RC 22072 (98989) 22 May 2001

Research Report

Hot-Carrier Distribution Functions and Hot-Luminescence Intensities at Energies Above 1eV From Future Si FETs Operating Near 1V

J. C. Tsang, J. A. Kash, M. V. Fischetti
IBM Research Division
Thomas J. Watson Research Center
P. O. Box 218
Yorktown Heights, NY 10598

Research Division
Almaden - Austin - Beijing - Haifa - India - T. J. Watson - Tokyo - Zurich

Hot-Carrier Distribution Functions and Hot-Luminescence Intensities at Energies Above 1eV From Future Si FETs Operating Near 1V

J. C. Tsang, J. A. Kash, and M. V. Fischetti

IBM T. J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598
jtsang@us.ibm.com

Abstract: The intensity in the near infrared of the hot-carrier light emission from advanced CMOS technologies will not decrease significantly below present levels in the next several years even as operating voltages approach 1V. This means that hot-luminescence techniques for circuit characterization such as PICA will remain viable over this period. The effect on the emission intensity of the decreasing operating voltages and device dimensions is dominated by the dependence of the hot-carrier distribution on voltage. Calculations of hot-electron distribution functions for small nFETs including Coulomb scattering from the heavily-doped source and drain show that the distribution functions for energies greater than 1 eV will only decrease slightly as devices scale from the present 1.8V technology to below 1V.

I. Introduction

Hot-carrier effects in sub 0.25 μm channel length FETs are now of both scientific and technical interest. Scientifically, they are a good test of our understanding of high field transport in semiconductors. Technically, the switching induced pulsed light emission due to the momentary presence of hot-carriers near the drains of field effect transistors (FETs) is used to measure when gates switch in CMOS integrated circuits (ICs)[1]. Detailed timing data have been obtained using Picosecond Integrated Circuit Analysis (PICA) from the backsides of normally operating, flip-chip packaged ICs. The successful diagnostic use of PICA on 0.18 μm devices technologies for a bias voltage V_{dd}= 1.8V has been reported[2] and work on state-of-the-art devices is continuing.

By the last third of this decade, CMOS circuits will feature gate lengths near 50 nm, and operating voltages below 1V.[3] It has been suggested that the intensity of the hot-carrier light emission in small FETs where V_{dd}<1V will be too weak for the practical use of light emission tools like PICA. This suggestion is based largely on the assumption that hot-carriers in FETs cannot have excess energies much larger than V_{dd}, and the fact that contemporary optical detectors capable of picosecond timing measurements on low-light-level signals have poor sensitivity at photon energies below 1 eV. This would produce a fatal mismatch between the spectra of the light emission and the response curves of the detectors. In addition, the switching currents due to the decreased dimensions of the devices will also decrease. Significant decreases of the intensity of the hot-carrier emission below present levels would pose a major challenge to techniques like PICA for future backside measurements of working chips.

The intensity of the hot-carrier emission from a FET depends on the source-drain current, and the number of carriers with enough kinetic energy to produce detectable light. Reductions in the source-drain current from existing to future technologies can be estimated from arguments based on generalized scaling theory for MOSFETs. The density of high energy carriers also depends on the current-voltage characteristics of the working device as well as the hot-carrier distribution function near the drain of the FET in saturation. The important role of the carrier distribution function produces a strong exponential dependence of the hot-carrier emission intensity on the voltage,V_{ds}, between the source and drain for saturated FETs.[4] This paper first describes results related to recent self-consistent Monte Carlo simulations[5,6] of the electrical

properties of small Si FETs. The simulations included long-range Coulomb interactions between carriers. The calculations generate near-drain hot-carrier distribution functions for nFETs with effective gate lengths, L_{eff} =0.15 μm operating at 1.8V, and 50 nm, at V_{dd} around 1V. Previously published simulations [5,6] of similar devices usually involved larger values of V_{dd} where the bias voltages were much greater than the minimum energy of detectable photons. The low voltage simulations show that several percent of the hot-carriers have more than 1.2 eV of kinetic energy for future, small, low voltage, CMOS technologies. Such spectra can be observed by existing fast, low-light-level photodetectors so that hot-carrier light emission techniques for IC characterization will continue to be applicable for CMOS technologies at the 50 nm, 1V, level. The simulations suggest future experiments since they highlight significant differences between the hot carrier distribution functions for micron scale devices, and sub 0.1 μm FETs.

Simple analytic approaches to the magnitudes of the electric fields near the drain of an FET in saturation, and the simulations, are combined to provide estimates of the relative intensities of switching induced hot-carrier light emission for CMOS technologies with a range of different parameters going beyond the simulations. The effects on the intensity of the hot luminescence of the variations of parameters such as the oxide thickness, and junction depths of the source and drain are discussed.

The scaling induced changes in the magnitudes of the switching currents and voltages in present and future FETs, in the context of the CMOS gates they will be part of, are also described. The switching currents in CMOS circuits depend on the loading of the gates. The increasing importance of circuit delays arising from interconnects can increase the hot-carrier emission. Short channel and ballistic transport effects also increase the maximum energies of hot carriers near the drain of a small FET, increasing the intensity of the switching induced emission. Unlike the long channel devices, in short channel and ballistic devices, the emission is strong when the gate and source-drain voltages are equal.

Previously reported measurements of hot-electron emission in 0.18μm nFETs connect theoretically derived hot-carrier distributions to existing detectors and provide an experimental basis for estimates of the ability to detect hot-carrier light emission in future CMOS technologies using these detectors. The experimental results verify a number of the predictions of the Monte Carlo simulations for the behavior of the high energy tails in current CMOS technologies. The

measurable hot-carrier light emission at photon energies above the 1eV bandgap of Si, for V_{dd} less than the Si bandgap, also reflects the fact that the emission shares the same origin as the excess substrate currents and defect generation rates in small FETs for V_{dd}<1V, i.e. a hot-carrier distribution function with an anomalous high energy tail due to Coulomb effects.[5,6]

II. Hot-carrier luminescence and hot-carrier distribution functions.

PICA emission from a normally functioning CMOS IC is due to the currents that occur during switching of its gates built from n- and pFETS. The evolution of CMOS circuits, and their FETs from today's devices to the 50 nm gate lengths anticipated after 2005 can be described in terms of the scaling of the dimensions of the devices and their operating voltages. This provides a framework for the description of future circuit and device properties. The hot-luminescence from an FET depends linearly on the amount of charge flowing through it so that if all other parameters affecting the emission are held constant, the changes in the current due to geometric scaling produce proportional changes in the emission intensity. Since the hot-luminescence has been used to measure the behavior of 0.18 μm devices operating at 1.8V, the simple scaling of this emission to 50 nm devices will not produce large changes in the switching currents and the current dependent emission intensities.

The linear dependence of the intensity of the hot-emission on the switching current is dwarfed by the intensity's exponential dependence on V_{ds}. Changing the bias voltage from 1.5 to 1V in a saturated FET at a gate voltage, V_{gs}, produces a several orders of magnitude decrease in the intensity of the emission, far greater than the change in the current. The populations in the high energy tails of the hot-carrier kinetic energy distribution in an FET can change exponentially with V_{ds} and V_{gs} in large, micron scale, devices.[4] The effects on the magnitude of the high energy tail of a hot-carrier distribution due to the scaling of CMOS gates to smaller lengths and lower operating voltages is the principle focus of this paper as it demonstrates that these tails do not disappear rapidly for kinetic energies greater than 1eV when the bias voltage decreases below 1V, contrary to recent speculation.

The emphasis on the magnitude of the hot-carrier distribution at energies near 1.2eV is due to the fact that the spectral response of fast photodetectors such as imaging-microchannel-plate-photomultipliers, and single-photon-avalanche-photodiodes, falls

off precipitously at energies below 1eV.[7,8] In the case of the imaging photomultipliers, the low energy threshold is about 1.25eV. These detectors are currently used to observe and temporally resolve the switching induced hot-carrier emission from CMOS circuits in backside measurements on working chips.[1,7] Because practical measurements are generally made through the substrates of flip-chip packaged parts, given the practical thicknesses (approximately 100 μm) to which chips can now be thinned, emission from FETs at energies above 1.5eV is strongly attenuated by the Si substrate. Therefore, it is unnecessary to know the behavior of the carrier distribution function at energies much larger than 1.5eV.

The hot-carrier distribution function near the drain of an nFET reflects a variety of factors including the Coulomb interaction between the gate induced charge carriers and the high densities of carriers in sources and drains. The detailed calculation of the hot-carrier emission intensity and its spectral distribution is difficult since it involves both the calculation of the band structure derived matrix elements for optical transitions as well as the statistical mechanics of the distribution functions.[9,10] However, the changes observed in the time resolved hot-carrier emission from FETs as a function of operating conditions or device characteristics arise mainly from changes in the number of hot-carriers and their energy distribution. The changes are largely independent of the matrix elements for the optical transitions over narrow ranges of energies. Therefore, once the emission is measured in a CMOS device technology over the spectral range of interest, where the hot-carrier kinetic energy distribution function has also been calculated, the intensity of the hot-carrier emission from any other CMOS technology can be determined if the appropriate hot-carrier distribution is known. In the following section, this is done for simulations and experimental results obtained on contemporary sub-0.25 μm metallurgical gate length FETs operating at 1.8V, and then applied to future 50 nm, 1V technologies through simulations.

III. Simulations of Hot-Carrier Distributions in 50 nm and 0.25 μm FETs.

Two distinct types of hot-carrier distributions have been generated by simulations of the electronic properties of small Si FETs.[5] The simulations are distinguished by their behavior at higher kinetic energies, and are strongly affected by whether long-range Coulomb interactions between the charge carriers and the heavily doped sources and drains are included in the

modeling. The distributions in Figure 1 were derived for a 50nm metallurgical channel length, L_{eff} =65 nm, nFET using a 2D Monte Carlo/Poisson simulator. The gate oxide thickness was 2.8 nm, the source and drain doping levels were $2.6x10^{20}$ As atom/cm^3, and the junction depth was 60 nm. $V_{ds}=V_{gs}$ = 1V. The previously published energy distributions for this device generally involved an operating voltages of 1.5V.[5] The dimensions and 1V operating voltage of the simulated device in Figure 1 are similar to those expected for future 50 nm CMOS technologies. The open circles in Figure 1 define a simple distribution whose average energy with respect to the bottom of the conduction band is 0.42 eV. The effects of long-range Coulomb interactions between the high carrier densities in the source and drain and the charge carriers in the channel are included in this self-consistent simulation. For kinetic energies between 0.3 and 1.5eV the functional behavior of the distribution is close to that of an exponential characterized by an effective temperature, T_e =2700K. This exponential is shown by the dotted line. About 28,000 particles were used for the full channel simulation. A related simulation on the same device, but for V_{ds}=1.5V, can be found in Figure 17 of Fischetti and Laux(FL) [5].

For hot-carriers in semiconductors, the high energy carrier distribution often takes the functional form of an exponential with an effective temperature, T_e, which is greater than the lattice temperature. For the L_{eff}=65 nm FET simulated in Figure 1, this functional form is just a fit for the actual energy distribution. This was explained by FL[5] to arise from phenomena very different from the randomized independent particle model that is normally used to derive the Maxwell Boltzman distribution. The fluctuations in the source and drain are characterized by their plasma frequency. The resulting high energy tail in the distribution has a characteristic energy that is the strength of the Coulomb interaction between the high density carriers. FL have used the expression "Coulomb temperature," to describe this characteristic energy. The hot carriers in the channel come into equilibrium with the carriers in the source and drain through the Coulomb interaction, producing the exponential, 2700 K-like, high energy tail for the channel carriers in Figure 1a. About 2% of all the hot carriers have kinetic energies above 1.1eV and would be detectable by an imaging-microchannel-plate-photomultiplier like that used by Tsang and Kash in their recent studies of switching induced hot carrier emission from working devices.[1,7]

The solid dots in Figure 1 show the calculated hot-carrier distribution for the same device, and biased in the same way, except that the Coulomb interactions involving the charge carriers in the FET's channel and the electrons in the heavily doped source and drain are not included in the simulation. The solid symbols show that in the absence of carrier-carrier scattering involving the source and drain, the maximum kinetic energy of charge carriers in the channel is strongly limited by the bias voltage. The high energy tail of the hot-carrier distribution in the absense of Coulomb interactions involving the heavily doped source and drain has a characteristic temperature of about 300 K and appears at a kinetic energy less than V_{ds}.[5] The 300K exponential tail represents the exchange of energy between the hot carriers and the room temperature phonons. The number of hot-carriers described by the solid dots of Figure 1 which can generate photons detectable by the imaging-photomultiplier used by Tsang and Kash is at least three orders of magnitude smaller than for the open circles in Figure 1. Failure to account in simulations of hot-carriers in the FET for the presence of long-range Coulomb interactions between the high densities of source and drain carriers and the charge carriers in the channel produces a severe underestimate of the number of hot-carriers with energies greater than V_{ds}. The incorrect hypothesis that there will be very little hot-carrier light emission from CMOS devices at photon energies above 1 eV for operating voltages below 1V is an artifact of single particle simulations of hot-carrier behavior in small FETs.

Other characteristics of the hot-carrier distributions in Figure 1 can be calculated from the Monte Carlo simulations. For the full simulation including the long range Coulomb interaction, the peak field near the drain is about $4x10^5$ V/cm, and the peak average energy, $<E>_{av}$, of the hot-carriers is about 0.42 eV with respect to the bottom of the conduction band. The spatial variation of these parameters are shown in Figure 2 for $V_{ds}=V_{gs}=1V$. The highest energy carriers are found within a few nanometers of the drain, and are associated with its high field regions or where the electron potential energy changes most rapidly in space. The non-zero average kinetic energy around the source and drain in Fig. 2 is due to the filling of the conduction bands by the heavy doping. Unlike the case of the high energy tail of the hot-electron distribution where the failure to properly include the Coulomb interaction produces order of magnitude changes at energies above V_{ds}, the Coulomb interaction has only a small impact on the average energy of the

carrier distribution and the peak electric field as shown previously by Fischetti, Laux and Crabbe.[6]

The hot-carrier energy distribution functions near the drain of the 50 nm device for V_{ds}=0.5, 0.75, 1 and 1.5V and V_{gs}=1V are shown in Figure 3. The average energy of the carrier distribution near the drain in the 50 nm FET depends on V_{ds}, varying from about .18eV (V_{ds}=0.5V and lower), to 0.25 eV (V_{ds}=0.75V), to 0.62 eV (V_{ds}=1.5V). T_e for the above values of V_{ds} are 0.16eV, 0.22eV, and 0.33 eV. The noise in the distributions seen at higher energies and lower densities is due to the finite number of particles used in the simulations. The interaction between the hot-carriers in the channel and the equilibrium carriers in the source and drain impose the "Coulomb temperature," of the heavily doped source and drain on the energy distribution of the hot carriers in the channel. This produces a high energy distribution function with a minimum T_e=2000K or 0.16eV. The temperature of the high energy tail of the distribution function for this device will not decrease to 300K as V_{ds} is reduced. FL have shown that increasing the distance between a particular point in the channel and the drain decreases the strength of the "hot" high energy tail of the electron distribution with respect to the 300K tail with orders of magnitude changes occuring on the 100nm scale. Increasing the average energy of the hot carriers in the channel can increase in their effective temperature and average energy. The minimum energy at which this "hot" tail first appears, and its relative intensity with respect to the low energy distribution function depends on the applied voltage, the proximity of the heavily doped regions, and other parameters as discussed in Ref. 5. At kinetic energies above 1eV, the dependence of the hot-carrier distribution function on V_{ds} for the 50 nm device is weaker than would be expected for an exponential distribution where the $T_e = E_{av}$. The existence of a minimum 2000 K "Coulomb temperature," describing the high energy tail of the hot-carrier energy distribution in sub 0.1 μm FETs is a unique characteristic of the simulations of FL.[5] Verification of the minimum temperature will be an important test of the simulations as appropriate devices become available.

The open circles in Figure 4 are the hot-carrier kinetic energy distribution near the drain of a 0.25 μm metallurgical gate length generated by a self consistent Monte Carlo simulation including Coulomb interactions. The oxide thickness is 6 nm, and the peak donor concentration in the drain is about $3x10^{20}$ cm^{-3}. L_{eff} = .15 μm nFET for this device. The junction depth is about

60nm for the drain extension. V_{dd} =V_g=1.8V in this simulation. These values are similar to those used in contemporary devices with drawn gate lengths in the range of a quarter-micron and from which PICA measurements have been reported.[2] The hot carrier distribution including Coulomb effects in Figure 4 is similar to that shown by the open circles in Figure 1. Its peak average energy is about .36 eV. In both cases, over 2% of the hot carriers are found with energies above 1.1eV. A dotted line characterized by a T_{eff} = 2700 K is also given in this figure and provides a good description of the high energy tail of the distribution. The filled dots in Figure 3 are obtained from a simulation which did not include Coulomb interactions for this device. As in the case of the Figure 1 when Coulomb effects are ignored, the distribution shows a rapid decrease in the number of carriers at energies at kinetic energies above 0.8 eV. Unlike the case of the 50 nm FET in Figure 1 where there are strong ballistic effects, the ballistic effects in the .25 μm gate length NFET are weaker. The decrease in the distribution above 0.8 V is more closely related to the 0.36 eV average energy of the hot-carrier distribution rather than V_{dd} as in the case of Figure 1 where Coulomb effects were neglected. Figure 4 shows that even for V_{dd}>>1V, single particle simulations of the hot-carrier distribution in FETs can produce undetectably low densities of carriers at energies near 1.2eV. In this case, failure to account for the Coulomb interaction between charges in the channel, source and drain will also produce several-order-of-magnitude errors in estimates of the number of hot-carriers with kinetic energies large enough to be observed by contemporary photon detectors.

Single particle Monte Carlo simulations by Higman et al. [9] have been carried out for 1 μm FETs biased at 5V. The hot-carrier distribution has a non-thermal shape, and is characterized by an average energy of above 0.6V. Between 1.2 and 2 eV, the hot-carrier density decreases with an effective temperature of several thousand degrees. At least 20% of the hot carriers have more than 1.2 eV of kinetic energy. This calculated distribution is consistent with hot-carrier luminescence measurements made on larger structures where the bias voltages are more than 1V above the range of photon energies which are detected and where the size of the FET is such that the long-range Coulomb effects are insignificant. [12]

The simulations of the hot-carrier distributions for 50nm and 0.25 μm and larger FETs, define when a measurable fraction of the hot-carriers can have kinetic energies greater than 1eV. This is predicted in all calculations to occur for micron-scale FETs in saturation when the bias

voltage above a few electron volts. For very small FETs with heavily doped sources and drains, failure to correctly include long range Coulomb effects between mobile charges in the channel, source and drain, leads a several orders of magnitude underestimate of the number of hot-carriers that can contribute to hot luminescence at photon energies above 1eV when V_{ds} is less than 1V. The long-range Coulomb interaction "thermalize" the channel hot-carriers near the drain, and produce the high energy tail needed for efficient measurements of hot-carrier emission in future CMOS technologies. The importance of the long-range Coulomb interaction for the high temperature "thermalization" of the hot-carrier distribution in small FETs is demonstrated in Figure 5 which shows the strong quantitative similarity of the hot-carrier distributions of the devices simulated in Figures 1 and 4 at energies above the respective average energies of 0.42 and 0.36 eV and less than 1.5 eV,

A PICA system based on a commercial imaging photomultiplier [7] has been used to study 0.18 μm FETs biased at 1.8V.[2] The hot-carrier distributions for this part should be quantitatively similar to the simulated distribution shown in Figure 3 and characterized by a T_e=2700 K with an average energy of .36 eV. The differences in the switching currents per transition for the L_{eff} = .15 μm and 65 nm devices biased at 1.8 and 1 V respective operating voltages should be less than 9X. Since the hot carrier distribution function for the 50 nm, 1V part in Figure 1 is almost identical to that for the larger part in Figure 3, which has been already measured with high signal to noise, the 50 nm, 1V part should produce enough light to be studied by existing PICA systems.

IV. Analytic Treatments of High Field Pinch off Regions in FETs

The Monte Carlo simulations of Fischetti and Laux for 0.25 μm and 50 nm FETs generate hot-carrier distributions and quantities including the electric field, and average electron energy in the FET as shown in Figure 2. Because of the small length scales involved in current and future CMOS technologies, the electrical behavior of device based on these technologies cannot be simply extracted from their local properties and electric fields. However the local electric fields still are critical quantities. In their simulations, FL showed that the peak fields, average energies of the hot-carrier distributions, and drift velocities[Figure 16, Ref. 5] along the channels are strongly correlated as seen in Figure 2. The highest fields, average energies, and drift velocities

are found in the high field region of an FET near the drain when the device is in saturation. The peak average electron energies and electric fields show similar dependences on V_{ds}.

Generalized scaling theory for FETs produces a set of relationships for the physical dimensions and operating voltages in CMOS devices which limit the changes in the shape and magnitude of the electric field as devices shrink. Simple analytic models for the magnitudes of the electric field near the drain in saturated FETs have been derived.[13] The analytic expressions show that the maximum electric field, E_{max} , in a saturated FET will occur near the drain. The maximum value of the field will be:

$$E_{max}= (((V_{ds}-V_{dsat})/l)^2 + E_{sat}^2)^{0.5} \quad (1)$$

where

$$l=((\varepsilon_{si}/\varepsilon_{ox})t_{ox}X_j)^{1/2}. \quad (2)$$

V_{ds} is the source drain voltage, V_{dsat} is the saturation voltage, E_{sat} is a saturation field which is the order of the critical field where velocity saturation effects become significant, ε_{si} and ε_{ox} are the dielectric constants for Si and the gate oxide, t_{ox} is the oxide thickness and X_j is the depth of the drain junction. E_{max} depends strongly on both the oxide thickness and the junction depth. If the lateral dimensions and V_{ds} for an FET are scaled by a constant factor, but t_{ox} and X_j are unchanged, the maximum field of the FET in saturation will decrease by the scaling factor. This is even though for operation of the FET in the linear range, the channel field will be constant. While the above expression are very simple approximations, it is interesting to note that the predicted values of E_{max} are in good agreement, within 20%, with those obtained from the simulations of Fischetti and Laux. Under these circumstances, the conservative values for the oxide thickness and junction depth used by Fischetti and Laux in their simulations for a 50 nm FET, as compared to the predictions for the 60nm node for 2008 in the "Year 2000" update to the International Technology Roadmap for Semiconductors (2.8 nm vs 1.2nm, and 60 nm vs 26 nm) mean that the near drain fields for the "Roadmap" 60 nm node will be about a factor of two larger than in the simulation. The strong correlation between the maximum electric fields near the drain, and the average energies of the hot-carriers shown by FL in their simulations means that the increase in the field will increase the average energy of the hot-carriers. Given the simple functional form of the hot carrier distribution, this increase in the average energy will produce a further increase in the density of hot-carriers with energy sufficient for their detection by

conventional fast optical detectors. This will help compensate for any decrease in the intensity of the hot-carrier emission due to the decrease in the switching current. This increase also means that if the bias voltage is reduced below 1V, there will still be a significant number of hot-carriers with kinetic energies above 1V. Conversely, if future CMOS technologies are unable to scale the oxide thicknesses and junction depths with decreasing gate lengths, or if the operating voltages is required to decrease faster than the gate lengths, there will be significant decreases in the intensity of the hot-carrier emission which may make it impossible to use it for circuit and device characterization. This argument is limited by the appearance of ballistic effects in the channel of the FET. If ballistic effects are dominate, then the hot-electron distribution will be defined by V_{ds} and the doping level of the source and drain, and be independent of the shape and magnitudes of the fields between the source and the drain. Ballistic transport means there is no energy loss in the channel until carriers reach the drain so that the energy of the hot-carriers near the drain is at a maximum which will maximize the number of electrons at high kinetic energies.

V. CMOS Circuit Scaling Effects on PICA Intensities.

The intensity of the hot-carrier emission due to switching in FETs depends very strongly on the details of the energy distribution of the hot-carriers near the drain. This quantity cannot be readily derived from the basic electrical characteristics of the device. The intensity of the emission used in PICA also shows a weaker dependence on the simple electrical properties of the FETs and their loads in CMOS circuits. These effects can be derived from conventional analyses of the effects of reductions in device dimensions and operating voltages on device behavior.

Reductions of CMOS device dimensions and operating voltages consistent with normal generalized scaling theory produce reductions in the device currents, switching times, and capacitances.[13] As device dimensions shrink, the density, total number of devices and die sizes increase in CMOS chips. The delays and time constants associated with the gates show increasing contributions from the interconnects. Additional interconnect related capacitances increase the amount of charge that a gate has to sink or source. This will increase the intensity of the hot carrier emission. Reductions in device dimensions also mean that the saturation of the drain current occurs at lower voltages due to velocity saturation. This also increases the intensity

of the switching-induce hot-carrier emission since it increases the difference between V_{ds} and V_{dsat}, a critical parameter in Eq. 1.

The total charge switched by a CMOS gate scales as the product of either the load capacitance and the operating voltage or the source-drain current and the switching time. For the devices in described in Figure 1 and 4, and scaling from the metallurgical gate lengths, the charge responsible for the switching current in the 1V, 50 nm device is about .11 of that in the 1.8V, 0.25 μm FET, assuming that the load capacitance scales with the gate dimensions. This would produce a comparable reduction in the emission intensity. This decrease is the decrease per switching event. Since it is associated with reduced gate to gate delays, such a decrease is associated with the ability of the chip to operate at higher speeds. The increased number of switching events-per-unit-time due to the higher speed operation can compensate for the decreased emission per switching event. Therefore, the emission-per-unit-time for a device operating at the highest speeds allowed by the reduced gate-to-gate delays remains constant.

If the capacitance of the load seen by the output of the gate in question does not scale with the device dimensions but is larger, then the light emission can be proportionally larger also as more charge will flow through the gate per switching event. For the case of an nFET, and an arbitrary input signal, if the switching time of the gate due to the size of the output capacitance is slower than the rise time of the input, the voltage between the source and drain will decrease more slowly. Such extra loading can arise due to the complexity of interconnects which increases with the number of gates on the chip, and the small physical size of the wires comprising the interconnects.

In Figure 6, we show a number of schematic switching waveforms for a simple inverter. The input waveform is shown by the broken line, and the output waveform, whose fall time is twice the rise time of the input, is shown by the solid line. The open circles represent the difference between the output voltage and the input voltage, or the voltage drop between the end of the channel and the drain for the nFET of the inverter. For small values of the threshold voltage, this is the key quantity in expression 1 above. The small, connected, solid squares represent the current through the nFET during the switching event indicated by the lines in Figure 6. Light is emitted only when there is current flowing through the FET, and when there is a substantial voltage drop between the end of the channel and the drain. Since the intensity of the

emission depends exponentially on the voltage drop near the drain for a large device, significant emission only occurs over a tiny fraction of the switching time. If the switching time is increased due to an increase in the loading of the circuit, more current will flow and there will be more emitted light. The increase will be linear in the switching time.

For the small FETs discussed already, the hot carrier distributions were calculated for bias conditions where $V_{gs}=V_{ds}$. The I-V characteristics of these devices show significant velocity saturation effects. Long channel devices show weak hot carrier luminescence when $V_{gs}=V_{ds}$. This is because the potential at the end of the channel is almost the same as at the drain so that the voltage drop across the pinch off region is small. As channel lengths decrease, the saturation of the drain current for $V_{ds}<<V_g-V_t$ where V_t is the threshold voltage produces the strong emission observed for $V_{gs}=V_{ds}$ in sub-micron nFETs. The appearance of current saturation for small values of V_{ds} means the carriers between the saturation point and the drain are not confined to the semiconductor-oxide interface region and cannot be described by 2 dimensional treatments. The voltage drop in eq. (1) between the end of the channel and the drain is the difference between V_{ds} and V_{dsat}. For very short channel devices[13],

$$V_{dsat} = (2v_{sat} L(V_g-V_t)/m\mu_{eff})^{0.5} \quad (3)$$

where v_{sat} is the saturation velocity, L is the gate length, and μ_{eff} is the electron mobility. For 50 nm devices, and 1V operating voltages, V_{dsat} is much smaller than V_g-V_t. As a result, $V_{ds}-V_{dsat}$, will be larger than in long channel devices. This is schematically shown in Figure 6 by the open squares. For ballistic devices, the limiting voltage drop near the drain is the full voltage across the nFET or V_{out}. In the short channel or ballistic nFET of a CMOS inverter, the voltage drop near the drain will decrease more slowly during switching than its long channel analogue. This will produce an increase in the intensity of the switching induced light emission. The magnitude of the increase will depend of the details of the devices and circuits. The capacitive and short channel effects can compensate for the decreased intensity of the switching induced light emission due to the scaling of the currents in these devices as the devices decrease in size, and the operating voltages are reduced. Therefore, the factor of 0.11 change mentioned earlier in this section is the worse case.

VI. Measured Hot Carrier Emission Intensities in Sub-Micron FETs

Switching induced hot carrier light emission has been reported in many different sub-micron CMOS technologies.[1,2,12,15] These now include 0.7 μm, and 0.25-0.15 μm technologies. The hot carrier emission obtained from the front side of 0.7 μm technology gates where the bias voltages are above 3.5 V have been characterized by effective temperatures between 2000-3000 K for photon energies between 1 and 3 eV.[12] This is consistent with the simulations of Higman et al., since the measured 0.7 μm gates include lightly doped drain extensions to reduce the maximum electric fields near the drain.[14]

Published switching induced hot carrier emission at photon energies between .9 and 1.3eV from the backsides of 0.25 μm nFETs can be characterized by an effective temperature of about 1350K.[15] This is close to the 1205K effective temperature predicted by Fischetti and Laux for the high energy tail of the electron distribution for a donor concentration of 10^{20} cm^{-3}.

The variation with operating voltage of the intensity of the switching induced hot carrier emission for 0.25 μm technologies with operating voltage and lateral dimensions is also consistent with simulations and simple modeling.[15,16] T_e of the hot carrier distribution shows a weak dependence on bias voltage[15,16], consistent with the fact that the high energy tail is a property of the high density of carriers in the source and drain of the devices, and that the increase in the average energy of the carrier distribution is associated with shifts of the high energy tail to higher energies as well as changes in its slope.[17] Finally, measurements of the light emission from families of quarter-micron, 1.8V, devices where the bias voltages, and lateral dimensions, but not the vertical dimensions have been scaled show behavior in agreement with the simple analytic treatment of the high fields near the drains in saturated FETs. The decrease in the bias voltage, without an accompanying decrease in the oxide thickness and junction depth produces an exponential decrease in the intensity of the hot carrier emission, even if the gate length is also reduced.[16]

VII. Conclusions.

The high energy tails of the hot carrier distribution functions determine the strength of the hot carrier luminescence in FETs. It has been shown previously that the high energy tail also

determines the substrate currents and defect generation rates at the Si-SiO_2 interface.[4,15] The distribution of hot carriers at kinetic energies above the 1 eV band gap in silicon in small FETs has been of interest since the observation of appreciable substrate currents for bias voltages below the bandgap of Si.[18] Fischetti and Laux explained the experimental measurements of both excess substrate currents and continued defect generation through the simulations described in this paper. The absense of either a V_{dd} related threshold for the disappearance of the substrate current or a reduction in the generation rate for interface defects, also explains in a phenomenological manner why switching induced hot carrier luminescence will be an effective tool for circuit and device characterization into the second half of this decade. This has been shown here combining results obtained on existing systems when the bias voltages near 1.8V, and through simulations of future sub 0.1 μm CMOS technologies for low operating voltages. These results all show the critical importance of Coulomb effects in understanding the properties of future advanced CMOS technologies.

References.

1. J. C. Tsang, J. A. Kash, and D. P. Vallett, Proceedings of the IEEE 85, 1440-1459 (2000).

2. W. Huott, M. McManus, D. Knebel, S. Steen, D. Manzer, P. Sanda, S. Wilson, Y. Chan, A Pelella and S. Polonsky, Proceeding of the International Test Conference, Atlantic City, NJ, 1999, 883.

3. "International Technology Roadmap for Semiconductors-2000 Update," Semiconductor Industry Association, 2000.

4. S. Tan and C. Hu, IEEE Trans. Electron Devices ED-31, 1264 (1984).

5. M. Fischetti and S. Laux, Journal of Applied Physics 89, 1205 (2001).

6. M. Fischetti, S. Laux, and E. Crabbe, Journal of Applied Physics 78, 1058 (1995).

7. J. C. Tsang, J. A. Kash and D. P. Vallett, IBM J Res. Develop. 44, 583 (2000).

8. F. Stellari, F. Zappa, S. Cova, C. Porta and J. C. Tsang (submitted for publication).

9. J. Bude, N. Sano, and A. Yoshii, Phys. Rev. B45, 5848 (1992).

10. S. Villa, A. Lacaita, and A. Pacelli, Phys. Rev. B42, 10993 (1995).

11. J. Higman, K. Hess, C. Hwang, and R. Dutton, IEEE Trans. Elec. Dev. 36, 930 (1989).

12. J. A. Kash and J. C. Tsang, Phys. Stat. Solidi(B) 204, 507 (1997).

13. Y. Taur and T. H. Ning, "Fundamentals of Modern VLSI Devices," Cambridge University Press, Cambridge, UK, 1998, p.156.

14. C. Koburger et. al., IBM J Res. Develop. 39, 215 (1995).

15. S. Rusu, S. Seidel, G. Woods, D. Grannes, H. Muljono, J. Rowlette and K. Petrosky, IEEE ISSCC-2001, 276 (2001).

16. J. C. Tsang (unpublished)

17. A. Pacelli and A. Lacaita, Semicond. Sci. Technol. 11, 1642 (1996).

18. M. V. Fischetti and S. E. Laux, Tech. Dig. IEDM 1995, 305 (1995).

Figure Captions:

Figure 1. Calculated electron energy distribution functions for carriers near the drain, where the average energy of the carriers is near its peak, for an L_{eff}=65 nm FET biased at 1V and described in the text. The open circles are the distribution calculated including the long range Coulomb interaction with carriers in the drain while the filled dots are the distribution neglecting this.

Figure 2. The dependence of the average electron energy and electric field strength on position in a 50 nm nFET for V_{ds}=V_{gs}=1.0 V. Other parameters for the FET are given in the text.

Figure 3. Calculated electron energy distributions for the device in Figure 1 for V_{ds}=0.5, 0 .75, 1.0 and 1.5V. V_{gs}=1V.

Figure 4. Calculated electron energy distribution functions for carriers near the drain, where the average energy of the carriers is near its peak, for an L_{eff} =0.15 μm channel length FET biased at 1.8V and described in the text. The open circles are the distribution function including Coulomb processes while the filled dots are the function neglecting Coulomb interactions.

Figure 5. A comparison of the hot carrier distribution functions derived for the L_{eff}=65 nm technology in Figure 1, and the 0.15 μm technology described in Figure 3. The long range Coulomb interaction between carriers in the channel and the high density of doped carriers in the drain of the FET was included in the simulations.

Figure 6. A schematic of the relationship between the input and output voltages in a CMOS gate and the difference between the output voltage and the saturation voltage for long channel, and short channel devices. The light emission measured in PICA measurements occurs when the current through the device is > 0, and either V_{out}-V_{in}, or V_{out} -V_{dsat} is large. In short channel devices, the latter will always be larger than the former

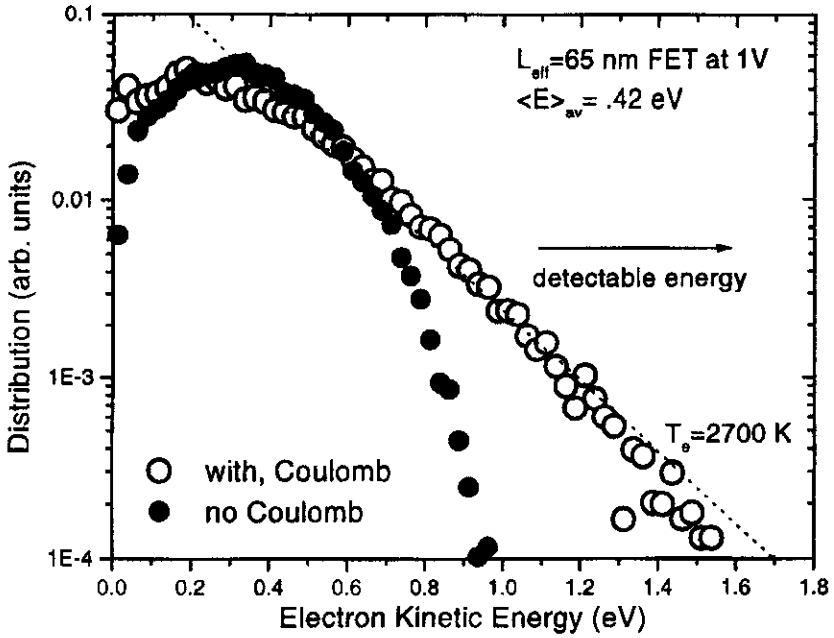

Figure 1

Figure 1. Calculated electron energy distribution functions for carriers near the drain, where the average energy of the carriers is near its peak, for an L_{eff}=65 nm FET biased at 1V and described in the text. The open circles are the distribution calculated including the long range Coulomb interaction with carriers in the drain while the filled dots are the distribution neglecting this.

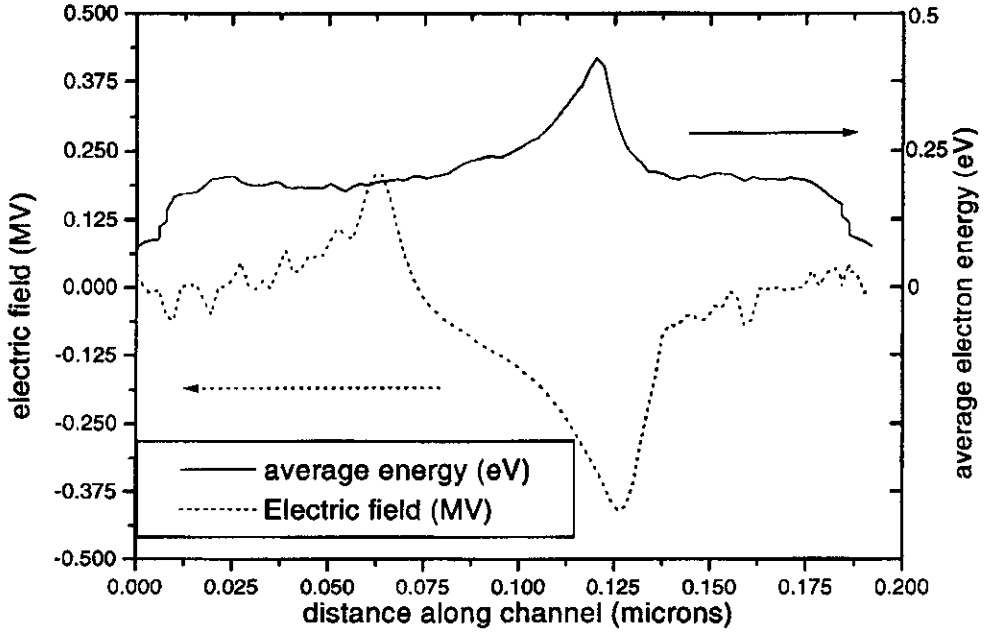

Figure 2

Figure 2. The dependence of the average electron energy and electric field strength on position in a 50 nm nFET for $V_{ds}=V_{gs}=1.0$ V. Other parameters for the FET are given in the text.

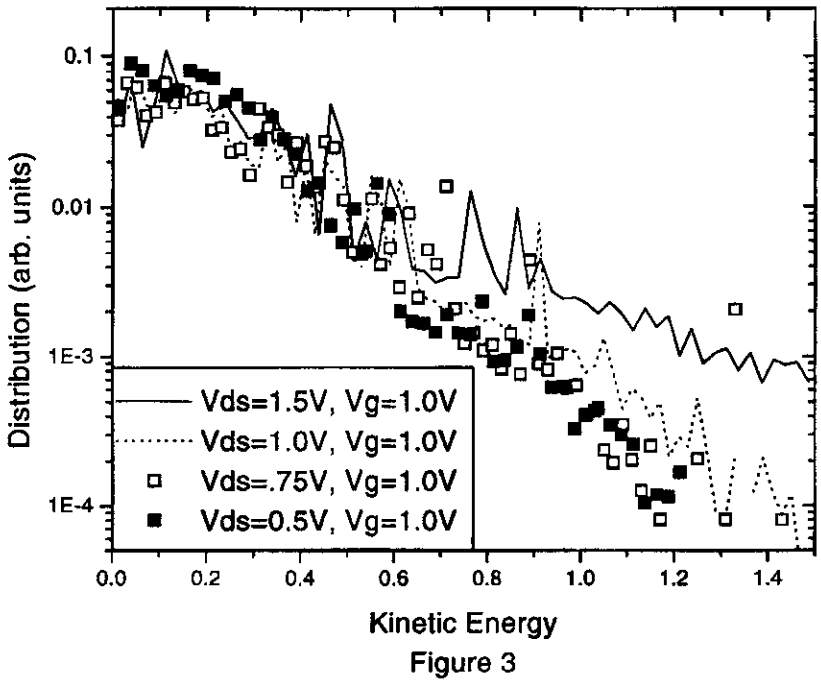

Figure 3

Figure 3. Calculated electron energy distributions for the device in Figure 1 for V_{ds}=0.5, 0 .75, 1.0 and 1.5V. V_{gs}=1V.

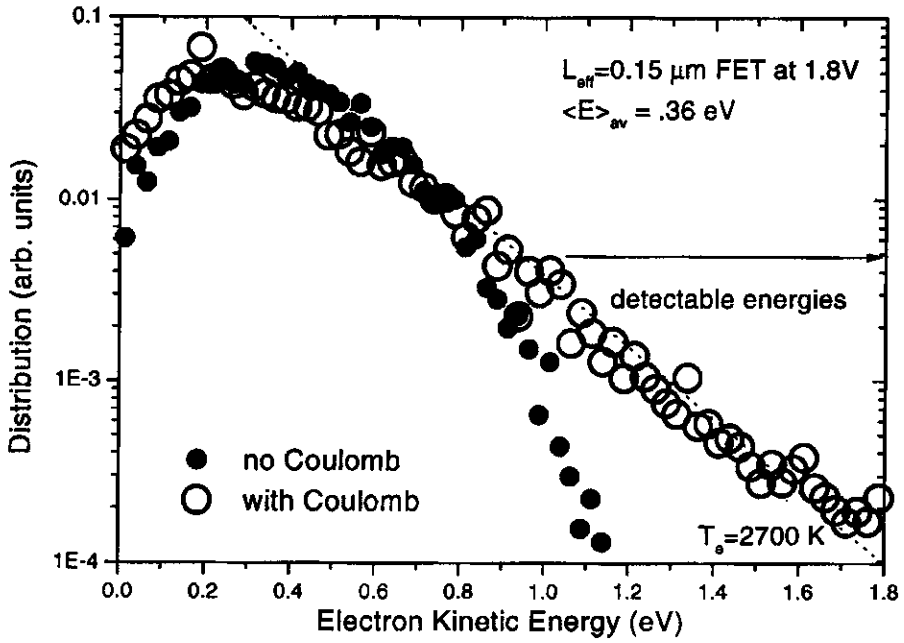

Figure 4

Figure 4. Calculated electron energy distribution functions for carriers near the drain, where the average energy of the carriers is near its peak, for an L_{eff} =0.15 μm channel length FET biased at 1.8V and described in the text. The open circles are the distribution function including Coulomb processes while the filled dots are the function neglecting Coulomb interactions.

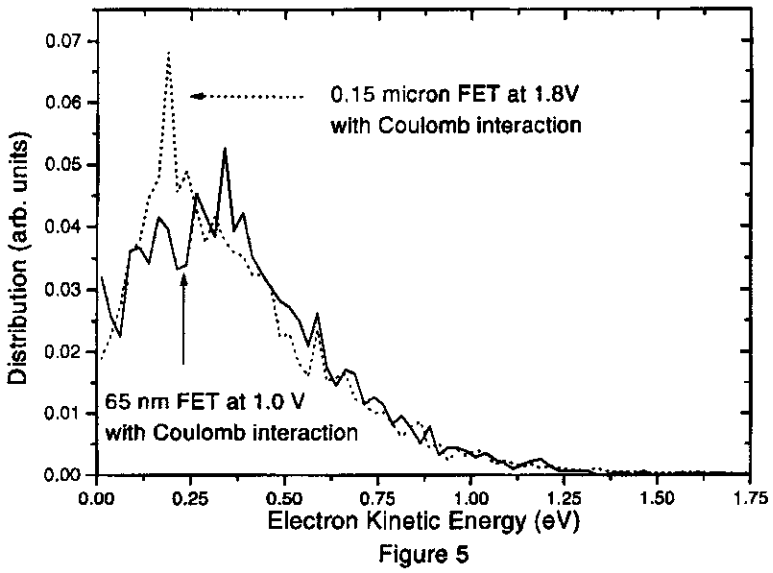

Figure 5. A comparison of the hot carrier distribution functions derived for the L_{eff}=65 nm technology in Figure 1, and the 0.15 μm technology described in Figure 3. The long range Coulomb interaction between carriers in the channel and the high density of doped carriers in the drain of the FET was included in the simulations.

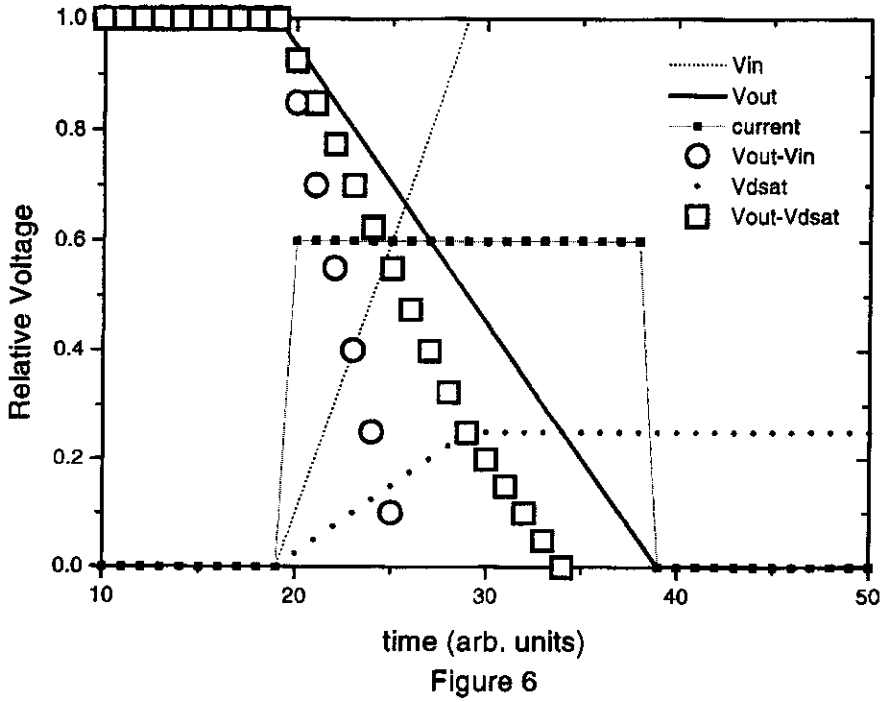

Figure 6

Figure 6. A schematic of the relationship between the input and output voltages in a CMOS gate and the difference between the output voltage and the saturation voltage for long channel, and short channel devices. The light emission measured in PICA measurements occurs when the current through the device is > 0, and either V_{out}-V_{in}, or V_{out} -V_{dsat} is large. In short channel devices, the latter will always be larger than the former