Numerical Confirmation of Inelastic Trap-Assisted Tunneling (ITAT) as SILC Mechanism

Ting-Kuo Kang, Ming-Jer Chen, *Senior Member, IEEE*, Chuan-Hsi (Jason) Liu, Yih J. Chang, and Shou-Kong Fan

***Abstract*—This paper presents a quite comprehensive procedure covering both the stress-induced leakage current (SILC) and oxide breakdown, achieved by balancing systematically the modeling and experimental works. The underlying model as quoted in the literature features three key parameters: the tunneling relaxation time τ, the neutral electron trap density N_t, and the trap energy level E_t. First of all, 7-nm thick oxide MOS devices with wide range oxide areas are thoroughly characterized in terms of the optically induced trap filling, the charge-to-breakdown statistics, the gate voltage developments with the time, and the SILC I–V. The former three are involved together with a percolation oxide breakdown model to build N_t explicitly as function of the stress electron fluence. Then the overall tunneling probability is calculated, with which a best fitting to SILC I–V furnishes τ of 4.0×10^{-13} s and E_t of 3.4 eV. The extracted τ is found to match exactly that extrapolated from existing data. Such striking consistencies thereby provide evidence that inelastic trap-assisted tunneling (ITAT) is indeed the SILC mechanism. Differences and similarities of the involved physical parameters between different studies are compared as well.**

***Index Terms*—Flash, gated-diode, inelastic tunneling, MOSFET, oxide breakdown, percolation, SILC, stress-induced leakage current, trap-assisted tunneling.**

I. Introduction

STRESS-induced leakage current (SILC) is one of the biggest reliability issues in MOS devices, especially the nonvolatile flash memory [1]. In nonvolatile flash memory, high-field or Fowler–Nordheim (F–N) tunneling during the erase cycle can produce a variety of defects within tunnel oxides, among which the most concerned for SILC are the neutral electron traps. The principal reasons are that these generated traps can serve as a stepping stone to effectively shorten the tunneling distance, causing more electrons leaking out of the floating gate. Such knowledge of the impact of SILC stems from an early series of studies [2]–[5]. Recent experimental demonstrations in terms of carrier separation technique [6] and oxide thickness dependent measurement [7] both witness the inelastic behavior of trap-assisted tunneling. That is, the tunneling electrons from the cathode side lose part of their energy via inelastic scattering and fall down to the underlying trap site from which they are instantly de-trapped out to the anode side. Despite few literature efforts made to address the scattering/capturing process microscopically, inelastic trap-assisted tunneling (ITAT) proposed as the origin of SILC is currently largely accepted. By analogy to the most efficient generation/recombination center in the mid-gap of a semiconductor p-n junction, there should exist a certain trap position featuring that a local tunneling probability P_1 from cathode to that site equals another probability P_2 from it to anode. Under such situation, a maximum overall tunneling probability P is created through $P = P_1P_2/(P_1 + P_2)$ [8]. Such a picture of the maximum likelihood had led to the following analytic model for the SILC current–voltage (I–V) [8]:

$$J_{\mathrm{SILC}} = \frac{qt_{ox}N_t}{2\tau}\,\mathrm{Exp}\left\{\left(-\frac{4}{3}\,\frac{(2m_{ox})^{0.5}}{\hbar}\,\frac{1}{qE_{ox}}\right) \times \left(E_t^{3/2} - [E_t - qE_{ox}(t_{ox} - x_t)]^{3/2}\right)\right\} \tag{1}$$

where

- t_{ox} oxide thickness;
- E_{ox} oxide field;
- m_{ox} ($0.42m_o$ [9]) effective electron mass in the oxide;
- τ tunneling relaxation time, x_t ($\approx$3.6 nm calculated [8] for 7-nm thick oxide, for example) lies at the most favorable position, and the trap properties at this specific location are usually described by the neutral electron trap density N_t and the trap energy level E_t.

This unique model had exhibited the comparable ability of tackling SILC I–V [8] as the complicated version [6], [7]; strictly speaking, however, reported agreement with singly SILC I–V unnecessarily means that the current understandings of underlying physical aspects (i.e., the maximum likelihood of tunneling to and from the traps located at the most favorable position, etc.) would have gotten clarification in the way. Part of the reasons is that the tunneling relaxation time τ was not fully explored in [8]. Here τ is better viewed a fundamental parameter and can be defined the tunneling time extrapolated at a zero oxide thickness [9].

This paper presents a quite comprehensive procedure of balancing systematically both the modeling and experimental works from SILC to oxide breakdown. Equation (1) is favored here due to its uniqueness as mentioned above. In contrast to the pioneering work [8] where N_t was treated as a fitting parameter, however, this work manages it explicitly in advance, achieved by extending to experimental Weibull distributions of charge-to-breakdown as well as with the aid of some percolation oxide breakdown models [10]–[15]. As a result,

Manuscript received March 19, 2001; revised May 24, 2001. This work was supported by the National Science Council under Contract 89-2215-009-049. The review of this paper was arranged by Editor C.-Y. Lu.

T.-K. Kang and M.-J. Chen are with the Department of Electronics Engineering, National Chiao-Tung University, Hsinchu, Taiwan, R.O.C.

C.-H. Liu, Y. J. Chang, and S.-K. Fan are with the United Microelectronics Corporation, Hsinchu, Taiwan, R.O.C.

Publisher Item Identifier S 0018-9383(01)09373-X.

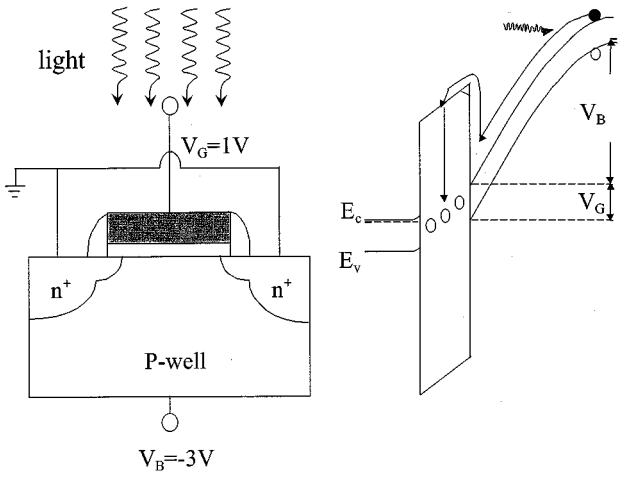

Fig. 1. Experimental setup and energy-band diagram for the optically induced trap filling method.

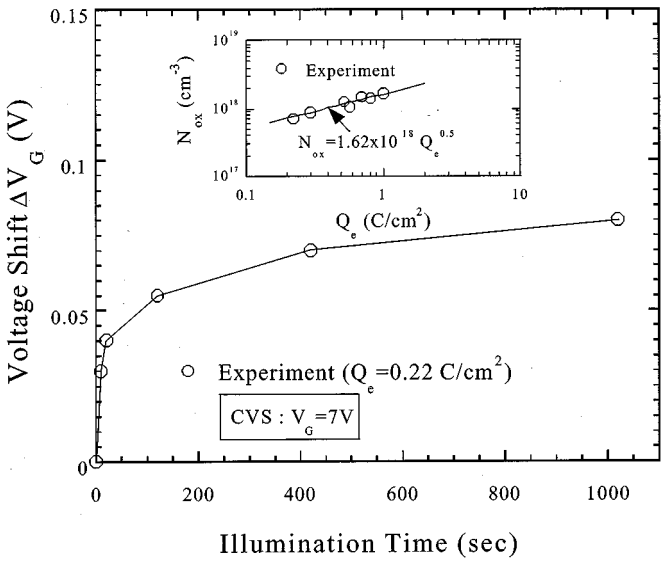

Fig. 2. Gate voltage shift versus illumination time. The inset shows the estimated trap density N_{ox} versus Q_e from seven samples. A power-law relation is drawn by best fitting data points. CVS is constant voltage stress.

the number of the unknown parameters in (1) reduces to two, namely, τ and E_t, whose values can further be accessed by fitting SILC I–V while simultaneously accounting for the maximum overall tunneling probability. Fortunately, there appears in the open literature [9] the experimental tunnel time versus oxide thickness, making possible direct verification on the extracted τ. Should an expected coincidence turn out, it can be acknowledged that current understandings involved in SILC mechanism indeed stand on the ground. Also addressed are differences and similarities of physical parameters as compared with the quotation [5] and [8].

II. Experimental Procedure

A variety of MOS devices were fabricated in the same process. The oxide film was thermally grown in dry oxygen ambient to 7-nm thick. The oxide areas were drawn in a wide range of four decade with aim to judge the present SILC theory that, whatever the areas used are, the traps generated during high field stress are spatially randomly allocated within the whole oxide space. This is valid until a percolation path for breakdown is formed locally. The first sample was n-channel MOSFETs having gate width-to-length ratio of 20 μm/0.3 μm (oxide area of 6×10^{-8} cm^2) to build a linkage between the trap generation density and the stress electron fluence. The stress condition was gate voltage $V_G = 7$ V with source, drain, and substrate tied to ground, which was followed by the optically induced trap filling. Fig. 1 schematically shows this optical injection method involving two distinct processes: the photo-generation process via a tungsten lamp to supply electron seed in the substrate; and the carrier heating process via a negative substrate bias of -3 V to raise these electrons up to the higher energy level enough to surmount the Si/SiO$_2$ barrier and fill the traps in oxide. Then operating in gated-diode forward mode [16] (i.e., a forward bias of 0.2 V was applied to the drain with source open and substrate grounded) can sensitively detect the filled traps.

The secondary sample was n$^+$-poly/p-substrate MOS capacitors having two very large oxide areas of 1.0×10^{-4} and 6.25×10^{-4} cm^2 for charge-to-breakdown test. The stress condition was constant current of 400 mA/cm^2 with positive gate voltage. The third sample was n-channel MOSFETs with oxide area of 5.2×10^{-6} cm^2 for monitoring the time evolution of gate voltage. The stress condition was constant current of 38.5 mA/cm^2 with positive gate voltage and with source, drain, and substrate connected to ground. The stressing was periodically interrupted to measure SILC I–V characteristics until oxide wear-out occurred.

III. Quantifying Neutral Traps

From the measured drain current in forward gated-diode mode versus gate voltage for the first sample, it is observed that the gate-voltage shift ΔV_G associated with the current peak in the depletion region increases with the filling or illumination time and gradually tends to saturate as depicted in Fig. 2. Assuming that the occupied traps are spatially distributed uniformly within the oxide as adopted elsewhere [12], [17], the saturation voltage shift $\Delta V_{G,\mathrm{SAT}}$ from the optical filling method can be directly linked to the occupied trap density N_{ox} through $\Delta V_{G,\mathrm{SAT}} = qt_{ox}^2 N_{ox}/2\varepsilon_{ox}$ where ε_{ox} is the oxide permittivity. The resulting N_{ox} in 1/cm^3 for different Q_e in C/cm^2 is displayed in the inset of Fig. 2, showing a power-law relation

$$N_{ox} = 1.62 \times 10^{18} Q_e^{0.5}. \tag{2}$$

A percolation oxide breakdown model [13] formulates explicitly the generated neutral electron trap density N_t as a function of three physical controlling factors: the possible minimum trap number $(= (t_{ox} - 2t)/2r)$ [14], the possible minimum area $(= \pi r^2)$ [10], [11], [14] of the locally conductive path, and the ultimate thickness limit of 2.5 nm for breakdown [18]. Here, r is the trap radius and t (= 0.5 nm [13]–[15]) is the transition layer thickness. N_t can be related directly to N_{ox} via filling fraction p: $N_t = N_{ox}/p$. The physical origin behind p is Coulomb repulsion [19]. p can be quantified from a Weibull

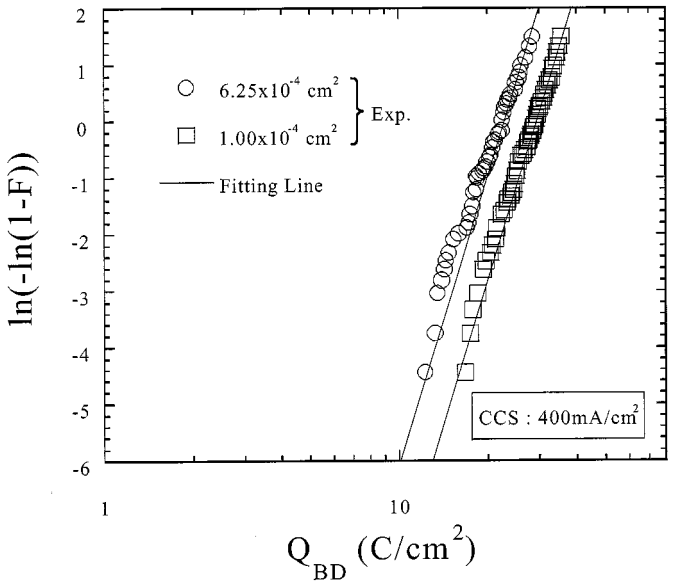

Fig. 3. Experimental charge-to-breakdown statistical distribution in Weibull plot. Two fitting lines of equal slope are shown for finding the Weibull slope. CCS is constant current stress.

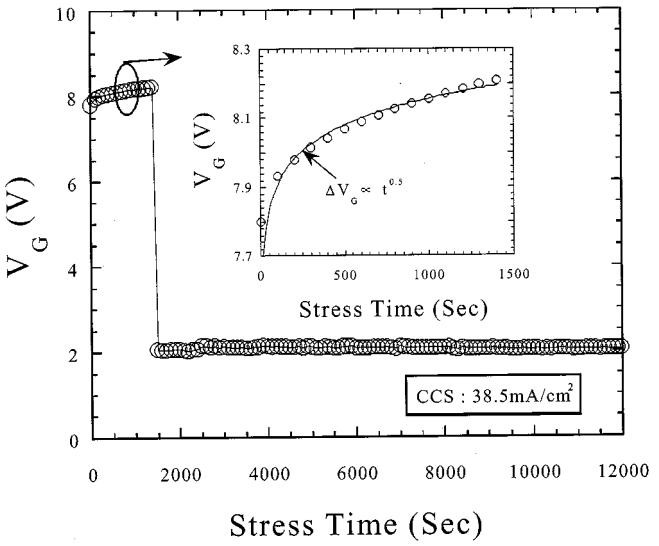

Fig. 4. Measured variation of the gate voltage versus stress time under the constant current stress condition. The evolution prior to breakdown is magnified in the inset, where a power-law line is shown.

plot of charge-to-breakdown Q_{BD} statistical data as shown in Fig. 3 from the secondary sample with two oxide areas. Here Q_{BD} is the Q_e at the onset of oxide breakdown. The Weibull slope s and the modal (63%) $Q_{BD}(m)$ of the charge-to-breakdown distributions can be expressed as [13]

$$s = 0.7821 \times \frac{t_{ox} - 2t}{r} \times 0.5 \tag{3}$$

$$Q_{BD(m)} = \left(\frac{pC_0}{1.62 \times 10^{18}}\right)^2 \left(\frac{A_o}{A_{ox}}\right)^{1/s}. \tag{4}$$

Here, C_0, the modal of the N_t distributions at a specific area A_o of 900 nm^2, reads [13]

$$C_0 = -1.5 \times 10^{20} + 7.611 \times 10^{19} \left(\frac{t_{ox} - 2.5}{r}\right)^{1/3} r^{-2}. \tag{5}$$

A_{ox} is the oxide area in nm^2. Extracting the values of s (through two fitting lines of equal slope) and $Q_{BD}(m)$ from Fig. 3 and substituting into the above expressions, a unique solution for two different areas yields $p = 4.75\%$ for $r = 0.42$ nm. Note that such a very low occupancy fraction of the total neutral trap density is essential in addressing ITAT. Therefore, we achieve the goal of quantifying in advance the amount of generated neutral electron traps for given stress electron fluence.

IV. Parameter Extraction and Comparisons

Fig. 4 shows the evolution of gate voltage for the third sample subject to constant F–N tunneling stressing. It can be seen that gate voltage gradually increases with time until at around 1400 s a large drop down to 2 V occurs, indicating a hard breakdown event. The built power-law relationship between N_t and Q_e reproduces excellently such breakdown event, regardless of areas used. This is achieved by substituting the stress current density of 38.5 mA/cm^2 and area of 5.2×10^{-6} cm^2 into (4) and (5). The resulting time to breakdown is 1332 sec, quite close to the spontaneous point in Fig. 4, as expected by Poisson area scaling [20]. The increment of gate voltage with respect to the initial value is magnified in the inset of Fig. 4 for stress time prior to breakdown, validating the power-law expression of $\Delta V_G \propto t^{0.5}$. Obviously, the traps generated are spatially randomly distributed within the whole oxide film, indicating that SILC magnitude obeys a linear relation with oxide area.

The measured SILC I–V before and after stress is displayed in Fig. 5. This figure reveals that the SILC I–V curves prevailing in the low voltage regime are raised up for increasing electron stress fluence, whereas in the high voltage region the I–V is intact, an indicative of F–N tunneling dominating. The latter property facilitates transformation from gate voltage to oxide field E_{ox}. F–N tunneling fitting [21] was applied to fresh I–V in Fig. 5 to access oxide field E_{ox}. The resulting $I - E_{ox}$ is depicted in Fig. 6. Prior to dealing with SILC I–V, a calculation work was carried out to furnish a set of τ and E_t to meet the condition of $P_1 = P_2$ for the maximum overall tunneling probability. Fig. 7 shows such calculation results using formula in [8]. From Fig. 7, a specific set of $\tau = 4.0 \times 10^{-13}$ s and $E_t = 3.4$ eV is rigorously selected since they are able to offer a best reproduction of SILC I–V or equivalently $I - E_{ox}$ in Fig. 6. Note that the main role of E_t and τ is to adjust the slope and magnitude of SILC I–V, respectively. The same power-law relation between N_t and Q_e is also involved in the way. This explains why we call the procedure of balancing systematically both the modeling and experimental works. This procedure comprises the quantification process for the trap density, the verification process for gate voltage evolution and time-to-breakdown, and the parameter extraction process. Finally, a fundamental parameter of interest is examined. Fig. 8 re-plots experimental tunnel time versus oxide thickness from [9], where a straight line drawn through all data points is extrapolated far to zero oxide thickness (the authors in [9] suggested that the most right side data would be adjusted upward for measurement reasons). As we place our extracted τ in the figure corresponding to zero oxide thickness,

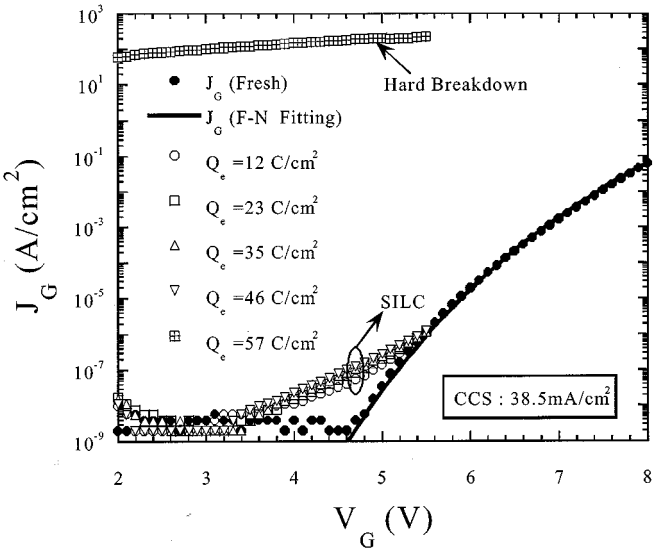

Fig. 5. Measured gate current versus gate voltage with electron fluence as a parameter. An F–N fitting line is plotted together.

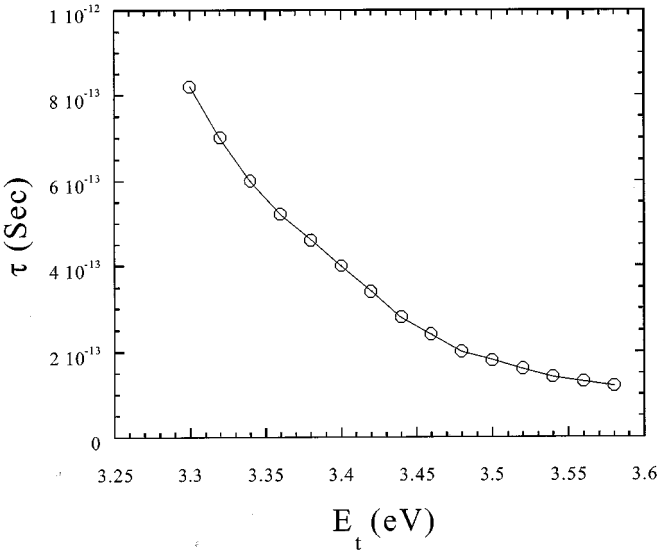

Fig. 7. Calculated set of the relaxation time and trap energy level at the trap site x_t of 3.6 nm to meet the maximum overall tunneling probability.

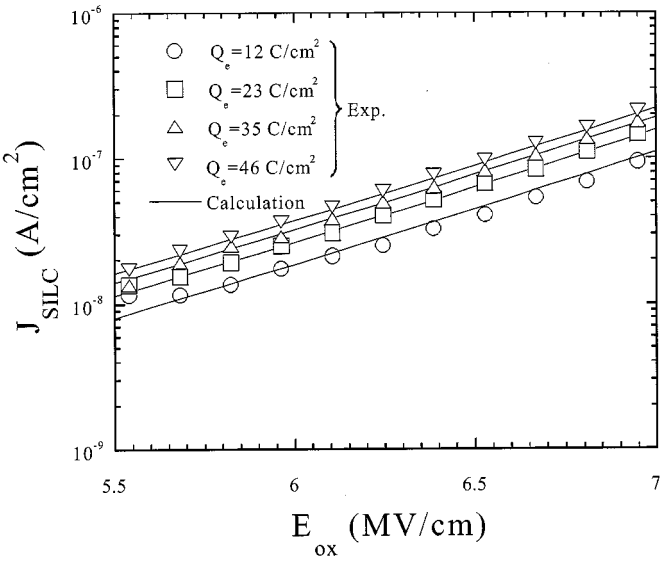

Fig. 6. Comparison of experimental and calculated SILC component versus oxide field corresponding to Fig. 5. The illustrated oxide field strength values were obtained by means of a fitting to F–N portion in Fig. 5.

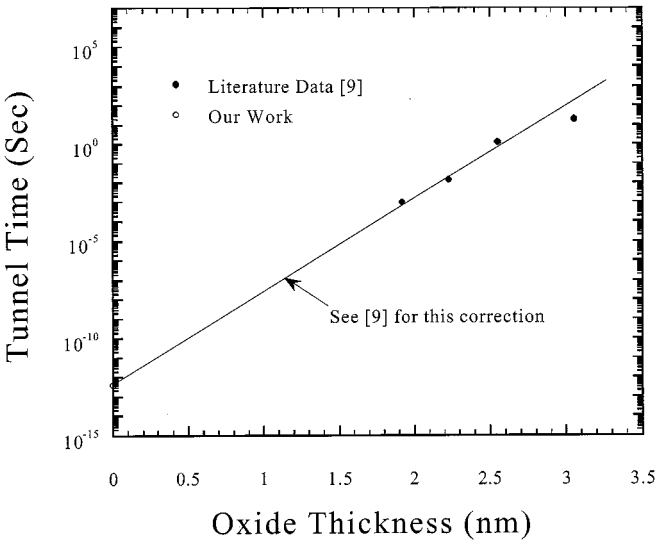

Fig. 8. Plot of tunnel time versus oxide thickness using literature data from [9]. Our extracted relaxation time is also shown and is found to fall on the same straight line with these data.

strikingly it matches exactly the line. Therefore, the expected coincidence turns out and it can be acknowledged unambiguously that current understandings involved in SILC mechanism indeed stand on the ground; for example, ITAT does favor the maximum likelihood of tunneling to and from the traps at the specific position; and the occupancy fraction of N_t is very low, ensuring the possibility of ITAT.

Several physical parameters involved in SILC I–V are also available in the open literature [5], [8], and are quoted here to explore differences and similarities of the involved parameters. Firstly, in the original work [8] pioneering the SILC model (1), τ was set at 1×10^{-15} s without particular reasons, far away from ours by about two orders of magnitude. However, our extracted E_t does not show such huge difference with that (3.6∼4.0 eV) in [8]. This is not strange as one can recognize from (1) that the exponent factor and the pre-exponent factor are essentially independent of each other, and only in the pre-exponent factor, τ and N_t are merged together and are thereby affected each other if extracted simultaneously. Secondly, a sophisticated modeling of SILC I–V curves in [5] reported the trap energy level = 2.3∼2.4 eV and the trap cross section = 10^{-15}∼10^{-16} cm^2. The latter is somewhat comparable with the trap sphere area ($=\pi r^2$) of 5.6×10^{-15} cm^2 in our work. Only the trap energy level is quite spread, which is likely attributed to different process technologies used. That is, the trap energy level would be considered a measure of the inelastic scattering property and different oxides from different process technologies reflect different scattering behaviors. Such arguments involved in inelastic scattering process certainly need further research.

V. Conclusion

A procedure of balancing systematically both the modeling and experimental works covering SILC and oxide breakdown has been comprehensively carried. Involved in this procedure are the quantification process for the trap density, the verification process for gate voltage evolution and time-to-breakdown, and the parameter extraction process for SILC I–V. Eventually the tunneling relaxation time viewed as a fundamental parameter has been exactly confirmed by existing data, thus clarifying unambiguously current physical aspects of SILC mechanism:

1) ITAT does favor the maximum likelihood of tunneling to and from the traps at the specific position;
2) the occupancy fraction of the total generated neutral density is very low ensuring the possibility of ITAT.

Differences and similarities of physical parameters between different studies have been compared as well, suggesting that the trap energy level would be considered a measure of the inelastic scattering property and is process technology dependent.

References

[1] K. Naruke, S. Taguchi, and M. Wada, "Stress-induced leakage current limiting to scale down EEPROM tunnel oxide thickness," in *IEDM Tech. Dig.*, 1988, pp. 424–427.

[2] R. Moazzami and C. Hu, "Stress-induced current in thin silicon dioxide films," in *IEDM Tech. Dig.*, 1992, pp. 139–142.

[3] D. J. Dumin and J. R. Maddux, "Correlation of stress-induced leakage current in thin oxides with trap generation inside the oxides," *IEEE Trans. Electron Devices*, vol. 40, pp. 986–993, 1993.

[4] D. J. DiMaria and E. Cartier, "Mechanism for stress-induced leakage currents in thin silicon dioxide films," *J. Appl. Phys.*, vol. 78, pp. 3883–3894, 1995.

[5] K. Sakakibara, N. Ajika, M. Hatanaka, and H. Miyoshi, "A quantitative analysis of stress induced excess current (SIEC) in SiO_2 films," in *Proc. IEEE IRPS*, 1996, pp. 100–107.

[6] S. Takagi, N. Yasuda, and A. Torium, "Experimental evidence of inelastic tunneling and new I–V model for stress-induced leakage current," in *IEDM Tech. Dig.*, 1996, pp. 323–326.

[7] E. Rosenbaum and L. F. Register, "Mechanism of stress-induced leakage current in MOS capacitors," *IEEE Trans. Electron Devices*, vol. 44, pp. 317–323, 1997.

[8] S. Kamohara, D. Park, and C. Hu, "Deep-trap SILC (stress induced leakage current) model for nominal and weak oxides," in *Proc. IEEE IRPS*, 1998, pp. 57–61.

[9] I. Lundstrom and C. Svensson, "Tunneling to traps in insulators," *J. Appl. Phys.*, vol. 43, pp. 5045–5047, 1972.

[10] J. Sune, I. Placencia, N. Barniol, E. Farres, F. Martin, and X. Aymerich, "On the breakdown statistics of very thin SiO_2 films," *Thin Solid Films*, vol. 185, pp. 347–362, 1990.

[11] D. J. Dumin, J. R. Maddux, R. S. Scott, and R. Subramoniam, "A model relating wearout to breakdown in thin oxides," *IEEE Trans. Electron Devices*, vol. 41, pp. 1570–1580, 1994.

[12] R. Degraeve, G. Groeseneken, R. Bellens, M. Depas, and H. E. Maes, "A consistent model for the thickness dependence of intrinsic breakdown in ultrathin oxides," in *IEDM Tech. Dig.*, 1995, pp. 863–866.

[13] H. T. Huang, M. J. Chen, J. H. Chen, C. W. Su, C. S. Hou, and M. S. Liang, "A trap generation statistical model in closed-form for intrinsic breakdown of ultrathin oxides," in *Proc. IEEE Int. Symp. VLSI-TSA*, 1999, pp. 70–73.

[14] M. J. Chen, H. T. Huang, J. H. Chen, C. W. Su, C. S. Hou, and M. S. Liang, "Cell-based analytic statistical model with correlated parameters for intrinsic breakdown of ultrathin oxides," *IEEE Electron Device Lett.*, vol. 20, pp. 523–525, 1999.

[15] M. J. Chen, T. K. Kang, C. H. Liu, Y. J. Chang, and K. Y. Fu, "Oxide thinning percolation statistical model for soft breakdown in ultrathin gate oxides," *Appl. Phys. Lett.*, vol. 77, pp. 555–557, 2000.

[16] M. J. Chen, T. K. Kang, H. T. Huang, C. H. Liu, Y. J. Chang, and K. Y. Fu, "Forward gated-diode measurement of filled traps in high-field stressed thin oxides," *IEEE Trans. Electron Devices*, vol. 47, pp. 1682–1683, Aug. 2000.

[17] T. H. Ning, C. M. Osburn, and H. N. Yu, "Emission probability of hot electrons from silicon into silicon dioxide," *J. Appl. Phys.*, vol. 48, pp. 286–293, 1977.

[18] D. J. DiMaria and J. H. Stathis, "Ultimate limit for defect generation in ultrathin silicon dioxide," *Appl. Phys. Lett.*, vol. 71, pp. 3230–3232, 1997.

[19] Y. Nissan-Cohen, J. Shappir, and D. Frohman-Bentchkowsky, "Trap generation and occupation dynamics in SiO_2 under charge injection stress," *J. Appl. Phys.*, vol. 60, pp. 2024–2035, 1986.

[20] D. R. Wolters and J. F. Verwey, *Instabilities in Silicon Devices*. Amsterdam, The Netherlands: Elsevier, 1986, pp. 332–335.

[21] A. Gupta, P. Fang, M. Song, M. R. Lin, D. Wollesen, K. Chen, and C. Hu, "Accurate determination of ultrathin gate oxide thickness and effective polysilicon doping of CMOS devices," *IEEE Electron Device Lett.*, vol. 18, pp. 580–582, 1997.

Ting-Kuo Kang was born in Kaohsiung, Taiwan, R.O.C., on November 12, 1971. He received the M.S. degree in electrical engineering from the DaYeh Institute of Technology, Chang-Hwa, Taiwan, in 1996 and is currently pursuing the Ph.D. degree in the Institute of Electronics, National Chiao-Tung University, Hsinchu, Taiwan.

His research interests are in the areas of ultrathin gate oxide reliability.

Ming-Jer Chen (S'78–M'79–SM'98) received the B.S. degree in electrical engineering (with highest honors) from National Cheng Kung University, Taiwan, R.O.C., in 1977, and the Ph.D. degree from National Chiao-Tung University (NCTU), Taiwan, in 1985.

Since 1985, he has been with the Department of Electronics Engineering, NCTU, where he is a Professor. From 1987 to 1992, he was a principal consultant for TSMC, where he led a team from NCTU and ERSO/ITRI to build process window and design rule. In 1996 and 1997, he critically enabled the ERSO/ITRI video A/D converters and the TSMC mixed-mode CMOS processes, respectively. From 2000 to 2001, he was a Visiting Professor at the Department of Electrical Engineering and Center for Integrated Systems, Stanford University, Stanford, CA. His current research interests are on technology reliability physics and nano-scale electronics. He has graduated six Ph.D. students and has been granted four U.S. patents and six Taiwanese patents in the above areas.

Dr. Chen received the 1992 and 1993 Chinese Young Engineer Paper Award and the 1996 Acer Distinguished Ph.D. Dissertation Award. He is a member of Phi Tau Phi.

Chuan-Hsi (Jason) Liu was born in Taiwan, R.O.C., in November 1965. He received the Ph.D. degree in electrical engineering from Arizona State University, Tempe, in 1997.

Afterwards, he joined United Microelectronics Corporation (UMC), Hsinchu, Taiwan, where he has been engaged in the research and development of gate oxide process and reliability, and was promoted to Section Manager in November 1999. In May 2000, he was sent to IBM, Hopewell Junction, NY, for a joint project (the development of 0.10 μm technology) between UMC, IBM, and Infineon, in which he has been dedicated to the growth and characterization of ultrathin gate dielectrics. He has published approximately 15 technical papers.

Yih J. Chang, photograph and biography not available at the time of publication.

Shou-Kong Fan received the Sc.D. degree in materials science from the Massachusetts Institute of Technology, Cambridge, in 1987.

From 1986 to 1988, he was a Process Engineer at the Semiconductor Sector, Texas Instruments (TI). From 1988 to 1994, he was Member of Technical Staff, TI's Central Research Lab, working on III-V based HBT microwave device. He also served as an Adjunct Professor at Southern Methodist University, Dallas, TX, from 1989 to 1990, teaching senior level electrodynamics. In 1994, he joined TI-Acer as Project Manager to work on a 16M DRAM design. Later he led a device design group to convert the 16M DRAM process to 0.5 μm logic process for foundry service using TCAD tool. From 1998 to 2000, he first worked as Deputy Director on TCAD and later on reliability as Specialist at WSMC. Since June 2000, he is with UMC as the Senior Reliability Assurance Department Manager. His responsibilities include the execution or implementation of reliability assessment, process/product qualification and the wafer-level reliability monitoring scheme among all UMC fabs.