.7$ μm at $V_D = 5$ V).

Chynoweth [19] followed similar assumptions as Wolff, but pursued lower fields more common in MOSFET's. His expression for the ionization coefficient is

$$\alpha \sim \exp\left(-\frac{\text{const}}{\mathcal{E}}\right). \tag{2}$$

Note that the substrate current is related to I_D by $I_B \sim \alpha I_D$.

Shockley [11] arrived at a result similar to [19], using physical arguments related to the probability that a carrier could be lucky enough to travel ballistically for many mean free paths without scattering—and have, at the end of travel, enough energy to break a Si–Si bond.

Baraff [20] solved the Boltzmann equation numerically, and was able to demonstrate the connection between the Wolff and Chynoweth–Shockley limits for the ionization rate, over the full range of field. His assumptions also included zero temperature and equilibrium field. His treatment was numerical, and so obscures the mechanisms of the impact ionization process.

It should be noted that none of the above four treatments allows for the possibility of measurable I_B at low V_D. Each assumes a carrier begins its travel through the equilibrium field at the minimum of the band energy. So, if the total potential drop drain-to-source is less than that needed to break a Si–Si bond, each would predict zero substrate current.

Hu [21] was the first to apply Shockley's LE model empirically to a MOSFET configuration, and used it to explore gate current. Subsequently, Tam, Hu, and their co-workers [22] explained the observation of I_B at low V_D by showing the connection between the LE model and the effective carrier temperature. This temperature characterizes the carrier distribution in energy beyond the band minimum, assuming the carrier is in quasi-thermal equilibrium with the lattice phonon background, and gives rise to the carrier temperature T_e defining an MB distribution of carrier energies beyond the minimum.

Ridley [23] was the first to show the dominant impact ionization processes in an analytical fashion, without obscuring the physics. The four important physical mechanisms are depicted in Fig. 5. Comparison of his analytical result with Baraff's numerical one is quite favorable. Again, assumptions of zero temperature and equilibrium field are implicit to his treatment. It should be noted that Ridley ascribes the dominant contribution to α over the entire range of field to the processes involving lucky drift, Fig. 5(b) and (d).

Wada *et al.* [24] applied the Chynoweth formula to the local fields in a MOSFET (as opposed to the global treatments of the previous researchers). However, since they did not allow the possibility of I_B for V_D less than the ionization threshold, the observation of Tam *et al.* [22] would not be predicted with this model. The temperature dependence of I_B was not investigated.

Chan and Thurgate [25], [26] were the first to use Schockley's LE model (in a sense, a ballistic formulation of Chynoweth's equation (2), above) in a 2-D simulator to describe I_B. They attempted to use this formulation to compensate for the true nonequilibrium field found in a MOSFET. Instead of calculating an ionization coefficient applied strictly at each point in the device as in [24], they followed a carrier's path in the MOSFET and calculated the chance of ionization at each point along the path—albeit according to an equation similar to (2). The total ionization thus became a function of contributions from

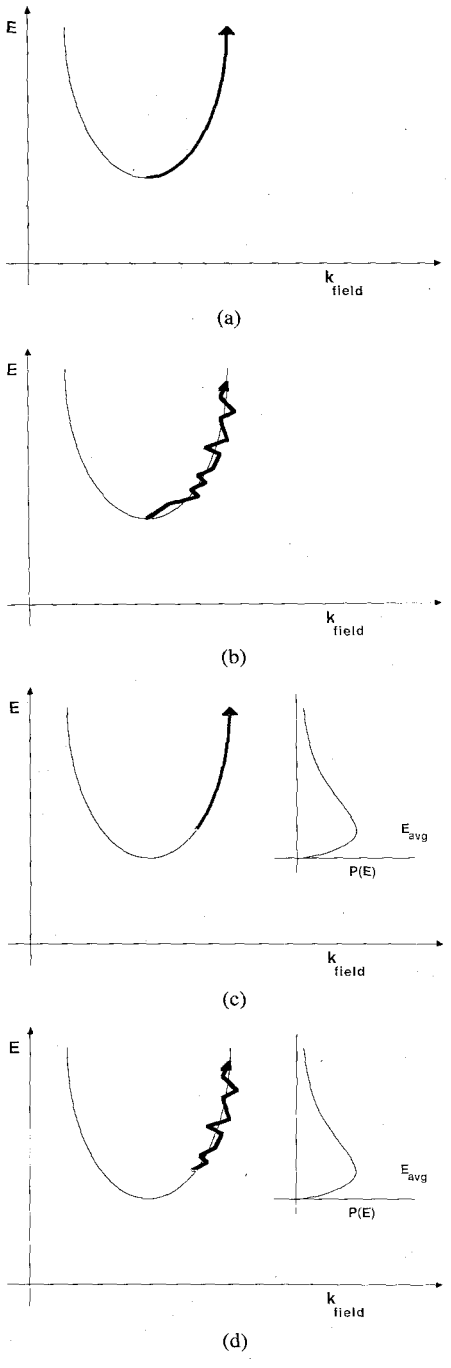


Fig. 5. Schematic of impact ionization processes after Ridley [20]. (a) Pure lucky ballistic. Carriers suffer no scattering in attaining the ionization threshold. (b) Pure lucky drift. The jagged path represents scattering to other states in E-k space, but off the ballistic or unscattered path parallel to the field. Thus, momentum may relax, but not energy. The momentum relaxation time is much less than the energy relaxation time. (c) Lucky-ballistic motion from $2k_BT_e$, the average energy after the carrier is "thermalized": after the carrier reaches a steady energy state between field gain and phonon loss. (d) Lucky drift from the thermalized average.

many carriers, each followed from a starting point, and monitored for its probability for ionization downstream. Equation (2) was essentially transformed into the form $\exp(-x/\lambda)$, where λ is the mean free path, as detailed below. However, one can see V_{xover} would not be predicted with this sort of a formulation: since λ is proportional to mobility, and since mobility increases monotonically as temperature decreases, one would only expect I_B to increase for a Shockley LE formulation alone. Also, their model specifically excluded the possibility of I_B if V_D was less than the ionization threshold.

Werner *et al.* [27] applied a more sophisticated version of the Shockley model to MOSFET's, in particular LDD devices. The mean free path, in particular, was enhanced, along the lines of [28]. Again, however, the dependence on λ would preclude the observation of V_{xover} with this model; and, I_B was again strictly cut off for V_D less than the ionization threshold.

Fukuma and Uebbing [29] added momentum and energy conservation equations to the semiconductor equations, along with formulations of the momentum and energy relaxation times, to complete a 2-D MOSFET simulator that could predict velocity overshoot (i.e., nonequilibrium) effects without reliance on a Monte Carlo method. They did not specifically look at substrate current, but did explore the average energy distribution above the band energy minimum.

Using CADDET [30], we have implemented our model as follows. We proceed from [26] by allowing carrier energies beyond the band minimum. The full rate equation is

$$I_B = \left(\frac{\lambda_{op}}{\lambda_{ii}}\right) \sum_p \sum_{i_p=1}^{N_p-1} \sum_{j_p=i_p+1}^{N_p} \frac{I_{D,p}}{\lambda} \Delta x_{i_p} \exp\left(-\frac{x_{j_p} - x_{i_p}}{\lambda}\right) \times \int_{E-qV_{x_{j_p}-x_{i_p}}}^{E_{\text{cut}}} dE \frac{E}{(kT_e)^2} \exp\left(-\frac{E}{kT_e}\right). \quad (3)$$

The pre-factor comes from [11] and is discussed in [26], and is the probability that impact ionization is the next inelastic collision. p is the index over current paths: CADDET breaks the channel current into roughly 200 equal-current paths after solving the appropriate semiconductor equations and assigns field, potential, and mobility values to each point along each path. $I_{D,p}/\lambda$ is the collision rate per unit length along p. The double sum adds up the impact ionization probability between all initial points x_{i_p} and final points x_{j_p} along the path p; the second sum has $j_p = i_p + 1$ as a lower limit to avoid double counting. N_p is the number of discrete grid points along p. Δx_{i_p} represents the current packet in the vicinity of x_{i_p}, which we follow downstream in exploring its ionization probability. The first exponential is just Shockley's LE probability.

This implementation follows carriers only downstream, and so does not immediately allow calculation of the local ionization rate at any point. However, this rate may be calculated easily by fixing j_p, then doing a backward sum along the path p over the index i_p.

The integral is the MB extension of the LE model and represents the probability that a carrier with energy beyond the band minimum contributes to impact ionization. The lower limit is the ionization threshold. $qV_{x_{j_p}-x_{i_p}}$ is the energy gained from the field between x_{i_p} and x_{j_p}. E_g is the ionization energy. Only carriers with $E > E_g - qV_{x_{j_p}-x_{i_p}}$ may contribute to impact ionization. The lower limit E_g

TABLE I
TEMPERATURE-INDEPENDENT MODEL PARAMETERS

	Parameter	holes	electrons
Mobility	γ ($\frac{cm}{V}$)	3.34×10^{-5}	1.23×10^{-5}
	β	1.00	1.25
I_B	y_0 (Å)	500	500
	λ_0 (Å)	51.5	105
	E_{cut} (eV)	0.22	0.22
	E_{op} (eV)	0.063	0.063

$- qV_{x_{j_p}-x_{i_p}}$ must be non-negative, indicating the carriers lie at or above the band minimum. The integral has a simple analytic form that speeds computation.

The upper limit of E_{cut} truncates the MB distribution at a value between three and four times the optical phonon energy (equilibrium between carrier energy and the optical phonon gas gives rise to the carrier temperature T_e). The physical basis for E_{cut} is discussed later.

T_e is the mean carrier temperature that defines the MB distribution according to [31]

$$T_e = (T/2)\left\{1 + \left[1 + \frac{3\pi}{8}\left(\frac{\mu \mathcal{E}_\parallel}{v_s}\right)^2\right]^{1/2}\right\}. \tag{4}$$

$\mathcal{E}_\parallel$ is the field along the current path. v_s is the velocity of longitudinal sound in the semiconductor. We note that (4) uses the full mobility rather than the low-field mobility in calculating T_e. This is a departure from [31], where the calculation assumes longitudinal acoustic—not optical—phonon scattering in establishing the carrier's energy steady state between field gain and phonon loss. However, the use of the full mobility allows T_e to approach a constant value in the high-field limit, as one expects. The model for μ is crucial to the energy distribution. It is treated in detail in the Appendix.

The normalization condition on the MB distribution is

$$1 = A \int_0^\infty E \exp\left(-\frac{E}{kT_e}\right) dE. \tag{5}$$

The carrier must be at or beyond the band minimum. Thus, $A = 1/(kT_e)^2$.

The bulk scattering mean free path used in the LE portion of (3) is given by [20]

$$\lambda_B = \lambda_0 \tanh\left(\frac{E_{op}}{2kT}\right). \tag{6}$$

As in [26], the total mean free path is also dependent on the surface reduction factor h

$$h = 1 - 0.46 \exp(-y/y_0). \tag{7}$$

Noting that $y_0 = 500$ Å, the total mean free path is then

$$\lambda = h\lambda_B. \tag{8}$$

As shown in Table I, our values for λ_0 agree closely with those of Tam *et al.* (106 Å) [28] and Ning *et al.* (105 Å) [32] for electrons, and with that of Crowell and Sze (55 Å) [33] for holes.

We interpret the full rate equation (3) as follows: the channel carriers are assumed to be in thermal equilibrium with the lattice phonons at every point along each current path p. Phonon emission is the dominant means assumed for carriers with high energy (obtained from the field) to equilibrate with the lattice; this gives rise to the parameter T_e, which describes the full carrier distribution function.

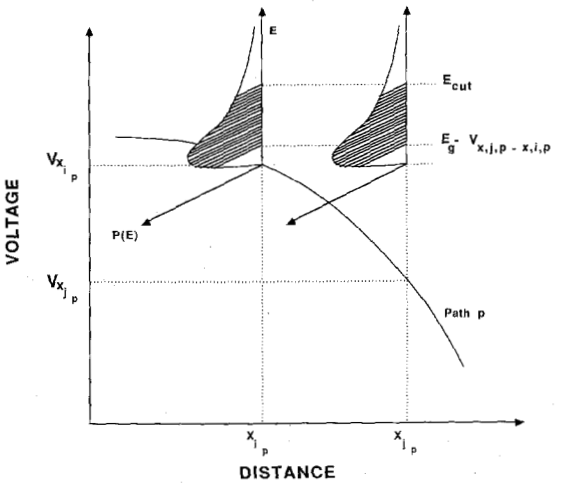


Fig. 6. Schematic of the impact ionization model. $qV_{x_{jp}-x_{ip}}$ is the energy acquired in traversing a path $x_{jp} - x_{ip}$, as defined in (1). $P(E)$ is the MB distribution, determined from the field, mobility, and temperature at x_{ip}. Only the cross-hatched portion of the MB distribution contributes to impact ionization: carriers with $E < E_g - qV_{x_{jp}-x_{ip}}$ have not reached the ionization threshold; carriers with $E > E_{\text{cut}}$ are either already in the drain when they impact ionize, or have low probability of their next collision at x_{jp} being and impact-ionizing one.

This MB distribution function is then propagated unscattered along the path p, from the point x_{ip} (see Fig. 6.) An implicit assumption is that λ is not a function of energy so that the MB distribution moves undistorted along the path: in other words, the carriers travel ballistically. λ is also assumed to be independent of the electric field, except as indirectly manifested through the factor h in (7).

Thus, the model employs the process of Fig. 5(c), though in greater detail since all carrier energies are considered, not just the average energy. The carriers are unaffected by their previous history—by what happened to them upstream of x_{ip}—except insofar as this history gives rise to their equilibrium distribution beyond the band minimum. The generation of I_B occurs due to downstream ballistic propagation of this equilibrium MB distribution.

V. MEASUREMENTS AND MODEL COMPARISON

Previously, a local LE model was used successfully to predict I_B [26], as well as other NMOS effects [27]. One drawback was an inability to predict the observation of $V_{x\text{over}}$ (see Fig. 3), or I_B at low V_D. To describe both the I_B versus V_G and I_{BMAX} versus T curves, our model requires knowledge of only two parameters for each device (λ_0 and E_{cut}) to cover the entire range of V_G, V_D, L_e, and T. The physical basis for λ_0 has been discussed above, while that of E_{cut} follows shortly.

Figs. 7–11 show comparisons between CADDET simulations using the model and our devices for a few of the temperatures and channel lengths measured. These re-

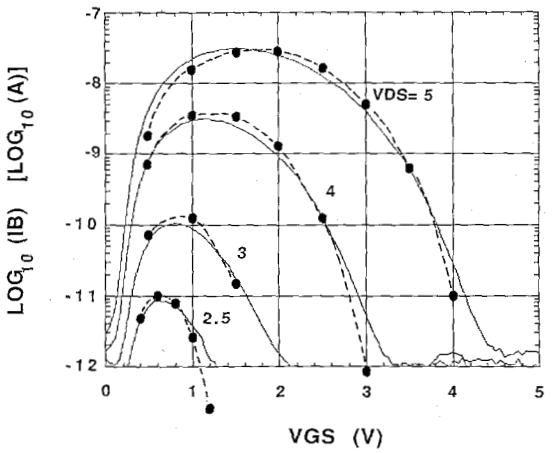


Fig. 7. Substrate current versus gate voltage for 25/25 n device at $T = 299$ K. Simulation results shown by •••.

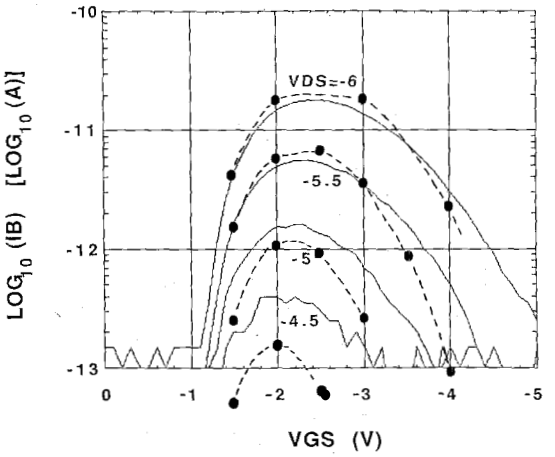


Fig. 10. Substrate current versus gate voltage for 25/25 p device at $T = 294$ K. Simulation results shown by •••.

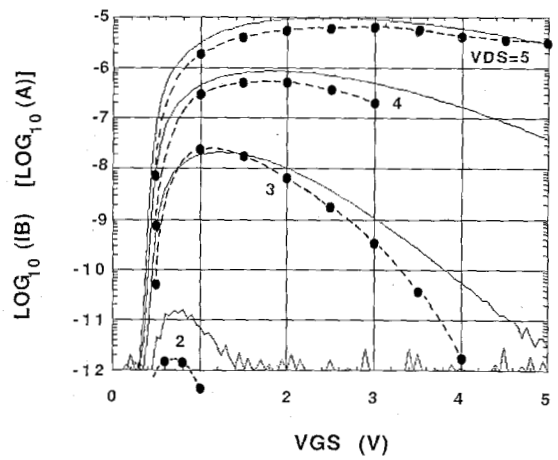


Fig. 8. Substrate current versus gate voltage for 25/3 n device at $T = 77$ K. $L_e = 2.15$ μm. Simulation results shown by •••.

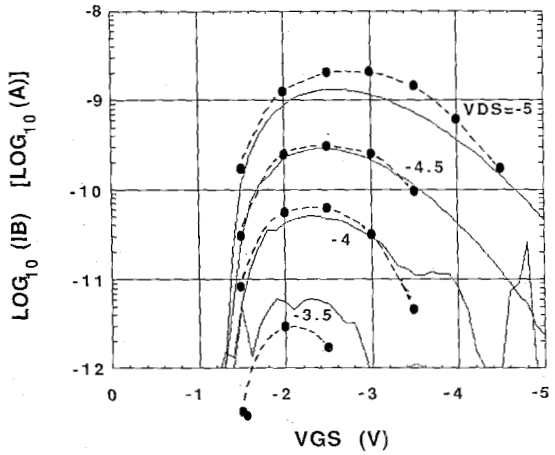


Fig. 11. Substrate current versus gate voltage for 25/2 p device at $T = 77$ K. $L_e = 1.17$ μm. Simulation results shown by •••.

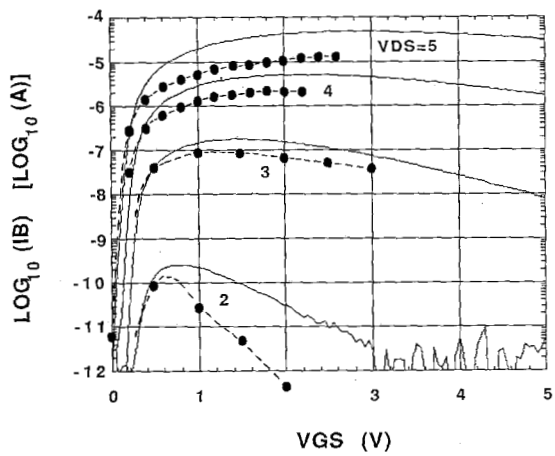
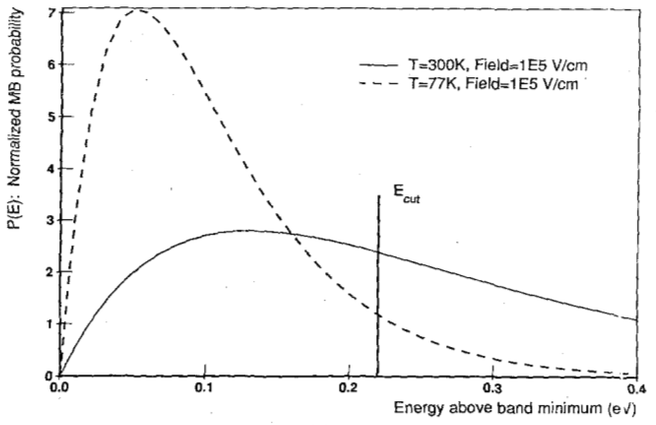


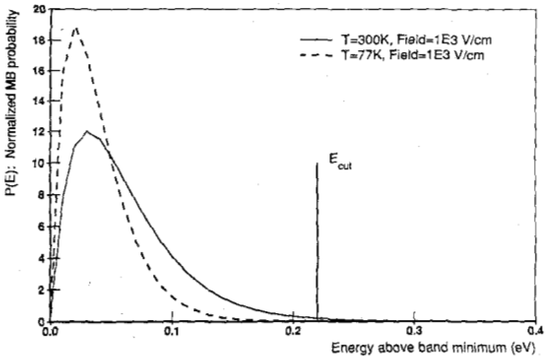
Fig. 9. Substrate current versus gate voltage for 25/1.75 n device at $T = 77$ K. $L_e = 0.85$ μm. Simulation results shown by •••. High-V_G points unavailable due to CADDET convergence problem in standard MOSFET solution.

sults—at the longest and shortest L_e, at the highest (300 K) and lowest (77 K) temperatures, for both device types—give an excellent description of the model's performance. In particular, V_{xover} is well predicted (see Fig. 4).

Fig. 7 shows the comparison for a long n-channel device at room temperature. The model performs well for these conditions, but does show the beginnings of underprediction for I_B as the device goes out of saturation for any particular V_D. Fig. 8 shows the liquid-nitrogen characteristic for a device of intermediate channel length. Here, while the value of V_G at I_{BMAX} is well-predicted, and while I_B is very accurately given for $V_D = 3$ V, we see some significant underprediction of I_B at the highest and lowest V_D. Again, I_B is also underpredicted as the device goes out of saturation. As the fields for a given V_D are increased by going to a submicrometer L_e device (Fig. 9), we see these trends continue. Somewhat similar behavior holds for the p-channel devices (Figs. 10 and 11). We note that some of the mismatch in the p-channel devices was due to CADDET's difficulties in predicting the I–V curve correctly. This may be attributed to CADDET's use of pure Boltzmann statistics, which assumes complète ionization of impurities. The PMOSFET's had a compensating implant in the device channel, so that incomplete ionization occurs at low temperatures, which cannot be modeled by CADDET. The reasons for the model's underpredictions are discussed in the next section.



(a)



(b)

Fig. 12. How V_{xover} is predicted by an energy distribution. The probability curves are normalized for unity area. Dashed lines are for 77 K; solid lines are for 300 K. $E_{cut} = 0.22$ eV is shown. (a) $\mathcal{E} = 10^5$ V/cm. (b) $\mathcal{E} = 10^3$ V/cm. See text for discussion.

Fig. 12 demonstrates how the MB distribution can explain the observed V_{xover}. In Fig. 12(a), the MB distribution for a field of 10^5 V/cm is compared for $T = 77$ and 300 K. As expected, the peak moves closer to the band energy minimum at lower temperatures. However, at both temperatures, significant portions of this normalized curve lie at or just below E_{cut}, so that the T dependence is not immediately clear. If we consider, though, that "high field" also means high V_D, with plenty of ionizing potential available in the channel over short path lengths $x_{j_p} - x_{i_p}$, it becomes apparent that most of the carriers below E_{cut} will participate in impact ionization. Thus, $E_g - qV_{x_{j_p} - x_{i_p}}$ will approach the origin. With this additional caveat, it can be seen that the area of the curve for 77 K up to E_{cut} is significantly greater than for 300 K, indicating I_B will be greater, as well—even apart from the increase in λ.

This may be seen by taking the expression for T_e, (4) in the limit $\mu\mathcal{E} >> v_s$, and putting this limiting form back into the MB distribution exponential. If we show only the temperature-dependent variables, this procedure yields

$$I_B \sim \exp\left(-\frac{E}{\text{const } T\mu}\right). \tag{9}$$

Even if the temperature change of μ in the high-field is not as rapid as T^{-1}, the combined temperature dependence of this term with the LE term will be enough to insure that, as T decreases, the impact ionization current will increase.

The second set of curves Fig. 12(b) demonstrates the situation for small V_D. Here, (at 10^3 V/cm, just to illustrate the limit) carriers hot enough to contribute to impact ionization lie only in the tail of the MB distribution. Combined with the fact that the available potential is low, so that only hot carriers near E_{cut} will ever meet the ionization threshold condition, we see the area under the tail decreases as T decreases. Thus, we have the prediction of V_{xover}, just as is measured experimentally.

This may be seen as well by looking at the expression for T_e in the limit $\mu\mathcal{E} << v_s$. This limit yields

$$I_B \sim \exp\left(-\frac{E}{\text{const } T\{1 + \text{const } \mu^2\}}\right). \tag{10}$$

Here, T is the dominant term, even more so than the tanh term for λ, so that as T decreases, I_B will also decrease.

Any model must explain the feature alluded to by Eitan *et al.* [15] (who used an ionization threshold of $1.5E_g$), and proven by Tam *et al.* [22]; that is, measurable I_B for V_D less than the ionization potential E_g/q. Because we augment the LE model with a MB distribution of carrier energies, we deal directly with the quasi-thermal equilibrium approach outlined in [22]. However, because we cut off the MB distribution sharply at $E_{cut} = 0.22$ eV, one would predict, with our model, that I_B would go to zero for $V_D < (E_g - E_{cut})/q$. Extrapolating for $V_D < 0.9$ V in [38, Fig. 2], we expect this prediction to be incorrect. One might propose another probability term in the I_B rate equation following the Fermi–Dirac form

$$I_B \sim \frac{1}{1 + \exp\left(\dfrac{E - E_{cut}}{kT_e}\right)} \tag{11}$$

which would allow for both continuous measurements of I_B below $(E_g - E_{cut})/q$, and the decreasing slope of log (I_{BMAX}/I_D) versus T for decreasing V_D (see Fig. 3). We are exploring ways of allowing I_B at $qV_D < E_g - E_{cut}$, while maintaining the model's computational compactness.

VI. Discussion

During the initial phase of comparing our model with the MB distribution to our measurements, it became clear that including the full area under the MB distribution all the way to infinity (letting the upper limit in the second integral of (3) go to ∞) over-predicted the observed I_B. We speculated that other inelastic scattering mechanisms, apart from impact ionization, became important for the carriers with highest energy. In particular, we suspected that the λ for the carriers in the LE expression was not energy independent, but grew large as the carrier energy grew beyond the band minimum. Thus, we relaxed the normalization condition in (5) and, in effect, truncated the

MB distribution at a value E_{cut}. $E_{cut} = 0.22$ eV proved to be the best value for both n- and p-channel devices.

This speculation saw some qualitative verification in the work of Mahan [34], who solved the Boltzmann equation in a novel fashion for optical phonon scattering (the dominant scattering mechanism related to impact ionization), taking into account upstream effects and nonequilibrium field on the carrier energy distribution. It was shown that, in the high-field region of a MOSFET ($\sim 10^5$ V/cm), many of the carriers had energy far in excess of that predicted by a T_e characterizing an equilibrium MB distribution. In effect, the highest energy carriers upstream of this decidedly non-Boltzmann distribution had a much longer effective mean free path, when combined with the high field. This indicated our assumptions of field- and energy-independence for our LE mean free path were incorrect, justifying E_{cut} at least qualitatively. In addition, such field and energy dependences for λ could mean the carriers at the highest energies physically reach the drain without scattering, so that impact ionized carriers would not be detected in the substrate.

Ridley [23] also addressed the issue of E_{cut}, in a more quantitative sense. He suggested, somewhat arbitrarily but with the relation of momentum relaxation time to energy relaxation time in mind, that $E_{cut} < 3q\lambda\mathcal{E}$ is the bound for considering whether carriers in the high field regime thermalize and participate in impact ionization. The factor of three is not critical and can vary between 2 and 4 without great departure from the result of Baraff. Thus, for a field of roughly 10^5 V/cm, we get a value for E_{cut} of between 0.2 and 0.4 eV, according to this reasoning.

Ridley has stated that "lucky-drift" processes (see Fig. 5(b) and (d)) are the most probable for impact ionization. This statement runs counter to our formulation, relying on the process of Fig. 5(c), which compares so well to experiment. The probable resolution to the contradiction, at least in the context of Ridley's work, is that a fixed E_{cut} targeted to give the best results at moderate fields (around $V_D = 3$ V), will under-predict I_B at high field where E_{cut} should increase well above 0.22 eV. It should also underpredict at low field where pure lucky drift is dominant and the concepts of T_e and quasi-thermal equilibrium are valid, so that the full distribution to $E = \infty$ should be included.

It is clear that some gaps still exist in a complete physical understanding of substrate current generation in MOSFET's due to impact ionization. First, regarding underprediction of I_B at high V_G and low V_D (see Fig. 8, for instance), an MB distribution of energies may not be the correct one to use [34]. However, abandoning the MB distribution would make the simulation task more difficult, as Monte Carlo techniques [35], novel but intricate Boltzmann equation solutions [34], or more careful considerations of upstream effects [36] become necessary to establish the energy distribution.

Second, low predictions for I_{BMAX} at the highest V_D could be the result of using a fixed E_{cut} when a field-dependent value is more appropriate, in Ridley's vernacular; the neglect of additional scattering mechanisms for carriers with energy greater than the ionization threshold [27], [28]; neglect of the carrier's history along the current path [36]; or the onset of significant avalanche, requiring treatment of both carrier types.

Third, underprediction of I_B at lower V_D, or in the lower field region (III in Fig. 2) of the substrate current characteristic where V_D approaches V_G, is probably again the result of a fixed E_{cut}. Inclusion of the full MB distribution (letting E_{cut} go to ∞) would not only increase I_B in this field region, but also allow prediction of I_B at values of qV_D less than $E_g - E_{cut}$, which is a demonstrated experimental result.

Hänsch [36] has arrived at an expression that takes into account finite-temperature nonequilibrium fields and the need to look at upstream as well as downstream processes in determining the full ionization rate. His derivation, along the lines of Keldysh [37], arrives at an expression which has both a LE term as well as a term similar to our MB distribution term

$$I_B \sim \exp\left(-S\frac{E_g - qV_{x_{j_p}-x_{i_p}}}{q\mathcal{E}\lambda}\right). \tag{12}$$

The denominator has been shown to be identical to the T_e form in (3) [28]. The factor S is field dependent and takes into account upstream as well as downstream considerations; near fields of 10^5 V/cm, S is constant and near unity. For low fields, S varies so that $T_e \sim T$; for the highest fields, S decreases, producing an effect similar to Ridley's cutoff.

Finally, we have asserted that dc device degradation under stress will be proportional to I_B at low T as well as 300 K. Despite some temperature studies of hot-carrier degradation [38], this has not been demonstrated experimentally. As stated above, verification of this assertion will depend on the temperature dependence of interfacial charge transport, and insulator trapping. Direct measurements of gate current and threshold voltage shifts are underway to investigate these effects.

VII. Conclusions

We have characterized the substrate current in a CMOS technology over a wide range of channel lengths and bias voltages. The characterization has demonstrated a possible limit on power supply voltages, inferred from the relationship between I_B and device degradation. It has done so by detailing the dependence of the peak substrate current on the drain and gate voltage, channel length, and temperature.

A physical model to explain the experimental observations has been developed based on the notion that carriers can have energies significantly beyond their band minima due to equilibrium between carrier kinetic energy and lattice phonons. The model has been implemented in the 2-D device simulator CADDET. The model requires a knowledge of only two parameters for each device type to describe the observed I_B, both of which are shown to have a physical basis (λ_0, in the sense of Baraff [20], and

E_{cut} in the sense of Ridley [23]). Furthermore, because the model is based on a local knowledge of channel fields, voltages, and current densities, it is expected to be equally applicable to other technologies (e.g., LDD and/or E/D NMOS). The close agreement between model and experiment indicates that design and realization of a reliable scaled low-temperature CMOS technology should not be hindered by an inordinately low power supply voltage.

Appendix
Mobility Model

As shown in (4), the mobility is crucial in determining the equilibrium energy distribution of the channel carriers. Thus, it is important from the standpoint of completeness that the full temperature-dependent mobility model be detailed. This model, implemented in CADDET, has been particularly successful in predicting I_D–V_D characteristics in our CMOS devices over the full ranges of T, L_e, V_D, and V_G explored during our substrate current characterization.

The mobility is given by

$$\mu = f[g * h * \mu_0] \tag{A1}$$

where

$$\mu_0 = \mu_{min} + \frac{\mu_{max} - \mu_{min}}{1 + \left(\frac{N_T}{N_{ref}}\right)^{\alpha}} \tag{A2}$$

accounts for the low-field bulk mobility and its temperature and doping dependences [39], and $N_T = N_A + N_D$ is the total doping. g is the vertical field reduction factor after [40] and is temperature-independent [2]

$$g = (1 + \gamma \mathcal{E}_{\perp})^{-1/2}. \tag{A3}$$

$\mathcal{E}_{\perp}$ is the field component perpendicular to the quasi-Fermi level gradient. The γ used by [40] has been updated to include the μ_0 of [39] (see Table I).

h is the same factor for surface roughness reduction as in [26]; μ and λ thus have the same surface roughness dependence, as one might expect.

The lateral field reduction [39] yields a final mobility of

$$\mu = gh\mu_0 \left[1 + \left(\frac{gh\mu_0 \nabla\phi}{v_{sat}}\right)^{\beta}\right]^{-1/\beta} \tag{A4}$$

where $\nabla\phi$ is the quasi-Fermi level gradient. μ is thus field, position, and temperature dependent.

The temperature dependence for saturation velocity is given by [41]

$$v_{sat} = \frac{2.4 \times 10^7}{1 + 0.8 \exp\left(\frac{T}{600}\right)} \text{ cm/s}. \tag{A5}$$

$\mu_{max,n}$ is given by

$$\mu_{max,n} = 1430 \left(\frac{T}{300}\right)^{-2.25} \frac{\text{cm}^2}{\text{V} \cdot \text{s}}, \quad T > 110 \text{ K} \tag{A6}$$

$$\mu_{max,n} = 1430 \left(\frac{110}{300}\right)^{-2.25} \left(\frac{T}{110}\right)^{-1.6} \frac{\text{cm}^2}{\text{V} \cdot \text{s}}, \quad T < 110 \text{ K} \tag{A7}$$

a result derived from the experiments of [42] and the theory of [43]. A similar expression for holes comes from comparisons of experiment [41] and theory [44] as well

$$\mu_{max,p} = 474 \left(\frac{T}{300}\right)^{-2.2} \frac{\text{cm}^2}{\text{V} \cdot \text{s}}. \tag{A8}$$

The temperature dependence of μ_{min} (heavy-doping limit) comes from [1] and [45]; the T_{room} value is from [42], [46]:

$$\mu_{min,n} = 80 \left(\frac{T}{300}\right)^{-0.45} \frac{\text{cm}^2}{\text{V} \cdot \text{s}}, \quad T > 200 \text{ K} \tag{A9}$$

$$\mu_{min,n} = 80 \left(\frac{200}{300}\right)^{-0.45} \left(\frac{T}{200}\right)^{-0.15} \frac{\text{cm}^2}{V \cdot s}, \; T < 200 \text{ K} \tag{A10}$$

$$\mu_{min,p} = 45 \left(\frac{T}{300}\right)^{-0.45} \frac{\text{cm}^2}{\text{V} \cdot \text{s}}, \quad T > 200 \text{ K} \tag{A11}$$

$$\mu_{min,p} = 45 \left(\frac{200}{300}\right)^{-0.45} \left(\frac{T}{200}\right)^{-0.15} \frac{\text{cm}^2}{\text{V} \cdot \text{s}}, \quad T < 200 \text{ K}. \tag{A12}$$

The temperature dependence of α and N_{ref} is taken from fitting (A2) to the theoretical expressions of [47]

$$N_{ref} = 1.12 \times 10^{17} \left(\frac{T}{300}\right)^{3.2} \text{cm}^{-3} \text{ (n-type Si)} \tag{A13}$$

$$N_{ref} = 2.23 \times 10^{17} \left(\frac{T}{300}\right)^{3.2} \text{cm}^{-3} \text{ (p-type Si)} \tag{A14}$$

$$\alpha = 0.72 \left(\frac{T}{300}\right)^{0.065}. \tag{A15}$$

No difference is assumed between holes and electrons for α. N_{ref} has a different $T = 300$ K value for n-type and p-type Si, based on [42], [46].

Table I summarizes the values of temperature-independent parameters.

Acknowledgment

The authors thank J. Woo for useful discussions and development of the temperature measurement system, W. Engeler of GE and Z. Shanfield of Northrop for pointing out informative references, and W. Hänsch of Siemens for a fruitful discussion on impact ionization physics.

References

[1] F. H. Gaensslen *et al.*, "Very small MOSFET's for low-temperature operation," *IEEE Trans. Electron Devices*, vol. ED-24, p. 218, 1977.
[2] S. Tewksbury, "N-channel enhancement-mode MOSFET characteristics from 10 to 300 K," *IEEE Trans. Electron Devices*, vol. ED-28, p. 1519, 1981.
[3] J. W. Schrankler *et al.*, "Cryogenic behavior of scaled CMOS devices," in *IEDM Tech. Dig.*, p. 598, 1984.

[4] S. Ogura *et al.*, "Submicron MOSFET performance at liquid nitrogen temperature," in *ISSC Dig. Tech. Papers*, p. 160, 1986.
[5] J. C. S. Woo and J. D. Plummer, "Short-channel effects in MOSFET's at liquid nitrogen temperatures," *IEEE Trans. Electron Devices*, to be published.
[6] I. Kato *et al.*, "1.5 μm gate CMOS operated at 77 K," in *IEDM Tech. Dig.*, p. 601, 1984.
[7] G. Gildenblatt *et al.*, "Investigation of cryogenic CMOS performance," in *IEDM Tech. Dig.*, p. 268, 1985.
[8] E. Takeda and N. Suzuki, "An empirical model for device degradation due to hot-carrier injection," *IEEE Electron Device Lett.*, vol. EDL-4, p. 111, 1983.
[9] T. Horiuchi *et al.*, "A simple method to evaluate device lifetime due to hot-carrier effect under dynamic stress," *IEEE Electron Device Lett.*, vol. EDL-7, p. 337, 1986.
[10] S. Tam *et al.*, "Correlation between substrate and gate currents in MOSFET's," *IEEE Trans. Electron Devices*, vol. ED-29, p. 1740, 1982.
[11] W. Shockley, "Problems related to p-n junctions in silicon," *Solid-State Electron.*, vol. 2, p. 35, 1961.
[12] J. R. Pfiester, "Performance limits of CMOS very large scale integration," Stanford Electronics Lab. Tech. Rep. G541-1, p. 100, 1984.
[13] R. C. Jaeger and F. H. Gaensslen, "Simple analytical models for the temperature dependent threshold behavior of depletion-mode devices," *IEEE Trans. Electron Devices*, vol. ED-26, p. 501, 1979.
[14] J. Y. C. Sun and M. R. Wordeman, "Degradation of 77-K MOSFET characteristics due to channel hot electrons," *IEEE Trans. Electron Devices*, vol. ED-31, p. 1970, 1984.
[15] B. Eitan *et al.*, "Impact ionization at very low voltages in silicon," *J. Appl. Phys.*, vol. JAP-53, p. 1244, 1981.
[16] S. Laux and F. Gaensslen, "A study of avalanche breakdown in scaled n-MOSFETs," in *IEDM Tech. Dig.*, p. 84, 1984.
[17] D. Lau *et al.*, "Low-temperature substrate current characterization of n-channel MOSFETs," in *IEDM Tech. Dig.*, p. 565, 1985.
[18] P. A. Wolff, "Theory of electron multiplication in silicon and germanium," *Phys. Rev.*, vol. 95, p. 1415, 1954.
[19] A. G. Chynoweth, "Ionization rates for electrons and holes in silicon," *Phys. Rev.*, vol. 109, p. 1537, 1958.
[20] G. A. Baraff, "Distribution functions and ionization rates for hot electrons in semiconductors," *Phys. Rev.*, vol. 128, p. 2507, 1962.
[21] C. Hu, "Lucky-electron model of channel hot electron emission," in *IEDM Tech. Dig.*, p. 22, 1979.
[22] S. Tam *et al.*, "Hot-electron currents in very short channel MOSFET's," *IEEE Electron Device Lett.*, vol. EDL-4, p. 249, 1983.
[23] B. K. Ridley, "Lucky-drift mechanism for impact ionisation in semiconductors," *J. Phys. C: Solid State Phys.*, vol. 16, p. 3373, 1982.
[24] M. Wada *et al.*, "Modelling of hot-carrier effects in small-geometry MOSFETs," *Electron. Commun. Japan*, vol. 67-C, p. 96, 1984.
[25] N. Chan and T. Thurgate, "An impact ionization model for 2D device simulation," *IEEE Trans. Electron Devices*, vol. ED-31, p. 1978, 1984.
[26] T. Thurgate and N. Chan, "An impact ionization model for two-dimensional device simulation," *IEEE Trans. Electron Devices*, vol. ED-32, p. 400, 1985.
[27] C. Werner *et al.*, "Optimization of lightly doped drain MOSFET's using a new quasiballistic simulation tool," in *IEDM Tech. Dig.*, p. 770, 1984.
[28] S. Tam *et al.*, "Lucky-electron model of channel hot-electron injection in MOSFET's," *IEEE Trans. Electron Devices*, vol. ED-31, p. 1116, 1984.
[29] M. Fukuma and R. Uebbing, "Two-dimensional MOSFET simulation with energy transport phenomena," in *IEDM Tech. Dig.*, vol. 84, p. 621, 1984.
[30] M. Mock, "A two-dimensional model of the insulating-gate field-effect transistor," *Solid-State Electron.*, vol. 16, p. 601, 1973.
[31] R. H. Bube, *Electronic Properties of Crystalline Solids.* New York: Academic, 1974, p. 296ff.
[32] T. Ning *et al.*, "Emission probability of hot electrons from silicon into silicon dioxide," *J. Appl. Phys.*, vol. 48, p. 286, 1977.
[33] C. R. Crowell and S. Sze, "Temperature dependence of avalanche multiplication in semiconductors," *Appl. Phys. Lett.*, vol. 9, p. 242, 1966.
[34] G. D. Mahan, "Hot electrons in one dimension," *J. Appl. Phys.*, vol. 58, p. 2242, 1985.
[35] R. K. Cook and J. Frey, "An efficient technique for two-dimensional simulations of velocity overshoot effects in Si and GaAs devices," *COMPEL*, vol. 1, p. 65, 1982.
[36] W. Hänsch, private communication.
[37] L. V. Keldysh, "Kinetic theory of impact ionization in semiconductors," *Sov. Phys.—JETP*, vol. 37, p. 509, 1960.
[38] J. A. Brachitta *et al.*, "Hot-electron-induced degradation in MOSFET's at 77 K," *IEEE Trans. Electron Devices*, vol. ED-32, p. 1850, 1985.
[39] D. M. Caughey and R. E. Thomas, "Carrier mobilities in silicon empirically related to doping and field," *Proc. IEEE*, vol. 55, p. 2192, 1967.
[40] K. Yamaguchi, "Field-dependent mobility model for two-dimensional numberical analysis of MOSFET's," *IEEE Trans. Electron Devices*, vol. ED-26, p. 1068, 1979.
[41] C. Jacoboni *et al.*, "A review of some charge transport properties of silicon," *Solid-State Electron.*, vol. 20, p. 77, 1977.
[42] W. R. Thurber *et al.*, "Resistivity-dopant density relationship for phosphorus-doped silicon," *J. Electrochem. Soc.*, vol. 127, p. 1807, 1980.
[43] P. Norton *et al.*, "Impurity and lattice scattering parameters as determined from Hall and mobility analysis in n-type silicon," *Phys. Rev.*, vol. B8, p. 5632, 1973.
[44] S. S. Li, "The dopant density and temperature dependence of hole mobility and resistivity in boron doped silicon," *Solid-State Electron.*, vol. 21, pp. 1109, 1978.
[45] G. L. Pearson and J. Bardeen, "Electrical properties of pure silicon and silicon alloys containing boron and phosphorus," *Phys. Rev.*, vol. 75, p. 865, 1949.
[46] W. R. Thurber *et al.*, "Resistivity-dopant relationship for boron-doped silicon," *J. Electrochem. Soc.*, vol. 127, p. 2291, 1980.
[47] S. S. Li and W. R. Thurber, "The dopant density and temperature dependence of electron mobility and resistivity in n-type silicon," *Solid-State Electron.*, vol. 20, p. 609, 1977.

*

Albert K. Henning (M'86) was born in Chicago, IL, on January 29, 1955. He received the A.B. degree *magna cum laude* with distinction in physics in 1977 and the A.M. degree in physics in 1979, both from Dartmouth College. From 1977 to 1979 he was a Dartmouth Fellow. His master's thesis was entitled, "DC Conductivity in Compensated p-Ge Between 0.35 K and 4.2 K."

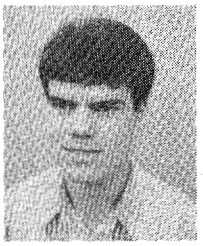
From 1979 to 1982, he was a Device Physicist for Intel Corp. in Santa Clara, CA, where he worked in both NMOS and CMOS process development for SRAM and microprocessor applications. From 1982 to 1983, he consulted with Intel in the areas of interconnection capacitance and CMOS latchup. Since 1982, he has been a Research Assistant at Stanford University, with emphases on hot carrier phenomena in CMOSFET's and cryogenic operation and reliability of CMOS technology.

Mr. Henning is a member of Sigma Xi.

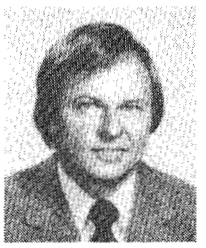
*

Nelson N. Chan (S'73–M'78) received the B.S./M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 1972 and 1979, respectively. His thesis research was on numerical simulation of integrated bipolar transistors for high-frequency applications.

Since joining the Intel Corporation, Santa Clara, CA, in 1979, he has been engaged in the enhancement of device simulation capabilities at Intel and the development of simulation models to aid technology design of short-channel MOSFET's. The areas of study have included punchthrough, impact ionization, and oxide injection of hot electrons.



Jeffrey T. Watt was born in North Bay, Canada, on October 9, 1961. He received the B.S. degree in electrical engineering from Queen's University, Kingston, Canada, in 1983 and the M.S. degree in electrical engineering from Stanford University, Stanford, CA, in 1984. He is currently working toward the Ph.D. degree in electrical engineering at the Stanford University Integrated Circuits Laboratory. His research interests are in the area of low-temperature device physics and process technology.



James D. Plummer (M'71-SM'82-F'85) was born in Toronto, Canada, on December 3, 1944. He received the B.S. degree in 1966 from the University of California, Los Angeles, and the M.S. and Ph.D. degrees, both in electrical engineering, from Stanford University in 1967 and 1971, respectively.

He is presently a Professor in the Electrical Engineering Department, and the Director of the Integrated Circuits Laboratory at Stanford University. His current research interests center on basic modeling of semiconductor processes including oxidation, epitaxy, and ion implantation, on the physics and technology of scaled bipolar and MOS devices, and on high-voltage devices and integrated circuits. He is the author of numerous technical papers in these areas.

Dr. Plummer is a member of Tau Beta Pi, Sigma Xi, American Physical Society, and the Electrochemical Society. He has received three best paper awards at the International Solid-State Circuits Conference, and was the Technical Program Committee Chairman for the 1980 ISSCC.