# 行政院國家科學委員會專題研究計畫　成果報告

## 高度微縮金氧半場效電晶體機械應力效應之研究(3/3)

計畫類別：個別型計畫
計畫編號：NSC94-2215-E-009-005-
執行期間：94 年 08 月 01 日至 95 年 07 月 31 日
執行單位：國立交通大學電子工程學系及電子研究所

計畫主持人：陳明哲

計畫參與人員：許義明，謝振宇，李建志，李韋漢，林以唐，梁惕華，呂立方，周佳宏，簡鶴年，陳彥銘，宋東壕

報告類型：完整報告

處理方式：本計畫可公開查詢

中　華　民　國 95 年 10 月 30 日

# 行政院國家科學委員會補助專題研究計畫成果報告

※※※※※※※※※※※※※※※※※※※※※※

※　　高度微縮金氧半場效電晶體機械應力效應之研究(3/3)　　※

※※※※※※※※※※※※※※※※※※※※※※※

計畫類別：■個別型計畫　　□整合型計畫

計畫編號： NSC 94-2215-E-009-005-

執行期間： 94/08/01 ~ 95/07/31


計畫主持人：陳明哲

計畫參與人員：許義明，謝振宇，李建志，李韋漢，林以唐，
梁惕華，呂立方，周佳宏，簡鶴年，陳彥銘，宋東壕


成果報告類型(依經費核定清單規定繳交)：□精簡報告　■完整報告

本成果報告包括以下應繳交之附件：

□赴國外出差或研習心得報告一份

□赴大陸地區出差或研習心得報告一份

□出席國際學術會議心得報告及發表之論文各一份

□ 國際合作研究計畫國外研究報告書一份

處理方式：除產學合作研究計畫、提升產業技術及人才培育研究計畫、列管計畫及下列情形者外，得立即公開查詢

□涉及專利或其他智慧財產權，□一年□二年後可公開查詢


執行單位：國立交通大學電子工程學系


中　華　民　國　95　年　10　月　28　日

# 行政院國家科學委員會補助專題研究計畫成果報告

## 高度微縮金氧半場效電晶體機械應力效應之研究(3/3)

## The Impact of Mechanical Stress in Highly Scaled MOSFETs (3/3)

執行期限: 94/08/01 ~ 95/07/31

計畫編號: NSC 94-2215-E-009-005-

主持人：陳明哲教授　　國立交通大學電子工程學系

## 一、中文摘要

本計畫探索下世代場效電晶體機械應力效應之嶄新領域，將配合當年度互補式金氧半場效電晶體製程技術演進同時執行八大項目:

1. 以當年度最先進製程製作 n- and p-型 高度微縮金氧半場效電晶體元件，以金氧半場效電晶體微縮尺寸和閘極至淺槽隔離邊緣距離二者為圖案變動參數。 進行電流電壓/電容電壓量測。以自行發展的一維量子力學模擬器與實驗比較以萃取重要製程參數。
2. 執行二維製程及機械應力之模擬以萃取張力之分佈, 大小並決定其性質。
3. 定義不同機械應力條件即將萃取得到的張力大小表達為不同製程參數, 不同場效電晶體微縮尺寸, 和不同閘極至淺槽隔離邊緣距離的函數。
4. 以自行發展的量子力學模擬器計算不同機械應力條件下反置通道二維電子(洞)氣及次能帶分佈，以獲得遷移率等重要參數。建立機械應力之理論架構並導出張力解析模式。
5. 低溫實驗萃取通道彈道傳輸係數以與機械應力作一關聯。
6. 二維量子力學彈道傳輸模擬器之程式撰寫,除錯及執行。
7. 低頻雜訊及電報雜訊量測以偵測機械應力之細微變化。
8. 機械應力效應置入電路模擬器場效電晶體模式並進行電路功能方塊模擬。

本計畫目的即為藉著上述執行項目以肇清機械應力對下世代場效電晶體元件性能之效應及介觀物理機制，另方面將機械應力效應置入電路模擬器場效電晶體模式以進行正確的積體電路設計，進而使得系統單晶片積體電路設計複雜化及高密度化能有效解除機械應力之威脅。

*關鍵詞：場效電晶體; 機械應力; 張力; 二維電子氣; 二維電洞氣; 通道彈道傳輸; 低頻雜訊; 電報雜訊; 模式; 積體電路設計。*

## 英文摘要

The project explores the mechanical stress issue on the highly scaled MOSFETs. Following are the eight main items to be conducted along with state-of-the-art CMOS scaled technologies:

1. Fabricate n- and p-type highly scaled MOSFETs by advanced process technologies, followed by I-V/C-V characterization as well as our developed 1-D quantum mechanical I-V/C-V simulation to extract key process parameters.
2. Perform a two-dimensional process/mechanical stress simulator to extract the magnitude and distribution of strain and determine its property.
3. Experimentally define mechanical stress conditions in terms of strain expressed as function of process parameters, FET scaling factor, and distance between gate to STI (shallow trench isolation) edge.
4. Perform our developed 1-D quantum mechanical simulators to assess 2-dimensional electron (hole) gas and sub-band distributions, leading to extraction of relevant parameters like mobility. Establish theoretical framework of mechanical stress and develop analytic model of strain.
5. Perform low-temperature experiment to extract channel backscattering coefficients and relate them to mechanical stress.
6. Establish 2-D quantum mechanical ballistic simulators.
7. Perform low frequency noise and random telegraph signal experiment to detect some change in mechanical stress.
8. Incorporate mechanical stress into SPICE MOSFET model and perform simulation on functional circuit blocks.

The goal of the project is to clarify the impact of mechanical stress on next generation MOSFETs and the underlying mesoscopic physical mechanisms. Meanwhile, incorporation of mechanical stress into SPICE MOSFET model enables correct IC design and as a result, we can reach a a complex, highly dense SOC design effectively free of the harmful mechanical stress.

***Key Words***：*MOSFET; Mechanical Stress; Strain; 2-Dimensional Electron Gas; 2-Dimensional Hole Gas; Channel Ballistic Transport; Low-Frequency Noise; Random Telegraph Signal; Model; IC Design*

## 二、緣由與目的

依照 International Technology Roadmap for Semiconductors，在下世代 Silicon CMOS 製程技術演進過程當中， MOSFET 尺寸持續微縮且 Gate 極亦持續更為靠近 STI(Shallow Trench Isolation)邊緣，衍生的機械應力(Mechanical Stress)已有文獻報告對元件推動電流產生衰減或其他影響，此問題對於積體電路設計複雜化及高密度化之趨勢(尤其是系統單晶片設計 SOC)不利，故以機械應力效應為議題的研究最近在國際上受到高度重視。

我們過去數年經驗的累積反映出一個有系統的嚴謹性的機械應力效應研究是十分緊迫的，此為本計畫提出之背景:

1 在我們最近發表的創見性的一超大型級之測試載具晶片，內含有一連串高敏感度之測試結構，用以偵測超大型晶片在不同製程技術製造下之可靠度分析，並提出一新的應力模組理論，來估算晶片故障分布與故障機構模式。結果呈現超大型晶片故障之韋博統計分布與晶片幾何大小，封裝材料相關。此外由應力分析模擬和針壓實驗，我們找到一最佳抗應力緩衝層鍍於輸出入墊片上方。再將靜電放電保護元件及環狀振盪電路置於輸出入墊片下方。在傳輸線脈衝高壓測試下，輸出入墊片下之靜電放電保護元件的二次崩潰電壓和電流曲線仍具優秀的特性。而在直流與交流實驗測試下，輸出入墊片下之環狀振盪電路的傳播延遲時間仍然保留在大約 20 微微秒左右。

2. 本研究群這幾年已自行發展出 1-D 量子力學(Quantum Mechanical) I-V/C-V numerical 模擬器， 能計算出 Inversion Channel 因 Quantum Confinement 效應引致的二維電子氣(2DEG)或二維電洞氣(2DHG)及其 Sub-bands 分佈大小等。本計畫將此基礎延伸至 2-D Quantum Mechanical Ballistic 模擬器程式撰寫,除錯及執行以反映最近趨勢。

3. 萃取 Channel Backscattering Coefficients 的低溫實驗方法已在 2002 IEDM 發表。

4. 本研究群亦已成功建立高度靈敏低頻雜訊量測系統及時域擾動量測系統。

本計畫目的即肇清機械應力對下世代 FET 元件性能之效應及介觀物理機制，另方面將機械應力效應置入 SPICE MOSFET Model 以進行正確的 IC Design，進而使得 SOC 積體電路設計複雜化及高密度化之同時，能有效解除機械應力之威脅。

## 三、研究方法與成果

1. 我們在 Highly Scaled MOSFETs (即以 MOSFET 微縮尺寸和 Gate 極至 STI 邊緣距離二者為 Layout 變動參數)上量測不同機械應力條件下之 I-V 特性，實驗發現機械應力確會降低 n-Channel MOSFETs 元件推動電流但增加 p-Channel MOSFETs 元件推動電流，與文獻上發表者 不同。
2. 完成量子力學 I-V/C-V 模擬與實驗比較並萃取重要參數如 poly doping, substrate doping, flatband voltage, 有效氧化層厚度。
3. 完成二維製程及機械應力之模擬並萃取決定 Strain 之分佈, 大小, 性質。
4. 完成萃取得到的 Strain 大小表達為不同製程參數, 不同 MOSFET 微縮尺寸, 和不同 Gate 極至 STI 邊緣距離的函數。
5. Currently we are calculating 量子力學模擬計算不同機械應力條件下 Inversion Channel 二維電子(洞)氣及 Sub-bands 分佈，並獲得 Mobility 等重要參數。
6. Currently we are improving Mechanical Stress 之理論架構建立並導出 Strain 解析模式。
7. Currently we are conducting scattering 實驗萃取 Channel Backscattering Coefficients 並與 strain 應力作一 Correlation。

## 四、結論與討論

(1) 低頻雜訊拿來使用在監控受不同的製程應力程度下的氧化層介面品質，在承受製程應力的金氧半電晶體中的低頻雜訊量測中，發現靠近表面的缺陷密度隨著通道寬度而變化。這個發現可以解釋為在矽和氧化矽的介面間，因為晶格長度不匹配所造成的 $P_b$ 中心可視為靠近表面的缺陷的主要來源。在低頻雜訊的實驗中，對於通道寬度的縮減，相對應於應力的提高，也降低晶格長度不匹配的程度。

(2) 在產製下世代受應力電子元件 Strain Engineering 領域最近作出了重大貢獻---- 針對 Uniaxial Strain 下 Impurities (特別是 Boron, 為目前高度挑戰卻也爭議性極大的題目) 在 Silicon 的高溫特殊擴散行為提出前所未有、嶄新物理模式並獲得實驗支持:

M. J. Chen and Y. M. Sheu, "Effect of uniaxial strain on anisotropic diffusion in silicon," *Applied Physics Letters*, vol. 89, pp. 161908-1-181908-3, Oct. 2006.

以下為 *APL* 國際評審給的 Evaluation Summary:

Paper Interesting: Yes
Original Paper: Yes
Sufficient Physics: Yes
Well Organized and Clear: Yes
Free From Errors: Yes
Conclusions Supported: Yes
Appropriate Title: Yes
Good Abstract: Yes
Satisfactory English: Yes
Adequate References: Yes
Clear Figures: Yes
OVERALL RATING: Excellent

(3) 本人已將奈米線研究(一篇有關研究電流流經奈米線狀的 percolation path 的 APL 論文被德國以及希臘的頂尖大學研究團隊所引用於發表在 2005 APL 和 2006 PRB 的論文之中)延伸至奈米碳管電晶體，發表了一篇 *Applied Physics Letters* 論文 (此為跨領域論文,本人為通訊作者):

M. P. Lu, C. Y. Hsiao, P. Y. Lo, J. H. Wei, Y. S. Yang, and M. J. Chen, "Semiconducting single-walled carbon nanotubes exposed to distilled water and aqueous solution: electrical measurement and theoretical

calculation," *Applied Physics Letters*, vol. **88**, pp. 053114-1—053114-3, Feb. 2006.

此論文探討生物樣本電性量測及應用量子力學理論計算，特別為 *APL* 國際評審高度認定。亦被選入兩本知名的學術網路期刊:

- Selected Articles in *Virtual Journal of Nanoscale Science & Technology*, Vol. **13**, Issue 6, 2006.
- Selected Articles in *Virtual Journal of Biological Physics Research*, Vol. **11**, Issue 4, 2006.

(4) 庫倫能量(Coulomb energy)是針對電子填滿在氧化層之內的奈米尺寸陷阱(Nanometer-Scale Trap)，傳統的庫倫能量只考慮在表面的陷阱，在我們的研究之中，我們針對1.7 奈米厚度的氧化層元件所做的實驗驗證：比較深入的缺陷會造成庫倫能量的增高，另外一個證據來自於multiphonon 理論，可以解釋電子被缺陷抓住和釋放的能量交換，我們更進一步闡述我們所描寫的庫倫能量和一般在記憶體元件中電子填充量子點(Quantum Dot)或量子晶體(Nanocrystal)記憶體所面對的庫倫能量在物理上(即能量系統圖)是相當有所關連的:

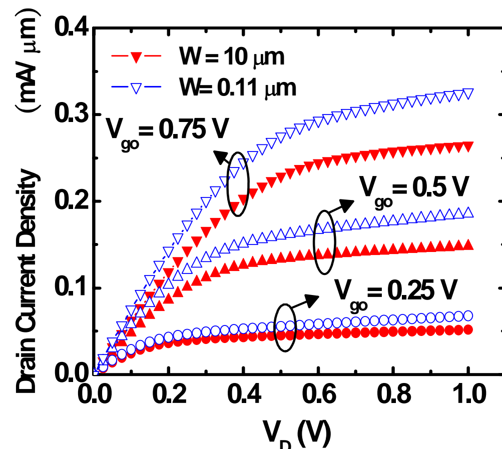
- M. P. Lu and M. J. Chen, "Oxide-trap-enhanced Coulomb energy in a metal-oxide-semiconductor system", *Physical Review B*, vol. **72**, pp. 235417-1—235417-5, December 2005.

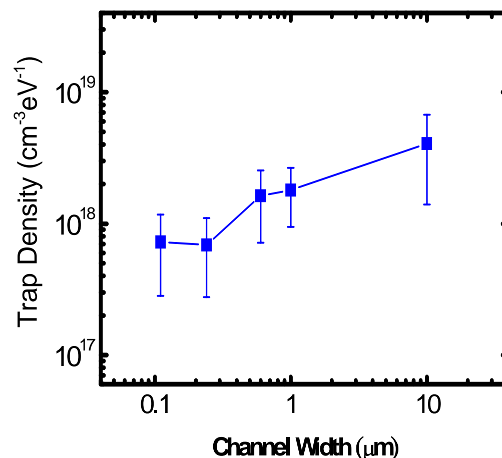
簡言之，此篇論文為相關領域二十幾年以來所作最好的，且為兩位國際評審認定，並被 American Journal Experts 所注意到，這是一個來自全美 Top 10 大學的研究生所成立的組織，其中有研究相關奈米領域的研究生，他在信中提及"*was impressed by the quality of the research and the thought behind your article*"。以嚴謹的物理探討延伸至量子點(Quantum Dot)或量子晶體(Nanocrystal)， 將能提出新的設計方法以因應奈米尺度的微觀世界。

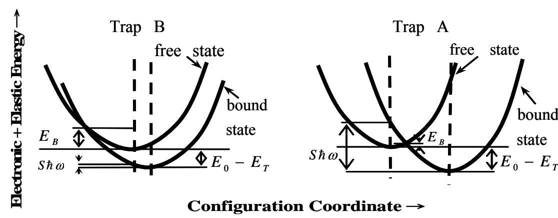
## 參考文獻(發表論文):


1. Y. M. Sheu, C. S. Chang, H. C. Lin, S. S. Lin, C. H. Lee, C. C. Wu, M. J. Chen, C. H. Diaz, "Impact of STI Mechanical Stress in Highly Scaled MOSFETs," *IEEE Symposium on VLSI-TSA, Technical Digest*, pp. 269-272, 2003.
2. Y. M. Sheu, Kelvin Y. Y. Doong, C. H. Lee, M. J. Chen, C. H. Diaz, "Study on STI Mechanical Stress Induced Variation in 90-nm CMOSFETs," *IEEE International Conference on Microelectronic Test Structures*, pp. 205-208, 2003.
3. M. J. Chen, H. T. Huang, Y. C. Chou, R. T. Chen, Y. T. Tseng, P. N. Chen, and C. H. Diaz, "Separation of Channel Backscattering Coefficients in Nanoscale MOSFETs," *IEEE Trans. Electron Devices*, vol. **51**, pp.1409-1415, September 2004.
4. Y. M. Sheu, S. J. Yang, C. C. Wang, C. S. Chang, L. P. Huang, T. Y. Huang, M. J. Chen, C. H. Diaz, "Modeling mechanical stress effect on dopant diffusion in scaled MOSFETs," *IEEE Trans. Electron Devices*, vol.52, pp. 30-38, January 2005.
5. M. J. Chen, R. T. Chen, and Y. S. Lin, "Decoupling channel backscattering coefficients in nanoscale MOSFETs to establish near-source channel conduction-band profiles," *Silicon Nanoelectronics Workshop*, pp. 50-51, June 2005. (Kyoto).
6. M. P. Lu and M. J. Chen, "Oxide-trap-enhanced Coulomb energy in a metal-oxide-semiconductor system", *Physical Review B*, vol. **72**, pp. 235417-1—235417-5, December 2005.
7. M. P. Lu, C. Y. Hsiao, P. Y. Lo, J. H. Wei, Y. S. Yang, and M. J. Chen, "Semiconducting single-walled carbon nanotubes exposed to distilled water and aqueous solution: electrical measurement and theoretical calculation," *Applied Physics Letters*, vol. **88**, pp. 053114-1—053114-3, Feb. 2006. Selected Articles in *Virtual Journal of Nanoscale Science & Technology*, Vol. **13**, Issue 6, 2006; and Selected Articles in *Virtual Journal of Biological Physics Research*, Vol. **11**, Issue 4, 2006.
8. M. P. Lu, W. C. Lee, M. J. Chen, "Channel-width dependence of low-frequency noise in process tensile-strained n-channel metal-oxide-semiconductor transistors," *Applied Physics Letters*, vol. **88**, pp. 063511-1—063511-3, Feb. 2006.
9. C. Y. S. Cho, M. J. Chen, C. F. Chen, P. Tuntasood, D. T. Fan, and T. Y. Liu, "A novel self-aligned highly reliable sidewall split-gate Flash memory," *IEEE Trans. Electron Devices*, vol. **53**, pp. 465-473, March 2006.
10. Y. M. Sheu, S. J. Yang, C. C. Wang, C. S. Chang, M. J. Chen, S. Liu, and C. H. Diaz, "Reproducing subthreshold characteristics of metal-oxide-semiconductor field effect transistors under shallow trench isolation mechanical stress using a stress-dependence diffusion model," *Japanese Journal of Applied Physics*, vol. **45**, pp. L849-L851, August 2006.
11. Y. M. Sheu, K. W. Su, S. Tian, S. J. Yang, C. C. Wang, M. J. Chen, and S. Liu, "Modeling the well-edge proximity effect in highly-scaled MOSFETs," *IEEE Trans. Electron Devices*, vol. 53, pp. 2792-2798, Nov. 2006.
12. M. J. Chen and Y. M. Sheu, "Effect of uniaxial strain on anisotropic diffusion in silicon," *Applied Physics Letters*, vol. 89, pp. 161908-1-181908-3, Oct. 2006.

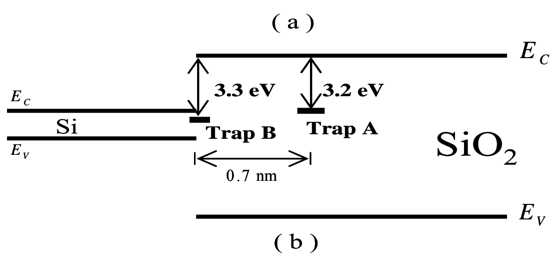

## Important Figures and Tables



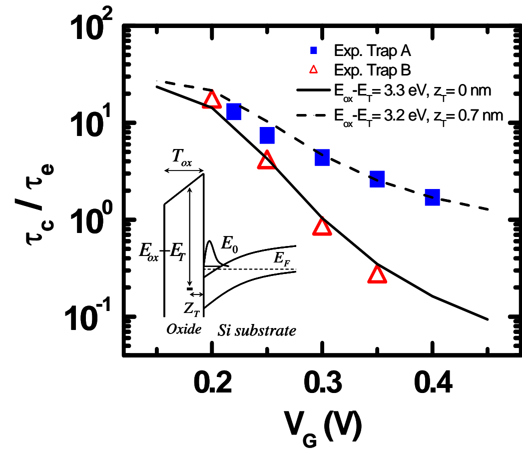



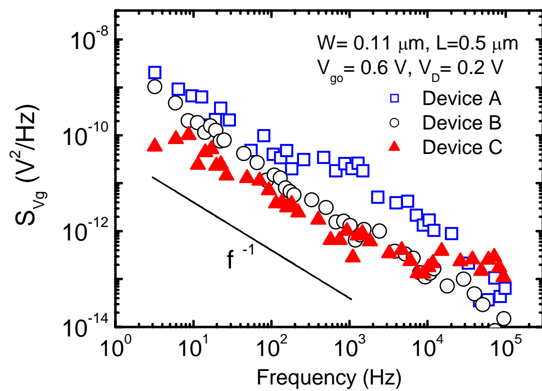


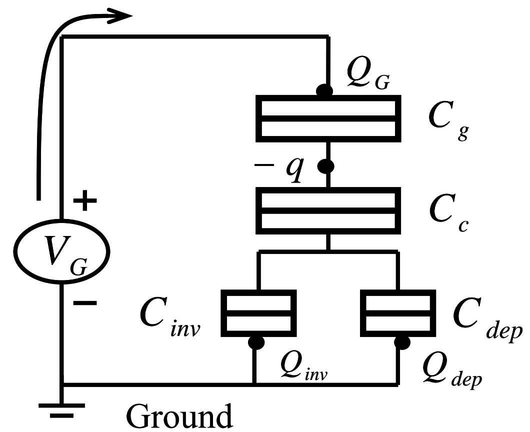

( a )

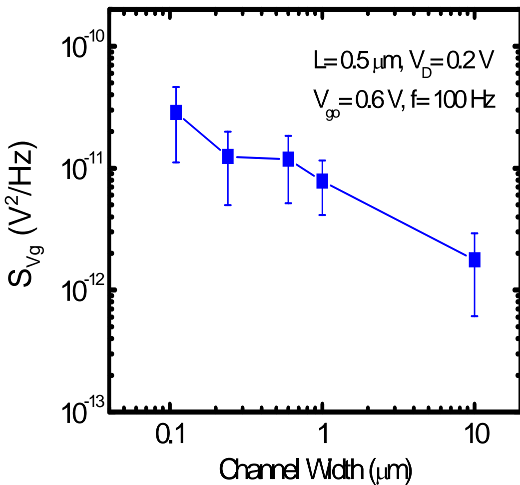


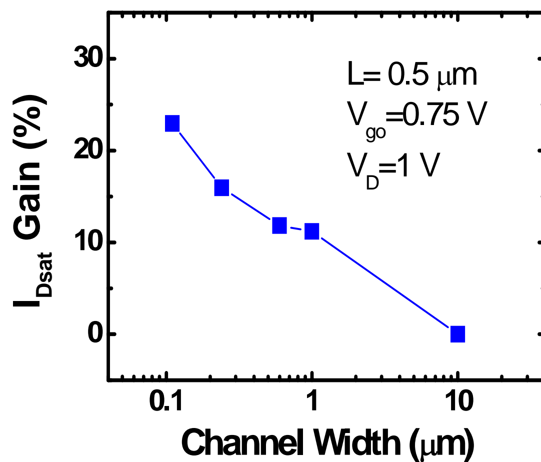

( b )

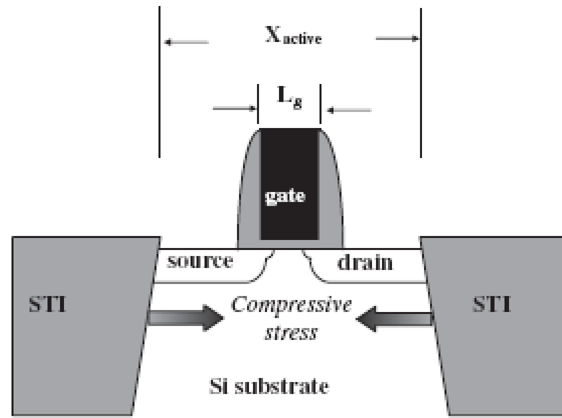


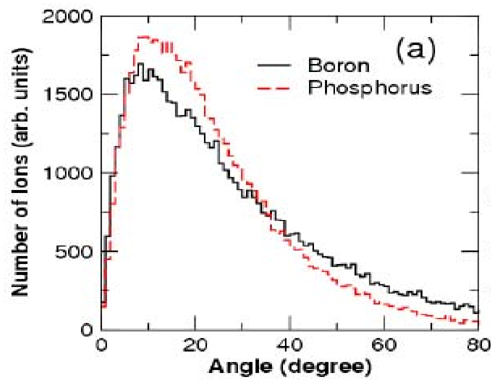



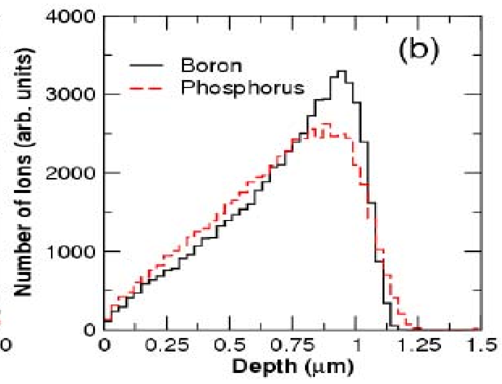



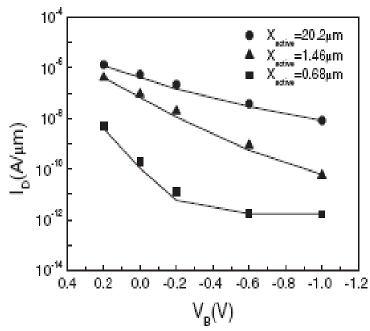



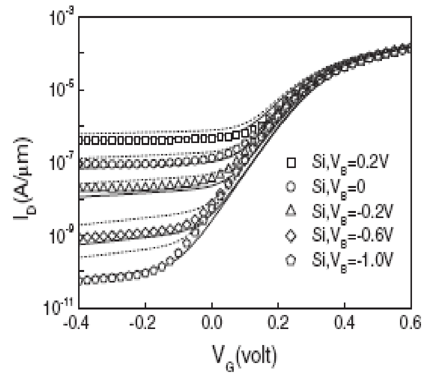





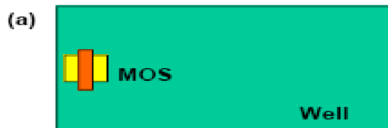



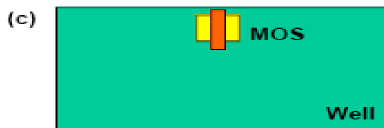



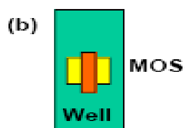

(a)



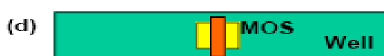

(c)

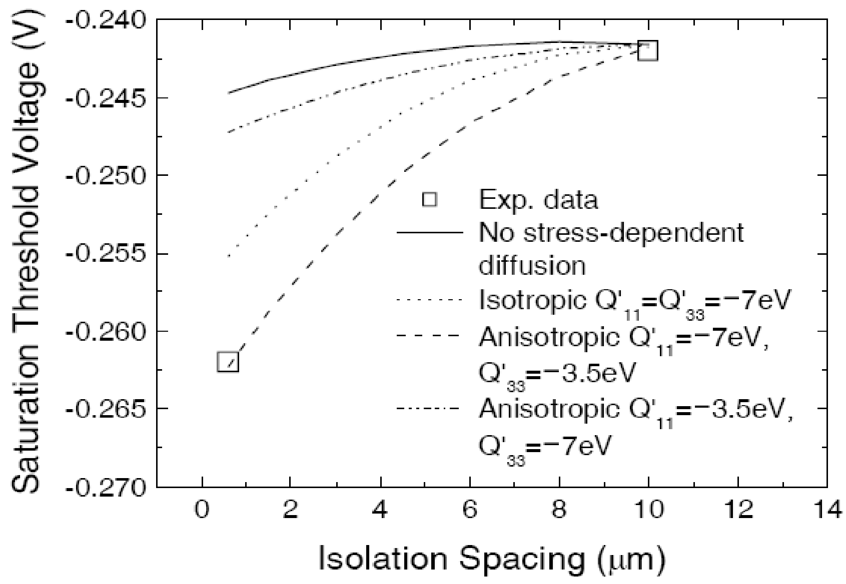



(b)

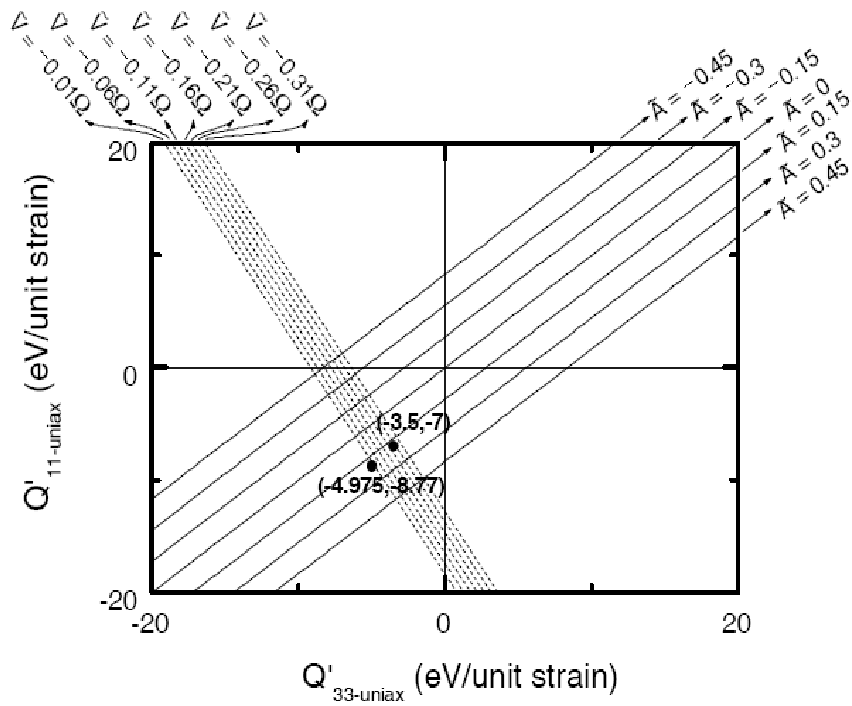



(d)





Saturation Threshold Voltage (V)
-0.240
-0.245
-0.250
-0.255
-0.260
-0.265
-0.270
0
2
4
6
8
10
12
14
Isolation Spacing (μm)
Exp. data
No stress-dependent diffusion
Isotropic Q'11=Q'33=−7eV
Anisotropic Q'11=−7eV, Q'33=−3.5eV
Anisotropic Q'11=−3.5eV, Q'33=−7eV

V̄ = −0.01Ω
V̄ = −0.06Ω
V̄ = −0.11Ω
V̄ = −0.16Ω
V̄ = −0.21Ω
V̄ = −0.26Ω
V̄ = −0.31Ω
Ā = −0.45
Ā = −0.3
Ā = −0.15
Ā = 0
Ā = 0.15
Ā = 0.3
Ā = 0.45
(-3.5,-7)
(-4.975,-8.77)
Q'11-uniax (eV/unit strain)
Q'33-uniax (eV/unit strain)
20
0
-20
-20
0
20

# Effect of uniaxial strain on anisotropic diffusion in silicon

Ming-Jer Chen[a] and Yi-Ming Sheu
*Department of Electronics Engineering, National Chiao Tung University, Hsin-Chu 300, Taiwan*

(Received 9 June 2006; accepted 30 August 2006; published online 17 October 2006)

A physical model is directly extended from the thermodynamic framework to deal with anisotropic diffusion in uniaxially stressed silicon. With the anisotropy of the uniaxial strain induced activation energy as input, two fundamental material parameters, the activation volume and the migration strain anisotropy, can be quantitatively determined. When applied to boron, a process-device coupled simulation is performed on a $p$-type metal-oxide-semiconductor field-effect transistor undergoing uniaxial stress in a manufacturing process. The resulting material parameters have been found to be in satisfactory agreement with values presented in the literature. © *2006 American Institute of Physics*. [DOI: 10.1063/1.2362980]

Strain engineering has been widely recognized as an indispensable performance booster in producing next-generation metal-oxide-semiconductor field-effect transistors (MOSFETs).[1,2] There have been two fundamentally different methods used to achieve this goal:[1,2] (i) biaxially strained silicon on a relaxed SiGe buffer layer and (ii) uniaxially strained silicon through the use of trench isolation, silicide, and cap layers during the manufacturing process. However, diffusion in strained silicon is essentially different from that of unstrained silicon. Thus, an understanding of strain dependent diffusion, as well as its control, is a challenging issue. So far, there have been significant studies in this direction covering a wide range of experimental findings and confirmations,[3–9] atomistic calculations,[10–13] physical models,[10–16] and technology computer-aided design.[17] Specifically, Cowern *et al.*[5] experimentally revealed a linear dependence of the activation energy on strain. Within the thermodynamic framework constructed by Aziz *et al.* (see Ref. 18, which is more recent and more thorough than the earlier works cited above), the activation volume ($\tilde{V}$) and the anisotropy of the migration volume ($\tilde{V}_{\parallel}^{m}-\tilde{V}_{\perp}^{m}$) exist in nature. The combination of the activation energy, the activation volume, and the anisotropy of the migration volume is remarkable, as demonstrated in a physical model[14–16,18] dedicated to both the hydrostatic pressure experiment and the in-plane biaxial stress experiment,

$$\tilde{V}+\frac{3}{2}\frac{Q'_{33-\text{biax}}}{Y_{\text{biax}}}=\pm\Omega+(\tilde{V}_{\parallel}^{m}-\tilde{V}_{\perp}^{m}), \tag{1}$$

$$\tilde{V}+\frac{3}{2}\frac{Q'_{11-\text{biax}}}{Y_{\text{biax}}}=\pm\Omega-\frac{1}{2}(\tilde{V}_{\parallel}^{m}-\tilde{V}_{\perp}^{m}), \tag{2}$$

where $Q'_{33-\text{biax}}$ is the biaxial strain induced activation energy in the direction normal to the silicon surface, $Y_{\text{biax}}$ is the biaxial modulus, $\Omega$ is the lattice site volume, and $Q'_{11-\text{biax}}$ is the biaxial strain induced activation energy in the direction parallel to the surface.

On the other hand, in the case of uniaxial stress as encountered while fabricating the MOSFET, without the use of a relaxed SiGe buffer layer, the stress is created through the trench isolation, silicide, or cap layers in a manufacturing process. Therefore, a straightforward extension to the uniaxial strain counterpart is essential. In this letter, one such model is derived and its linkage to the case of biaxial strain, Eqs. (1) and (2), is established. When applied to boron, a process-device coupled simulation is performed on a $p$-type MOSFET undergoing uniaxial stressing during the manufacturing process, followed by a systematic assessment of the fundamental material parameters.

According to Aziz[14] and Aziz *et al.*,[18] in the case of equilibrium or a quickly equilibrated point defect, the effect of stress on the dopant diffusivity in the direction normal to a (001) surface can be written as

$$\frac{D_{33}(\sigma)}{D_{33}(0)}=\exp\left(\frac{\sigma[V^{f}+\tilde{V}_{33}^{m}]}{k_{B}T}\right). \tag{3}$$

Here the product of the stress tensor $\sigma$ and the formation strain tensor $V^f$ is the work done against the stress field in defect formation, the product of the stress tensor $\sigma$ and the migration strain tensor $\tilde{V}_{33}^{m}$ is the work required for the transition in the migration path, $k_B$ is Boltzmann's constant, and $T$ is the diffusion temperature. The tensor $V^f$ involves the creation or annihilation of a lattice site, followed by a relaxation process,[14,18]

$$V^{f}=\pm\Omega\begin{bmatrix}0&&\\&0&\\&&1\end{bmatrix}+\frac{V^{r}}{3}\begin{bmatrix}1&&\\&1&\\&&1\end{bmatrix}. \tag{4}$$

The + sign denotes vacancy formation and the − sign represents interstitial formation. The relaxation volume propagates elastically to all surfaces, resulting in a change in the volume of the crystal by an amount $V^r$. $\tilde{V}_{33}^{m}$ is expected to have the form[14,18]

$$\tilde{V}_{33}^{m}=\begin{bmatrix}\tilde{V}_{\perp}^{m}&&\\&\tilde{V}_{\perp}^{m}&\\&&\tilde{V}_{\parallel}^{m}\end{bmatrix}. \tag{5}$$

In Eq. (5), $\tilde{V}_{\perp}^{m}$ and $\tilde{V}_{\parallel}^{m}$, respectively, reflect the dimension changes perpendicular and parallel to the direction of the net transport when the point defect reaches its saddle point.[14,18] Aziz further defined the activation volume as the sum of the

[a]Electronic mail: chenmj@faculty.nctu.edu.tw

 © 2006 American Institute of Physics
**Downloaded 17 Oct 2006 to 140.113.11.175. Redistribution subject to AIP license or copyright, see http://apl.aip.org/apl/copyright.jsp**

three diagonal elements of the formation strain tensor and the migration strain tensor, as expressed by

$$\tilde{V} = \pm\Omega + V^r + 2\tilde{V}_\perp^m + \tilde{V}_\parallel^m. \tag{6}$$

It is well recognized[12] that when applying a uniaxial stress in a certain direction parallel to the silicon surface, the solid will modify its shape in order to minimize the energy of the system. In other words, the solid will deform in such a way that each surface perpendicular to the applied stress direction becomes stress-free. The underlying stress tensor therefore is

$$\sigma = \sigma_{\text{uniax}} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}. \tag{7}$$

On the basis of Hooke's law, $\sigma_{\text{uniax}}$ in the linear elastic regime can be related to the uniaxial strain $\varepsilon_{\text{uniax}}$ induced in the same direction: $\sigma_{\text{uniax}}=Y_{\text{uniax}}\varepsilon_{\text{uniax}}$, where the uniaxial modulus $Y_{\text{uniax}}=(C_{11}-2\nu C_{12})$ with Poisson's ratio $\nu=C_{12}/(C_{11}+C_{12})$. $C_{11}$ and $C_{12}$ are the elasticity constants. Analogous to previous work,[5] the uniaxial strain induced activation energy in the direction normal to the (001) surface, $Q'_{33-\text{uniax}}$, can be linked to the underlying diffusivity,

$$\frac{D_{33}(\varepsilon_{\text{uniax}})}{D_{33}(0)} = \exp\left(-\frac{Q'_{33-\text{uniax}}\varepsilon_{\text{uniax}}}{k_B T}\right). \tag{8}$$

By combining Eqs. (4), (5), and (7) and equalizing Eqs. (3)–(8), one obtains $Q'_{33-\text{uniax}}/Y_{\text{uniax}} = -V^r/3-\tilde{V}_\perp^m$. Again, by incorporating Eq. (6), the following expression is produced:

$$\tilde{V} + 3\frac{Q'_{33-\text{uniax}}}{Y_{\text{uniax}}} = \pm\Omega + (\tilde{V}_\parallel^m - \tilde{V}_\perp^m). \tag{9}$$

It is then a straightforward task to derive the uniaxial strain induced activation energy $Q'_{11-\text{uniax}}$ in the applied stress direction: $Q'_{11-\text{uniax}}/Y_{\text{uniax}}=-V^r/3-\tilde{V}_\parallel^m$. Consequently, a similar model is achieved,

$$\tilde{V} + 3\frac{Q'_{11-\text{uniax}}}{Y_{\text{uniax}}} = \pm\Omega - 2(\tilde{V}_\parallel^m - \tilde{V}_\perp^m). \tag{10}$$

Obviously, the uniaxial strain version is closely related to its biaxial counterpart: by comparing Eqs. (1) and (9), $Q'_{33-\text{uniax}}=(Y_{\text{uniax}}/2Y_{\text{biax}})Q'_{33-\text{biax}}$ is obtained. Another relation can then be readily derived: $Q'_{11-\text{uniax}} = -(Y_{\text{uniax}}/2Y_{\text{biax}})Q'_{33-\text{biax}}+(Y_{\text{uniax}}/Y_{\text{biax}})Q'_{11-\text{biax}}$.

To produce the experimental parameters in terms of the anisotropy of the uniaxial strain induced activation energy, a uniaxial stress experiment was carried out in terms of a *p*-channel MOSFET in a state-of-the-art manufacturing process.[17] The channel length was maintained at 65 nm while changing the spacing in the channel length direction between the two trench isolation sidewalls. The topside layout is detailed elsewhere.[17] Under such a situation, the channel zone encounters a compressive stress from the nearby trench isolation regions in the channel length direction. The devices used are quite wide (10 $\mu$m), meaning that the strain in the channel width direction is relatively negligible. The (001) silicon surface is supposed to be stress free. This hypothesis has been validated using the sophisticated simulations detailed in Ref. 17, which revealed that in the proximity of the silicon surface, the stress in the channel length

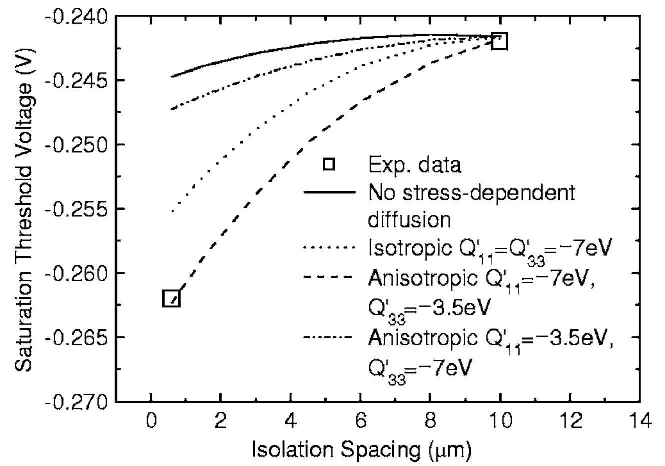



FIG. 1. Measured *p*-MOSFET saturation threshold voltage vs the spacing between the nearby trench isolation sidewalls in the channel length direction. Also shown are those (lines) from the process-device coupled simulation with and without the strain induced activation energies. The reason that the "no stress-dependent diffusion curve" is not entirely horizontal is due to dopant segregation near the edges of the source/drain regions. Specifically, the nonuniformity is caused by boron segregation occurring close to trench isolation oxide during the thermal process. Although the affected profile is not in vicinity of the MOSFET core region, a minor threshold voltage difference (~3 mV) between large and small active areas can still be observed, even without the stress-dependent diffusion model.

direction is much larger in magnitude than that in the direction normal to the surface. Therefore, the proposed physical model can be adequately applied. The effect of changing the spacing between the two trench isolation regions in the channel length direction is reflected in the measured saturation threshold voltage, as displayed in Fig. 1. The negative shift in the saturation threshold voltage with increasing stress (via decreasing spacing between the trench isolation regions) shown in Fig. 1 can be attributed to the retarded boron diffusion.

A two-dimensional process-device coupled simulation, as detailed in Ref. 17, was slightly modified by taking the anisotropy of the boron diffusivity into account,

$$\frac{D_{33}(\varepsilon_t)}{D_{33}(0)} = \exp\left(-\frac{Q'_{33-\text{TCAD}}\varepsilon_t}{k_B T}\right), \tag{11}$$

$$\frac{D_{11}(\varepsilon_t)}{D_{11}(0)} = \exp\left(-\frac{Q'_{11-\text{TCAD}}\varepsilon_t}{k_B T}\right). \tag{12}$$

According to the work in Ref. 17 the total strain $\varepsilon_t$ is the sum of the three strain components: $\varepsilon_{xx}$ in the channel length direction, $\varepsilon_{yy}$ in the channel width direction, and $\varepsilon_{zz}$ in the direction normal to the silicon surface. From the simulated strain distributions, $\varepsilon_t \sim \varepsilon_{xx}$, leading to $Q'_{33-\text{TCAD}} \approx Q'_{33-\text{uniax}}$ and $Q'_{11-\text{TCAD}} \approx Q'_{11-\text{uniax}}$. The simulated saturation threshold voltages for different values of $Q'_{33-\text{uniax}}$ and $Q'_{11-\text{uniax}}$ are plotted in Fig. 1 for comparison. The figure clearly exhibits that (i) the largest deviation occurs at $Q'_{33-\text{uniax}}=0$ and $Q'_{11-\text{uniax}}=0$, the case of no stress dependencies; (ii) the most accurate reproduction is achieved with the anisotropic activation energies, rather than the isotropic variety; and (iii) the anisotropy of the activation energy must be adequate, that is, $Q'_{11-\text{uniax}}=-7$ eV per unit strain and $Q'_{33-\text{uniax}}=-3.5$ eV per unit strain are more favorable than $Q'_{11-\text{uniax}}=-3.5$ eV per unit strain and $Q'_{33-\text{uniax}}=-7$ eV per unit strain.

Downloaded 17 Oct 2006 to 140.113.11.175. Redistribution subject to AIP license or copyright, see http://apl.aip.org/apl/copyright.jsp

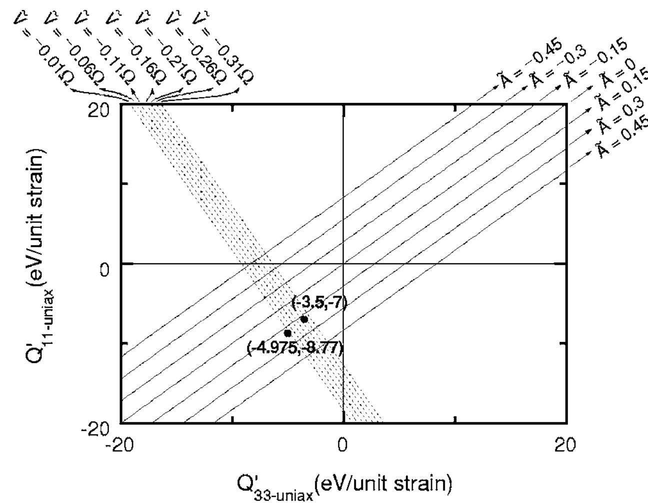



FIG. 2. Uniaxial strain induced activation energy in the applied stress direction (parallel to the silicon surface) vs that normal to the silicon surface. The lines are from Eqs. (9) and (10) for a literature range (Refs. 15, 16, and 18) of the activation volume and the migration strain anisotropy. Also plotted are the data points from the underlying experiment and the existing *ab initio* calculations (Refs. 12 and 13).

Prior to determining the fundamental material parameters, a systematic treatment, such as that indicated in Fig. 2, is demanded. In Fig. 2 a series of straight lines of $Q'_{11-\text{uniax}}$ vs $Q'_{33-\text{uniax}}$ are from Eqs. (9) and (10) for a literature range[15,16,18] of $\tilde{V}$ and the migration strain anisotropy $\tilde{A}$ $(\equiv(\tilde{V}^m_{\parallel}-\tilde{V}^m_{\perp})/\Omega)$.[18] In the calculation procedure, the following literature values were employed:[19] (i) $C_{11}=168$ GPa and $C_{12}=65$ GPa, giving rise to $Y_{\text{uniax}}=131$ GPa and $\nu=0.28$; (ii) $\Omega=2.26\times10^{-23}$ cm$^3$. The above experimental parameters are also added to the figure. From the figure a set of $\tilde{V}$ and $\tilde{A}$ can be clearly located around the data point. On the other hand, uncertainties exist based on a series of literature data: $\tilde{V}=-0.16\pm0.05$ Ω.[18] Taking such uncertainties into account, Fig. 2 reveals that the data point does match the upper limit, that is, $\tilde{V}=-0.21$ Ω. The corresponding $\tilde{V}^m_{\parallel}-\tilde{V}^m_{\perp}$ in the vicinity of 0.15 Ω is determined accordingly, falling within the reasonable range.[15,16,18] Such corroborating experimental evidence further indicates that the transient enhanced diffusion effect is relatively insignificant when compared to the long-term diffusion times in the underlying manufacturing process. Under such circumstances, the point defect is rapidly equilibrated relative to the entire diffusion time.

Finally, we quoted the existing *ab initio* calculations:[12,13] $Q'_{11-\text{biax}}=-19.2$ eV per unit strain and $Q'_{33-\text{biax}}=-13.9$ eV per unit strain, which were transformed via the aforementioned relationship into the equivalent $Q'_{11-\text{uniax}}$ of $-8.77$ eV per unit strain and $Q'_{33-\text{uniax}}$ of $-4.975$ eV per unit strain. In this process, the $Y_{\text{biax}}$ used was equal to 183 GPa according to $Y_{\text{biax}}=(C_{11}+C_{12}-\nu C_{12})$ with its Poisson's ratio $\nu=2C_{12}/C_{11}$. Evidently, the two data points are quite comparable to each other, as displayed in Fig. 2.

A physical model dealing with anisotropic diffusion in uniaxially stressed silicon is derived and is quantitatively connected to the biaxial case. A process-device coupled simulation is performed on a *p*-type MOSFET undergoing uniaxial stress during the manufacturing process. A systematic treatment is conducted and the resulting fundamental material parameters are in satisfactory agreement with literature values.

This research is supported by the National Science Council of Taiwan under Contract No. NSC94-2215-E-009-005.

[1]S. E. Thompson, M. Armstrong, C. Auth, M. Alavi, M. Buehler, R. Chau, S. Cea, T. Ghani, G. Glass, T. Hoffman, C. H. Jan, C. Kenyon, J. Klaus, K. Kuhn, Z. Ma, B. Mcintyre, K. Mistry, A. Murthy, B. Obradovic, R. Nagisetty, P. Nguyen, S. Sivakumar, R. Shaheed, L. Shifren, B. Tufts, S. Tyagi, M. Bohr, and Y. El-Mansy, IEEE Trans. Electron Devices **51**, 1790 (2004).
[2]C. H. Ge, C. C. Lin, C. H. Ko, C. C. Huang, Y. C. Huang, B. W. Chen, B. C. Perng, C. C. Sheu, P. Y. Tsai, L. G. Yao, C. L. Wu, T. L. Lee, C. J. Chen, C. T. Wang, S. C. Lin, Y. C. Yeo, and C. Hu, Tech. Dig. - Int. Electron Devices Meet. **2003**, 73.
[3]N. Moriya, L. C. Feldman, H. S. Luftman, C. A. King, J. Bevk, and B. Freer, Phys. Rev. Lett. **71**, 883 (1993).
[4]P. Kuo, J. L. Hoyt, J. F. Gibbons, J. E. Turner, R. D. Jacowitz, and T. I. Kamins, Appl. Phys. Lett. **62**, 612 (1993).
[5]N. E. B. Cowern, P. C. Zalm, P. van der Sluis, D. J. Gravesteijn, and W. B. de Boer, Phys. Rev. Lett. **72**, 2585 (1994).
[6]F. H. Baumann, J. H. Huang, J. A. Rentschler, T. Y. Chang, and A. Ourmazd, Phys. Rev. Lett. **73**, 448 (1994).
[7]P. Kuo, J. L. Hoyt, J. F. Gibbons, J. E. Turner, and D. Lefforge, Appl. Phys. Lett. **66**, 580 (1995).
[8]P. Kringhoj, A. Nylandsted Larsen, and S. Y. Shirayev, Phys. Rev. Lett. **76**, 3372 (1996).
[9]N. R. Zangenberg, J. Fage-Pedersen, J. Lundsgaard Hansen, and A. Nylandsted Larsen, J. Appl. Phys. **94**, 3883 (2003).
[10]M. S. Daw, W. Windl, N. N. Carlson, M. Laudon, and M. P. Masquelier, Phys. Rev. B **64**, 045205 (2001).
[11]M. Laudon, N. N. Carlson, M. P. Masquelier, M. S. Daw, and W. Windl, Appl. Phys. Lett. **78**, 201 (2001).
[12]M. Diebel, Ph.D. Dissertation, University of Washington, 2004.
[13]S. T. Dunham, M. Diebel, C. Ahn, and C. L. Shih, J. Vac. Sci. Technol. B **24**, 456 (2006).
[14]M. J. Aziz, Appl. Phys. Lett. **70**, 2810 (1997).
[15]Y. Zhao, M. J. Aziz, H. J. Gossmann, S. Mitha, and D. Schiferl, Appl. Phys. Lett. **74**, 31 (1999).
[16]Y. Zhao, M. J. Aziz, H. J. Gossmann, S. Mitha, and D. Schiferl, Appl. Phys. Lett. **75**, 941 (1999).
[17]Y. M. Sheu, S. J. Yang, C. C. Wang, C. S. Chang, L. P. Huang, T. Y. Huang, M. J. Chen, and C. H. Diaz, IEEE Trans. Electron Devices **52**, 30 (2005).
[18]M. J. Aziz, Y. Zhao, H. J. Gossmann, S. Mitha, S. P. Smith, and D. Schiferl, Phys. Rev. B **73**, 054101 (2006).
[19]H. F. Wolf, *Semiconductors* (Wiley, New York, 1971), p. 44.

Downloaded 17 Oct 2006 to 140.113.11.175. Redistribution subject to AIP license or copyright, see http://apl.aip.org/apl/copyright.jsp

APPLIED PHYSICS LETTERS **88**, 063511 (2006)

# Channel-width dependence of low-frequency noise in process tensile-strained *n*-channel metal-oxide-semiconductor transistors

Ming-Pei Lu
*Department of Electronics Engineering, National Chiao Tung University, Hsin-Chu, Taiwan, Republic of China*

Wen-Chin Lee
*Exploratory Technology Development-1 Department, Taiwan Semiconductor Manufacturing Company, Hsin-Chu, Taiwan, Republic of China*

Ming-Jer Chen[a)]
*Department of Electronics Engineering, National Chiao Tung University, Hsin-Chu, Taiwan, Republic of China*

(Received 13 October 2005; accepted 29 December 2005; published online 8 February 2006)

Low-frequency noise measurement in process tensile-strained *n*-channel metal-oxide-semiconductor field-effect transistors yields the density of the interface states, exhibiting a decreasing trend while decreasing the channel width. This finding corroborates the group of $P_b$ centers caused by the lattice mismatch at (100) Si-$SiO_2$ interface as the origin of the underlying interface states. The inverse narrow width effect appears to be insignificant, substantially confirming the validity of the noise measurement. The present noise experiment therefore points to the enhancement of the tensile strain in the presence of channel narrowing, which in turn reduces the lattice mismatch. © *2006 American Institute of Physics*. [DOI: 10.1063/1.2172287]

Channel strain engineering is currently recognized as an indispensable performance booster in producing next-generation metal-oxide-semiconductor field-effect transistors (MOSFETs).[1] To achieve this goal, two fundamentally different methods have been proposed:[1] (i) Strained silicon (SSi) on a relaxed SiGe buffer layer; and (ii) process strained silicon (PSS) through the trench isolation, silicide, and cap layer. On the other hand, in the areas of unstrained counterparts, low-frequency noise has been extensively utilized since it can provide the opportunity to examine the interfacial physics.[2–4] Thus, it is a challenging issue for the low-frequency noise measurement to find further potential applications in the strain case. Recently, one such study[5] has been demonstrated that an improved noise performance can be achieved on biaxial tensile-strained substrates. In the present work, we conduct a channel-width-dependent low-frequency noise experiment on a process tensile-strained *n*-channel MOSFET. The resulting noise data are useful in addressing the effect of enhanced tensile strain in the channel width direction.

The device under test was an *n*-channel MOSFET fabricated using the concept of process tensile strain, mainly through the trench isolation.[6] The physical gate oxide thickness was 1.4 nm as determined by capacitance-voltage fitting. The channel length was 0.5 $\mu$m while the channel width spanned a wide range of 0.11, 0.24, 0.6, 1, and 10 $\mu$m. Here, a reduction in channel width means an enhancement in tensile strain in the channel width direction. This can be easily understood by means of the current drive enhancement factor against channel width as shown in Fig. 1. The inset of the figure displays measured drain current per unit channel width versus drain voltage with the gate overdrive as a parameter. As expected, the drain current per unit channel width increases as the channel width is decreased. The increased drain current can be well related to the mobility enhancement; that is, the tensile stress causes subbands energy shift, which in turn suppresses the intervalley phonon scattering while reducing the effective conductivity mass, thereby enhances the mobility.[1,7]

The low-frequency noise measurement setup used was the same as that detailed elsewhere.[4] The measurement frequency ranged from 3 Hz to 100 kHz while operating the devices at a drain voltage of 0.2 V. Here, the noise experiment was carried out in terms of the input-referred noise voltage spectral density $S_{V_g}$. Figure 2 depicts measured $S_{V_g}$ versus frequency for a gate overdrive of 0.6 V, where three devices, as labeled A, B, and C with the same channel width

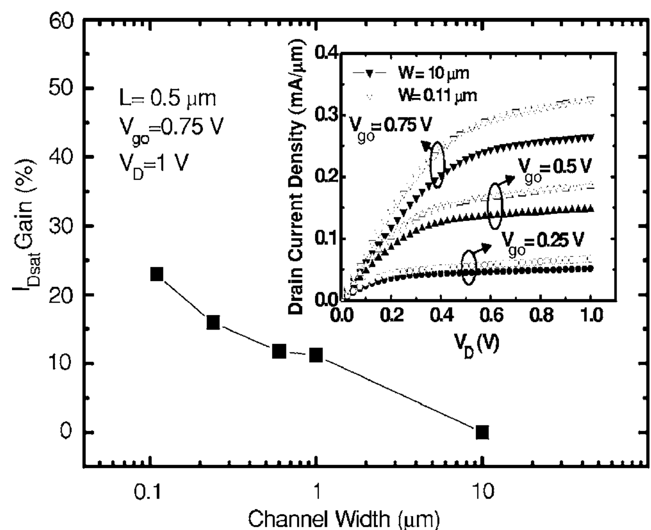



FIG. 1. The experimental drain current enhancement factor vs channel width for gate overdrive $V_{go}$=0.75 V and drain voltage $V_D$=1 V. The inset shows measured drain current per unit channel width vs drain voltage with gate overdrive as a parameter.

a)Electronic mail: chenmj@faculty.nctu.edu.tw

 © 2006 American Institute of Physics
Downloaded 09 Feb 2006 to 140.113.11.175. Redistribution subject to AIP license or copyright, see http://apl.aip.org/apl/copyright.jsp

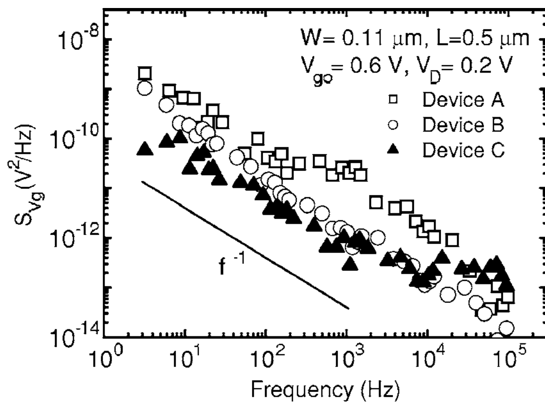



FIG. 2. The measured input-referred noise voltage spectral density at $V_{\mathrm{go}}=0.6$ V and $V_{\mathrm{D}}=0.2$ V vs frequency for three different positions on the wafer.

of 0.11 $\mu$m, represent three different positions on the wafer. Apparently, a considerable variation of low-frequency noise exists between devices, which can be attributed to statistical fluctuations of the number of the interface traps.[8] Hence, it is argued that the measured noise data essentially can follow the $1/f^{\gamma}$ relationship with the power coefficient $\gamma$ close to unity.

Figure 3 shows measured $S_{\mathrm{Vg}}$ against channel width for a certain gate overdrive of 0.6 V at a specific frequency of 100 Hz. In this figure, each of the five error bars represents the standard deviation of the distribution created from a total of fifteen samples; and the data point stands for the mean of the distribution. This figure reveals that the low-frequency noise, on the average, increases with decreasing channel width. We also further conducted the case of varying gate overdrive and found that no significant deviation from that in Fig. 3 can be observed, provided that the gate overdrive of more than 0.4 V is applied.

Also shown in Fig. 3 is the corresponding average threshold voltage shift with respect to the wide structure (i.e., 10 $\mu$m) such as to address the possibility of the inverse narrow width effect (INWE) caused by the impurity segregation

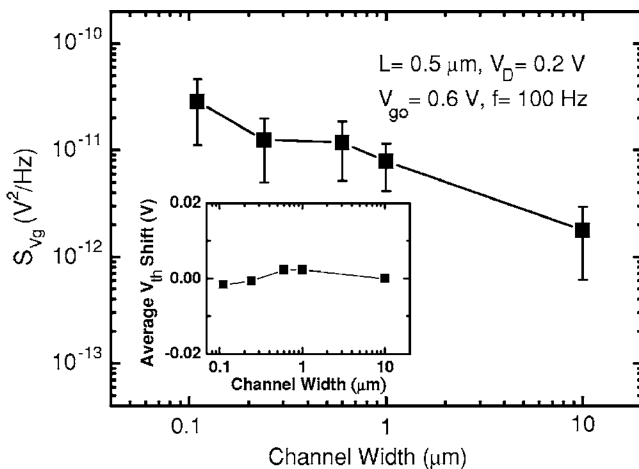



FIG. 3. The measured input-referred noise voltage spectral density at a specific frequency of 100 Hz vs channel width for $V_{\mathrm{go}}=0.6$ V and $V_{\mathrm{D}}=0.2$ V. The error bar represents the standard deviation of the distribution and the data point the mean of the distribution. The inset shows the corresponding average threshold voltage $V_{\mathrm{th}}$ shift with the wide structure (i.e., 10 $\mu$m) as a reference point. Here, the threshold voltage was determined at $V_{\mathrm{D}}=0.025$ V.

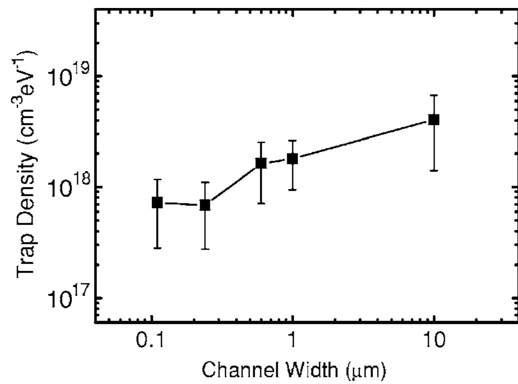



FIG. 4. The extracted effective interface state density corresponding to Fig. 3.

and the fringing electric field at the isolation sidewall.[9,10] It can be seen that the threshold voltage shift is rather small ($<5\%$), indicating that the inverse narrow width effect is not a significant issue in the undertaken devices. Furthermore, since the noise data are obtained in strong inversion, the channel current part along the edge of the channel is not significant relative to the overall channel one. As a result, the inverse narrow width effect can be substantially weakened. Therefore, it is reasonably drawn that the present low-frequency noise is a good tool to monitor the Si-$SiO_2$ interface over the whole channel area.

The weak dependence of low-frequency noise on gate overdrive ($>0.4$ V) as mentioned above suggests that the carrier number fluctuations prevail in the strong inversion mode. In other words, under such situations the Coulomb scattering can be ignored due to the screening of the trapped charge by the gate electrode and the inversion-layer charge. Hence, the following input-referred noise voltage spectral density expression can be adequately cited[3,11]

$$S_{\mathrm{Vg}}=\frac{q^2 k_B T\lambda}{C_{\mathrm{eff}}^{\;2}}\frac{N_t}{WL}\frac{\mathbf{1}}{f^{\gamma}}, \tag{1}$$

where $q$ is the elementary charge, $k_B$ is Boltzman's constant, $T$ is the absolute temperature, $\lambda$ is the tunneling distance ($\sim$0.1 nm), $W$ is the channel width, $L$ is the channel length, $C_{\mathrm{eff}}$ is the effective gate oxide capacitance per unit area, and $N_{\mathrm{t}}$ is the effective near-interface oxide trap density. With known $C_{\mathrm{eff}}(\approx 1.75\ \mu\mathrm{F/cm}^2)$ from the undertaken manufacturing process, fitting of all the $S_{\mathrm{Vg}}$ data using Eq. (1) led to the distribution of $N_t$ as shown in Fig. 4 versus channel width. Again on the average, the interface state density decreases with decreasing channel width. Specifically, a reduction in channel width by a factor of about 100 produces a tenfold reduction in interface state density. Once again, a reduction in channel width means an enhancement of tensile strain in that direction; therefore, the present noise experiment points to a reduction in interface state density in the presence of enhanced tensile strain in the channel width direction.

It is interesting to further examine the physical origin of the underlying interface states. Analogous to the electron-spin resonance (ESR) experiment on a (111) Si-$SiO_2$ interface,[12–14] the interface states investigated in our low-frequency noise work can be attributed to the group of $P_b$ centers (or equivalently the dangling-bond defects as characterized in terms of $Si_3\equiv Si\cdot$) caused by the lattice mismatch

Downloaded 09 Feb 2006 to 140.113.11.175. Redistribution subject to AIP license or copyright, see http://apl.aip.org/apl/copyright.jsp

at the (100) Si-$SiO_2$ interface. Here, the lattice-mismatch stress is primarily related to the thermal oxidation process since in this process about 2.2 unit volumes of oxide are produced for each unit volume of silicon consumption. Obviously, the lattice mismatch can be reduced using a tensile strain, thereby leading to a reduction in $P_b$ centers.[14] Note that there were few studies on the usage of the ESR technique to detect $P_b$ centers in the case of (100) Si-$SiO_2$ interface. The noise experiment on the (100) Si-$SiO_2$ interface in this work again corroborates the action of applying a tensile strain: *Enlarging the Si-Si interatomic distance before the silicon oxidation process is carried out, which leads to reduced lattice-mismatch stress during the subsequent thermal oxidation process.*

Channel-width-dependent low-frequency noise measurement has been applied on a process tensile strained *n*-channel MOSFET. One important finding has been straightforwardly created: *Enhanced tensile strain in the channel narrowing direction can reduce the lattice-mismatch defects.*

This work was supported by the National Science Council of Taiwan under Contract No. NSC 93-2215-E-009-002.

[1] S. E. Thompson, M. Armstrong, C. Auth, M. Alavi, M. Buehler, R. Chau, S. Cea, T. Ghani, G. Glass, T. Hoffman, C. H. Jan, C. Kenyon, J. Klaus, K. Kuhn, Z. Ma, B. Mcintyre, K. Mistry, A. Murthy, B. Obradovic, R. Nagisetty, P. Nguyen, S. Sivakumar, R. Shaheed, L. Shifren, B. Tufts, S. Tyagi, M. Bohr, and Y. El-Mansy, IEEE Trans. Electron Devices **51**, 1790 (2004).
[2] M. J. Kirton and M. J. Uren, Adv. Phys. **38**, 367 (1989).
[3] K. K. Hung, P. K. Ko, C. Hu, and Y. C. Cheng, IEEE Trans. Electron Devices **37**, 654 (1990).
[4] M. J. Chen, T. K. Kang, Y. H. Lee, C. H. Liu, Y. J. Chang, and K. Y. Fu, J. Appl. Phys. **89**, 648 (2001).
[5] E. Simoen, G. Eneman, P. Verheyen, R. Delhougne, R. Loo, K. De Meyer, and C. Claeys, Appl. Phys. Lett. **86**, 223509 (2005).
[6] C. H. Ge, C. C. Lin, C. H. Ko, C. C. Huang, Y. C. Huang, B. W. Chen, B. C. Perng, C. C. Sheu, P. Y. Tsai, L. G. Yao, C. L. Wu, T. L. Lee, C. J. Chen, C. T. Wang, S. C. Lin, Y. C. Yeo, and C. Hu, Tech. Dig. - Int. Electron Devices Meet. 73 (2003).
[7] S. Takagi, J. L. Hoyt, J. J. Welser, and J. F. Gibbons, J. Appl. Phys. **80**, 1567 (1996).
[8] M. J. Uren, D. J. Day, and M. J. Kirton, Appl. Phys. Lett. **47**, 1195 (1985).
[9] M. Sugino and L. A. Akers, IEEE Electron Device Lett. **4**, 114 (1983).
[10] S. K. H. Fung, M. Chen, and P. K. Ko, Solid-State Electron. **41**, 1885 (1997).
[11] H. M. Bu, Y. Shi, X. L. Yuan, J. Wu, S. L. Gu, Y. D. Zheng, H. Majima, H. Ishikuro, and T. Hiramoto, Appl. Phys. Lett. **76**, 3259 (2000).
[12] E. H. Poindexter, P. J. Caplan, B. E. Deal, and R. R. Razouk, J. Appl. Phys. **52**, 879 (1981).
[13] A. Stesmans, Phys. Rev. B **48**, 2418 (1993).
[14] A. Stesmans, D. Pierreux, R. J. Jaccodine, M. T. Lin, and T. J. Delph, Appl. Phys. Lett. **82**, 3038 (2003).

Downloaded 09 Feb 2006 to 140.113.11.175. Redistribution subject to AIP license or copyright, see http://apl.aip.org/apl/copyright.jsp

# Oxide-trap-enhanced Coulomb energy in a metal-oxide-semiconductor system

Ming-Pei Lu and Ming-Jer Chen*
*Department of Electronics Engineering, National Chiao-Tung University, Hsin-Chu, Taiwan*
(Received 15 April 2005; revised manuscript received 12 October 2005; published 16 December 2005)

Coulomb energy is essential to the charging of a nanometer-scale trap in the oxide of a metal-oxide-semiconductor system. Traditionally the Coulomb energy calculation was performed on the basis of an interfacelike trap. In this paper, we present experimental evidence from a 1.7-nm oxide: Substantial enhancements in Coulomb energy due to the existence of a deeper trap in the oxide. Other corroborating evidence is achieved on a multiphonon theory, which can adequately elucidate the measured capture and emission kinetics. The corresponding configuration coordinate diagrams are established. We further elaborate on the clarification of the Coulomb energy and differentiate it from that in memories containing nanocrystals or quantum dots in the oxide. Some critical issues encountered in the work are addressed as well.

DOI: 10.1103/PhysRevB.72.235417 PACS number(s): 72.20.Jv, 73.40.Qv, 73.50.Td, 73.61.Ng

## I. INTRODUCTION

In a metal-oxide-semiconductor (MOS) system, a Coulomb barrier arises during the charging of a nanometer-scale trap in the oxide. Thus, a critical energy to overcome the barrier, namely, Coulomb energy, plays a vital role in the capture kinetics.[1,2] Traditionally the Coulomb energy was calculated on the basis of an interfacelike trap. This treatment essentially remains valid if the oxide used is much thicker. However, with the currently aggressive downscaling of the oxide thickness, the oxide trap is likely situated deeper into the oxide from the $SiO_2/Si$ interface and therefore, the Coulomb energy is expected to be affected due to enhanced image charge. However, little work has been done in this direction since the introduction of the Coulomb energy concept.[1,2] On the other hand, it is noteworthy that the definition of the Coulomb energy in the case of the oxide trap[1,2] is significantly different from that in memories containing nanocrystals or quantum dots in the oxide.[3–6] However, such a confusing issue has not yet been clarified.

In this paper, we exhibit experimental evidence for the Coulomb energy enhancement in the presence of a deeper oxide trap. The other corroborating evidence is achieved based on a multiphonon theory with the configuration coordinate diagrams taken into account. We further elaborate on the clarification of the Coulomb energy in a MOS system containing a nanometer-scale trap in the oxide and differentiate it from that in a MOS memory containing a nanocrystal or dot in the oxide, followed by a concrete discussion on the critical issues encountered in the work.

## II. EXPERIMENT

The $n$-channel metal-oxide-semiconductor field-effect transistors (MOSFETs) with varying channel lengths and widths (60 nm to 600 nm) were fabricated in a state-of-the-art manufacturing process. The key process parameters as obtained by capacitance-voltage ($C-V$) fitting were $n^+$ polysilicon doping concentration $=1.3\times10^{20}$ cm$^{-3}$, gate oxide thickness$=1.7$ nm, and channel doping concentration $=8\times10^{17}$ cm$^{-3}$. To detect a potential oxide trap with fluctuating occupancy, the random telegraph signals (RTS) measurement is a good means.[1,2,7–9] The RTS measurement-equipment and method used were the same as that described elsewhere.[10] The operating conditions at room temperature were $V_D=10$ mV and with $V_G$ ranging from 0.2 to 0.4 V. The purpose of the low voltage operation is twofold: (i) it can ensure no extra trap created during the long-term RTS measurement; and (ii) the devices under study can readily reduce to a near-equilibrium one-dimensional(1D)MOS system. We conducted extensive RTS measurement across the whole wafer and found that as expected, the occurrence probability of RTS events in underlying devices is extremely low. For those devices identified with RTS, it was found that (i) the same abrupt transitions between two distinct states in drain current also simultaneously occur in source current; and (ii) no such noticeable changes can be observed in gate or bulk current, opposed to the recent literature[11] with a smaller oxide thickness ($\sim$1.3 nm). Therefore, the RTS events encountered in our work are due to the transfer of a single electron between a certain process-induced defect in the oxide and the underlying conductive channel layer. The capture time associated with the upper level of RTS current and the emission time associated with the lower current level both were exponentially distributed. The mean of the capture time distribution, designated $\tau_c$, divided by the mean of the emission time distribution $\tau_e$ is given in Fig. 1 against gate voltage for two devices labeled Traps $A$ and $B$. The inset of Fig. 1 shows the corresponding time evolutions of RTS drain current at a certain gate voltage. Figure 1 reveals that while initially the $\tau_c/\tau_e$ ratio is comparable between Traps $A$ and $B$, with gate voltage increasing further, the Trap $B$'s $\tau_c/\tau_e$ drops with a faster rate than Trap $A$.

## III. ANALYSIS AND PHYSICAL INTERPRETATIONS

The size of the trap under study must be significantly less than the oxide thickness used (1.7 nm) since no noticeable change in the gate current was observed. Hence, the trap responsible for the measured RTS in drain current is a nanometer-scale trap. To explore the measured $\tau_c/\tau_e$, it is necessary to know in advance the amount of the image or

 ©2005 The American Physical Society

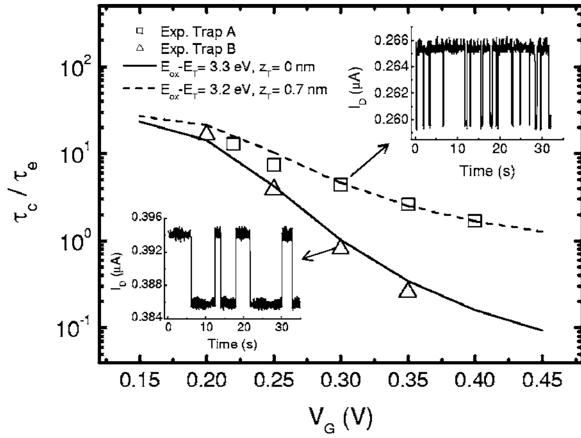



FIG. 1. Measured mean capture time to mean emission time ratio versus gate voltage for two devices labeled Traps *A* and *B*. The inset shows the time records of RTS drain current at a fixed gate voltage of 0.3 V. The fitting lines from Eq. (2) are also shown.

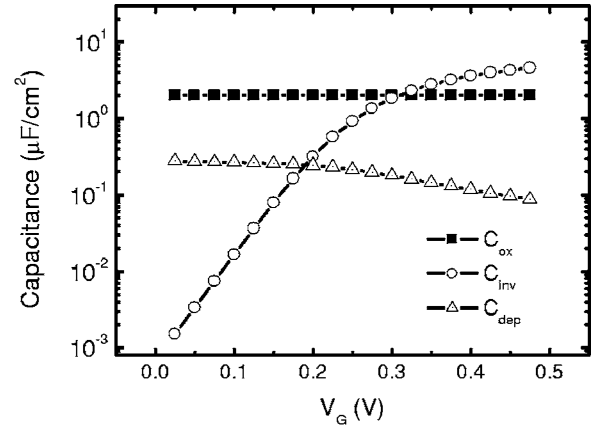



FIG. 3. Simulated results of the key capacitance components versus gate voltage.

induced charge on the gate as a single electron is inserted into the oxide trap. First of all, it is well recognized that once a single electron is inserted into the oxide, the Debye screening length of a single electron (~70 nm) (Refs. 4 and 6) develops laterally around a *negatively charged* nanometer-scale trap in the oxide. Here, the Debye screening length is the effective size of the "cloud" of the induced charges on the electrodes. Thus, only within the Debye screening length can the plate capacitor approximation readily apply, leading to a capacitive coupling equivalent circuit as shown in Fig. 2. The capacitance model accounts for the effect of the trap depth and the charge sharing between gate, inversion layer, and silicon depletion region. Owing to the insertion of one electron into a depth $z_T$ from the $SiO_2/Si$ interface, the gate oxide capacitance per unit area $C_{ox}$ associated with the oxide thickness $t_{ox}$ can be separated into two distinct components: the trap to anode (near the gate) capacitance per unit area $C_g = C_{ox}t_{ox}/(t_{ox} - z_T)$ and the trap to cathode (near the channel) capacitance per unit area $C_c = C_{ox}t_{ox}/z_T$. The other capacitances such as the inversion-layer capacitance per unit area $C_{inv}$ and the silicon depletion capacitance per unit area $C_{dep}$ can be quantified using a self-consistent Schrödinger-Poisson equations solver with the process parameters mentioned above as input. Figure 3 shows the simulated results of the key capacitance components versus gate voltage. The proposed capacitance model exactly reduces to that by Schulz[1] for the case of $z_T = 0$. Indeed, the calculated results on a 17-nm oxide are consistent with those in the literature.[2]

While a single electron is inserted into the trap, the potential change $\Delta V$ in the trap reads as $\Delta V = q/(A_{DB} \times C_{eff})$ where $A_{DB}$ is the effective Debye screening area and $C_{eff}$, the equivalent capacitance per unit area seen from the trap to the ground, can be derived from the model. Then the image charge (positive) $Q_G$ developed on the gate electrode can be expressed as $Q_G = \Delta V \times (A_{DB} \times C_g)$. Combining both equations while eliminating the common factor (i.e., Debye screening area), one achieves $Q_G (= qC_g/C_{eff})$

$$Q_G = q \times \frac{z_T \times (C_{inv} + C_{dep}) + C_{ox}t_{ox}}{t_{ox}C_{ox} + t_{ox}(C_{inv} + C_{dep})}. \quad (1)$$

The calculated gate image charge as depicted in Fig. 4 remains constant until a 2DEG (2D electron gas) layer critically appears (at $V_G \approx 0.1$ V), and then due to increasing screening by the inversion-layer charge, the gate image charge decreases with increasing gate voltage. Specifically, the figure reveals that an increase in the trap depth can substantially increase the gate image charge. In the presence of a 2DEG layer, the source and drain are electrically tied to-

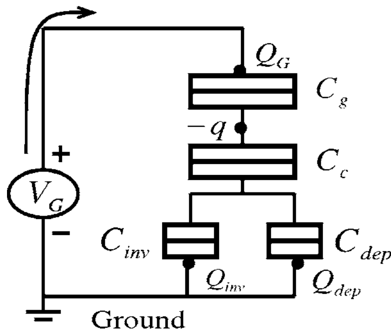



FIG. 2. Capacitive coupling equivalent circuit, accounting for the effect of the trap depth and the charge sharing between gate, inversion layer, and silicon depletion region.

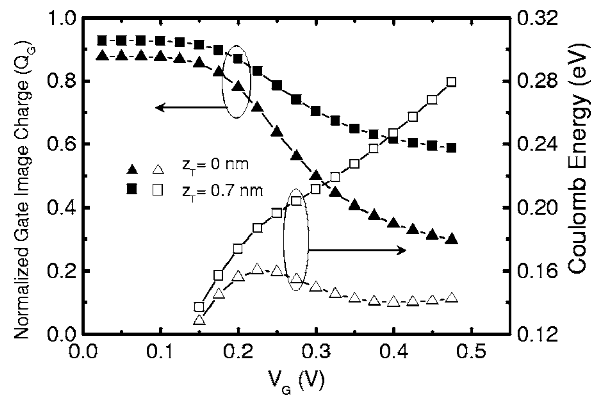



FIG. 4. Calculated gate image charge and Coulomb energy versus gate voltage for two trap depths in the oxide.

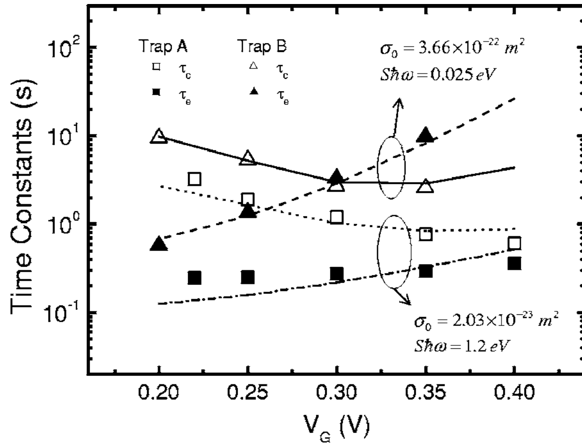



FIG. 5. Comparison of the measured and calculated capture time constants and emission time constants versus gate voltage.

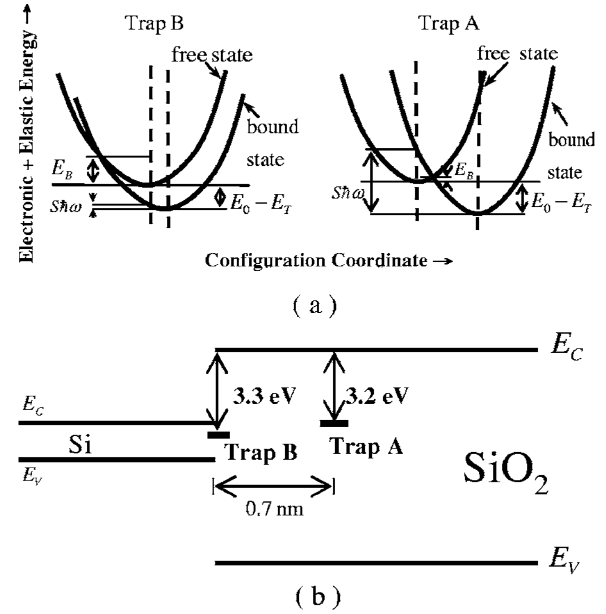



FIG. 6. Schematic configuration coordinate diagrams used for a phenomenological description of the capture and emission kinetics encountered in Traps $A$ and $B$. The corresponding energy band diagrams in flatband conditions are also given, schematically showing the trap depth and its energetic level in the oxide.

gether and thereby the Coulomb energy can readily be written as $\Delta E \approx Q_G V_G$.[1,2] The calculated Coulomb energy is together plotted in Fig. 4, showing that the Coulomb energy associated with the interface trap increases with gate voltage until encountering a certain peak. However, such a peak point disappears in the case of nonzero trap depth and the Coulomb energy instead piles up over the conventional value.

According to the principle of detailed balance with the Coulomb energy included, the $\tau_c/\tau_e$ ratio can read as[1]

$$\frac{\tau_c}{\tau_e} = e^{(E_T - E_F + \Delta E)/k_B T}. \tag{2}$$

In Eq. (2), the trap level $E_T$ relative to the quasi-Fermi level $E_F$ is a function of gate voltage and can readily be quantified using the Schrödinger-Poisson solver. The best fitting results achieved using Eq. (2), with $z_T = 0.7$ nm and $E_{OX} - E_T = 3.2$ eV for Trap A and $z_T = 0$ nm and $E_{OX} - E_T = 3.3$ eV for Trap $B$, are shown in Fig. 1. Here $E_{OX}$ denotes the oxide conduction band edge. Evidently, the fitting quality is fairly good. The extracted $E_{OX} - E_T$ values are close to the $SiO_2/Si$ interface barrier height, as expected due to the low voltage operation. It is hence argued that an interface trap exists in the Trap $B$ device while a 0.7-nm deep trap in the oxide prevails in Trap $A$. In other words, the conventional Coulomb energy appears to work well for the Trap $B$ device but leads to poor quality in fitting the Trap $A$ data. Such a remarkable difference in $\tau_c/\tau_e$ between Traps $A$ and $B$ can therefore serve as experimental evidence of the Coulomb energy enhancement.

Other corroborating evidence can be obtained through the fitting of the measured mean capture time versus gate voltage as shown in Fig. 5. Since the capture kinetics involve the thermal activation process at room temperature of operation, a multiphonon emission theory was utilized to calculate the capture time

$$\frac{1}{\tau_c} = \sigma \nu_{th} \frac{n_s}{z_{qm}} e^{-\Delta E/k_B T} \tag{3}$$

where $\nu_{th}$ is the carrier thermal velocity ($\approx 1.23 \times 10^5$ m/s), $n_s$ is the inversion-layer electron density per unit area, and $z_{qm}$ is the average thickness of the inversion layer. $\sigma$ is the multiphonon capture cross section and can be written as

$$\sigma = \sigma_0 e^{-E_B/(k_B \times T)}. \tag{4}$$

The prefactor $\sigma_0$ involves the interaction between the trap state and free electron wave function.$E_B$ is the thermal activation barrier height and according to multiphonon emission theory the thermal activation barrier height at high temperature ($k_B T > \hbar\omega/2$) can reduce to[12,13]

$$E_B = \frac{(E_0 - E_T - S\hbar\omega)^2}{4S\hbar\omega}, \tag{5}$$

where $E_0$ is the energy level of the lowest subband for unprimed valley and $S\hbar\omega$ is the lattice relaxation energy ($S$ is the Huang-Rhys factor). Fitting the $\tau_c$ data in Fig. 5 to Eq. (3) yielded the lattice relaxation energy $S\hbar\omega$ of 1.2 and 0.025 eV for Traps $A$ and $B$, respectively; and $\sigma_0$ of $2.03 \times 10^{-23}$ and $3.66 \times 10^{-22}$ m$^2$ for Traps $A$ and $B$, respectively. The fitting quality is again good and the same parameters readily reproduced the $\tau_e$ data as depicted in Fig. 5. Specifically, the extracted $\sigma_0$ values are physically reasonable from the viewpoint of the penetration of the wave function into the oxide: the capture cross section decreases with increasing trap depth from the $SiO_2/Si$ interface. The extracted values of the lattice relaxation energy also correctly reflect the status of the trap: A deeper trap (i.e., Trap A in our work) is accompanied with a higher lattice relaxation energy.[14,15] Using the above extracted results, we constructed a configuration coordinate diagram of the underlying electron-lattice system as schematically shown in Fig. 6 for both devices. Also plotted in Fig. 6 are the MOS energy band diagrams (removing the polysilicon part) in the flatband case, showing the spatial

distance and energetic level of the trap. The calculation results show that the thermal activation barrier $E_B$ of Trap *A* is substantially smaller than Trap *B*, as is clearly indicated in Fig. 6.

## IV. FURTHER CONSIDERATIONS

### A. On the definition of Coulomb energy in trap case

Good reproduction of the measured time constants over gate voltage range, such as those in Figs. 1 and 5, is essential and crucial in the areas of MOSFET RTS. This means that the Coulomb energy involved must quantitatively follow that in Fig. 4. The corresponding Coulomb energy lies between 120 and 280 meV, comparable with that (250 meV) in the similar RTS measurements by Schulz.[1]

As a single electron is inserted into the oxide trap, the total energy of the MOS system will change. The change in energy of the system can be divided into two parts: one is the storage energy and the other is the work done by the voltage source. The change in the storage energy term is

$$\Delta E_S = \frac{q^2}{2 \times A_{DB} \times C_{eff}}. \qquad (6)$$

$\Delta E_S$ was calculated to have a value of around 1 meV for the Debye screening length of 70 nm, which is negligibly small in magnitude. This means that the Coulomb energy in terms of the work ($\approx Q_G V_G$) done by an external voltage source dominates. Therefore, the definition of $\Delta E \approx Q_G V_G$ as adopted in the areas of MOSFET RTS[1,2] is valid.

### B. On the nanocrystals case

There are several fundamental differences between a MOS system with a nanometer-scale trap in the oxide and a MOS system with a nanocrystal or dot in the oxide. First, the self-capacitance of a nanocrystal dot in the oxide can be well linked to the actual dot diameter (this promises applications as a nanoscale floating gate) whereas from the MOS electrostatics point of view, it is the Debye screening length prevailing in the trap case. Second, in our RTS measurement the gate voltage was fixed such as to ensure a quasiequilibrium MOS system; and different gate voltages under such quasiequilibrium conditions produced different RTS data. However, during typical Coulomb blockade experiments on nanocrystalline memories, the gate voltage must continuously change in order to produce a series of Coulomb blockade events. Third, once captured, the electrons essentially remain in the dots (unless a potential leakage is present or the retention time is exceeded); however, this is not the case for the oxide trap, as evidenced by the fluctuating occupancy.

The experimentally determined Coulomb energy in the nanocrystalline dots memories[3–6] ranged from 46 to 168 meV. However, the definition of the Coulomb energy is significantly different from that in Refs. 1 and 2. Instead, an alternative treatment on the basis of the Coulomb blockade theory was widely adopted in the areas of nanocrystalline dots memories. For example, the product of the gate voltage shift between two subsequent Coulomb blockade events and the gate-to-dot coupling coefficient can be directly connected to the critical energy required to overcome the barrier due to the single electron storage energy and the quantum confinement induced energy separation. The single electron storage energy is defined as the Coulomb energy $\Delta E \approx q^2/2C_{dot}$ where $C_{dot}$ is the self-capacitance of the dot. Obviously, different situations encountered can lead to different definitions on the Coulomb energy.

## V. CRITICAL ISSUES

### A. Screening length

Due to the usage of a heavily doped $n^+$ polysilicon gate, one may consider the Thomas-Fermi screening length instead as employed in the metal case. However, a self-consistent Schrödinger-Poisson solving over the range of gate voltage under study reveals a band bending across a polydepletion region near the oxide. The corresponding electron density at the interface is found to be about one order of magnitude less than the immobile positively charged impurity concentration. Hence, in the presence of the polydepletion in our work, the Debye-Hueckel screening length considerably applies, which should be much larger than the Thomas-Fermi screening length (of the order of 1 nm) in the metal gate case. To further support this argument, from the measured RTS relative amplitude at $Vg=0.2$ V, we estimate the amount of the affected area to be *at least* 28 and 35 nm across the charged trap for $z_T=0.7$ nm and $z_T=0$ nm, respectively. Thus, the cited 70 nm for the Debye screening length remains reasonable. Even the replacement with a lower value of 28 or 35 nm causes little error.

### B. Silicon depletion charge

The Coulomb energy also includes the contribution by the charge induced at the edge of the semiconductor depletion region. The corresponding amount of energy is the product of the induced charge at the edge of the semiconductor depletion region times the difference ($\sim$0.07 eV) between Fermi level and valence band edge at the bulk part of the substrate. The depletion image charge at $Vg=0.2$ V is found to be $0.1e$ and $0.07e$, respectively, for $z_T=0$ and 0.7 nm, and each decreases with increasing gate voltage. As a result, the Coulomb energy due to the depletion image charge becomes of the order of a few milli-electron-volts and drops with increasing gate voltage. Obviously, the role of the charge induced at the edge of the semiconductor depletion region is so insignificant that the depletion image charge can be neglected in the present work.

### C. Electron tunneling

First of all, a deeper oxide trap may not always dictate a longer time. According to the configuration coordinate diagrams that describe the electron-lattice coupling, our data point to the opposite case: A deeper oxide trap produces a smaller time constant. This is reasonable since all the extracted parameters can find their physical origins as detailed above. If the electron tunneling were involved only, then the capture time would be the sum of the tunneling time from the

channel conduction band edge to certain oxide depth $z_T$ plus the subsequent multiphonon emission time such as to lower the energy of the tunneling electrons to the same level as the trap. One can estimate the tunneling time of around $10^{-9}$ sec across $z_T$ of 0.7 nm (Ref. 16) and can reasonably hypothesize that the multiphonon emission time is a spontaneous event (as can be easily understood from the configuration coordinate diagrams in Fig. 6; the hypothesis also works well for the areas of the trap assisted tunneling), leading to a capture time of the order of $10^{-9}$ sec. Obviously, the possibility of the electron tunneling must in principle be removed since the measured capture times fall within 0.5 to 6 sec. On the other hand, once trapped the electrons may instantly tunnel to the gate electrode, contributing to the gate current. In other words, under such situations, no RTS in drain or gate current can be detected due to the extremely slow detection process in measurement setup. Moreover, in our work the gate current was found to be several orders of magnitude less than the drain current, indicating the absence of the electron tunneling in determining the experimental RTS drain current.

Note that the high and low levels of RTS current represent the different stable states as denoted the free and bound state in the configuration coordinate diagrams in Fig. 6. The detailed balance essentially applies only to two such states, rather than the abrupt transitions between the two. The capture and emission time constants represent the critical times required to overcome the barrier height and reach the crossing point, then instantly entering into the other stable state.

Eventually, the measured discrete switching RTS drain current indicates that the transit time between the high and low levels is substantially less than the integration time in measurement setup.[10] In other words, the abrupt transition between two stable states represents a spontaneous event with respect to the measurement setup. Hence, the corresponding transient displacement current through the gate electrode may escape detection. This explains why we saw only a flat gate current level (with typical thermal or shot fluctuations around it) over the whole observation time.

## VI. CONCLUSION

We have presented experimental evidence concerning the Coulomb energy enhancement in a MOS system with a nanometer-scale oxide trap. Other corroborating evidence based on a multiphonon theory has elucidated the measured capture and emission kinetics. The corresponding configuration coordinate diagrams have been established. We have further elaborated on the clarification of the Coulomb energy and have differentiated it from that in memories containing nanocrystals as a floating gate. Some critical issues encountered in the work have been addressed as well.

## ACKNOWLEDGMENT

The authors would like to thank Professor Supriyo Datta and Professor Mark Lundstrom of Purdue University for access to the self-consistent Schrödinger-Poisson simulation program. This work was supported by the National Science Council of Taiwan under Contract No. NSC93-2215-E-009-002.

*Electronic address: chenmj@faculty.nctu.edu.tw

[1] M. Schulz, J. Appl. Phys. **74**, 2649 (1993).
[2] H. H. Mueller, D. Wörle, and M. Schulz, J. Appl. Phys. **75**, 2970 (1994).
[3] S. Tiwari, F. Rana, H. Hanafi, A. Hartstein, E. F. Crabbe, and K. Chan, Appl. Phys. Lett. **68**, 1377 (1996).
[4] L. Guo, E. Leobandung, and S. Y. Chou, Science **275**, 649 (1997).
[5] M. Saitoh, N. Takahashi, H. Ishikuro, and T. Hiramoto, Jpn. J. Appl. Phys., Part 1 **40**, 2010 (2001).
[6] S. Huang, S. Banerjee, R. T. Tung, and S. Oda, J. Appl. Phys. **93**, 576 (2003).
[7] K. S. Ralls, W. J. Skocpol, L. D. Jackel, R. E. Howard, L. A. Fetter, R. W. Epworth, and D. M. Tennant, Phys. Rev. Lett. **52**, 228 (1984).
[8] K. R. Farmer, C. T. Rogers, and R. A. Buhrman, Phys. Rev. Lett. **58**, 2255 (1987).
[9] M. J. Kirton and M. J. Uren, Adv. Phys. **38**, 367(1989).
[10] M. J. Chen and M. P. Lu, Appl. Phys. Lett. **81**, 3488 (2002).
[11] A. Avellan, W. Krautschneider, and S. Schwantes, Appl. Phys. Lett. **78**, 2790 (2001).
[12] D. V. Langin, in *Deep Centers in Semiconductors*, edited by S. T. Pantelides (Gordon and Breach, Yverdon, 1992).
[13] B. K. Ridley, *Quantum Processes in Semiconductors*, 3rd ed. (Clarendon, Oxford, 1993).
[14] A. Palma, A. Godoy, J. A. Jiménez-Tejada, J. E. Carceller, and J. A. López-Villanueva, Phys. Rev. B **56**, 9565 (1997).
[15] W. B. Fowler, J. K. Rudra, M. E. Zvanut, and F. J. Feigl, Phys. Rev. B **41**, 8313 (1990).
[16] I. Lundström and C. Svensson, J. Appl. Phys. **43**, 5045 (1972).

Japanese Journal of Applied Physics
Vol. 45, No. 32, 2006, pp. L849–L851
©2006 The Japan Society of Applied Physics

# Reproducing Subthreshold Characteristics of Metal–Oxide–Semiconductor Field Effect Transistors under Shallow Trench Isolation Mechanical Stress Using a Stress-Dependent Diffusion Model

Yi-Ming SHEU[1,3], Sheng-Jier YANG[2], Chih-Chiang WANG[1], Chih-Sheng CHANG[2], Ming-Jer CHEN[3], Sally LIU[1] and Carlos H. DIAZ[2]

[1]*Device Engineering Division, Taiwan Semiconductor Manufacturing Company, 8 Li-Hsin Sixth Road, Science-Based Industrial Park, Hsin-Chu 300, Taiwan, R.O.C.*
[2]*Logic Technology Division, Taiwan Semiconductor Manufacturing Company, 8 Li-Hsin Sixth Road, Science-Based Industrial Park, Hsin-Chu 300, Taiwan, R.O.C.*
[3]*Department of Electronics Engineering, National Chiao-Tung University, Hsin-Chu 300, Taiwan, R.O.C.*

(Received June 27, 2006; accepted July 13, 2006; published online August 11, 2006)

N-channel metal–oxide–semiconductor field effect transistors (MOSFETs) with a lightly doped well exhibit subthreshold current versus voltage ($I$–$V$) characteristics that are sensitive to shallow trench isolation (STI) mechanical stress. Such striking dependencies offer the opportunity to validate a proposed two-dimensional (2D) process model that relates the impurity diffusion to the mechanical stress throughout the substrate. With the assistance of sophisticated process/device simulations, the model appears to satisfactorily reproduce subthreshold $I$–$V$ characteristics for different active area lengths and different substrate biases. The stress-dependent point defect equilibrium concentration and diffusion model are also implemented to evaluate the stress effect on transient enhanced diffusion. [DOI: 10.1143/JJAP.45.L849]

KEYWORDS: dopant diffusion, mechanical stress, strain, shallow trench isolation, MOSFET, modeling, simulation

Shallow trench isolation (STI) induced mechanical stress in scaled metal–oxide–semiconductor field effect transistors (MOSFETs) is currently an issue of concern[1–6] because the stress magnitude increases tremendously with the reduction in the device active area. Experimental work and numerical simulations have been conducted to examine the impact of STI mechanical stress on device performance.[3–5] Impurity diffusivity difference caused by mechanical stress, as one plausible origin for the threshold voltage shift, was first investigated by Scott *et al.*[6] Thus, it is essential to quantitatively deal with the mechanical stress-dependent dopant diffusion while evaluating scaled MOSFET performance under STI mechanical stress. In addition, a great deal of work has been devoted to addressing dopant diffusion behavior in silicon under the influence of mechanical stress,[7–10] leading to the mechanical strain-dependent impurity diffusion models in an Arrhenius form. Recently, a modified Arrhenius-type diffusion model that accounts for strain-dependent diffusion has been proposed[11] for the purpose of the numerical modeling.

This work has been conducted to corroborate the validity of the STI mechanical stress-dependent diffusion model[11] in a MOSFET device with an underlying lightly doped well, which exhibits a significant mechanical stress effect on the subthreshold current versus voltage ($I$–$V$) characteristics. The stress-dependent point defect equilibrium concentration and diffusion, which dominates the transient enhanced diffusion (TED), has also been taken into account in this study.

The dopant diffusion change due to mechanical stress has been derived from point defect (interstitials and vacancies) changes.[9] Mechanical strain influences the point defect formation and migration, while the microscopic volume change and the pressure both contribute to the Gibb's free energy change. Thus, the dopant diffusivity ratio with and without strain can be expressed in an Arrhenius form; and the strain-induced point defect energy change can be translated to the dopant diffusivity change. To deal with the general non-uniform mechanical stress-dependent diffusion, we developed a volume-change-ratio induced diffusion activation energy shift model (VIDAESM):[11]

$$D_{\mathrm{S}}(T,x,y) = D_{\mathrm{A}}(T)\exp\left[-\frac{\Delta E_{\mathrm{S}} V_{\mathrm{cr}}(T,x,y)}{kT}\right] \quad (1)$$

where $D_{\mathrm{S}}$ is the dopant diffusivity under strain, $D_{\mathrm{A}}$ is the dopant diffusivity without strain, $T$ is the temperature, $V_{\mathrm{cr}}$ is the volume change ratio due to stress and is a function of position, and $\Delta E_{\mathrm{S}}$ is the activation energy per volume change ratio depending on the dopant species. When the strain is small in magnitude, the volume change ratio can be expressed as

$$V_{\mathrm{cr}}(T,x,y) \cong \varepsilon_{\mathrm{t}}(T,x,y) \equiv \varepsilon_{xx}(T,x,y) + \varepsilon_{yy}(T,x,y) + \varepsilon_{zz} \quad (2)$$

where $\varepsilon_{xx}$ is the strain along the channel length direction, $\varepsilon_{yy}$ is the strain in the direction perpendicular to the silicon surface, $\varepsilon_{zz}$ is the strain along the channel width direction, and $\varepsilon_{\mathrm{t}}$, total strain, is the summation of $\varepsilon_{xx}$, $\varepsilon_{yy}$, and $\varepsilon_{zz}$. $\varepsilon_{zz}$ is zero in this work as a result of the wide structures adopted. Therefore, eq. (1) becomes

$$D_{\mathrm{S}}(T,x,y) = D_{\mathrm{A}}(T)\exp\left[-\frac{\Delta E_{\mathrm{S}}\varepsilon_{\mathrm{t}}(T,x,y)}{kT}\right] \quad (3)$$

The equation is consistent with experimental data[7,9] and theoretical calculations[12] showing that a linear dependence of the mechanical strain to dopant diffusivity ratios in a log scale. Recently, Diebel[12] studied stress-dependent point defect equilibrium concentration and diffusion by means of *ab initio* calculations. The results also show a linear dependence of both point defect equilibrium concentration and diffusivity in a log scale on the mechanical strain. Thus, it is reasonable to express the point defect equilibrium concentration and diffusion in an Arrhenius form.

$$C^*_{\mathrm{S}}(T,x,y) = C^*_{\mathrm{A}}(T)\exp\left[-\frac{\Delta E_{\mathrm{C}}\varepsilon_{\mathrm{t}}(T,x,y)}{kT}\right] \quad (4)$$

where $C^*_{\mathrm{S}}$ is the point defect equilibrium concentration under

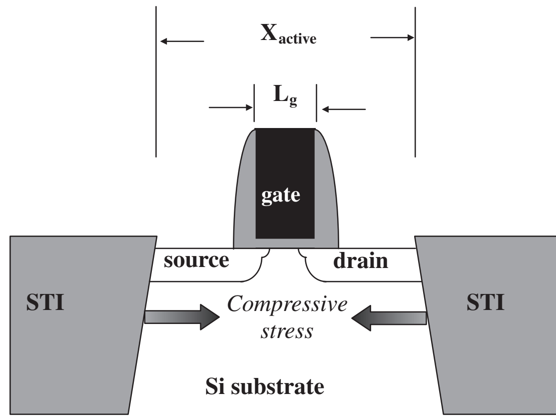



Fig. 1. Schematic cross-section of the device under study along the channel length direction.

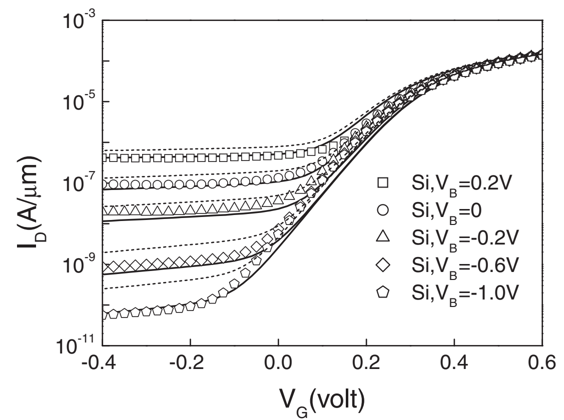



Fig. 2. Experimental and simulated $I_D$–$V_G$ curves for $X_{\text{active}} = 1.46\,\mu\text{m}$. Symbols represent experimental data, dashed lines are the simulation results without considering stress-dependent diffusion models, and solid lines are the final simulated results including stress-dependent diffusion models.

strain. To investigate the transient enhanced diffusion, only strain-dependent interstitial diffusion is needed. $\Delta E_C$ for the vacancy extracted from calculation results[12] is $+7.9$ eV/unit strain. Extracted $\Delta E_C$ and $\Delta E_S$ values for the interstitial are $-7.0$ and $+0.99$ eV/unit strain, respectively. Furthermore, interstitial diffusivity and equilibrium concentration product, $D_I C_I^*$, is reduced under the compressive strain conditions. Two-dimensional (2D) process/device simulators, TSUPREM4 and MEDICI, were employed. The stress-dependent diffusion models were incorporated into TSUPREM4 through its user-specified equation interface.

A series of n-channel MOSFETs were fabricated using state-of-the-art process technology. Test structures had three active area length ($X_{\text{active}}$ in Fig. 1) values: 0.68, 1.46, and 20.2 μm. $X_{\text{active}}$ is the design parameter to modulate mechanical stress. The minimum $X_{\text{active}}$ dimension of 0.68 μm takes the presence of one contact window area in the source/drain into account. The gate length and width were 0.17 and 10 μm, respectively. The retrograde well implantations are omitted so as to enhance the sensitivity of the subthreshold characteristics to STI mechanical stress, offering the opportunity to verify the validity of the above mentioned diffusion model. The measured subthreshold $I$–$V$ characteristics at $V_D = 1.2$ V are depicted in Fig. 2, with the substrate bias as a parameter. Previous work[13] revealed that the substrate bias measurement is a suitable verification index of the MOSFET doping profile because of the high sensitivity of carrier diffusion current to the dopant profile. The procedure for obtaining $\Delta E_S$ values for various impurities began by calibrating the one-dimensional dopant profiles in blanket control wafers, using processes that covered the range of device wafer process conditions. The results were taken as stress-free dopant profiles and used to calibrate the dopant diffusion parameters without considering stress-dependent diffusion effect. 2D MOSFET structures were then simulated in conjunction with the mechanical stress model. Calibrated diffusion parameters were employed to simulate a large $X_{\text{active}}$ case, where the stress level is negligible. All major front-end process steps from the STI to the source/drain anneal were considered. The corresponding simulation geometries were calibrated using TEM cross-sectional images. Device simulations were performed and the device model parameters were calibrated to fit the $I$–$V$ of the large $X_{\text{active}}$ MOSFET. Next, the process simulations based on VIDAESM for all $X_{\text{active}}$ values were conducted with an initial set of $\Delta E_S$ values. Then, the device simulations with smaller $X_{\text{active}}$ values were performed and compared with the $I$–$V$ data. The above procedure was iterated until a satisfactory reproduction of subthreshold $I$–$V$ data was achieved in all cases.

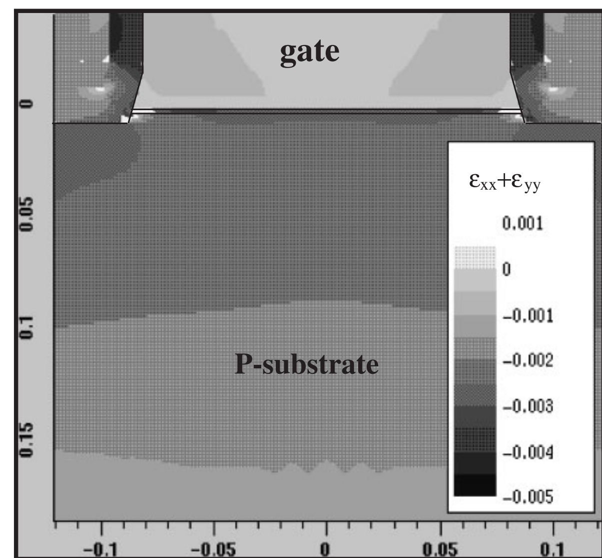



Fig. 3. Simulated strain distribution in the silicon after entire front-end process for $X_{\text{active}} = 0.68\,\mu\text{m}$. The total strain is in the MOSFET core region are compressive due to thermal gate oxidation and thermal mismatch between STI oxide and silicon.

The simulated strain distribution results for $X_{\text{active}} = 0.68$ μm are given in Fig. 3. It can be seen that the magnitude of the total strain, or volume change ratio, $\varepsilon_{xx} + \varepsilon_{yy}$, is negative in the MOSFET core area, meaning that the device experiences compressive stresses during the process. In addition, the $\varepsilon_{xx} + \varepsilon_{yy}$ of $X_{\text{active}} = 0.68\,\mu\text{m}$ was found to be much larger in magnitude than $X_{\text{active}} = 20.2\,\mu\text{m}$. The compressive stress stems mainly from the lower thermal expansion rate of the STI oxide when compared to silicon, as

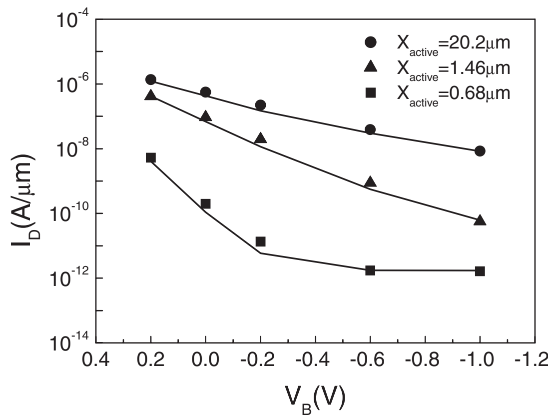



Fig. 4. $I_D$ values at $V_G = -0.4$ V for $X_{active} = 20.2$ μm, $X_{active} = 1.46$ μm, and $X_{active} = 0.68$ μm. Symbols represent experimental data and solid lines are the final simulated results including stress-dependent diffusion models.

well as the thermal gate oxidation induced volume expansion at the STI edge. Thus, as $X_{active}$ decreases, the STI approaching the MOSFET core region increases the magnitude of the compressive stress. The extracted the $\Delta E_S$ for phosphorus, arsenic, and boron in previous work[11] are $-30$, $-14$, and $-7$ eV/unit strain, respectively. The negative sign of $\Delta E_S$ denotes diffusion retardation caused by the compressive stress in pure silicon for these impurities, and is in agreement with the literature[7,10,14] for boron and phosphorus. Note that, so far, no conclusive argument has been reached regarding the arsenic diffusion behavior under a general non-uniformly compressive stress in pure silicon.

In the MOSFET structure used this paper, source and drain phosphorus diffusion is much more sensitive to the mechanical stress than boron and arsenic because of the absence of the high concentration retrograde P-type doping well. The experimental silicon and simulation results are shown in Figs. 2 and 4. In the absence of the VIDAESM, the simulated leakage current (that is, the flat region in Fig. 3) was found to be much higher for MOSFETs with smaller $X_{active}$ values as illustrated in dashed lines in Fig. 2. The $I_D$ value for $X_{active} = 1.46$ μm at $V_B = -1$ V and $V_G = -0.4$ V is $2.5 \times 10^{-10}$ A/μm without using VIDAESM, which is much larger than the result obtained using VIDAESM ($6.3 \times 10^{-11}$ A/μm). The corresponding silicon experimental data is $5.6 \times 10^{-11}$ A/μm. The simulations that incorporated VIDAESM revealed that the punchthrough between the deeper part of the phosphorus source and drain is responsible for the leakage current. This means that the dopant diffusion becomes less as $X_{active}$ is decreased, and is consistent with the results indicated in Fig. 4: a decrease in $X_{active}$ produces a substantially large reduction in leakage current. In Fig. 4, the gate-edge tunneling current of $X_{active} = 0.68$ μm MOSFET prevails in the background current, regardless of high negative substrate biases.

To investigate the impact of the mechanical stress on transient enhanced diffusion, the stress-dependent point defect diffusion and equilibrium concentration models described in eqs. (3) and (4) are applied to the numerical simulator. The simulation results show slightly higher subthreshold leakage current, which implies that the dopant diffusion is the stronger. The $I_D$ value for $X_{active} = 1.46$ μm at $V_B = -1$ V and $V_G = -0.4$ V increase from $6.3 \times 10^{-10}$ to $7.1 \times 10^{-10}$ A/μm. The explanation for this phenomenon is that the interstitial equilibrium concentration $C_I^*$ decreases under the compressive stress and therefore the interstitial supersaturation factor, $C_I/C_I^*$, increases after impurity ion implantation, which results in the TED enhancement. The effect is not significant because high ramp rate rapid thermal anneals were applied after ion implantations. To further fit the experimental data after taking the stress-dependent TED effect into account, the final value $\Delta E_S$ of phosphorus is fine tuned from $-30$ to $-33$ eV/unit strain. Figure 4 depicts the corresponding results of experimental data fitting.

In conclusion, lightly doped well MOSFETs with subthreshold characteristics that are sensitive to STI mechanical stress have been fabricated to study the stress-dependent diffusion effect. Utilizing the proposed stress-dependent diffusion model together with the extracted stress-dependent diffusion coefficients has greatly improved the accuracy of reproducing subthreshold *I–V* for different active area lengths and substrate biases in the numerical 2D process/device simulations. Adding stress-dependent point defect models shows a higher subthreshold leakage current, implying a stronger TED effect due to lower interstitial equilibrium concentration under compressive stress. After incorporating the stress-dependent TED effect, the stress-dependent diffusion coefficient for phosphorus has been slightly updated.

**Acknowledgements**

This work was supported by the National Science Council of Taiwan under Contract No. NSC94-2215-E-009-005. The authors would like to thank colleagues of Taiwan Semiconductor Manufacturing Co. for their contribution on the wafer fabrication.

1) R. A. Bianchi, G. Bouche and O. Roux-dit-Buisson: Int. Electron Devices Meet. Tech. Dig., 2002, p. 117.
2) K. W. Su, Y. M. Sheu, C. K. Lin, S. J. Yang, W. J. Liang, X. Xi, C. S. Chiang, J. K. Her, Y. T. Chia, C. H. Diaz and C. Hu: Proc. Custom Integrated Circuits Conf., 2003, p. 245.
3) B. Dietrich, V. Bukalo, A. Fischer, K. F. Dombrowski, E. Bugiel, B. Kuck and H. H. Richter: Appl. Phys. Lett. **82** (2003) 1176.
4) Y. M. Sheu, C. S. Chang, H. C. Lin, S. S. Lin, C. H. Lee, C. C. Wu, M. J. Chen and C. H. Diaz: Int. Symp. VLSI-TSA, 2003, p. 269.
5) V. Senez, A. Armigliato, I. De Wolf, G. Carnevale, R. Balboni, S. Frabboni and A. Benedetti: J. Appl. Phys. **94** (2003) 5574.
6) G. Scott, J. Lutze, M. Rubin, F. Nouri and M. Manley: Int. Electron Devices Meet. Tech. Dig., 1999, p. 827.
7) N. E. B. Cowern, P. C. Zalm, P. van der Sluis, D. J. Gravesteijn and W. B. de Boer: Phys. Rev. Lett. **72** (1994) 2585.
8) S. Chaudhry and M. E. Law: J. Appl. Phys. **82** (1997) 1138.
9) M. J. Aziz: Appl. Phys. Lett. **70** (1997) 2810.
10) N. R. Zangenberg, J. Fage-Pedersen, J. L. Hansen and A. Nylandsted Larsen: J. Appl. Phys. **94** (2003) 3883.
11) Y. M. Sheu, S. J. Yang, C. C. Wang, C. S. Chang, L. P. Huang, T. Y. Huang, M. J. Chen and C. H. Diaz: IEEE Trans. Electron Devices **52** (2005) 30.
12) M. Diebel: Ph. D. Thesis, University of Washington, Seattle, WA, 2004.
13) Z. K. Lee, M. B. McIlrath and D. A. Antoniadis: IEEE Trans. Electron Devices **46** (1999) 1640.
14) H. Park, K. S. Jones, J. A. Slinkman and M. E. Law: Int. Electron Devices Meet. Tech. Dig., 1993, p. 303.