Revista Brasileira de Aplicação de Vácuo, Vol. 10, nº 1, 1991

MOSFET OPERATION AT CRYOGENIC TEMPERATURES

R. L. ANDERSON
Cryoelectronics Laboratory

University of Vermont
Burlington VT 05405
U.S.A.

ABSTRACT

Device improvements obtained from exploiting the dependence of physical characteristics of silicon at low temperature are above and beyond those improvements obtained from the usual geometric scaling of device dimensions. As device geometries continue to shrink into the deep submicrometer regime, second order effects begin to limit further increases in device speed from scaling alone. Temperature scaling provides an additional variable for system optimization. The gain in performance at cryogenic temperatures, however, isat the expense of inconvenience and additional cost for system refrigeration ,and of increased susceptibility to hot carrier degradation. Optimizing CMOS technology and design for low temperature applications can increase performance and reduce power dissipation without increasing hot carrier degradation.

INTRODUCTION

Operation of semiconductors at cryogenic temperatures, particularly atliquid nitrogen (LN) temperature, has received considerable attention duringthe past 25 years [1-7]. We can distinguish three possible motivations for the increased research in this area.

One reason is to obtain an understanding of the general properties of materials and device physics at low temperatures. Some properties exhibit substantial changes when compared to the more familiar behavior at room temperature. Understanding these changes is of great importance to low temperature as well as room temperature applications. In particular, the temperature dependence of certain parameters gives insight to the physical mechanisms responsible for the parameter values at room temperature and togive understanding to some of the failure mechanisms.

The second reason is that some devices can operate only at low temperature, as is the case, for example, in superconducting structures and Josephson devices. Additionally, MOSFET's having channel lengths below about 0.15 m may not be operable at room temperature.

Finally, there is the optimization of characteristics and performance improvement of a device which otherwise can operate at room temperature. Thisis the case, for example, in CMOS circuits where an appreciable reduction inpower delay product is achieved at cryogenic temperatures when compared to room temperature operation. In this case, however, the performance gain must justify the cost and inconvenience of refrigeration.

The following discussion focuses on bulk silicon CMOS applications at cryogenic temperatures. Improvements in device and circuit performance and power are described in terms of changes in the physical properties of the semiconductor. This is followed by a discussion of reliability considerations.

To be specific, we compare the design criteria and power supply considerations for CMOS circuits (LNCMOS) designed for liquid nitrogen temperatures (LNT or 77 K) compared to those (RTCMOS operating at room temperature (RT or 300 K). For illustration, a digital CMOS invertor circuitis discussed Figure 1 shows the schematic of a typical CMOS invertor.

While MOSFET's have been observed to operate down to liquid helium temperature [8,9] because of carrier freezeout effects the devices behave appreciably differently than at room temperature [10,11]. Because their operation cannot be predicted from tests made at room temperature, there islittle current work at such low temperatures. This paper does not consider operating temperatures below

77 K.

SWITCHING SPEED

Switching speed is determined by the time required to charge and discharge circuit capacitance. This is dependent on the magnitude of the capacitance, the values of the contact resistance, the MOSFET series resistance, the conducting line resistance and the MOSFET transconductance. All of these parameters favor increased switching speed at 77 K compared to 300 K.

The oxide capacitance is virtually temperature independent. The junction or depletion region capacitance, however, decreases slightly with decreasing temperature because of the increase in built in voltage and thusan increased junction width. This capacitance can be further decreased at low temperature because the doping beneath the gate can be reduced to obtain asmaller threshold voltage in the LNCMOS devices as is discussed later. This reduced capacitance is a relatively small effect, however, and will not be considered further.

The conductivity of aluminum is about an order of magnitude larger at 77 K than at room temperature[4,12]. For a given cross section of aconducting line the line resistance is reduced by this same factor. Atradeoff can be made between line resistance and chip area covered by thelines. Reduced line area permits a higher packing density. We note that ifthe conducting lines are degenerate polycrystalline Si or ion implanted Si,the conductivity increase is less than about 20% [4].

The MOSFET series resistance is also expected to be smaller at 77 K compared to room temperature because of the increased carrier mobility in thesource and drain diffusions. We have measured series resistance in p channel and in n channel MOSFET's and have found a decrease in resistance with decreasing temperature as shown in Fig. (2). In both cases, the 77 K value of series resistance is about one half its room temperature value.

Many modern MOSFET's are contacted by silicides. It is found that the resistivity of $TiSi_2$ is less by a factor of 3-4 at LNT compared to RT [13].

For a device of a given geometry, the transconductance is proportionalto the carrier drift velocity in the channel. At low channel fields (longchannels, small drain voltage) the transconductance is proportional to thelow field carrier mobility. This mobility is observed to increase with decreasing temperature to about 60 100 K because of the reduced carrier scattering due to lattice vibrations. At lower temperatures yet, the mobility tends to decrease because of scattering by ionized impurities. The temperature corresponding to the maximum mobility is dependent on the total impurity concentration, decreasing with reduced concentration. Fig. (3) showsthe temperature dependence of channel mobility for both electrons (NMOS) andholes (PMOS) for a given technology. The mobility is typically a factor offour to six larger at LNT than at RT. The exact increase is technology~dependent. For these long channel devices the transconductance, and the switching speed, is increased by this same factor for a given supply voltage.

At high channel fields (short channels, large drain voltage), however,the drift velocity is limited by the carrier interaction with optical phonons. There is some controversy in the literature about the value of this carrier saturation velocitv and its temperature dependence [14,15]. We have measured v_s for holes to be on the order of 1.3×10^7 cm/s at 77 K, some 40% greater than at room temperature.

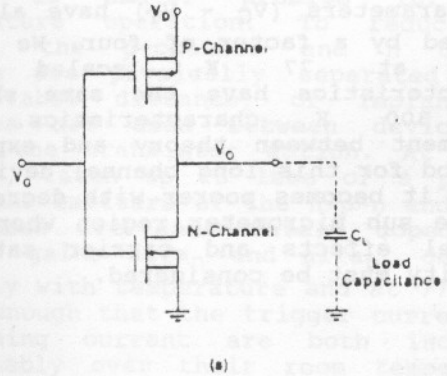
The switching time is related to the time it takes to turn the devices OFF and ON. Since in the OFF condition The FET is operating well below threshold, the variation of drain current I_D with gate voltage V_G in this subthreshold region is important. Fig. (4) shows the log I_D - V_G characteristics for a PMOS device at several temperatures. It can be seen that the subthreshold slope increases with decreasing temperature thusreducing the switching time. It is also noted that the extrapolated subthreshold characteristics merge at a common point corresponding to thethreshold voltage at zero kelvins.

DEVICE TEMPERATURE CONSIDERATIONS

In many integrated circuits, operation is limited by the temperature of the chip. The chip temperature can be reduced by increasing the heat transfer rate or by decreasing the power dissipation per transistor. Both of these conditions can be met at 77 K. First, the thermal conductivity of Siis a factor of 6 larger than at room temperature so that dissipated heat canbe more easily removed from the transistors. Further, if the chip is immersedin liquid nitrogen the chip liquid heat transfer rate is larger than the chip-air transfer rate

associated with room temperature operation. The net heat transfer rate in this case is about a factor of ten larger than at roomtemperature [2].

Further, at reduced temperatures, the devices can be operated at lower supply voltages such that the power dissipation per device can be reduced[16,17]. To illustrate the reduced supply voltage permitted we take the caseof a digital CMOS invertor of Fig.(1) in which the gate to source voltage as well as the drain to source voltage is



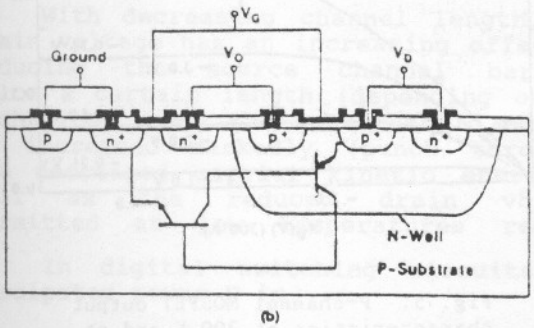


Fig. 1. (a) Schematic of a CMOS invertor and (b) cross-section indicating typical parasitic bipolar structures.

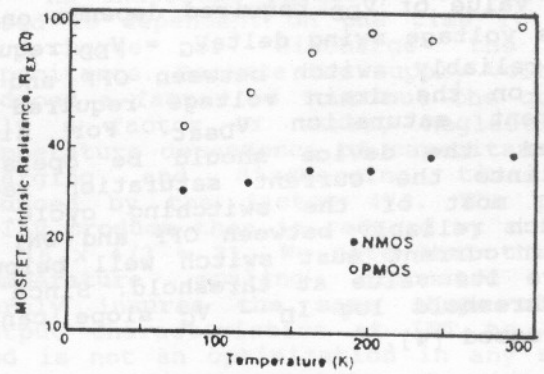


Fig. 2. Temperature dependence of MOSFET extrinsic resistance.

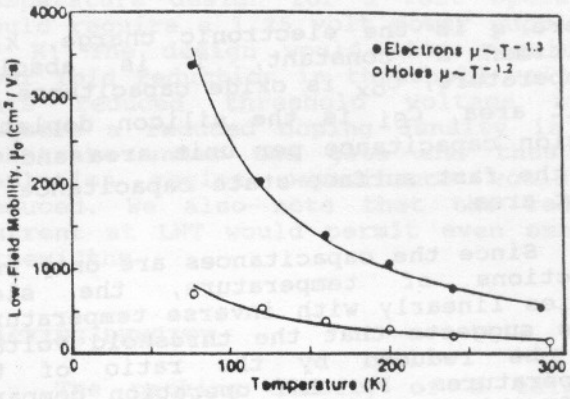


Fig. 3. Temperature dependence of low field channel mobility for electrons and for holes.

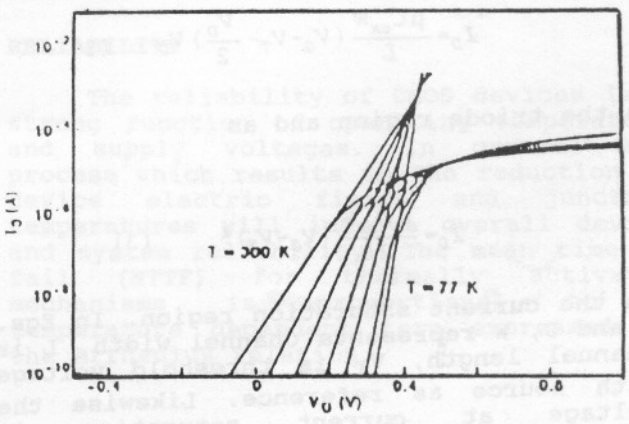


Fig. 4. Semi-log plot of drain current-gate voltage characteristics at constant V_D = 0.05 V. The subthreshold slope increases with reducing temperature.

either zero or the supply voltage V_{DD}. The value of V_{DD} required depends on the gate voltage swing deltaV$_G$ = V_{DD} required to reliably switch between OFF and ON, and on the drain voltage required for current saturation V_{Dsat}. For higher speed, the device should be operating wellinto the current saturation region over most of the switching cycle. To switch reliably between OFF and ON, the drain current must switch well belowand above its value at threshold. Since the subthreshold log I_D - V_G slope can be expressed [4],

$$slope = \frac{qC_{ox}}{2.3kt(C_{ox}C_{si}+C_{fs})} \quad (1)$$

where q is the electronic charge, k is Boltzmann's constant, T is absolute temperature, C_{ox} is oxide capacitance per unit area, C_{Si} is the silicon depletion region capacitance per unit area and C_{fs} is the fast surface-state capacitance per unit area.

Since the capacitances are only weak functions of temperature, the slope varies linearly with inverse temperature. This suggests that the threshold voltage can be reduced by the ratio of the temperatures. For LNT operation compared to RT this is a factor of four reduction.

Above threshold the current can be expressed approximately as

$$I_D = \frac{\mu C_{ox} W}{L}\left(V_G - V_T - \frac{V_D}{2}\right) V_D \quad (2)$$

in the triode region and as

$$I_D = \frac{\mu C_{ox} W}{2L}(V_G - V_T)^2 \quad (3)$$

in the current saturation region. In Eqs. 2 and 3, W represents channel width, L is channel length, V_T is threshold voltage with source as reference. Likewise the voltage at current saturation is approximately

$$V_{Dsat} = (V_G - V_T) \quad (4)$$

Thus if V_G = V_{DD} and V_T are both reduced by a factor of four, V_{Dsat} will be reduced by this same factor and the operation at 77 K will be similar to thatat room temperature.

Since all voltages are reduced by a factor of four, the current is reduced by the factor 77/16 300 (Eqs (2) and (3)). As indicated earlier, itis found that is a factor of 4 to 6 larger at LNT. Choosing an average value of 5.3 results in a reduction in current by a factor of 3. [16].

This behavior has been experimentally confirmed. Fig. (5) shows the output characteristics for a p channel MOSFET at 300 K and at 77 K. The voltage and current scales are reduced by factors of four and three respectively. The parameters (V_G - V_T) have also been reduced by a factor of four. We can see that at 77 K, scaled output characteristics have the same shape as the 300 K characteristics. While agreement between theory and experiment is good for this long channel device (L = 9 m) it becomes poorer with decreasing L in the sub micrometer region where short channel effects and carrier saturation velocity must be considered.

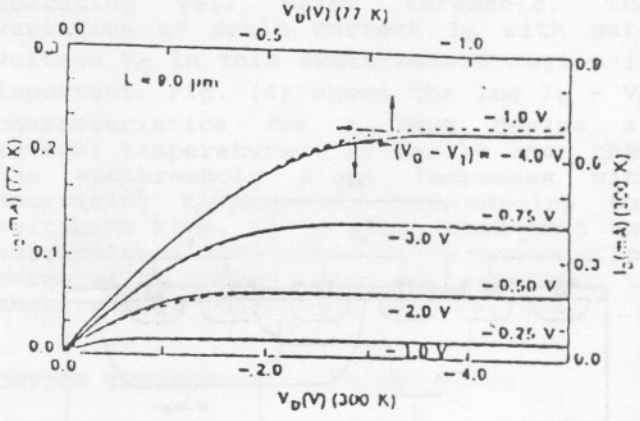


Fig. 5. P-channel MOSPET output characteristics at 300 K and at 77 K. V_D and (V_G - V_T) are scaled by the factor 1/4 at 77 K while the current is scaled by the factor 1/3.

connections; by the spacing between device sand by the operating temperature of the MOSFET's.

While the increased conductivity of aluminum lines can reduce the surface area associated with these lines, the major restrictions on packing density results from the requirement to separate devices to minimize latch-up [18,19] and to avoid exceeding a specified operating temperature.

Latch-up occurs in a CMOS transistor by the 4-layer bipolar transistoraction between the source of the n channel transistor and the drain of the p-channel device (Fig 1). The 4-layer device has a tendency to switch into a low resistance state (latch up) if the sum of the common base current gains of the two transistors n and p exceed unity. This latch-up effect is amajor problem for room temperature operation. To reduce the effect, the p-channel and n channel devices are physically separated by a considerable distance, or implants or trenches are used between devices to reduce the transistor action. At 77 K, however, latch up is less of a problem since the emitters of the n-p-n and p-n-p structures are degenerately doped, the current gains $alfa_n$ and $alfa_p$, decrease markedly with temperature and at 77 K are small enough that the trigger current and sustaining current are both increased appreciably over their room temperature values. While latchup has been observed at 77 K, it is not expected to be a serious problem.

PUNCH-THROUGH

With decreasing channel length, the drain voltage has an increasing effect on reducing the source channel barrier. Below a certain length (depending on the technology used) the subthreshold current is increased markedly (punch through). The decreased carrier kinetic energy as well as the reduced drain voltage permitted at low temperatures reduces

In digital switching circuits the dissipated power P is

$$P = CV^2f, \qquad (5)$$

where C is the load capacitance, V is the capacitor voltage when fully charged and f is switching frequency. A reduction of supply voltage by a factor offour results in a reduction in power dissipation by a factor of 16.

As indicated earlier, the switching speed is dependent on the time it takes to charge or discharge the load capacitance. Because the supply voltageis reduced a factor of four but the current only a factor of three, neglectingthe temperature dependence of capacitance the charging and discharging times are reduced by the factor 4/3. The power-delay product then is reduced by a factor of 16 x 4/3 = 21. We note that the room temperature scaling discussed above merely insures the same shape of the output characteristics at LNT as at RT and is not an optimization in any sense. As for room temperature operation, of course, there exists a tradeoff between switchingspeed and power dissipation.

We note here that converting a room temperature design for 5 volt operation would require a 1.25 volt power supply at 77 K. The design would also require a four fold reduction in threshold voltage. This reduced threshold voltage would require a reduced doping density in the substrate under the gate and thus the depletion region capacitance would be reduced. We also note that the reduced current at LNT would permit even smaller linewidths

PACKING DENSITY

The packing density of a chip is limited by the sizes of the devices, the conducting lines, and the pad this punch througheffect. We believe that the use of devices with channel length less thanabout 0.15 m will require low temperature operation to minimize this punch-through effect.

RELIABILITY

The reliability of CMOS devices is a strong function of operating temperature and supply voltages. In general any process which results in the reduction of device electric fields and junction temperatures will improve overall device and system reliability. The mean time to fail (MTTF) for thermally activated mechanisms is proportional to a temperature dependent term expressed by the Arrhenius relation

$$MTTF \propto \exp\left[\Delta H\left(\frac{1}{T_o} - \frac{1}{T_R}\right)\right. \qquad (6)$$

where T_O and T_R are the operating temperature and reference temperature

respectively, and deltaH is a parameter related to the activation energy of agiven thermal process. Typically deltaH is in the range 0.3 to 1.2 eV. When operating at 77 K and assuming a weak temperature dependency (deltaH = 0.3 eV) equation (6) suggests an improvement in MTTF of 10^{15} over its value at 300 K. It should be emphasized however that the values of deltaH were determined from experiments centered around 300 K to 450 K and may not be valid or extendableto the 77 K region. To date there has been little published data on reliability of CMOS devices at 77 K with the exception of degradation due tohot carriers.

GATE OXIDE INTEGRITY

Dielectric breakdown is a strong function of applied electric fields and a much weaker function of operating temperature. Experimental values of deltaH are controversial but 0.3 eV is often used as a conservative estimate. Thereduction of device voltages possible with low-temperature operation should greatly improve dielectric reliability although there is little published experimental data at low-temperatures.

IONIC INSTABILITY

The motion of ionic species (primarily Na^+) in the active device regions alters electrical parameters and can affect circuit functionality and performance. The degradation is strongly temperature dependent with deltaH typically 0.7 eV. Although less of a concern with standard NMOS processes,the re-emergence of CMOS and the incorporation of silicided diffusions has led to an increased sensitivity of devices to ionic degradation (18). At 77 K the motion of the heavier ions (e.g. sodium) is negligible and not of concern.

HOT CARRIERS

Electrons and holes in the inversion layer gain kinetic energy from the electric field and lose this energy to lattice collisions. If the fields are high enough the carriers become "hot" and their energy exceeds the thermal energy of the lattice. These "heated" carriers are able to surmount the oxide-silicon barrier and are injected into the gate where they contribute toa gate current. A small percent become "trapped" in the oxide and alter device parameters. Electrons are more of a concern than are holes due to the lower barrier energy for injection. As temperature is lowered the mean free path of the electrons increase due to the reduction in thermally generated lattice vibrations. This results in increased susceptibility to hot carrier degradation. Recent studies have suggested that the observed increase indegradation rate at 77 K is not due to enhanced trapping in the oxide but rather the increased influence of the trapped charge on device operation at 77 K [20]. Scaled CMOS device designs have migrated towards Lightly Doped Drain (LDD) structures to alleviate hot carrier problems. The structures reduce the peak electric field by grading the source and drain regions by multiple ion implant steps. Operation of MOSFET's with lower supply voltage,as is possible at low temperatures as indicates above, is expected to eliminate the need for these LDD structures.

ELECTROMIGRATION

Electromigration is the formation of metal opens and shorts in wiring levels resulting from the movement of metal atoms under conditions of high current density. The commonly used model assumes a thermal process with anactivation energy of 0.7 eV. Not only does the electromigration rate decrease with given temperature for a given current density, but the current itself is decreased by operating at lower voltages. Electromigration at 77 K is not expected to be a problem.

THERMAL CYCLING

It has been postulated that the large temperature excursions (300 K to 77 K) associated with cooling for operation at LNT may result in fatigue and failure. The severity of this potential problem is expected to depend on system use conditions and on packaging. Little or no failure do to cycling of devices packaged for room temperature operation has been reported.

REFRIGERATION CONSIDERATIONS

Efficient cooling of circuit modules in a low temperature system argues for use of a liquid cryogen. The cooling may be effected by either direct immersion of the circuit modules in the cryogen or by the flow of the cold liquid through channels in the modules themselves. The latter design issimilar to that now used in water cooled computer systems.

There would appear to be only three

suitable candidates for the liquid cryogen when hazardous and corrosive liquids are eliminated. These are neon which boils at 27 K, nitrogen which boils at 77 K, and argon which boils at 87 K at standard atmospheric pressure. Unless there are substantial technical advantages to use neon or argon, cost considerations will argue strongly in favor of liquid nitrogen.

Candidate cryogenic refrigerator types have been described by Longsworth, et al [21]. For the thermal loads of 400 W or more at 77 K reverse Brayton cycle refrigerator may be preferred. However, for smaller thermal loads either Gifford McMahon or Joule- Thompson refrigerators may beused. Recently a 250 W Stirling cycle refrigerator has been announced. It was designed specifically for cryogenic computers using CMOS technology.

SUMMARY AND CONCLUSIONS

The operation of CMOS systems at 77 K (immersed in liquid nitrogen) has many advantages compared to room temperature operation. These include increased operating speed, reduced power dissipation, increased packing density and increased reliability. It is not a question of if CMOS systems will be operating at 77 K but when. We predict that there will several such systems in operation by 1995.

ACKNOWLEDGMENTS

The authors acknowledge the helpful discussions with W. F. Clark, S. L.Titcomb, G. Pellegrini, C. Croteau, B. El Kareh, and F. Gaensslen. This work was supported in part by NSF, IBM, SRC, and CNPq.

REFERENCES

1. K. Jonscher, "Semiconductors at Cryogenic Temperatures", Proc. IEEE,vol. 52, pp. 109-1104, 1964.

2. R.W. Keyes, Erik P. Harris and K.L. Konnerth, "The Role of LowTemperatures in the Operation of Logic Circuitry", Proc. IEEE, vol. 58,pp. 1914-1932, 1970

3. R.W. Keyes, "Low Temperature High Mobility Transistor Materials",Comments on Solid State Phys., vol. 8, no. 2, pp. 37 46, 1977.

4. F.H. Gaensslen, V.L. Rideout, E.J. Walker and J.J. Walker, "Very SmallMOSFET's for Low Temperature Operation", Trans. Electron Devices, vol.ED 24, pp.218 229, 1977.

5. S.K. Tewksbury, "N Channel Enhancement Mode MOSFET Characteristics from10 to 300 K", Trans, Electron Devices, vol. ED 28, pp. 1519 1529, 1981.

6. S. Hanamura, M. Aoki, T. Masuhara, O. Minato, Y. Sakai and T. Hayashida, "Operation of Bulk CMOS Devices at Very LowTemperatures", Symp. VLSI Tech., 1983, pp. 46 47, 1983.

7. R.L. Anderson and J.L. Hill, "Low Temperature Electronics",Microelectronics Journal, vol. 19, pp. 7- 12, 1988.

10. R.C. Jaeger and F.H. Gaensslen, "Simulation of Impurity FreezeoutThrough Numerical Solution of Poisson's Equation with Application toMOS Device Behavior", IEEE Trans, Electron Devices, vol. ED 27 pp. 914-920, 1980.

11. R.G. Pires, R.M. Dickstein, S.L. Titcomb and R.L. Anderson, "Carrier Freezeout in Silicon", Cryogenics, vol. 30, pp. 1064 1068, 1990.

8. C.G.Rogers, "MOST's at Cryogenic Temperatures", Solid State Electron.,vol. 11, no. 11, pp. 1079 1091, 1968

9. S.H.Wu and R.L.Anderson, "MOSFET's in the 0K Approximation", Solid State Electronics, vol. 17, pp. 1125-1137, 1974.

12. R.K.Kirschman, "Cold Electronics:an Overview", Cryogenics., vol. 25,no. 3, pp. 115-122, 1985.

13. Lia Krusin-Elbaum, Jack Yuan-Chen Sum and Chung-Yu Ting, "On the Resistivity of $TiSi_2$: The Implications for Low Temperature. Applications", IEEE Trans. Electron Devices, vol. ED-34, pp. 58-62, 1987.

14. S.Y.Chou, D.A.Amtoniadis and H.I.Smith, "Observation of ElectronVelocity Overshoot in Sub 100 nm channel MOSFET's in Silicon" IEEEElectron Device Letters, vol. EDL 6, pp. 665 667, 1985.

15. J.S.T. Huang and J.W. Schrankler, "Switching Characteristics of ScaledCMOS Circuits at 77 K", IEEE Trans. Electron Devices, vol. ED 34, pp.101 106, 1987.

16. J. L. Hill, R. G. Pires and R. L. Anderson, "Scaling Silicon MOSFET's For 77 K Operation", IEEE Trans. Electron Devices, in press

17. R. G. Pires and R. L. Anderson, "Scaling Deep Submicrometer MOSFET's for Low Voltage, Low Temperature Operation" Presented at SBMICRO 91, Belo Horizante, July, 1991

18. J.G. Dooley and R.C. Jaeger, "Temperature Dependence of Latch up inCMOS Circuits", IEEE Electron Device Lett., vol. EDL 5, pp. 41 43,1984.

19. E. Sangiorgi, R.L. Johnson, M.R. Pinto, P.F. Bechtold and W. Fichtnes,"Temperature Dependence of Latch up Phenomena in Scaled CMOSStructures", IEEE Electron Devices Lett., vol. EDL 7, pp. 28 31, 1986.

20. P. Heremans et al, "The Dependence of Channel Hot Carrier

Degradation on Temperature in the range 77 K to 300 K.", presented at 19th ESSDERC, Berlin, pp. 727-731, 1989.

21. R.C. Longsworth, W.A. Steyert, "Technology for Liquid-Nitrogen-CooledComputers", IEEE Trans. Electron Devices Vol. ED 34, pp. 4 7, Jan.1987