Low Temperature Effects on CMOS Circuits

Jin Hu
Electrical Engineering and Computer Science
Northwestern University
Evanston, IL 60208, USA

***Abstract*–As technology scaling becomes increasing more difficult, other venues such as decreasing operation temperature have offered further improvement in circuits. This paper presents the advantages of low (liquid nitrogen or LN) temperature have on CMOS circuits as well as a survey of current applications. Motivation for low temperature electronics (LTE) has risen further from the prominent drawbacks of technology scaling. By running circuits at 77°K (LN range), issues such as interconnect delay and latchups show significant improvement. However, LTEs have disadvantages as well, including freeze-out and costs involved in maintaining the temperature. Despite these drawbacks, LTEs display great potential for current and future commercial products.**

***Index Terms*–LTE, low temperature, VLSI, CMOS.**

I. INTRODUCTION

WITH current technology scaling trends, the typical MOSFET has undergone a significant reduction in sizing. While the individual transistor has been improved over time, the problem to further advance the technology, i.e. decrease the size, is becoming an increasingly difficult one to solve. Moreover, problems that were initially deemed insignificant have become more prominent, if not dominate. Lastly, while sizing down transistors have improved transistor characteristics, it offers no improvement on key issues such as interconnect delay. Thus, to rely solely on scaling is infeasible; a potential alternative is to operate circuits at a lower temperature. In particular, significant attention has been given to cryogenic or liquid nitrogen (LN) temperature electronics [1] [9].

Further motivation to study low temperature circuits (LTEs) stem from the following three reasons [2]. First and foremost, it is clear that temperature has a strong impact on several important properties such as conductance, drift velocity, and in regards to integrated circuits (ICs), noise margins. Studying the effect of temperature on such properties can give insight as to the behavior of LN and room temperature (RT) circuits. This could potentially lead to solving intrinsic problems such as reliability that would have otherwise gone unknown.

The second reason to further delve into LTEs is to explore that which is not possible for RT circuits. That is, certain electronics only run at low temperatures. Phenomena such as superconductance and the Josephson Effect only occur at significantly low - 77°K (LN) – temperatures. Furthermore, temperature-related transistor characteristics such as threshold voltage do not scale properly. Thus, understanding the effects of low temperatures is upgraded from an option to a requirement.

Lastly, existing CMOS technology can be further improved if transformed into LTEs. That is, performance is greatly increased if the IC is done in low temperature conditions as opposed to RT [6]. Low temperature conditions also have eliminated hazardous problems such as latchup and significantly improved the performance of certain applications such as dynamic RAM (DRAM) [3]. The drawback, however, is the cost to maintain the temperature at a low enough level.

The remainder of this paper focuses on semiconductors with an emphasis on CMOS ICs, in particular the CMOS inverter shown in Fig. 1. The format goes as follows: Section II discusses the effect of low temperatures on electrical resistance and current mobility. Section III discusses the effects on different types of semiconductors. Section IV discusses reliability issues. Section V discusses two currently implemented applications of LTEs, namely the improved DRAM [3] and high-performance computer systems [7] [8]. Finally, Section VI offers future work and conclusions.

II. ELECTRIC RESISTANCE AND MOBILITY

Low temperatures have numerous positive effects on CMOS ICs. This section covers the most notable improvements in terms of circuit conductance and electron mobility as well as some issues that were initially present in RT but no longer in LN. Transconductance and resistance correlate strongly with switching speed [4], one issue not affected by technology scaling. Latchup is also discussed in this section.

A. MOSFET Transconductance

Transductance (g_m) is defined as the variation of drain current that corresponds to a variation of gate voltage above threshold [2]. For low channel fields, i.e. long channel, small drain voltage, transconductance is directly proportional to the carrier mobility. The equation for the drain current in the linear region is given by

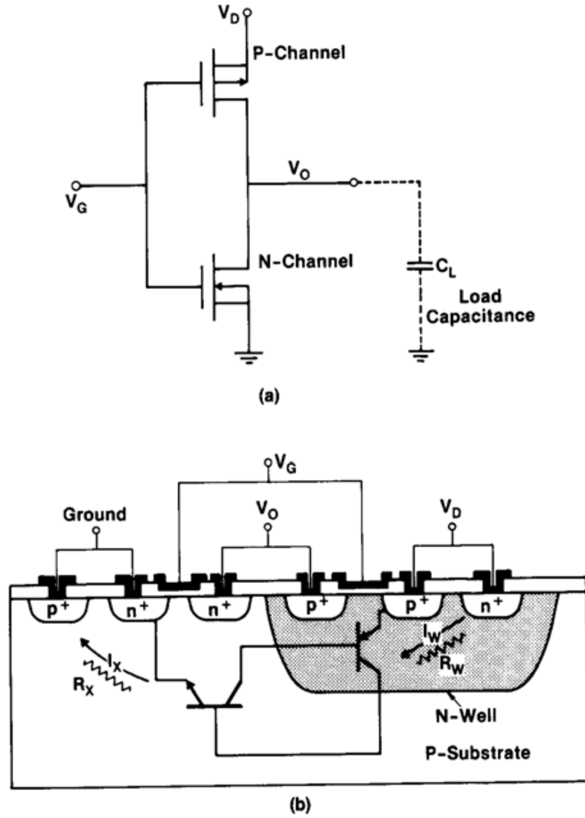

Fig. 1. (a) Schematic of a CMOS inverter. (b) Cross-section indicating typical parasitic bipolar structures. I_W = well current, R_W = well resistance, I_X = substrate current, R_X = substrate resistance [2].

$$I_D = \mu C_{OX} \frac{W}{L}\left(V_G - V_T - \frac{V_D}{2}\right)V_D \qquad (1)$$

and in the saturation region by

$$I_D = \mu C_{OX} \frac{W}{2L}\left(V_G - V_T\right)^2 \qquad (2)$$

where

- I_D drain current,
- μ effective low field surface mobility,
- C_{OX} gate-oxide capacitance,
- W/L effective device width to length ratio,
- V_D drain voltage,
- V_T threshold voltage referenced to GND.

Mathematically, transconductance is given by

$$g_m = \left.\frac{\partial I_D}{\partial V_G}\right|_{V_D} = \mu C_{OX} \frac{W}{L}(V_G - T_T). \qquad (3)$$

Because transconductance is dependent upon MOSFET geometry, it is more useful to define transconductance per unit device width (g_m/W).

As shown in Fig. 2, the mobility of carriers, both holes and electrons, is plotted versus temperature. Using CMOS 0.5-μm technology, the dependencies are shown when the MOSFETs are in the linear region. As temperature decreases, the mobility increases significantly, a factor of four to six times due to a reduction in atomic lattice vibrations. For saturation region, the relationship between mobility and temperature is similar and is directly proportional to the carrier drift velocity caused by the longitudinal field E_y .

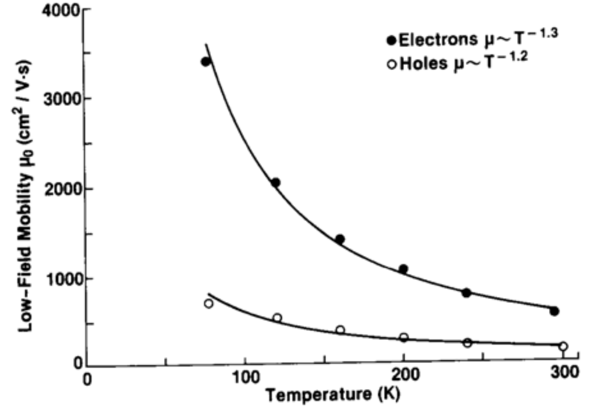

Fig. 2. Carrier mobility plotted versus temperature.

When E_y increases to or above a critical value, the channel mobility begins to decrease due to a phenomenon known as optical phonon scattering. Assuming silicon-based transistors, at RT conditions, the drift velocity v_s reaches its saturation point. The modified drain current equation is now given by

$$I_D = WC_{OX} v_s \left(V_G - V_T\right) \qquad (4)$$

Note that the drain current no longer depends on the channel length. This leads to the transconductance being dependent on only the saturation velocity and the gate-oxide capacitance, given by

$$\frac{g_m}{W} = C_{OX} v_s \qquad (5)$$

The factor of increase in transconductance under these conditions is the ratio between the saturation velocity at LN and at RT. While it is not fully accepted that the velocity saturation is dependent upon temperature [10], experimental results have shown that at LN conditions, the increase in carrier mobility is approximately 1.4 times greater than that at RT. The results are given in Fig. 3.

Note that this impact is most significant for long channel transistors. The transconductance increase is lower for submicron channel lengths due to the (assumed) smaller magnitude of E_y . At lower temperatures, saturation velocity occurs at a lower magnitude than RT because of the lowered field mobility. Fig. 4. shows the normalized results between RT and LN conditions.

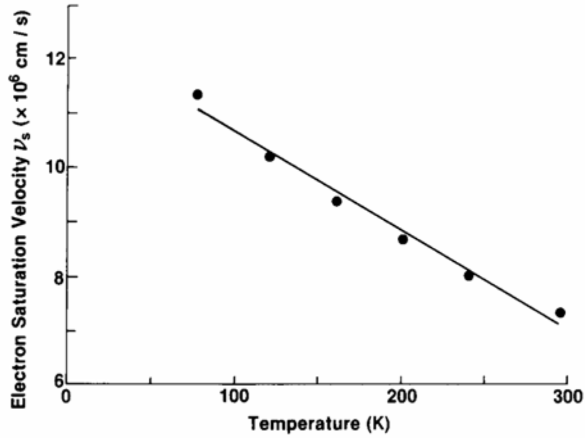

Fig. 3. Increase in electron saturation velocity as a function of temperature. v_S increases by around 1.4 times at LN above RT [2].

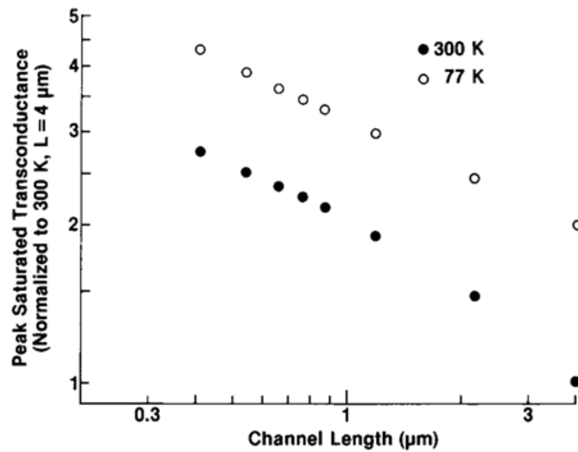

Fig. 4. Normalized peak NMOS saturated transconductance as a function of channel length, normalized to 300 K and $L = 4\ \mu m$ [2].

B. Interconnect Resistance

Interconnects are affected by two primary traits: sizing and temperature. For the purposes of this paper, interconnect sizes will remain constant such the only variable factor is temperature. With respect to resistance, low temperatures affect AC and DC resistances differently [1]. This is covered in greater detail below.

As the number of free carriers in a conductor does not increase with the decrease of temperature, the mobility does, thereby leading to an increase in conductivity. At RT conditions, a conductor's DC resistance is directly proportional to the dimension of the contact. At LN conditions, if the contact's material is greater than a critical size, then the performance improvement upgrades from proportional to a faster rate. This implies that the optimal LTE transformation is not to merely cool an existing circuit – depending on the contact sizes, some ICs perform better when started at LT conditions rather than brought to that stage.

One new phenomenon is the concept of superconductors: if a material is brought down to a low enough critical temperature T_C, the material's DC resistance is automatically brought down to (or close to) 0. This reduction in resistance is independent of the contact dimensions. The caveat, however, is that in order to maintain this condition, the temperature generally is less than 10°K. Other hybrid superconductors (non-elemental) that exhibit T_C as high as 70°K exist but are not used as contacts.

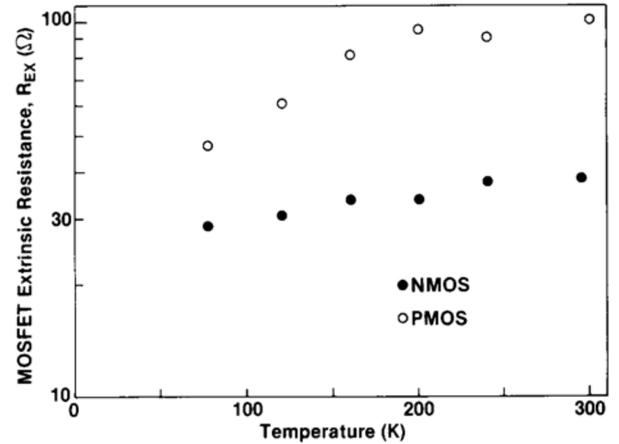

Fig. 5. Decrease in extrinsic resistance at low temperatures. Resistance decreases due to an increase in effective mobility [2].

In regards to AC, one main difference from DC is that the contact dimensions and size do not affect the resistance. If brought down to LN conditions, the improvement trend for current CMOS ICs is similar to that of DC. The amount of improvement levels off, however, right around 77°K; no significant AC resistance is found if cooled further. Keyes et al. have also shown that as ICs are made faster, the reduction amount of AC resistance will decrease and the optimal temperature for which this occurs will increase [1].

These trends, found experimentally or theoretically, are based on ideal contacts – there are no impurities, no dislocations, no grain boundaries, etc. Practically, any conductor for DC or AC) or superconductor (for AC) will have higher resistances. Furthermore, the usage of superconductors relies heavily on two factors. First, the material must remain at extremely low temperatures to maintain 0 DC resistance. Secondly, consideration must be given to the "critical" current density, i.e. the amount of current it can stand without having its self-field switch back to its normal conductive state. This depends then on the thickness and the material of the contact. For Type I superconductors (elemental but having extremely low T_C s), their critical current density is on the order of 10^7 Å $\cdot$cm^2 and can typically handle up 0.15 amperes [1]. For Type II superconductors (hybrid but having higher T_C s), the current density (and current tolerance) is substantially higher.

C. Latchup

The problem of latchup occurs when electrons take a path that was not intended. As transistor manufacturing is not an ideal process, there potentially could be impurities that could cause improper functionality. Latchups can be caused by a number of mechanisms, including but not limited to, current spikes, voltage spikes, and capacitive coupling during signal switching. Though not the most common obstacle regarding

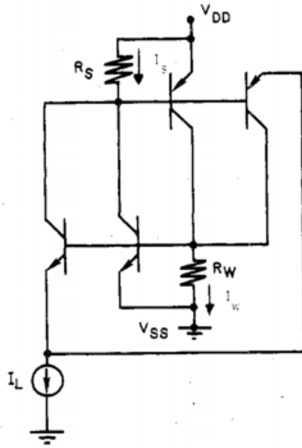

Fig. 6. Circuit model for latchup study [5].

transistors, it most certainly is not an issue to ignore.

The silver lining to this problem is that in order for latchups to occur, they must be sustained. This implies that only temporary damage is done. If the supply current (provided by V_{DD}) is below that of a certain holding current I_H, then latchup will no longer occur [11]. The holding current equation is given by

$$I_H = \frac{\alpha_n I_W + \alpha_p I_S}{\alpha_n \alpha_p - 1} \qquad (6)$$

where α_n, α_p are the current gains and based on Fig. 6. Typical results for a common CMOS chip are shown in Fig. 7. As temperature decreases the holding current consistently increases. From Fig. 8, the source and well resistances decrease substantially with the decrease of temperature. From (6), it is clear that with the decrease of temperature, the holding current will increase. But, if the sum of the current gains α_n, α_p is below unity, latchup will not occur [5].

The dip in temperature prevents many parasitic effects from occurring but does not completely eliminate latchup problems. As noted from Fig. 6, the holding current increases rapidly with the linear decrease of temperature. This trend implies latchup can potentially be eliminated if the temperature were lowered below LN.

III. LOW TEMPERATURE EFFECTS ON SEMICONDUCTORS

When cooled down to LN conditions, semiconductors exhibit an increase in mobility. Moderate increases (20%) in the speed have been achieved with germanium circuits [1]. However, lowering the temperature only yields a limited amount of improvement. The doping amount of the semiconductors plays an active role. In heavily doped semiconductors, the temperature's effect is lessened, thus severely restricting the amount of increase in mobility. In less heavily doped semiconductors, the charge carriers become trapped at lower temperatures, thus slowing down the transistor. However, if the proper higher lower temperatures,

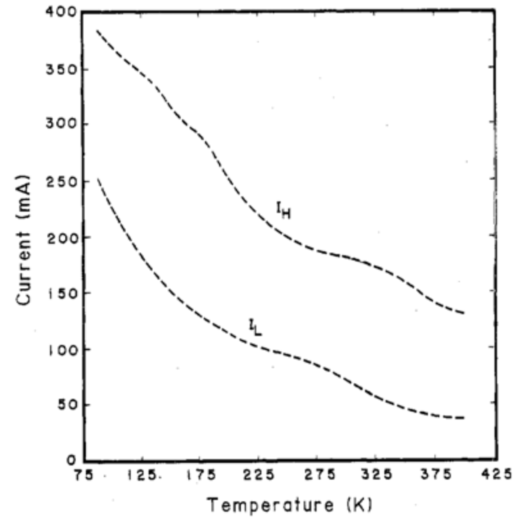

Fig. 7. Holding and trigger current variation as a function of temperature [5].

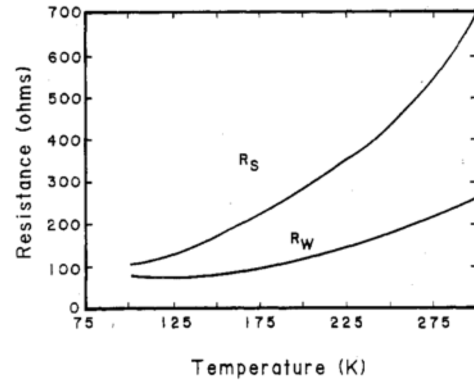

Fig. 8. Calculated temperature dependence of substrate and well shunt resistances for a bulk CMOS process [5].

carrier trapping is not significant, thereby making less heavily doped semiconductors attractive.

Another method to increase transistor speed is to increase the mobility of the carriers. Mobility is generally associated with semiconductors with low energy gaps. Thus, it seems reasonable to use this type of material to build transistors. However, this view is too simplistic (and optimistic). In general, these high-mobility low-energy-gap semiconductors have high atomic mass. Though the lower temperature does cause a decrease in thermal conductivity, it increases the density of phonon states, thereby increasing the phonon-phonon scattering. Experimental results show that the positive effects of thermal conductivity cancel out the negative effects of phonon scattering [1]. Low-gap semiconductors also have higher dielectric constants. This then leads to higher capacitance values and slower switching times.

IV. RELIABILITY

The operation reliability of CMOS circuits is dependent upon the voltage and current values. In general, any type of reduction in temperature or electric field will improve the circuit's stability. Its mean-time-to-fail (MTTF) is modeled by the Arrhenius relation

$$MTTF \; \alpha \; e^{[(\Delta H / k)(1/T_0 - 1/T_R)]} \qquad (7)$$

where

T_0	*Operating Temperature*,
T_R	Re *ference Temperature*,
ΔH	*Parameter* ~ 0.3 - 1.2eV

At LN conditions, $\Delta H = 0.3eV$ and there has been a suggested improvement of about 10^{15} in MTTF over the RT value [2]. With $\Delta H = 0.7eV$, problems such as electromigration and metal creeping are not expected to occur at LN conditions.

However, lowering the temperature has not been proven to improve the overall reliability of CMOS circuits; it has only been shown to resolve thermally activated issues. Transferring the IC to LN conditions may have solved a subset of problems but it also introduces a new set of problems as well.

A common issue for LTEs is circuit breakdown due to thermal stresses. When two differing materials are connected, they are usually caused by the different reactions due to the coefficients of thermal expansion (CTEs). As temperatures decrease, thermal stress usually increases.

Another problem of LTEs is the reparability factor. If a LTE breaks down, the repair process would need to involve heating the circuit back up to RT, fixing the problem, and then cooling the circuit back down to LN. Not only does this process involve extra time, this is far more complex than repairing a circuit in RT. To go about this issue, two methods have been proposed [1]. The first method is prevention – take the proper time and resources needed to ensure that the circuit will function correctly. The second is redundancy – if one component is weak, put another in parallel such that if one breaks down the other is still available. Both solutions, however, are expensive. The first method is not only costly at the ensuring stage but also if the circuit (unluckily) breaks down, more capital will be spent doing repairs. The second method may not be feasible, as reproducing and storing an entire component in LN conditions may be more costly.

V. Applications

The following discusses the results of two areas for which LTEs are currently employed. The first describes a 50% improvement over of a standard dynamic RAM (DRAM) cell. The second application showcases cooling methods to increase the performance of high-end computers. Provided that the trade-off of maintaining the temperature was worth it, the improvement over the previous RT model is significant.

A. Improved DRAM

Henkels et al developed a high-speed CMOS DRAM cell [3]. Fig. 9 shows the results of the improved DRAM's performance at lower temperatures. In experiments, when used in bulk, their 512K DRAM, done with 1-μm technology, still outperforms its 1-Mbit DRAM by at least a factor of two.

B. ETA10 Supercomputer

Developed using CMOS technology by ETA Systems Inc., the ETA10 is one of the first commercialized computers to be used under low temperatures. The results show that the clock speed doubled from that at RT conditions [7]. The design includes a built-in self-check system both at the board and chip level to ensure proper manufacturing and testing. Proper design, thermal control, and proper material selection also limited interconnect breakdowns.

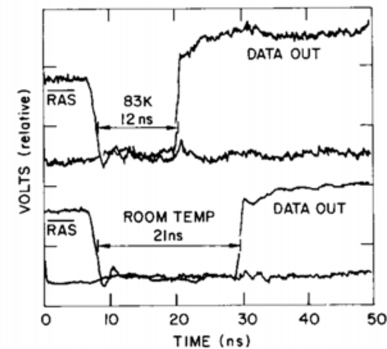

Fig. 9. Comparison of room-temperature and low-temperature access times with V_{DD} = 5 V. As the temperature is reduced from room temperature to 83°K, the access time decreases from 21 to 12 ns, an improvement of over 1.7 times [3].

C. High-End Server Cooling

Developed by IBM, this commercial product also utilizes low temperature to increase system performance. The downside of monetary cost and power consumption are heavily outweighed by the reduction of leakage current and increase in functionality and reliability [8]. The refrigeration unit designed to handle large applications include compensation for pipe cracks, condensation, and temperature.

VI. Conclusion

With technology scaling becoming exponentially difficult, more attention has been spent on alternatives such as lowering the operating temperature to increase circuit performance. At LN conditions, ICs show significant improvement over RT. Latchup, electromigration, and metal creeps have shown a significant amount of decrease. On the other hand, lower temperatures also introduce several new problems. Reliability, reparability, and choice of transistor material all need to be reevaluated. The effects of low temperature do not always scale well (linearly) with decreased technology scaling. However, several commercial applications have already been introduced; they show great promise and potential for future applications and research.

References

[1] R. W. Keyes, E. P. Harris, and K. L. Konnerth, "The role of low temperatures in the operation of logic circuitry," *Proc. IEEE*, vol. 58, pp. 1914 – 1932, 1970.

[2] W.F. Clark, B. El-Kareh, R. G. Pires, S. L. Titcomb, and R. L. Anderson, "Low temperature CMOS – a brief review," *IEEE Transactions on Components, Hybrids, and Manufacturing Technology*, vol. 15, pp. 397 – 403, June 1992.

[3] W. H. Henkels, N. C. C. Lu, W. Hwang, T. V. Rajeevakumar, R. L. Franch, K. A. Jenkins, T. J. Bucelot, D. F. Heidel, and M. J. Immediato, "A 12-ns low-temperature DRAM," *IEEE Transactions on Electronic Devices*, vol. 36, no. 8, pp. 1414 – 1422, August 1989.

[4] R. C. Jaeger and F. H. Gaensslen, "Low temperature semiconductor electronics," in *Proc. Thermal Phenomena in the Fabrication and Operation of Electronic Components: I-THERM '88 InterSociety Conference*, pp. 106 – 114, May 1988.

[5] J. G. Dooley and R. C. Jaeger, "Temperature dependence of latchup in CMOS circuits," *IEEE Electronic Device Letters*, vol. 5, no. 2, pp. 41 – 43, February 1984.

[6] J. W. Peeples, W. Little, R. Schmidt, and M. Nisenoff, "Low temperature electronics workshop," *Sixteenth IEEE SEMI-THERM Symposium*, pp. 107 – 108, 2000.

[7] D. M. Carlson, D. C. Sullivan, R. E. Bach, and D. R. Resnick, "The ETA 10 liquid-nitrogen-cooled supercomputer system," *IEEE Transactions on Electronic Devices*, vol. 36, no. 8, pp. 1404 – 1413, August 1989.

[8] R. R. Schmidt and B. J. Notohardjono, "High-end server low-temperature cooling," *IBM J. Res. & Dev.*, vol. 46, no. 6, pp. 739 – 751, November 2002.

[9] R. K. Kirschman, "Low temperature electronics," *IEEE Circuits and Devices Magazine*, vol. 6, issue 2, pp. 12 – 24, March 1990.

[10] J. S. T. Huang and J. W. Schrankler, "Switching characteristics of scaled CMOS circuits at 77 K," *IEEE Transactions on Electronic Devices*, vol. 34, pp. 452 – 462, 1987.

[11] D.B. Estreich, "The physics and modeling of latch-up and CMOS integrated circuits," Stanford Electronics Library, Stanford, CA, Tech. Rep. G-201-9, November 1980.