On-Chip Switched-Capacitor DC-DC Converter in Memory Technology : State of the Art and Challenges

This is the accepted version of the paper presented at IEEE ECCTD on Sep. 5, 2017.

On-Chip Switched-Capacitor DC-DC Converter in Memory Technology: State of the Art and Challenges

Toru Tanzawa, *Fellow, IEEE*

***Abstract*— This paper presents the state-of-the-art of integrated switched capacitor DC-DC converters in memory technology from NMOS to SOI DRAM and from 2D to 3D NAND focusing on capacitor structures with high capacitance density and small parasitic capacitance based on memory technology. The paper then proposes design for manufacturing to increase yield with redundant arrays of converter and design for reliability by making a balance between the capacitance density and the electric field across the capacitor. Design challenges are also discussed for emerging 3D cross-point memory.**

***Index Terms*— Switched capacitor, DC-DC converter, Charge pump, Memory technology, Design for manufacturing**

I. Introduction

SWITCHED-CAPACITOR (SC) DC-DC voltage up-converters are becoming one of the most critical integrated circuit blocks for energy harvesting in wireless sensor nodes to generate a voltage high enough for micro-watt sensing and computing LSIs in nano-meter Complementary Metal-Oxide-Semiconductor (CMOS) from environmental energy sources. SC DC-DC voltage down-converters have also been gathering circuit designers' interests to generate a voltage low enough and a current high enough for scaled CMOS transistors to run at a GHz clock with no overstress. Various on-chip SC voltage up- and down-converter circuits have been designed in memory technology since Falkner qualitatively showed that a voltage multiplier topology in a parallel configuration could be less sensitive to parasitic capacitance in 1973 [1] and Dickson demonstrated an integrated charge pump (CP) for programming data into Metal-Nitride-Oxide-Semiconductor (MNOS) memory cells in 1976 [2], where *CP* can mean both charge pump and DC-DC converter below.

Lee et al. integrated substrate bias generator in dynamic random access memory (DRAM) to bias the access transistors at a negative voltage based on switched-capacitor technique in 1979 [3]. More recently, growing numbers of papers discuss DC-DC converters using integrated capacitors fabricated in memory technology. Dong et al. designed CP using Oxide-Nitride-Oxide (ONO) film as the pump capacitor for NOR Flash memory [4], where the ONO film is placed between the floating gate and control gate of the memory transistor. Kuang et al. and Anderson et al. respectively proposed voltage up- and down-converters utilizing deep trench capacitors fabricated in Silicon-On-Insulator (SOI) embedded DRAM technology to increase output current density per silicon area and power efficiency due to lower parasitic capacitance in [5] and [6]. Ferroelectric random access memory (FeRAM) stores data with binary 0s and 1s as two possible polarizations [7]. When the ferroelectric film is always biased in a predetermined polarization, it can act as a non-linear capacitor with no hysteresis. Thus, it can be used as a pump capacitor of a DC-DC converter, as El-Damak et al. presented in [8].

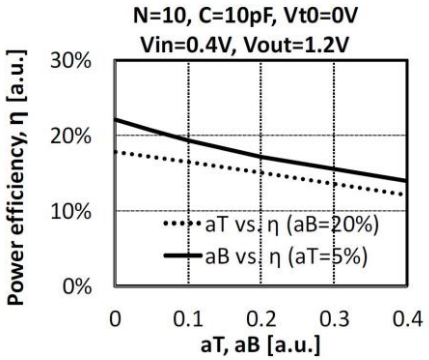

Fig. 1 Power efficiency vs. top (aT) or bottom (aB) plate parasitic capacitance relative to the capacitance of the pump capacitor [23].

Memory technology is transitioning from 2D to 3D to keep the trend on bit cost reduction beyond the end of Moore's law for memory [9]-[17]. Because the decoding transistors are still fabricated in 2D on a silicon substrate even though the array can be formed in 3D, the block size tends to increase from one technology to the next. As a result, word-line (WL) CP needs to have a larger driving current at every technology transition. Tier capacitors are proposed in [18] to reduce the CP area by a factor of 4.8. Cross-point memory technology for resistive RAM, phase change memory and magnetic RAM has been also proposed [19]-[22]. It will need a scalable area- and power-efficient DC-DC converter per technology transition as well as 3D NAND. Fig. 1 shows power efficiency vs. top and bottom plate parasitic capacitance relative to the pump capacitor of a 10-stage CP [23]. Thus, it is key that the pump capacitors have sufficiently low parasitic capacitance for high power efficiency.

This paper overviews on-chip SC DC-DC converters in memory technology focusing on pump capacitor structures. It also proposes designs for manufacturing (DFM) and reliability (DFR) when fine memory structures are used as the pump capacitors. Finally, it discusses challenges in designing CPs for cross-point non-volatile memory.

T. Tanzawa is with Faculty of Engineering, Shizuoka University, Hamamatsu 432-8011, Japan (e-mail: toru.tanzawa@shizuoka.ac.jp).

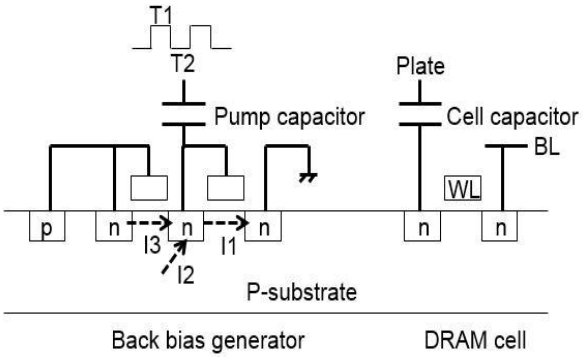

Fig. 2 Back bias generator integrated in DRAM [3]

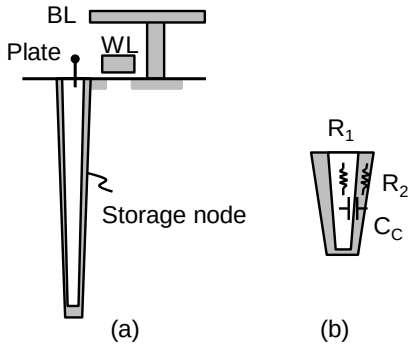

Fig. 3 Trench DRAM cell (a) and parasitic resistance of two terminals (b).

II. State of the Art

A. DRAM

DRAM has integrated a substrate bias generator since 1979 to reduce the cost [3]. Voltages at WLs and BLs were toggled between 0 V and 5 V during operations when the supply voltage was 5 V. Such a huge voltage swing could make PN junctions of NMOS transistors into forward bias regime locally due to capacitive coupling where it is far from body contacts if the p-type substrate is grounded. If this happens, stored charges could flow into the substrate, resulting in degradation in data reliability. To avoid it, another external negative voltage of −5 V was needed in addition to the power supply voltage of +5 V. The negative voltage was supplied to the substrate to have a sufficient operation margin with such a potential for eliminating the forward biasing of the PN junctions. The −5 V power supply was eliminated by implementing the back bias generator allowing reduction of the system cost and complexity having the negative voltage supply.

Fig. 3(a) illustrates a DRAM cell composed of an access transistor and a trench capacitor [28]. The aspect ratio of the trench has been increased to over 20. The plate is biased at $V_{DD}/2$ whereas the storage node is at 0V or V_{DD} when the stored data is "0" or "1", where V_{DD} is the BL voltage for data "1". As a result, the trench capacitor is always subjected to $V_{DD}/2$ regardless of the data. If the trench capacitors for CP are used for a DC-DC voltage down-converter [6] whose output voltage is about $V_{DD}/2$, the stress voltage and time for the trench capacitor of CP are well matched with those for the trench capacitor of the memory cell. In this case, there should be little concern on reliability of CP, assuming the cell array doesn't have any reliability issues during its lifetime. Conversely, if the trench capacitors in CP are used for a DC-DC voltage up-converter [5] whose output voltage is greater than V_{DD} for WL overdrive, multiple trench capacitors must be connected in series between V_{OUT} and ground in order not to overstress the trench capacitor of CP. When the capacitance in each stage needs to be C, each of two series-connected capacitors has to be *2C*. As a result, the total capacitance in each stage needs to be *4C*, which is 4 times larger than the case where only a single capacitor is required for each stage. Fig. 3(b) shows that each terminal of the trench capacitor has parasitic resistance. Unlike planar capacitors which can have contacts to interconnect at both ends, plate and storage nodes of a trench capacitor have contacts only at a single end. As a result, an intrinsic RC time constant of $(R_1+R_2)\,C_C$ for the trench capacitor is 4 times larger than that for the planar capacitor. That could limit the operating clock frequency. Pei et al showed that the cut-off frequency of a decoupling trench capacitor was below 1 GHz [24].

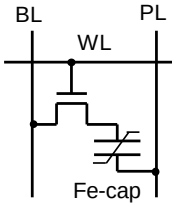

Fig. 4 FeRAM cell

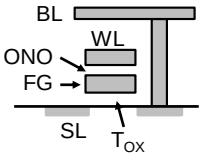

Fig. 5 NOR Flash cell

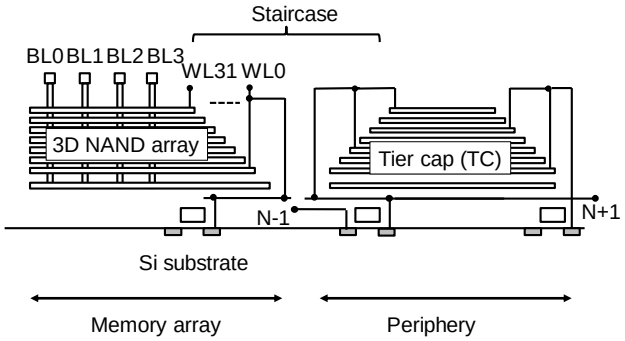

Fig. 6 Tier capacitor in 3D NAND technology [18]

B. FeRAM

Fig. 4 shows a circuit diagram of a one-transistor one-capacitor (1T-1C) FeRAM cell [7] [25], where PL is a plate line. Only while the memory cell is selected for read or program operation, the ferroelectric film is stressed by V_{DD}. Therefore, it may be used as a capacitor of a Dickson voltage doubler to overdrive WL. However, the stress time can be largely different between the cell capacitor and the pump capacitor. When the memory cell is randomly accessed in byte basis, the duty ratio on the stress time T_{DCDC}/T_{CELL} is as large as #WL X #Col, where T_{DCDC} (T_{CELL}) is the total stress time on DC-DC converter (every memory cell) and #WL (#Col) is the number of WLs (columns). For example, when each of #WL and #Col is 1k, the duty ratio is 10^6. Thus, such a significant difference in the stress time needs to be taken into consideration on reliability. When a DC-DC converter operates to generate a voltage lower than V_{DD} [8], a relaxed voltage stress on the ferroelectric film may compensate the duty ratio. When the process steps allow the formation of a

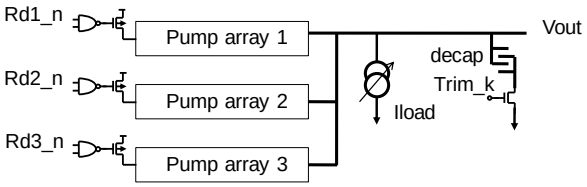

Fig. 7 CP with redundancy where at least two arrays are required.

Table I Test flow

1) Enable each array with I_load on at probe test
2) Measure Vout by array
3) Check if Vout >V_target
4) If array 1 doesn't meet the above, but array 2,3 meet, array 1 is marked as bad. ROM for rd1_n is programmed to be L.
5) When this die is enabled, only array2,3 work whereas array1 is not connected to Vin and Vout.
6) Similar things need to be done for decoupling cap based on the measured data on the leakage current when it's connected to Vout.

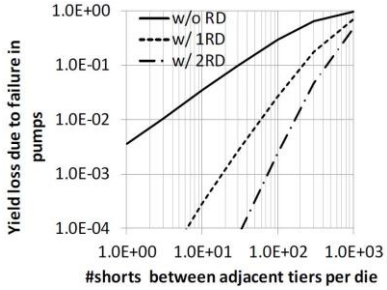

Fig. 8 Yield loss vs. the number of defects per die with 0, 1 or 2 redundant CP arrays

ferroelectric film over the access transistor, one may combine the ferroelectric film with the gate oxide to increase the capacitance density of the pump capacitor for squeezing DC-DC converters.

C. Flash

Fig. 5 shows NOR flash cell where FG is a floating gate, T_{OX} is tunnel oxide, and SL is a source line. When the gate coupling ratio is 50% and the source is grounded, the voltages across the ONO film and tunnel oxide are half of the gate voltage V_{PP} during programming. One may consider that each of the ONO and tunnel oxide films can be used in CP to generate $V_{PP}/2$. However, the stress time of the charge pump capacitors (T_{CP}) can be much longer than that of the Flash cell transistor (T_{CELL}). For example, when the number of WLs is 1k, the factor of T_{CP} to T_{CELL} is as large as 10^3. As a result, the maximum voltages for the ONO and tunnel oxide films need to be reduced, taking the stress time factor into account.

Fig. 6 depicts a tier capacitor together with a 3D NAND array [18]. Because the tier capacitor doesn't have pillars, it is considered as an ideal multiple plate capacitor with huge capacitance density and small parasitic capacitance.

III. Design for Manufacturing and Reliability

A. Design for manufacturing (DFM)

As reviewed in Section II, the memory cell capacitor and the parasitic capacitor in the memory array are becoming candidates for the pump capacitor in both stand-alone and embedded memory to squeeze the CP area and to improve power efficiency. However, one needs to design CP to be highly yielded even though fine-pitch structures can have potentially

Table II Voltage across the dielectric film of capacitor and the duty ratio of the stress time

Cap structure	ONO	Trench	FeRAM	Tiercap
V_{CAP}	Vpp/2	Vdd/2	Vdd	Vpp, Vpp/2
T_{CP}/T_{CELL}	#WL	1	#WL	#block

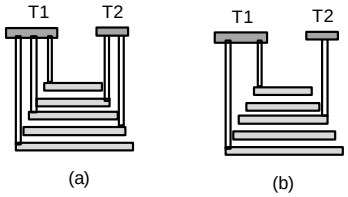

Fig. 9 Tier capacitor with no floating tiers (a) and one floating tier between two terminals T1 and T2 (b)

high defect density. Redundancy is a solution to that issues. A memory array can be yielded when it has spare rows and columns and when the defective memory cells are replaced with some of the spares. Likewise, if CP is designed so that one or a few arrays of a CP unit are added to the minimum number of CP arrays, it can be yielded even with a small number of detects in the pump capacitors. Because CP is designed to be composed of multiple arrays, each array can run at a clock with a different phase to average the peak current noise [26]. Fig. 7 illustrates CP with redundancy. Table I shows a test flow to select healthy CP arrays. It can run only at probe test, in every power-up sequence or in calibration as needed. Fig. 8 shows the effectiveness of redundancy when defects are distributed randomly when the minimum number of CP arrays is 10. When the yield loss due to unrecovered defective CP needs to be as low as 1%, an acceptable number of defects per die is increased from 3 to 70 with one redundant array or 150 with two redundant arrays. When those numbers are greater than the criteria on memory yield, the defects in the CP wouldn't affect overall yield loss.

B. Design for reliability (DFR)

Intrinsic time-dependent dielectric breakdown (TDDB) is modeled as $T_{BD} \sim exp(G/E_{OX})$, where T_{BD} is the time to breakdown, G is a constant, and E_{OX} is the electric field across the dielectric film [29]. When the reliability of the CP capacitor is not sufficient under the condition shown in Table II, the capacitor structure must be reconsidered. One possible solution is a series connection of two capacitors, which can reduce the electric field by half. Fig. 9 illustrates the case of a tier capacitor. By leaving one tier between T1 and T2 with no interconnection as shown in Fig. 9(b), the electric field is halved. As the number of floating layers is increased, reliability would exponentially improve with a reduced electric field, however, capacitance density would decrease at a rate inversely proportional to the number of floating tiers. Thus, the capacitor structure needs to be considered in terms of both reliability and capacitance density.

IV. Challenges

Now that NAND technology has transitioned from 2D to 3D, the other non-volatile memories (NVMs) are also moving into 3D technology. Cross-point memory technology for resistive RAM, phase change memory and magnetic RAM has been

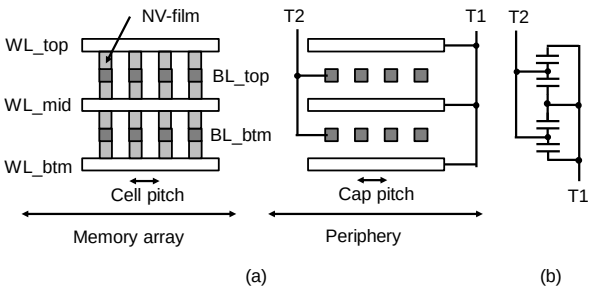

Fig. 10 Cross-sectional view of a cross point memory (a) and equivalent circuit of a capacitor structure (b).

proposed [19]-[22]. Two terminal cell structure is favorable for multiple stacking layers for WLs/BLs to increase bit density. Program throughput can be enhanced with simultaneous multi-bits program as page mode. It requires high current at a given program voltage condition. As a result, area- and power-efficient DC-DC converters will be required for the cross-point NVMs.

The multiple conductive layers as WLs and BLs in cell arrays may contribute to squeezing the CP area by not placing NV material between the conductive layers for the pump capacitor, as shown in Fig. 10 (a). Fig. 10(b) depicts its equivalent circuit, which is similar to a wire capacitor fabricated in standard CMOS as discussed in [27]. It may better have wider pitch in wires of the capacitor than the cell pitch to increase capacitance density, to decrease the parasitic resistance in wires and to decrease short failure probability. A design rule of the wires used for the pump capacitor needs to be determined by design for manufacturing and reliability.

V. Summary

DC-DC converters in memory technology were reviewed with a focus on integrated capacitor structures to increase the output current per silicon area and power efficiency. Taking advantage of the capacitive element of memory cell or the parasitic component, the converter area can be reduced. Designers need to take manufacturability and reliability into consideration on DC-DC converter design in memory technology.

References

[1] A. H. Falkner, "Generalised Cockcroft–Walton voltage multipliers," *Electron. Lett.*, vol. 9, no. 25, pp. 585–586, Dec. 13, 1973.

[2] J. F. Dickson, "On-chip high-voltage generation in MNOS integrated circuits using an improved voltage multiplier technique," *IEEE Journal of Solid-State Circuits*, vol. 11, No. 3, pp. 374–378, Jun. 1976.

[3] J. Lee, J. Breivogel, R. Kunita, C. Webb, "A 80ns 5V-only dynamic RAM," *ISSCC Dig. Tech. Papers*, pp 142–143, Feb. 1979

[4] W. Dong, P. Liyang, D. Zhigang, Z. Jun, "Charge pump system sharing the coupling capacitors for NOR flash memory," *Proceedings of International Conference on ASIC*, pp. 466 – 469, 2003.

[5] J. B. Kuang, A. Mathews, J. Barth et al., "An on-chip dual supply charge pump system for 45nm PD SOI eDRAM, *European Solid-State Circuits Conference*, pp. 66-69, Sep. 2008.

[6] T. M. Andersen, F. Krismer, J. W. Kolar, et al., "A sub-ns response on-chip switched-capacitor DC-DC voltage regulator delivering 3.7W/mm2 at 90% efficiency using deep-trench capacitors in 32nm SOI CMOS", *ISSCC Dig. Tech. Papers*, pp. 90-91, Feb. 2014.

[7] H. P. McAdams, R Acklin, T Blake et al., "A 64-Mb Embedded FRAM Utilizing a 130-nm 5LM Cu/FSG Logic Process", *IEEE Journal of Solid-State Circuits*, vol. 39, pp. 667-677, April 2004.

[8] D. El-Damak, S. Bandyopadhyay, and A.P. Chandrakasan, "A 93% Efficiency Reconfigurable Switched-Capacitor DC-DC Converter Using On-Chip Ferroelectric Capacitors," *ISSCC Dig. Tech. Papers*, pp. 374-375, Feb. 2013.

[9] M. Crowley, A. Al-Shamma et al., 512 Mb PROM with 8 layers of antifuse/diode cells, *ISSCC Dig. Tech. Papers*, pp. 284-5, Feb. 2003.

[10] C. J. Chevallier, C. H. Siau et al., A 0.13µm 64Mb multi-layered conductive metal-oxide memory, *ISSCC Dig. Tech. Papers*, pp. 260-1, Feb. 2010.

[11] H. Tanaka, M. Kido, K. Yahashi, et al, "Bit cost scalable technology with punch and plug process for ultra high density Flash memory," *Symposium on VLSI Technology Dig. Tech. Papers*, pp. 14-15, Jun. 2007.

[12] K. T. Park, J. M. Han, D. H. Kim, et al., "Three-dimensional 128Gb MLC vertical NAND Flash-memory with 24-WL stacked layers and 50MB/s high-speed programming", *ISSCC Dig. Tech. Papers*, pp. 334-5, 2014.

[13] J. Im, W. Jeong, D. Kim et al, A 128Gb 3b/cell V-NAND flash memory with 1Gb/s I/O rate, *ISSCC Dig. Tech. Papers*, pp. 130-1, Feb. 2015.

[14] D. Kang; W. Jeong, C. Kim et al, "256Gb 3b/cell V-NAND flash memory with 48 stacked WL layers, *ISSCC Dig. Tech. Papers*, pp. 13-1, Feb. 2016.

[15] T. Tanaka, M. Helm, T. Vali, et al., "A 768Gb 3b/cell 3D-floating-gate NAND flash memory," *ISSCC Dig. Tech. Papers*, pp. 142-144, Feb. 2016.

[16] R. Yamashita, S. Magia, T. Higuchi et al., "A 512Gb 3b/Cell Flash Memory on 64-Word-Line-Layer BiCS Technology", *ISSCC Dig. Tech. papers*, pp. 196-7, Feb. 2017.

[17] C.Kim, J. Cho, W. Jeong et al, "A 512Gb 3b/cell 64-Stacked WL 3D V-NAND Flash Memory", *ISSCC Dig. Tech. papers*, pp. 202-3, Feb. 2017.

[18] T. Tanzawa, T. Murakoshi, T. Kamijo et al., "Design challenge in 3D NAND technology: A 4.8X area- and 1.3X power-efficient 20V charge pump using tier capacitors," *IEEE Asian Solid-State Circuits Conference*, pp. 165 – 168, Nov. 2016.

[19] I. G. Baek, D. C. Kim, M. J. Lee, et al., "Multi-layer cross-point binary oxide resistive memory (OxRRAM) for post-NAND storage application," *IEEE InternationalElectron Devices Meeting*, pp. 750 – 753, Dec. 2005.

[20] A. Kawahara, R. Azuma, Y. Ikeda et al., "An 8Mb multi-layered cross-point ReRAM macro with 443MB/s write throughput," *IEEE International Solid-State Circuits Conference*, pp. 432 – 434, 2012.

[21] D. Kau, S. Tang, I. V. Karpov et al., "A stackable cross point Phase Change Memory," *IEEE International Electron Devices Meeting (IEDM)*, pp. 27.1.1-27.1.4, Dec. 2009.

[22] W. Zhao, S. Chaudhuri, C. Accoto et al., "Cross-Point Architecture for Spin-Transfer Torque Magnetic Random Access Memory," *IEEE Transactions on Nanotechnology*, Vol. 11. No. 5, pp. 907 – 917, May 2012.

[23] T. Tanzawa, "Innovation of switched capacitor voltage multiplier—Part 2: Fundamentals of charge pump," *IEEE Solid-State Circuits Mag.*, vol. 8, no. 2, Jun. 2016.

[24] C. Pei, R. Booth, H. Ho et al., "A novel, low-cost deep trench decoupling capacitor for high-performance, low-power bulk CMOS applications," *International Conference on Solid-State and Integrated-Circuit Technology*, pp. 1146 - 1149, 2008.

[25] T. Sumi, N. Moriwaki, G. Nakane et al., "A 256 kb nonvolatile ferroelectric memory at 3 V and 100 ns," *ISSCC, Dig. Tech. Papers*, pp. 268 - 269, 1994.

[26] T. Tanzawa, Y. Takano, K. Watanabe, S. Atsumi, "High-voltage transistor scaling circuit techniques for high-density negative-gate channel-erasing NOR flash memories," *IEEE Journal of Solid-State Circuits*, Vol. 37, No. 10, pp. 1318 - 1325, Oct. 2002.

[27] D. Somasekhar, B. Srinivasan, G. Pandya et al., "Multi-phase 1GHz voltage doubler charge-pump in 32nm logic process, *Symposium on VLSI Circuits*, pp. 196-197, 2009.

[28] H. Sunami, T. Kure, N. Hashimoto et al., "A corrugated capacitor cell (CCC) for megabit dynamic MOS memories," *International Electron Devices Meeting Dig. Tech. Papers*, pp. 806 - 808, 1982.

[29] J. C. Lee, I.-C. Chen, and C. Hu, "Modeling and characterization of gate oxide reliability," *IEEE Trans. Electron Devices*, vol. 35, pp. 2268.78, Dec. 1988.