ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

THE IMPACT OF CHANNEL ENGINEERING ON HOT-ELECTRON INJECTION IN THE DEEP-SUBMICROMETER FLASH MEMORY CELL

ZHANG YU

School of Electrical and Electronic Engineering

A thesis submitted to the Nanyang Technological University
in partial fulfillment of the requirement for the degree of
Master of Engineering

2008

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

ACKNOWLEDGMENT

First and foremost, I would like to express my deepest gratitude to my project supervisor, *Dr. Ang Diing Shenp*, for his patient guidance, invaluable advice, persistent support and encouragement throughout the entire course of the project, without which the completion of the project would not have been possible. His devotion and determination towards scientific research have greatly inspired me to strive not only for the research project but also for my professional career in future.

I also wish to offer my sincere thanks to *Mr. Tan Kok Tong*, the project co-supervisor in Systems on Silicon Manufacturing Company (SSMC) Pte. Ltd, for giving me this great opportunity to pursue a higher degree in the field of advanced semiconductor technology. Special appreciation goes to my SSMC mentor, *Mr. Kuan Hing Poh*, who had given me very helpful suggestions and willingly shared his knowledge on Flash memory technology. I am also grateful to *Mr. Glen Foo* and *Mr. Derek Sim*, for the enlightening discussion we had as well as their kindness in sharing the testing machine with me.

In addition, I am indebted to my family members, especially my parents and grandparents, for their continuous love and moral support all along. I also want to take this opportunity to thank my dearest Tony for his great understanding as well as encouragement during the course of the project.

Last but not least, I would like to extent my sincere thanks to all my colleagues in the Sensors & Actuators Lab II (NTU), as well as the engineers in the Section of Failure Analysis and Reliability (SSMC), for their help in making this project successful. Finally, I am grateful to SSMC for providing the research scholarship and the test devices used in this research project.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

TABLE OF CONTENTS

ACKNOWLEDGMENT i

ABSTRACT vi

LIST OF SYMBOLS viii

LIST OF FIGURES xv

LIST OF TABLES xxi

Chapter 1: Introduction 1

1.1 Background of Semiconductor Memory 1

1.2 Historical Evolution of Non-Volatile Memory 2

1.3 Programming/Erasing Mechanisms of Flash Memory 8

1.3.1 Programming and Erasing by Fowler-Nordheim (FN) Tunneling 9
1.3.2 Programming by Hot-Carrier Injection (HCI) 11

1.4 Reliability and Scaling Issues in Flash Memory........................14

1.4.1 Reliability Concerns.. 14
1.4.2 Scaling Issues of Flash Memory 16

1.5 Motivation and Scope of the Project...17

1.6 Major Contributions of this Thesis...19

1.7 Organization of the Thesis 20

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Chapter 2: Literature Review 22

2.1 A General Introduction to Hot-Carrier Effect 22

2.2 Classical Hot-Carrier Injection Mechanisms and Related Hot Carrier Effect 25

2.2.1 Channel hot-electron (CHE) injection 26
2.2.2 Drain avalanche hot-carrier (DAHC) injection..................................29
2.2.3 Substrate hot-electron (SHE) injection 32
2.2.4 Secondary generated hot-electron (SGHE) injection..........................33
2.2.5 Dependence of device degradation on HCI stress condition 36

2.3 Techniques to Suppress Hot-Carrier Effect 39

2.4 Hot-Electron Gate Current Models 40

2.4.1 Classical lucky-electron model (LEM) 40
2.4.2 Lattice-temperature assisted hot-carrier injection model....................44
2.4.3 Electron-electron scattering (EES) model and the refined effective electron-temperature (EET) model 45
2.4.4 Impact-ionization feedback (IIF) model 47

2.5 Summary 53

Chapter 3: Test Devices and Experimental Method 57

3.1 Equipment Set-up and Device under Test (DUT).....................57

3.2 Experimental Method 60

3.3 Summary 62

Chapter 4: Fundamental Device Characterization........... 64

4.1 Introduction...64

4.2 Basic electrical characteristics..64

4.2.1 Comparison of vertical channel doping profiles in cell A and cell B ..65
4.2.2 Comparison of short channel effect (SCE) in cell A and cell B 69

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

4.2.3 Summary of channel doping profile comparison 72

4.3 Hot-Carrier Generation and Injection Characteristics 73

4.4 Investigation on CHISEL injection 76

4.4.1 Experimental signatures of CHE and IIF-induced gate currents 77
4.4.2 Impact of channel engineering 79
4.4.3 Comparison of FN, CHE and CHISEL injections 84

4.5 Summary 87

Chapter 5: Non-Classical Hot-Electron Gate Current Injection 89

5.1 Observation of Non-Classical Gate Current Component.........89

5.2 Experimental Feature of the Non-Classical Hot-Electron Effect in Conventional CHE Injection Mode 93

5.3 A Phenomenological Model based on Impact-Ionization Feedback Mechanism 95

5.4 Summary 99

Chapter 6: Influence of Non-Classical Injection on Device Degradation.. 101

6.1 Introduction..101

6.2 Experimental 104

6.3 Results and Discussion 105

6.4 Summary 116

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Chapter 7: Conclusion and Recommendations for Future Work 118

7.1 Conclusion 118

7.2 Recommendations for Future Work 124

Publication List 126

Bibliography 127

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

ABSTRACT

Basic mechanisms governing the generation and injection of hot electrons in N-channel MOSFET are of fundamental importance to non-volatile memory application and reliability. Besides the common Channel Hot-Electron (CHE) injection, a relatively new CHannel-Initiated Secondary ELectron (CHISEL) injection, which entails the use of a reverse body bias and governed by the injection of tertiary electrons, is also adopted for memory programming nowadays. On the other hand, aggressive channel doping scheme is usually employed in advanced semiconductor memory fabrication, in order to suppress undesired short-channel effect which is severe in deep-submicrometer memory device especially when a relatively thick tunnel oxide (~ 7 – 8 nm) is also involved in such scaled cell for good data retention. Over the years, there has been limited study on how the channel doping engineering could affect the hot-electron injection characteristics in scaled memory cells, despite the general consensus that hot-electron effect is closely associated with the channel/drain dopant profile. This may primarily be due to the two-dimensional nature of the problem, which renders an experimental approach generally difficult.

In this thesis, we report the direct observation of a non-classical hot-electron gate current in scaled MOSFET memory cell under conventional CHE biasing condition, i.e. $V_{gs} \sim V_{ds}$, but at reduced voltage levels. Through detailed experimental study, we have clearly delineated that for devices with steeper vertical channel doping profile, the magnitude of non-classical gate current under reduced voltage condition is more

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

prominent. The similarity between the experimental signature of the non-classical gate current and that of the CHISEL induced gate current implies that the feedback heating of hot holes is the main driving force for the non-classical gate electron injection. A relatively high vertical built-in electric field in devices with steeper vertical channel profile can significantly enhance the gate injection of the subsequently generated tertiary electrons.

In addition, unlike the conventional CHE effect in which the spatial extent of hot-electron induced oxide/interface damage is relatively localized, the tertiary hot-electron injection is inherently non-local in nature. This could result in a substantial spread of hot-electron induced oxide and interface damage into the channel region. In fact, it is observed that even for decreased V_{ds}, the scalability of the hot-electron induced oxide damage region is still suppressed for cell exhibiting the non-classical hot-electron injection phenomenon. This observation suggests that the device lifetime projection based on the accelerated stress condition may grossly overestimate the parametric lifetime concerned.

Last but not least, since non-classical hot-electron effect induced oxide damage is distributed more towards the source side, it has an important implication on the scalability of the high-κ dielectric based trapped-charge memory, which is targeted for single-cell multi-bit memory application using separate source/drain side hot-electron injection. Because of the spatially greater distribution of non-classical hot-electron injection, the effect of source-side and drain-side hot-electron injection on the MOSFET characteristics may no longer be distinguishable in aggressively scaled memory cells.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

LIST OF SYMBOLS

Common Terminologies

CHE	Channel hot-electron
CHISEL	Channel-initiated secondary electron
CMOS	Complementary metal-oxide-semiconductor
DAHC	Drain avalanche hot-carrier
DDD	Double-diffused drain
DIBL	Drain-induced barrier lowing
DRAM	Dynamic random access memory
DUT	Device under test
EEDF (or DF)	Electron energy distribution function
EEPROM	Electrically erasable and programmable read-only memory
EPROM	Electrically programmable read-only memory
EES	Electron-electron scattering
EET	Effective-electron-temperature
EHP	Electron-hole pair
FET	Field effect transistor
FG	Floating-gate
FN tunneling	Fowler-Nordheim tunneling

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

HCI	Hot-carrier injection
HCE	Hot-carrier effect
HEI	Hot-electron injection
HP 4156B SPA	Hewlett-Packard 4156B Semiconductor Parameter Analyzer
IC	Integrated circuit
II	Impacted-ionization
IIF	Impact-ionization feedback
LDD	Lightly-doped drain
LEM	Lucky-Electron Model
LIIF (or lateral IIF)	Lateral impact -ionization feedback
MFP	Mean-free-path
MLDD	Moderately doped LDD
MOS	Metal-oxide-semiconductor
NVM	Non-volatile memories
PCM	Process-control monitoring
PII	Primary impact-ionization
PLDD	Profile LDD
RNO	Reoxidized-nitrided-oxide
SCE	Short-channel effect
SGHE	Secondary generated hot-electron
SHE	Substrate hot-electron
SII	Secondary impact ionization

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

SRAM	Static random access memory
TLDD	Inverse-T gate LDD
TTL	Transistor-transistor logic
VIIF (or vertical IIF)	Vertical impact-ionization feedback
1D	One-dimensional
2D	Two-dimensional

Parameters

C	Capacitance
D_{it}	Interface trap density (cm^{-2} eV^{-1})
E_{ch} , E	Channel electric field
$E_{//}$	Lateral channel field
$E_{\perp}$	Vertical injection field in the channel
E_{ox}	Electric field for carrier injection into oxide
E_e	Electron energy
G_d	Channel conductance
G_m	Transconductance
I_b, I_{SUB}, I_{sub}	Substrate current
I_{br} (= I_{sub}/I_d or I_b/I_d)	Normalized substrate current or the channel heating efficiency
I_{coll}	Photo-generation current
I_{cp}	Charge pumping current

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

I_d, $I_{ds}(\sim I_s)$	Device (or drain) current
I_{do}	Drain current of the fresh device measured at a specific gate voltage
I_g^e, I_G, I_g, I_{gs}	Hot-electron gate emission/ programming current
I_{gr}, η, $I_{g\ r}^{e}$ $(= I_g^e/I_d)$	Normalized gate current or the hot-electron gate injection efficiency
I_{off}	Off-state current
I_{on}	On-state current
J_y	Channel current density
L or L_g	Drawn channel length of a transistor
L_{eff}	Effective channel length of a transistor
M	Hot-carrier multiplication factor in an impact ionization process
M_1	Current multiplication factor of the primary impact ionization process at the drain-end
M_2	Current multiplication factor of the secondary impact ionization process deep inside the substrate
M_3	Current multiplication factor of the secondary impact ionization process along a sub-surface path in the channel
m	Characteristics exponent of the hot-carrier lifetime prediction model
$m*$	Effective mass of an electron
$N_a(x_d)$	Dopant concentration at a distance x_d from the Si-SiO_2 interface
N_{eff}	Average effective doping concentration (cm^{-3})
N_{it}	Interface trap density (cm^{-2})

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

N_{it0}	Initial interface state density before hot-carrier stress
N_{ot}	Oxide trap density (cm^{-2})
N_{ot0}	Initial oxide trap density before hot-carrier stress
N_{sub}	Substrate doping concentration (cm^{-3})
n	Characteristics exponent of the hot-carrier injection model
n_s	Surface electron density
$n(y)$	Electron concentration at depth y and position x in the channel
P_{inj}	Probability that a tertiary electron will be accelerated to the oxide interface and inject into the gate
P_1	Probability for an electron to acquire enough kinetic energy as well as sufficient normal momentum
P_2	Possibility for a hot electron traveling to the Si-SiO_2 interface without suffering any inelastic collision after the re-direction
P_3	Probability of an electron suffering no collision in the oxide image potential well when transits from the interface to floating gate
Q_{it}	Interface trapped charge
q	Electronic charge (1.6×10^{-19} C)
S_t	Sub-threshold swing (mV/dec)
T	Absolute temperature (Kelvin)
t	Stress time
T_L	Lattice temperature
T_e	Electron effective temperature
V_b, V_B, V_{sb}, V_{sub}, V_{bs} (< 0)	Reverse substrate bias; back bias

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

$V_{channel}(x)$	Channel potential along the horizontal direction (i.e. from the source to the drain end)
V_d^s, V_g^s, V_s^s V_{ds}^s,	Stress voltage (applied at drain, gate and source terminals, and across the channel region) during a hot-carrier stress experiment
V_{db}	Drain-substrate potential difference
$V_{d,}$ V_{ds}	Drain voltage; potential drop over the channel
$V_{gs,}$ $V_{G,}$ V_g	Gate voltage
V_g^m	Selected gate bias voltage to measure the post-stress device degradation
V_s	Source voltage
V_t, V_{th}	Threshold voltage
V_{tho}	Threshold voltage of a fresh device
W	Drawn channel width of a transistor
X_{ox}, t_{ox}	Tunnel-oxide or gate oxide thickness
X_j	Source/Drain junction depth
x_d	Depth of depletion region
$\alpha(E_{OX})$	Ratio of the Si-SiO_2 barrier height with respect to the impact ionization threshold (~ 2.5 – 3.5)
$\Delta D(t)$	Absolute parameter shift during aging
ΔI_d	absolute change in the measured drain current at that specific V_g during stress
ΔL	Length of the drain side region where most significant CHE injection occurs
ΔN_{it}	Increase of the interface trap density normalized for the entire channel
$\Delta N_{it}'$	Increase of the interface trap density in a very small

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

	degraded region δL
ΔN_{ot}	Increase of the oxide trap density
ΔV_{th}, ΔI_d, ΔG_m or ΔN_{it}	Hot-carrier stress induced variations parametric degradation
δL	Length of the degraded channel region
ε_s	Silicon dielectric constant
ε_{ox}	Gate oxide dielectric constant
κ	Relative gate dielectric constant
λ	Scattering mean-free-path
ξ_y	Channel electric field
τ_ε	Energy relaxation time
μ_{if}	Low filed mobility
τ_L τ,	Hot-carrier lifetime
v_s , v_{sat}	Saturated electron velocity
Φ_B, Φ_{Beff} , ϕ_{ox}	Si-SiO_2 barrier height (~ 3.2 eV) for electron injection towards the gate
Φ_{II}, Φ_{IIeff}, ϕ_{ii}	Impact ionization (II) threshold
Φ_{ITG}	Threshold energy for interface trap generation
$\Delta I_d/I_{do}$ x 100%	Percentage degradation of the linear drain current during the hot-electron injection stress

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

LIST OF FIGURES

Chapter 1

Fig. 1.1 A brief summary of the evolution of Non-Volatile Memory. 3

Fig. 1.2 A typical floating-gate structure of a non-volatile semiconductor memory cell [3]. 5

Fig. 1.3 Change of cell state and reading operation for floating-gate memory [58]. 5

Fig. 1.4 Energy band diagrams of an N-channel floating-gate memory cell when it is (a) being programmed and (b) being erased, respectively, by FN-tunneling mechanism. (Figures are from [3].) 10

Fig. 1.5 Schematic diagram of a 0.18-μm N-channel Flash memory cell when it is programmed by (a) channel-hot electron injection and (b) channel-initiated secondary electron injection. Biasing condition in the two programming schemes to yield comparable programming efficiency is indicated. Drawn cell dimension W/L = 0.35/0.18 μm/μm. Tunnel oxide thickness ≈ 5.5 nm [11]. 12

Chapter 2

Fig. 2.1 A conceptual view of hot-carrier reliability issue [27] [28]. 24

Fig. 2.2 A schematic diagram of an N-channel MOSFET showing various hot-carrier mechanisms that can induce the hot-carrier effect [31]. 26

Fig. 2.3 An illustration of channel hot-electron injection in an n-MOSFET. The channel electrons are deflected from the inversion layer towards the gate oxide near the drain-end [32]. As a consequence, a non-negligible gate current is measured. 27

Fig. 2.4 Gate currents due to CHE injection in an n-MOSFET [32]. 29

Fig. 2.5 Schematic illustration of the drain avalanche hot-carrier injection (DAHC) mechanism in an N-channel MOSFET. The hot electrons here originate from the primary impact ionization process at the drain-end [32]. 30

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Fig. 2.6 Gate current characteristics of an n-MOSFET. In addition to channel hot-electron injection, this plot also features the much smaller injection current attributed to drain-avalanche hot-carrier injection [32]. 31

Fig. 2.7 A schematic diagram of an N-channel MOSFET, showing uniform hot-electron injection over the entire channel region under the SHE injection mode [32]. 32

Fig. 2.8 (a) Substrate current induced hot-electron injection mechanism. $V_g<V_d$, $|V_{sub}|>0$. (b) Secondary hot-electron injection due to photon emission and re-absorption [32]. 35

Fig. 2.9 (a) A cross-sectional view of an N-channel MOSFET showing the three scattering possibilities in the LEM. (b) The energy band diagram of the three-step CHE injection process in the LEM [57][58]. 41

Fig. 2.10 Schematic representation of CHISEL injection in an n-MOSFET by means of impact-ionization feedback of hot holes [80] ($E_{//}$ and $E_{\perp}$ denote the lateral field along the channel and vertical field at the drain end, respectively.) 48

Fig. 2.11 Characteristic plots of I_{gr} versus I_{br} for different V_b. A large slope exists at V_b = 0 implying this biasing condition is dominated by CHE injection. As V_b becomes more negative, the slope gradually approaches 1, indicating the CHISEL injection becomes dominant at reversed-V_b condition [85]. 52

Chapter 3

Fig. 3.1 Block diagram of the experimental setup for common *I-V* measurement. (S, G, D, B respectively denotes the source, gate, drain and substrate pads of the test cell.) 57

Fig. 3.2 (a) Front view of the Suss MicroTec probe station used in this project with closed prober shield. (b) The test wafer is carefully placed on the thermal chuck of the probe station with tight suction provided by the vacuum pump. 58

Chapter 4

Fig. 4.1 (a) Extracting threshold voltage (V_{th}) in the linear mode (when V_{ds} ~ 0.1 V) by 'maximum-G_m' method. (b) Extracting the on-/off- state current and the sub-threshold swing in the saturation mode (when V_{ds} = V_{dd} = 1.8 V). 65

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Fig. 4.2 Output and transfer characteristics of test cell A and B with drawn W/L = 1 0/0.176 μm/μm. Cell A possesses a higher V_{th} than cell B as expected. 66

Fig. 4.3 Variation of the linear threshold voltage V_{th} (measured at V_d = 0.1 V) due to the change of substrate bias V_b in two types of 0.14-μm test devices, suggesting a larger channel doping concentration present in cell A. 67

Fig. 4.4 Extracted dopant concentration as a function of the vertical distance x_d from the Si-SiO_2 interface in the two different sample devices. 69

Fig. 4.5 Reduction of linear threshold voltage (V_{th}) as the drawn channel length (L_g) decreases from 0.4 μm to 0. 16 μm for cells A and B. 70

Fig. 4.6 Drain-induced barrier lowing effect in cells A and B at different drawn c hannel length as V_d changes from 0.05 V to 1.8 V. 71

Fig. 4.7 Hot carrier characteristics of a cell A device, taking drain bias (V_d) as the parameter for V_b = *0* V. 74

Fig. 4.8 Hot carrier characteristics of a type A cell, taking the drain (V_d) and the substrate (V_b) biases as parameters. (Solid circle: V_b = 0 V; open circle: V_b = -2 V.) 75

Fig. 4.9 Electron injection current (I_g) versus gate sweep voltage (V_g) characteristics for a typical N-channel test cell with drawn channel dimension (W/L) = 10/0.176 μm/μm and tunnel-oxide thickness (t_{ox}) = 8.1 nm. 76

Fig. 4.10 Hot-electron injection efficiency $I_{g\,r}^{\,e}$ versus primary impact ionization efficiency I_{br} of a typical 0.14-μm generation type A cell. Device dimension (W/L) = 10/0.176 μm/μm, tunnel-oxide thickness t_{ox} = 8.1 nm. 78

Fig. 4.11 $I_{g\,r}^{\,e}$ versus I_{br} characteristics (with parameter V_b) for an N-channel memory cell fabricated by commercial 0.18-μm technology. The gate bias V_g is fixed at 4 V, while the drain bias V_d is varied from 2.8 to 4.8 V. The drawn channel dimension (W/L) = 10/0.176 μm/μm and the tunnel-oxide thickness t_{ox} = 8.1 nm. 80

Fig. 4.12 $I_{g\,r}^{\,e}$ versus I_{br} characteristics (with parameter V_b) for an N-channel type A cell fabricated by commercial 0.14-μm technology. V_g is fixed at 2.5 V, while V_d is varied from 1.2 to 4.5 V (small $I_{g\,r}^{\,e}$ noise at low V_d range is not shown here). The drawn channel dimension (W/L) = 10/0.176 μm/μm and the tunnel-oxide thickness t_{ox} = 8.1 nm. 81

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Fig. 4.13 Comparison of the IIF-induced tertiary electron injection in two types of 0.14-µm devices (test cell A and cell B). V_g is fixed at 3.5 V, while V_d is varied from 1.2 to 4.5 V (small $I_{g\ r}^{e}$ noise at low V_d range is not shown here). The drawn channel dimension (*W/L*) for both cells = 10/0.176 µm/µm and the tunnel-oxide thickness t_{ox} = 8.1 nm. 82

Fig. 4.14 Another experimental result showing the more abrupt and heavier channel doping in cell A can enhance the significance of IIF and promote tertiary electron injection. V_d is varied from 1.2 to 4.5 V (small $I_{g\ r}^{e}$ noise at low V_d range is not shown here). The drawn channel dimension (*W/L*) for both cells = 10/0.176 µm/µm and the tunnel-oxide thickness t_{ox} = 8.1 nm. 83

Fig. 4.15 Programming current (I_g^e) measured in three different electron injection schemes for N-channel transistors fabricated by 0.18-µm and 0.14-µm technology (cell A). The drawn channel dimension (*W/L*) for both cells = 10/0.176 µm/µm and the tunnel-oxide thickness t_{ox} = 8.1 nm. 85

Fig. 4.16 Programming current (I_g^e) measured in three different electron injection schemes for N-channel transistors fabricated by 0.14-µm technology (cell B). The drawn channel dimension (*W/L*) = 10/0.176 µm/µm and the tunnel-oxide thickness t_{ox} = 8.1 nm. 86

Chapter 5

Fig. 5.1 $I_{g\ r}^{e}$ versus I_{br} characteristics for Cell A (14-µm generation) when V_g was fixed while V_d was increased from 3 to 4.5 V. Despite $V_g < V_d$, the hot-electron gate current continues increasing with V_d. Drawn channel dimension of the test cell (*W/L*) = 10/0.176 µm/µm, and tunnel-oxide thickness t_{ox} = 8.1 nm. 90

Fig. 5.2 $I_{g\ r}^{e}$ versus I_{br} characteristics are modulated by varying V_s from 0.2 to -0.2 V. Evident non-classical hot-electron injection is shown at V_b = 0 V. The test cells are type A cells fabricated by commercial 0.14-µm CMOS technology. 92

Fig. 5.3 $I_{g\ r}^{e}$ versus I_{br} characteristics of type B cell as V_s is varied from 0.2 to -0.2 V. Typical CHISEL injection signature appears at high reverse V_b bias (-2 and -3 V); however, no non-classical HCI component at V_b = 0 V. 92

Fig. 5.4 Modulating the injection efficiency ($I_{g\ r}^{e}$) by a wider V_s variation (from 0.5 to –0.5 V) on a 0.14-µm type A cell (drawn channel dimension *W/L* = 10/0.176 µm/µm). Dash lines (with n ~ 1) are the best fit of low I_{br} data points and correspond to the $I_{g\ r}^{e}$ component contributed by the non-classical hot-electron effect. Subtracting this non-classical $I_{g\ r}^{e}$

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

component from the total $I_g{}^e{}_r$ in high I_{br} regime can yield the characteristics of high-field electron injection solely (open circles: for $V_b = -1$ V, solid squares: for $V_b = 0$ V). 94

Fig. 5.5 Cross-sectional view of an N-channel MOSFET under different hot-electron injection mechanisms. 96

Fig. 5.6 Schematic diagram of the drain region illustrating the primary electron path (*solid arrow*) and the lateral feedback heating of holes (*open arrow*) at biasing condition $V_g \geqslant V_d$. Event *(1)*: primary impact ionization at drain side. Event *(2)*: secondary impact ionization in the channel. 98

Fig. 5.7 Schematic diagram of the drain region at biasing condition $V_g < V_d$. The corresponding primary electron path is deeper and the feedback hole path is shallower as compared to the condition when $V_g \geqslant V_d$. 99

Chapter 6

Fig. 6.1 Degradation of the linear drain current I_{do} (when $V_d{}^m = 0.1$ V) versus the hot-electron stress time for cell A. Left figure depicts $\Delta I_d/I_{do}$ at $V_g{}^m \sim V_{tho}$, indicating a condition where the oxide damage is extended into channel region. Right figure depicts $\Delta I_d/I_{do}$ at $V_g{}^m = 3$V, which is mainly due to the damage induced at the drain side. 107

Fig. 6.2 Degradation of the linear drain current I_{do} (when $V_d{}^m = 0.1$ V) versus the hot-electron stress time, presenting a classical evolution of the main hot-electron damage region from the drain-end towards the middle-channel region. 108

Fig. 6.3 The lateral spread of hot-electron damage region in cell B decreases drastically as $V_{ds}{}^s$ reduces for a positive increase in $V_s{}^s$. 108

Fig. 6.4 A substantial lateral spread of the hot-electron damage region into device channel is observed for cell A even when $V_{ds}{}^s$ is decreased. 109

Fig. 6.5 At a much higher gate overdrive, the effective hot-electron damage region in cell A appears to be reduced and the I_{do} degradation characteristic follows that of cell B. 110

Fig. 6.6 Hot-electron lifetime characteristic when the degradation metric ($\Delta I_d/I_{do}$ x 100% = 10%) is measured at $V_g{}^m = 3$ V. Damage in the channel region is not probed and thus the lifetime characteristic of cell A appears similar to that of cell B. The thick-solid line is the best-fit for the open symbols, while the

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

dotted line is the best-fit for the filled symbols. The thin-solid line is an extrapolation of the thick-solid line, showing almost the same slope (m ~3.6) over the whole I_{br} range under high gate-drive condition. 110

Fig. 6.7 Lifetime characteristic of cell A and cell B when the degradation metric ($\Delta I_d/I_{do}$ x 100%= 10%) is measured at $V_g^m \sim V_{tho}$ to probe the oxide damage spread in the channel region. The thick-solid (dotted) line is the best-fit for the open (filled) symbols. The thin-solid line is the extrapolation of the thick-solid line. 112

Fig. 6.8 Lifetime characteristic of cell A extracted at (a) V_g^m = 1 V and (b) V_g^m = 1.3 V. Degradation monitored at higher V_g^m mainly probes the oxide damage due to CHE injection. The thick-solid (dotted) line is the best-fit for the open (filled) symbols. The thin-solid line is the extrapolation of the thick-solid line. 112

Fig. 6.9 Extraction of lifetime characteristics for type A and type B cells (both with W/L = 10 µm /0.24 µm) at (a) $V_g^m \sim V_{tho}$ and (b) V_g^m = 3 V. Similar trend as for cells with channel dimension W/L = 10 µm / 0.176 µm is also observed. The thick-solid (dotted) lines are the best-fit for data measured in high (low) I_{br} region, while the thin-solid lines are the extrapolation of the corresponding thick-solid lines. 114

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

LIST OF TABLES

Table 3.1 Main process variations involved in the two different lots of test devices fabricated by the commercial 0.14-µm CMOS technology. 60

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Chapter 1: Introduction

1.1 Background of Semiconductor Memory

Since the 1970s, the semiconductor industry has been growing at a dramatic rate as predicted by the famous Moore's law [1]. To a large extent, this tremendous progress has been made possible by the great success in the development of the MOS (Metal-Oxide-Semiconductor) transistor technology. In the pursuit of higher drive current, lower power consumption, and greater packing density, the dimension (gate length) of the transistor is being continuously scaled down. However, along with this downscaling are new issues on process integration and device reliability, which have somewhat slowed down the 'stride' of technology evolution in recent years.

By virtue of its ability to realize high packing density as well as superior device performance, the MOS technology has fueled the rapid development of the semiconductor memory industry. The memory devices used in modern computational system can be divided into two basic categories, that is, Volatile Memory (VM) and Non-Volatile Memory (NVM). Volatile memory refers to a type of memory device which loses the data stored once the system power is switched off; hence, it needs a continuous power supply to stay viable. Most RAMs are volatile memories, including the static RAM (SRAM) and the dynamic RAM (DRAM). Compared with volatile memory, non-volatile memory can virtually retain the information stored forever if no erasing is carried out. In other words, the data stored in NVMs will not disappear even

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

when the system is powered off. The family of non-volatile memories can be further divided into two groups, the floating-gate memory and the charge-trapping memory, according to the way by which the charges are stored in the memory cell. If charges are stored in a semiconducting layer that is completely surrounded by a dielectric, this kind of NVM is known as the 'floating-gate (FG)' memory. In contrast, if charges are stored in the discrete trapping centers of a dielectric layer, this kind of NVM is referred to as the 'charge-trapping' memory.

1.2 Historical Evolution of Non-Volatile Memory

The historical development of NVMs comprises several main stages, as illustrated in Fig. 1.1. The following briefly describes each stage of the development, from the Masked Read-Only-Memory (Masked ROM), Programmable Read-Only-Memory (PROM) to modern programmable and erasable non-volatile memory devices adopting the control-gate/floating-gate structure, like EPROM, EEPROM and Flash memory.

- **Masked ROM (Read-Only-Memory)**

The very first ROMs were the Masked ROMs where the data or instructions were written permanently in the memory array by mask definition during IC design and manufacturing. If any design error occurs, the IC must be re-designed and

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

re-fabricated as the stored bits in a Masked ROM cannot be altered. The presence of an N-MOSFET at the cell-site (i.e. the intersection point of a Word-line (WL) and a Bit-line (BL)) stores bit '0' while its absence denotes bit '1'.

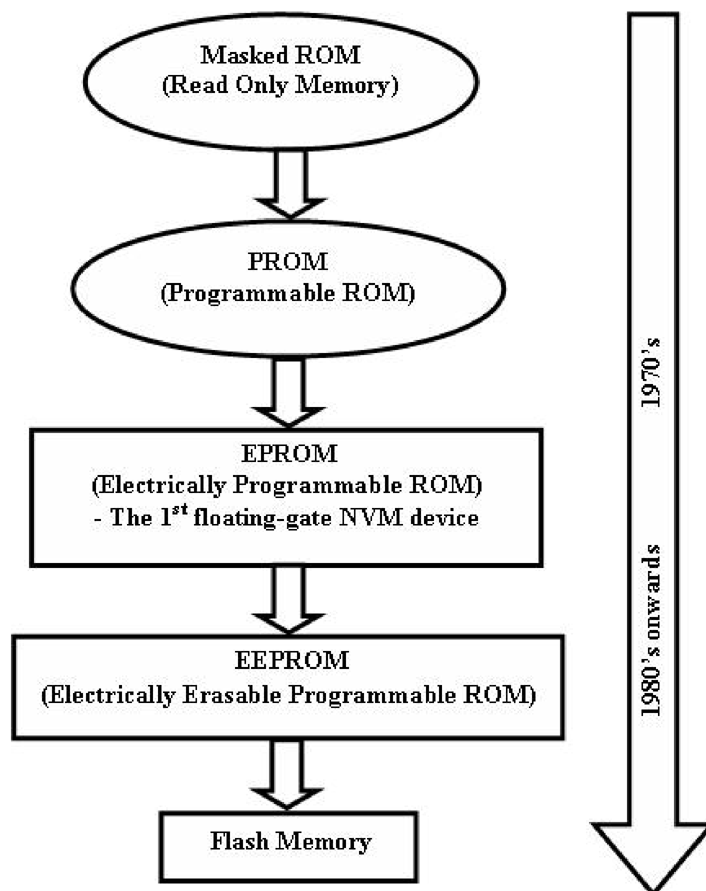


Fig. 1.1 A brief summary of the evolution of Non-Volatile Memory.

- **PROM (Programmable ROM)**

One step up from the Masked ROM is the PROM (programmable ROM). Its cells are all in an unprogrammed state when fabricated. There is a fusible link at each cell site, maintaining the bit at '0'. The end user can program cells as desired by sending a

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

high current pulse from the device programmer to the input pins of the chip so that fusible links are blown and the bit states are switched to '1'. However, the user can only program the PROM cell once. Thus, PROM is also known as one-time programmable (OTP) memory.

- **Floating-Gate NVMs: From EPROM, EEPROM to Flash Memory**

The idea of using floating gate (FG) to create a non-volatile memory was first suggested by D. Kahng and S. M. Sze [2] in 1967. In a FG memory, all basic cells have the same generic structure, as shown in Fig. 1.2, consisting of a stacked poly-silicon gate MOS transistor. An external voltage applied directly at the control gate (CG) is partially coupled to the floating gate because of the capacitive effect. The coupled FG voltage is then responsible for the injection/repulsion of electrons during memory programming/erasing operation.

The basic working principle of the floating-gate NVM is the transfer of electronic charges stored in a floating gate, by which the threshold voltage (V_t) of a cell is modified to switch between two or more distinct levels. From the amount of V_t-shift, the quantity of stored charges can be deduced**.** Conventionally, bit '1' is defined as the erased-state, which corresponds to a lower or negative V_t (i.e. V_{te}); bit '0' is defined as the programmed-state, which refers to a higher or positive V_t (i.e. V_{tp}). The information stored in a cell can be detected by applying a gate voltage V_{read} with a value between V_{te} and V_{tp}. As illustrated in Fig. 1.3, for a programmed cell, $V_{read} < V_{tp}$,

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

the transistor is in the 'cut-off' mode, registering a bit '0'. On the contrary, if a cell is erased, $V_{read} > V_{te}$, a channel is formed and current flows through the cell; hence, a bit '1' is detected.

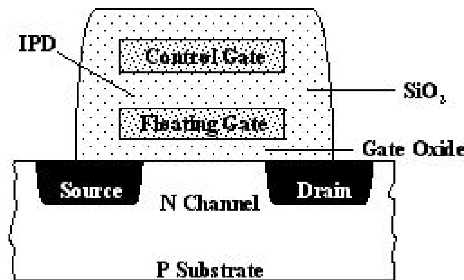


Fig. 1.2 A typical floating-gate structure of a non-volatile semiconductor memory cell [3].

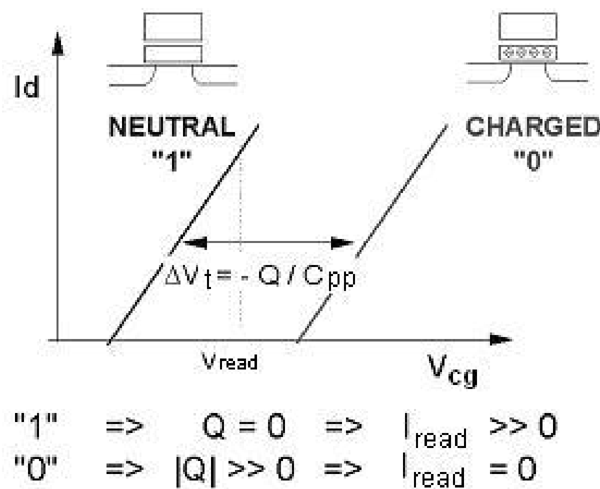


Fig. 1.3 Change of cell state and reading operation for floating-gate memory [58].

EPROM (Erasable and Programmable Read-Only-Memory) is the first semiconductor memory that relies on the floating-gate structure for data storage. It can be erased and re-programmed many times by the end user. Channel hot-electron

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

(CHE) injection or Fowler-Nordheim (FN) tunneling is used for programming while ultraviolet (UV) radiation is used to erase the EPROM array. Due to the requirement of UV-erase, a quartz window has to be incorporated in the memory package, leading to a high price of ERPOM; moreover, the chip needs to be taken out of the circuit in order for the stored information to be erased and all data will be erased under the UV radiation. In view of these complexities and rigidity, EPROM is not preferred in memory application, although its small cell size (1-transistor structure) is very attractive.

The drawback of EPROM in terms of erasability is obviated in a new generation of NVMs, that is, the EEPROM (Electrically Erasable Programmable ROM) devices. The first EEPROM was the Stacked-gate Avalanche-injection MOS (SAMOS) memory [4] invented in the 1970s. For this type of memory, all cell operations (programming/erasing/read) are controlled by an electrical signal and are performed in a byte-addressable way. Moreover, the memory chip can remain on the circuit board to realize ‘in-system erasability’ by electrical erasure. However, a higher level of functionality comes with a price of a larger cell size. Each EEPROM cell consists of two transistors (2-T), one is the floating-gate memory transistor and the other is the select transistor. Due to the 2-T structure, an EEPROM cell tends to have a size which is about two to three times larger than an EPROM cell, leading to a smaller packing density and hence a higher cost per bit.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

The latest generation of NVM is the floating-gate Flash memory. It was first introduced in the 1980's and was referred to as the Flash EEPROM or Flash memory in short [5]. It adopts the 1-T stacked-gate (CG-FG) cell structure and can be programmed byte-by-byte by either hot-carrier injection or FN tunneling. The erasing of Flash memory is achieved by FN tunneling and a large number of cells (by block, sector or page) can be erased simultaneously. Sometimes, in order to retain byte-addressable erasability, a 2-T cell structure (a memory transistor plus a select transistor) as in the case of the EERPOM is adopted.

Flash memory combines the advantage of a high packing density of EPROM and the in-system erasability and re-programmability of EEPROM. In addition, it integrates the ability of non-volatile storage with an access time comparable to DRAM, allowing direct execution of micro-codes. Therefore, it can be used for data accumulation and embedded code storage. These excellent properties have led to a large market success, spurred by the recent expansion of electronic portable systems (e.g. cellular phone, MP3 player etc.) requiring high-density NVM for both data storage (with NAND architecture) and embedded (with NOR architecture) applications. It is predicted that Flash memory will be the main driving force for the next stage of device revolution as it satisfies most of the characteristics of an ideal memory.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Nevertheless, the Flash memory has inherited reliability issues from EPROM and EEPROM as well, with some additional problems unique to this new NVM generation. All these have set stringent restriction on the scaling of Flash memory and are to be discussed in Section 1.4. In view of its largest share in modern non-volatile memory market and its susceptibility to common reliability problems, Flash memory has attracted a lot of research attention and has become a hot topic ever since its first application.

1.3 Programming/Erasing Mechanisms of Flash Memory

In the last three decades, various mechanisms to transfer the electronic charges to or from the floating gate have been introduced, of which four have been proven to be viable. They are:

- Fowler-Nordheim (FN) tunneling through the thin tunnel-oxide (< 12 nm)
- Enhanced FN-tunneling through the poly-oxide
- Source-side injection (SSI)
- Channel hot-electron (CHE) injection

The first two mechanisms are based on the concept of quantum mechanical tunneling across a thin oxide layer, while the latter two are hot-electron injection mechanisms due to field-heating of electrons. Although the use of hot-hole injection as a erasing mechanism had been attempted [6], it was never applied due to the low

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

injection efficiency of hot holes. Hence, for non-volatile memory application, N-channel transistors are generally used, in view of a smaller mass, higher mobility and lower barrier height of electrons.

For Flash memory, the most commonly used programming mechanisms are CHE injection [7] (for NOR Flash) and FN-tunneling [8] (for NAND Flash). The latter is also employed for Flash memory erasing. In the following sections, we shall discuss how the two charge transferring mechanisms perform basic programming and/or erasing operation for the floating-gate Flash memory.

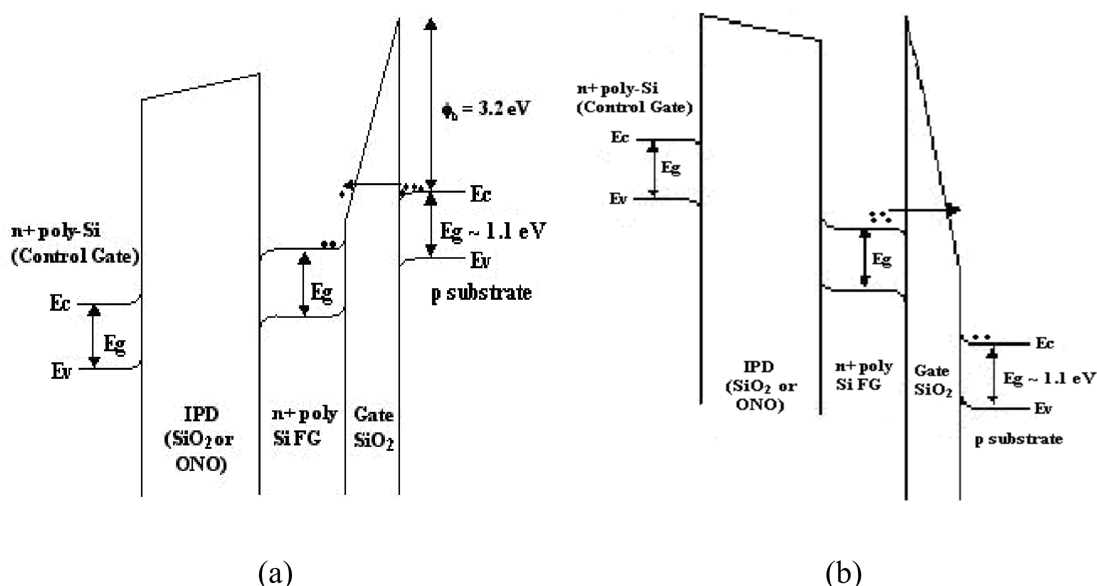
1.3.1 Programming and Erasing by Fowler-Nordheim (FN) Tunneling

FN-tunneling is a field-assisted electron tunneling mechanism whereby electrons are transported through a ‘narrowed’ oxide layer. It was first identified by Fowler and Nordheim in the study of electrons tunneling through a vacuum barrier, and was later quoted by Lenzlinger and Snow for oxide tunneling [9]. It is an important mechanism for both programming and erasing operations of Flash memory. The corresponding energy band diagrams are shown in Fig. 1.4.

When a cell is programmed by FN-tunneling (Fig. 1.4(a)), a large positive control-gate voltage is applied across the stacked-gate N-MOS structure. This will result in a high electric field (>10 MV/cm) across the gate (tunnel) oxide and

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

electrons in the silicon conduction band see a triangular energy barrier with narrowed tunnel width. With a sufficiently high electric field, the barrier width becomes so narrow that electrons can tunnel through from the silicon conduction band into the oxide conduction band. The electrons charge up the poly-silicon floating gate and thus the cell gets programmed. When FN-tunneling is used to erase a memory cell (Fig. 1.4(b)), a large negative voltage is applied at the control gate. The corresponding reverse electric field creates a path for electrons to tunnel through the gate oxide similar as in the case for programming. Therefore, electrons are swept from the poly-Si floating gate to the substrate, and the cell is erased.



(a) (b)

Fig. 1.4 Energy band diagrams of an N-channel floating-gate memory cell when it is (a) being programmed and (b) being erased, respectively, by FN-tunneling mechanism. (Figures are from [3].)

The simplest form of FN-tunneling current density is given by equation (1.1). It is simple but accurate enough to be used to explain the FN-tunneling mechanism in

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

non-volatile floating-gate memory.

$$J = \alpha E_{inj}^2 \exp[\frac{-E_c}{E_{inj}}], \text{ or } J[A/m^2] = 1.15\times10^{-6} E_{inj}^2 \exp[\frac{-2.54\times10^{10}}{E_{inj}}] \quad (1.1)$$

As we can see, an injection field of 10 MV/cm can lead to a current density of approximately 10^7 A/m^2. In order to reach this high field with limited gate voltage, very thin tunnel oxide has to be used in Flash memory device. However, yield as well as reliability consideration restricts the practical oxide thickness to be around 8 to 10nm [10], below which, deteriorative direct tunneling mechanism becomes significant, degrading the memory reliability.

1.3.2 Programming by Hot-Carrier Injection (HCI)

Unlike FN-tunneling, hot-electron injection is only able to inject electrons into the floating gate but not the reverse. Therefore, it can only be used for programming but not erasing Flash memory cells. Moreover, as the negative charges on the FG could establish an electric field opposing further injection of electrons, HCI is a more self-limiting process as compared to FN-tunneling.

There are two types of HCI programming schemes for floating-gate Flash memory cells. One is the traditional channel hot-electron (CHE) injection and the other is the channel-initiated secondary electron (CHISEL) injection scheme.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Fig. 1.5 (a) and (b) show the schematic diagrams of an N-channel memory cell programmed by the CHE and CHISEL schemes [11], respectively. Typical terminal voltages adopted by these two programming schemes to yield comparable programming efficiency are indicated.

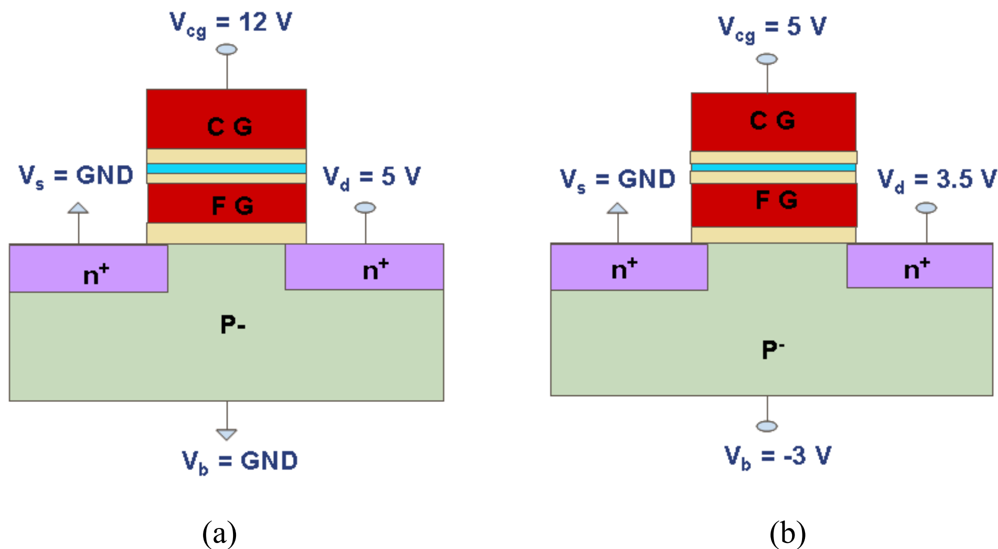


Fig. 1.5 Schematic diagram of a 0.18-μm N-channel Flash memory cell when it is programmed by (a) channel-hot electron injection and (b) channel-initiated secondary electron injection. Biasing condition in the two programming schemes to yield comparable programming efficiency is indicated. Drawn cell dimension W/L = 0.35/0.18 μm/μm. Tunnel oxide thickness ≈ 5.5 nm [11].

As we can see, the voltages applied in these two schemes are quite different. This is because in CHE programming, it is the deflection of channel electrons that contributes to charge injection to the floating gate; whereas in the CHISEL mode, it is the injection of energetic tertiary electrons that forms the programming current (I_g). Due to the distinct origins of the injected electrons, CHE injection needs a higher drain voltage to impart sufficient kinetic energy to channel electrons, and an even

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

higher gate bias to drive channel electrons towards the floating-gate. Therefore, in actual CHE programming application, a pair of comparable gate and drain voltages (i.e. V_{fg} (=1/2V_{cg}) ~ V_{ds}) are often employed, to achieve the maximum injection efficiency of channel electrons.

Fig. 1.5(a) depicts the CHE injection scheme in Flash memory. It is achieved by applying a high drain voltage (V_{ds}) and a high control-gate voltage (V_{cg}) to the cell to be programmed. Other cells which share the same word-line or bit-line with the target cell are not programmed, since no hot electrons can be generated when only one of the V_{ds} and V_{cg} is high. In CHE injection, electrons traveling along the channel are ‘heated up’ by lateral electric field resulting from the high drain bias. Some channel electrons can gain sufficient energy to surmount the Si-SiO_2 energy barrier (Φ_B ~ 3.2 eV for electrons). Under a favorable condition, such as elastic Si-lattice scattering, energetic electrons will deflect towards the floating-gate. This gives rise to a hot-electron gate current [12]; meanwhile, the memory cell is programmed by the storage of the injected electrons on the floating gate.

Fig. 1.5(b) illustrates the CHISEL injection scheme, which has also been adopted for Flash memory programming, considering its superior programming efficiency at low voltages. In this injection mode, a reverse substrate bias (V_b) is used, i.e. a negative V_b is applied to the substrate contact of an N-channel cell. It can enhance the vertical impact-ionization feedback of hot holes as well as the injection of generated

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

tertiary electrons. Hence, a pair of lower voltage biases can be applied to the gate and drain terminals in CHISEL programming, as compared to those in CHE injection. The reverse V_b can accelerate secondary holes generated in the primary impact ionization (PII) process while they are drifting towards the substrate contact. Secondary impact ionization (SII) process is thus induced in the bulk region, yielding tertiary electrons which are extremely energetic as they result from two consecutive impact ionization processes. These electrons can inject back to the floating gate under the influence of a high vertical field established by the reverse V_b, and the injection efficiency is reported to be much higher than that of primary electrons in CHE injection. Hence, CHISEL provides a faster programming speed at equal or less power as compared to CHE. Moreover, the lower biases entailed in CHISEL programming can be easily derived from a single scaled power supply, thus, no high voltage transistors are needed for writing or V_t-convergence. The above advantages have made the CHISEL-based Flash memory cell an ideal candidate for giga-byte memory application.

1.4 Reliability and Scaling Issues in Flash Memory

1.4.1 Reliability Concerns

Reliability issues in Flash memory are generally related to the following aspects:

- Dispersed distribution of threshold voltages in erased cells;
- Program/read disturb to unselected cells which share the same word-line or

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

bit-line as the cell to be programmed/erased;

- Degradation of data retention and cycling endurance due to the deterioration of gate oxide and Si-SiO_2 interface integrity.

Except for the first one, we can see that all the above issues are more or less associated with hot-electron injection, either introduced by a deliberate programming operation or by an undesired electrical stress. It is also found that the hot-electron induced degradation continues to be a reliability concern even when the supply voltage is scaled down below 3 V. And the manifestation of cell degradation is invariably an increase in interface trap density ($\Delta N_{it} \equiv \Delta Q_{it} /q$) and corresponding changes in threshold voltage, device current and transconductance [13][14].

Recent studies have shown that, for a given technology, as the transistor channel length is scaled down towards 0.1 μm, the worst case DC stress condition for the N-MOSFET is gradually changed from $I_{b,max}$ (peak substrate current bias condition) to $V_g \sim V_d$, under accelerated test conditions [15][16]. This implies that the stress condition to create the maximum interface trap generation has switched from $I_{b,max}$ to $V_g \sim V_d$ stress condition [17]. As the latter biasing condition is often adopted in CHE programming scheme for Flash memory to maximize the electron injection efficiency, the impact of a change in the worst-case stress condition on the reliability of deep submicron memory devices needs to be carefully studied.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

1.4.2 Scaling Issues of Flash Memory

With the advancement of semiconductor technology, the feature size of Flash memory cell is scaled down continuously. This is to integrate more functionality into memory chips so that they can perform complicated tasks as required in modern electronic systems. Moreover, a scaled memory cell increases packing density and lowers the manufacturing cost per bit.

However, unlike the CMOS baseline transistor whose oxide thickness can be scaled down by almost the same ratio as its channel length, Flash memory has stringent requirement on its tunnel-oxide thickness. To comply with the rigorous requirements on data retention (normally up to 10 years) and on cycling endurance (10k ~ 100k cycles) [18], the minimum physical thickness of tunnel-oxide has been restricted to be around 7 to 8 nm for state-of-the-art [19] floating-gate Flash memory device. Therefore, while the channel length keeps on scaling down, the vertical dimension of the cell cannot be scaled so aggressively. This imbalance in vertical and horizontal scaling can lead to severe short-channel effect (SCE) in aggressively scaled Flash memory cells.

In order to alleviate the problem, proper transistor engineering is necessary, either through modification of the cell structure or optimization of the channel doping profile. The most direct way is by increasing the channel dopant, that is, applying heavier V_t-adjust and/or halo/pocket implant dose in the channel region so as to mitigate the

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

short-channel effect due to tunnel-oxide inscalability.

Unfortunately, an abrupt channel doping profile will significantly influence the channel electric-field and its local variation, which suppresses SCE but aggravates the hot-carrier injection related reliability problem. The latter is very sensitive to the local channel doping concentration, which determines the location of the peak channel electric field and the main location of oxide and interface traps generated in a hot-carrier injection process. The creation of oxide defects will in turn impose restrictions on the physical scaling of channel length and thus the memory cell size. As a result, a careful optimization on the channel doping profile becomes extremely important for modern Flash memory device for which the typical channel length has scaled into the deep sub-micrometer regime.

1.5 Motivation and Scope of the Project

Flash memory has taken the largest share of the semiconductor NVM market. Due to its susceptibility to technology scaling and hot-carrier effect, an in-depth study on Flash memory with hot-carrier programming scheme is of utmost importance to the subsequent development and application of advanced semiconductor memory. Nevertheless, as the topic itself is too wide, we only focus on the following aspects to

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

investigate the hot-carrier injection phenomenon in aggressively scaled Flash memory.

Firstly of all, as HCI plays a dual role in memory operation, that is, useful in cell programming but harmful in generating oxide defects, a comprehensive literature review on various HCI mechanisms is conducted in our work, together with the study of the influence of hot-carrier effect on device reliability.

Next, our study is focused on CHE and CHISEL injections, the two common programming schemes in non-volatile memory. The generation and injection of tertiary electrons is carefully examined. A non-classical impact-ionization feedback (IIF) mechanism possessing a signature similar to that of the CHISEL injection but occurring under typical CHE biasing condition is observed. An extensive experimental investigation is thus carried out, which eventually reveals the origin of this non-classical component. In view of the tight correlation between channel doping profile with conventional HCI mechanisms, the influence of channel profile on the magnitude of this non-classical injection component is also examined.

Last but not least, the spatial distribution of oxide damage created by this non-classical injection is carefully studied, by investigating the device degradation characteristics. Device lifetime is found to be limited by the non-classical electron injection as well, which is discussed at the end of the work.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

1.6 Major Contributions of this Thesis

In this work, hot-electron injection mechanism is investigated in commercial Flash memory cells. For the first time, we have shown experimentally the existence of a non-classical hot-electron gate current component in the conventional CHE injection dominated regime (i.e. when $V_g \sim V_d$). As V_d decreases, the non-classical electron injection is shown to dominate over the CHE component. The experimental signature of the non-classical component is also presented by comparing it to those of CHE and CHISEL injections. Injection of energetic tertiary electrons, which are generated by the lateral impact-ionization feedback of hot holes, is proposed to be the governing mechanism.

Furthermore, based on the correlation study between the channel doping profile and the non-classical electron injection, we have shown a strong dependence of the non-classical hot-electron injection on the vertical doping profile; traditionally, only the horizontal dopant profile is believed to be effective in influencing hot-electron injection.

We have also shown by experimental evidence that the oxide and interface defects induced by the non-classical hot-electron injection are greater and more widely distributed in the channel region, as compared to that created by CHE injection. The limited scalability of the damage region in turn leads to a greater effect on parametric shifts as the gate overdrive decreases. This effect is manifested in a marked reduction in the slope of the lifetime plot at reduced V_{ds} (where the non-classical hot-electron effect dominates).

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Therefore, lifetime projection from accelerated stress data may not capture this effect at low V_{ds}, resulting in a grossly overestimated device lifetime.

1.7 Organization of the Thesis

There are seven chapters in this thesis and they are organized as follows.

Chapter 1 is an introduction chapter presenting the background information of semiconductor memory as well as the motivation and the scope of this research project.

Chapter 2 presents the literature survey on various hot-carrier injection mechanisms and previously proposed models. Two common memory programming schemes are also reviewed followed by the techniques to suppress the adverse hot-carrier effect.

Description of sample devices and equipment setup used in this project is given in Chapter 3. Typical test devices employed in this project are fabricated by the commercial 0.18-µm and 0.14-µm twin-well CMOS Flash technology. Different channel engineering scheme was adopted in the fabrication of the two generation of devices as well as within the 0.14-µm generation, which leads to differences in

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

channel doping profiles and thus different hot-carrier injection behavior in these devices.

In Chapter 4, fundamental electrical characteristics, common hot-carrier injection behavior and the dependence of hot-electron injection on channel doping profile are demonstrated. In addition, experimental features of the vertical-IIF induced CHISEL injection and its low-voltage programming capability are verified.

A direct observation of a non-classical hot-electron effect in the traditional CHE dominated regime (i.e. $V_g \sim V_d$ and $V_b = 0$ V) is presented in Chapter 5. A phenomenological model based on the concept of sub-surface lateral feedback heating of hot holes is proposed, which could satisfactorily explain this non-classical effect.

Chapter 6 presents the influence of the non-classical tertiary-electron injection on cell reliability, followed by an elaboration of its influence on device scalability and hot-carrier lifetime estimation.

Finally, Chapter 7 summarizes the main results and conclusion of this research project, as well as the recommendations for future work.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Chapter 2: Literature Review

As in the case of the baseline technology, there are many reliability problems associated with the development of the Flash memory. One of the most important issues is hot-carrier injection (HCI), which can set stringent limitations on memory reliability. This chapter provides an introduction and overview of the hot-carrier effect (HCE) in MOS transistors. Several hot-electron injection mechanisms and associated HCE reported in the literature are first presented. The dependence of device degradation on hot-electron stress condition is briefly discussed. Some useful techniques which can be used to suppress the hot-electron effect induced device degradation are subsequently introduced. This is followed by a review of common hot-electron gate current models proposed by using different degradation-monitoring parameters and computer simulation. A summary for the above discussion is given at the end of this chapter.

2.1 A General Introduction to Hot-Carrier Effect

The supply voltage of an integrated circuit was often fixed at 5 V in the past. Although it has been scaled down recently, the rate of reduction is still slower than that of the transistor gate length. As a result, a high channel electric field exists in an ultra-short channel (L) transistor, especially at the drain region of the transistor in the saturation mode. This high electric field accelerates carriers to energy much higher

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

than their normal thermal energy and these carriers are therefore termed as 'hot carriers' [20]. Hot electrons are usually referred to as 'lucky electrons' or impact-ionization (II) generated electrons [21] which gain sufficient energy ('hot' enough) to surmount the Si-SiO_2 energy barrier without suffering or just suffering very few energy-losing collisions along their paths. Various phenomena caused by the generation and injection of hot carriers in scaled MOSFETs are collectively referred to as 'hot-carrier effect' [20] [21].

HCE can be monitored by a shift in threshold voltage (V_t), degradation in drain current (I_d) or in transconductance (G_m). Besides device degradation, there are many other manifestations of hot carriers in a Flash memory cell, such as the substrate current (I_{sub}) induced by primary impact ionization (PII), the light emission (I_{coll}), the hot-electron gate current (I_g) as well as the creation of oxide (N_{ot}) and interface traps (N_{it}). This is illustrated in Fig. 2.1.

Any of the above hot-electron effect could set scaling limit for a Flash memory cell, but some could provide useful functions. For example, device lifetime or degradation can be predicted by I_{sub}; the programming current and defects generated in the tunnel oxide and at the Si-SiO_2 interface of a floating-gate memory cell can be estimated by I_g. Moreover, HCI can be used for floating-gate memory programming. Several mechanisms for hot-carrier generation and injection have been reported in the literature. These include channel hot-electron injection [20], drain avalanche

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

hot-carrier (DAHC) injection [22], secondary generated hot-electron (SGHE) [23] and substrate hot-electron (SHE) injection [24]. In spite of the numerous HCI mechanisms available, only CHE and CHISEL (a derivative of SGHE) injections are commonly utilized for Flash memory programming [25] [26], in view of their good programming efficiency and less severe defects introduced during the injection process.

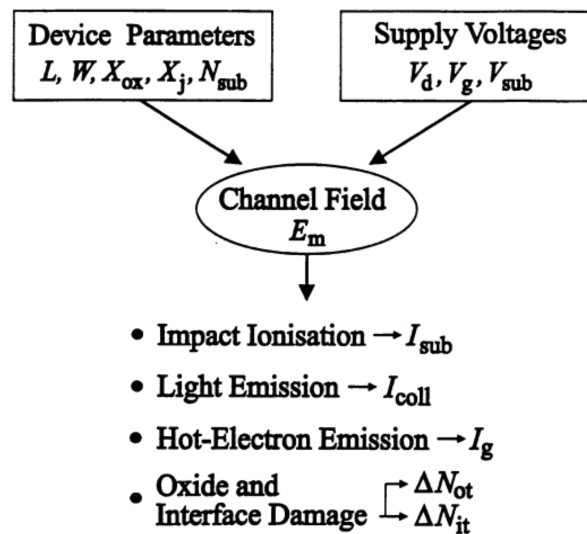


Fig. 2.1 A conceptual view of hot-carrier reliability issue [27] [28].

Extensive research has been carried out to understand, characterize and reduce the hot-carrier effect and the susceptibility of transistor to hot-carrier injection induced degradation since the early 1980s. This topic is still very important today due to the mass production and application of deep sub-micrometer floating-gate memories, in which the hot-electron injection acts as a ‘double-sided sword’. This is because in a HCI programming process, a cell is programmed by hot-electron injection. Meanwhile, the injection of electrons can lead to trap generation, which deteriorates the integrity of tunnel oxide and Si-SiO_2 interface, and decreases device lifetime.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

The dependence of HCE on various design parameters, such as channel and source/drain doping profile [29] [30], has also been investigated. Several empirical or analytical HCI models are proposed, most of which are attempting to simulate the hot-carrier injection behavior and/or to provide a way to estimate device hot-carrier stress lifetime. Advances in this research field will be discussed in the following sections.

2.2 Classical Hot-Carrier Injection Mechanisms and Related Hot Carrier Effect

Fig. 2.2 illustrates several mechanisms which give rise to hot-carrier effect in a conventional N-channel MOSFET. As we can see, these mechanisms include generation, injection and trapping of hot electrons and holes.

Device failure caused by HCE can be generally classified into two groups. The first group can be ascribed to the generation and injection of hot holes. Electrons are easily heated up when traveling along the channel, gaining enough energy to cause impact ionization at the drain-end. Electron-hole pairs (EHPs) are generated in this high field region, which not only increases the drain current by providing excess electrons but also enhances the substrate current (I_{sub}) since more hot holes are collected at the back contact. I_{sub} can overload the substrate-bias generator and cause local substrate potential fluctuations, inducing CMOS latch-up and/or snap-back breakdown.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

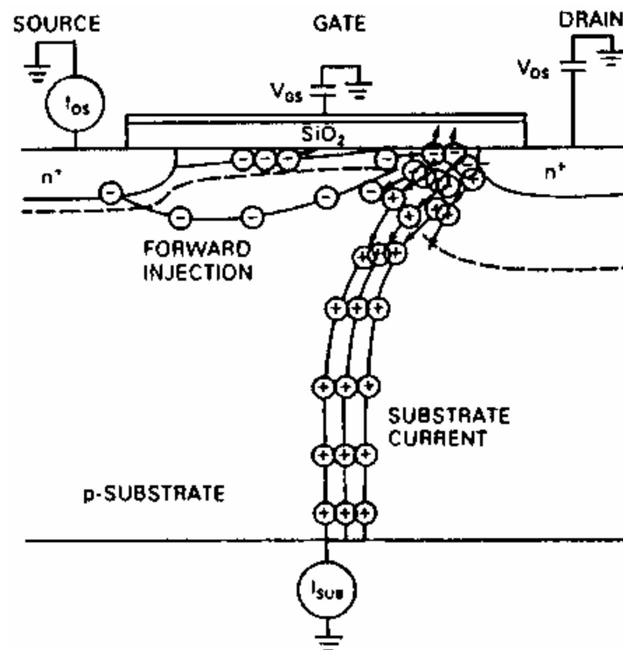


Fig. 2.2 A schematic diagram of an N-channel MOSFET showing various hot-carrier mechanisms that can induce the hot-carrier effect [31].

The second failure group arises as hot electrons in the channel surmount the Si-SiO_2 potential barrier. Hot-electron injection to the gate electrode results in negative charge trapping in the gate oxide. Interface states are also created which can degrade the quality of the Si-SiO_2 interface. As a result, an increment in gate leakage occurs, which may eventually lead to oxide breakdown. For the abovementioned failure modes, some are induced by a single HCI mechanism while the others may be the combined result of several concurrent degradation mechanisms.

2.2.1 Channel hot-electron (CHE) injection

As mentioned earlier, channel hot-electron (CHE) injection is due to the emission

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

of lucky electrons from the channel into the gate oxide, causing a significant degradation of the oxide and the Si-SiO_2 interface. This is as depicted in Fig. 2.3.

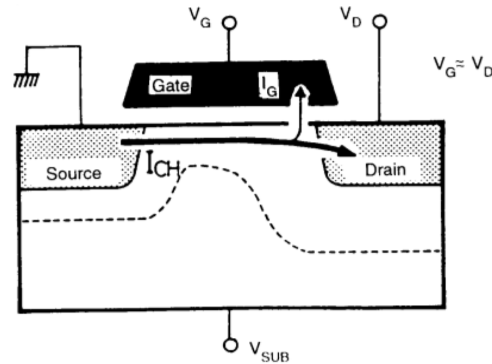


Fig. 2.3 An illustration of channel hot-electron injection in an n-MOSFET. The channel electrons are deflected from the inversion layer towards the gate oxide near the drain-end [32]. As a consequence, a non-negligible gate current is measured.

CHE injection will dominate the HCI process when a device is biased at the same gate and drain voltages ($V_g \sim V_d$). This biasing condition actually corresponds to a maximum injection of lucky electrons, which give rise to a peak electron gate current. As this is an optimum biasing scheme for lucky-electron injection [33], $V_g \sim V_d$ is widely adopted for memory programming, in order to generate the highest electron gate current (I_g) and injection efficiency (I_g/I_d) for a given V_d. Under this condition, channel electrons are often distributed, on average, closer to the interface than biasing at peak substrate current ($I_{b,max}$) [34]. In addition, the positive oxide field in the channel region attracts channel electrons towards the interface, enhancing the injection of electrons towards the gate [34]. However, a rapid decrement of electron injection is observed as moving towards the source side, due to the reduction of average electron energy.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

The Lucky-Electron Model (LEM) has been applied to simulate the gate current over a wide range of channel lengths and voltages. A simple interpretation of this model would imply a complete annihilation of charge injection for V_{ds} < 3.2 V. However, recent experimental studies have shown that I_g can still be measured for V_{ds} as low as 1.75 V [35]. This means that channel electrons can possess energy higher than qV_{ds}, and the strict requirements as claimed in lucky-electron concept are not accurate. Therefore, alternative models have been proposed to explain the sub-3 V hot-electron injection, and will be discussed in Section 2.4.2 to Section 2.4.4.

Fig. 2.4 depicts typical gate current characteristics with the drain voltage V_{ds} as the parameter. As we can see, peak I_g always occurs at $V_{gs} \sim V_{ds}$, indicating the maximum injection of lucky electrons in the CHE injection mode. At small gate bias ($V_{gs} < V_{ds}$), only a small I_g can be measured. This may be attributed to the opposing oxide electric field (E_{ox}), which can strongly suppress electron injection in the drain end, although a large lateral field (due to high V_{ds}) exists. As V_{gs} increases, the average lateral field between the mid-channel region and the drain decreases, and the spatial distribution of hot electrons is also shifted towards the drain side. As a result, part of the distribution of hot electrons is positioned in the high field region, leading to an exponential increase in the injection current. Eventually at $V_{gs} \sim V_{ds}$, the injection of hot electrons reaches its peak value. When $V_{gs} > V_{ds}$, E_{ox} at the drain end becomes favorable to electron injection as it transits from negative (repulsive) to positive (attractive). However, the high energy tail of the hot-electron distribution is subjected to the

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

average lateral field, which decreases as V_{gs} becomes larger. Therefore, a fast reduction of I_g with increasing V_{gs} is observed for $V_{gs} > V_{ds}$.

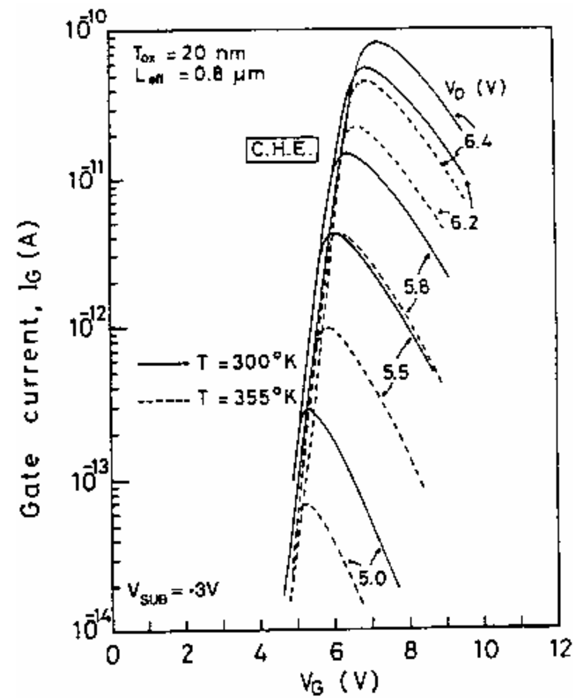


Fig. 2.4 Gate currents due to CHE injection in an n-MOSFET [32].

2.2.2 Drain avalanche hot-carrier (DAHC) injection

The gate current in regime $V_{gs} < V_{ds}$ down to $V_{gs} \sim V_{th}$ is often considered to be dominated by drain avalanche hot-carrier (DAHC) injection. Fig. 2.5 depicts the physical mechanism behind DAHC injection in an N-channel MOS transistor.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

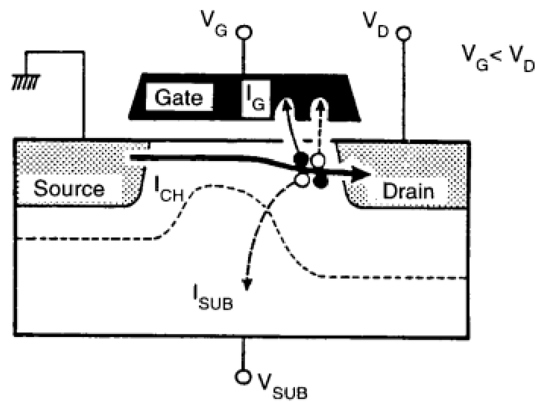


Fig. 2.5 Schematic illustration of the drain avalanche hot-carrier injection (DAHC) mechanism in an N-channel MOSFET. The hot electrons here originate from the primary impact ionization process at the drain-end [32].

Fig. 2.6 shows the gate current (I_g) which results from the injection of avalanche carriers when an N-channel device is operated in the deep-saturation regime. At $V_g \sim V_d$, there is only one peak I_g associated with the CHE injection mode. At $V_g < V_d$, however, there are two peaks of I_g present in the I_g versus V_g characteristics. These two peaks essentially correspond to the maximum injection of hot holes and hot electrons generated by the avalanche multiplication at the drain side.

As two types of hot carriers take part in the DAHC injection, more interface and oxide traps are generated [36]. Moreover, a direct analysis for DAHC injection becomes difficult as it is hard to tell whether a trap created inside the gate oxide or at the interface is caused by the injection of hot holes or electrons or both. As V_g is reduced further, hot-hole injection increases while hot-electron injection decreases. When the two effects balance out, the net result is almost zero I_g.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

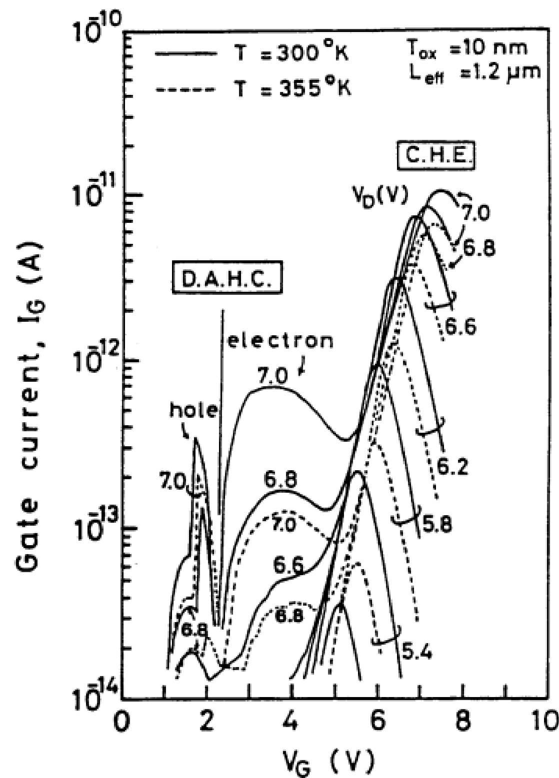


Fig. 2.6 Gate current characteristics of an n-MOSFET. In addition to channel hot-electron injection, this plot also features the much smaller injection current attributed to drain-avalanche hot-carrier injection [32].

When V_g is approaching the threshold voltage (V_t), carrier heating increases while electron injection is suppressed by the opposing E_{ox}, only the hole current (negative) is measured at the gate electrode. It is often orders of magnitude smaller than the CHE current because of: *1)* a larger barrier height for hole injection (~ 4 – 5 eV) as compared to that for electron injection (~3.2 eV); *2)* a larger scattering rate (short mean-free-path) that limits hole heating; and *3)* a sudden drop in the hole density of states at energies comparable to the hole barrier height.

In contrast, when V_g increases and approaches V_d, the hole current will vanish

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

quickly and the electron current begins to dominate due to the increase of an attractive vertical field.

2.2.3 Substrate hot-electron (SHE) injection

In contrast to CHE and DAHC injection where hot-carrier injection is localized at the drain end, substrate hot-electron (SHE) injection occurs more uniformly across the entire channel region. When an N-channel transistor is biased at $V_{gs} > V_t$ and $V_{ds} = 0$ under a large negative substrate voltage ($V_{bs} < 0$), a wide depletion region below the gate is created. Under this condition, a one-dimensional (1D) electric field that is perpendicular to the interface is formed below the gate region. The transport of hot electrons within Si substrate also becomes 1D, as illustrated in Fig. 2.7.

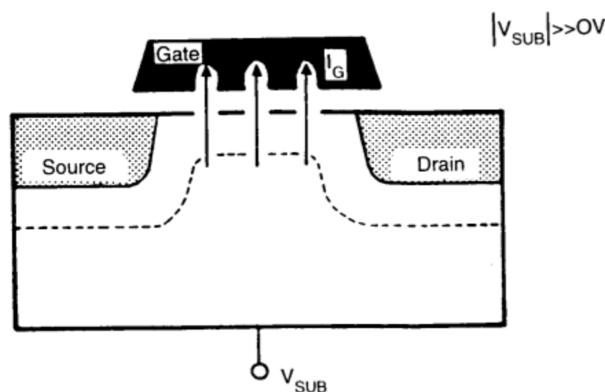


Fig. 2.7 A schematic diagram of an N-channel MOSFET, showing uniform hot-electron injection over the entire channel region under the SHE injection mode [32].

In view of a simpler analysis for uniform electron and hole trapping in SiO_2, SHE injection is often employed to evaluate dielectric quality and to study hot-carrier

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

effect even though SHE is not present in most circuits other than the bootstrap circuit.

In the SHE injection mode, electrons for injection can be produced by band-to-band tunneling or by forward-biasing a buried junction [37]. In the latter case, a buried N-well is situated inside a P-type substrate for the N-MOS transistor. By forward biasing the N-well/P-substrate junction, electrons injected into the P-type substrate are driven to the junction interface through diffusion. Upon entering the depletion region, they are accelerated towards the Si/SiO_2 by the substrate electric field. Eventually, they will either attain enough energy to be injected into the gate oxide, or be collected at the source and drain terminals.

Since the substrate bias controls the potential drop in the substrate depletion region, a larger V_{sb} can improve the injection efficiency (I_g/I_d) as the high energy tail of the distribution is enhanced. For a given V_{sb}, the hole current is smaller than the electron current due to a shorter mean-free-path and a higher Si-SiO_2 barrier for hole injection.

2.2.4 Secondary generated hot-electron (SGHE) injection

Secondary generated hot-electron (SGHE) injection previously is a collective term for a few mechanisms by which electrons created inside the depletion region below the gate are injected into the gate oxide to contribute an additional (sometimes dominant) component of the gate current. The mechanisms which may be responsible

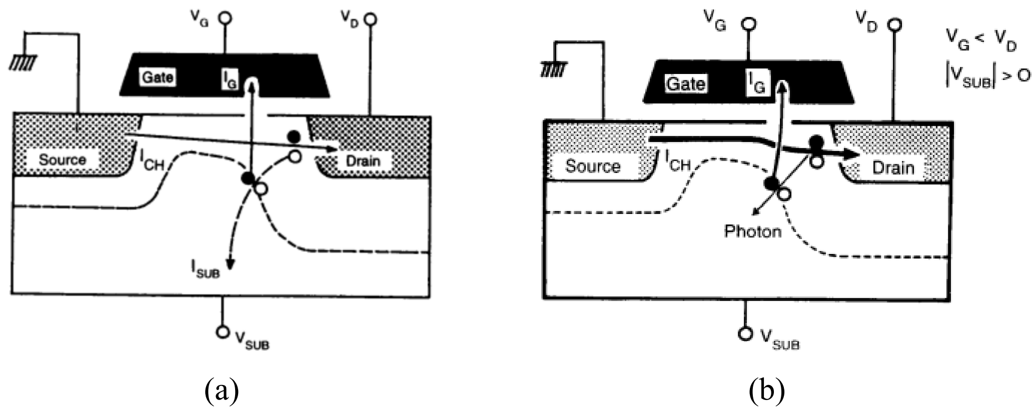
ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

for SGHE injection are listed as follows [38]:

- ***Bremsstrahlung radiation*** [39] (Fig. 2.8(a)): The photons emitted by the relaxation of channel hot carriers are re-absorbed in the bulk, creating additional hot-carrier pairs. Some of the electrons can diffuse to the edge of depletion region where they are accelerated towards the gate oxide.
- ***Band-to-band tunneling*** at the drain junction in heavily doped substrates [40].
- ***Thermal generation*** inside the depletion region or over the junction barrier [41].
- ***Secondary impact ionization*** [78] (Fig. 2.8(b)): holes generated by impact ionization in the proximity of the interface can re-ionize while drifting towards the substrate, creating tertiary electrons deep inside the depletion region. This mechanism can be considered as the ***precursor for the CHISEL injection***, which is basically the injection of energetic tertiary electrons induced by the vertical feedback heating of holes upon the application of a reverse-biased substrate voltage.

Unlike CHE and DAHC injections, which are more localized at the point of maximum channel field near the drain end, SGHE injection can occur within the substrate depletion region. Consequently, the gate current (I_g) which results from the SGHE injection is different from that generated in DAHC or CHE mode. It strongly depends on the back-bias voltage (V_{sb}) and will become prominent even for small V_g (< 5V).

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library



(a) (b)

Fig. 2.8 (a) Substrate current induced hot-electron injection mechanism. V_g<V_d, $|V_{sub}|$>0. (b) Secondary hot-electron injection due to photon emission and re-absorption [32].

In the SGHE injection regime, the drain-substrate potential difference (V_{db}) is larger than the potential drop over the channel (V_{ds}) by roughly an amount of V_{sb}, thus the maximum energy available to the substrate-generated carriers is larger than that of the channel electrons. Consequently, SGHE effects are best observed at low bias condition. The SGHE injection will become progressively more important as the supply voltages reduce and the drop of available potential energy along the channel becomes the limiting factor for CHE injection.

The importance of SGHE and its influence on device reliability performance are expected to increase as the technology scales down. This is because the triggering mechanisms of SGHE as listed above can be readily achieved in future scaled CMOS devices, especially for modern Flash memory, where higher electric field, heavier channel dopant concentration and more abrupt doping gradient are its typical characteristics.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

2.2.5 Dependence of device degradation on HCI stress condition

The classical concept of hot-carrier effect induced device degradation was that hole/electron trapping in the gate oxide and the interface states resulted in experimentally observable changes in threshold voltage and transconductance. Thus, it was believed that by monitoring the efficiency for charge trapping and interface-state generation, the degradation characteristics of MOSFET could be determined. Hot-carrier induced degradation of the N-channel transistor has been studied much more intensively than that for its P-channel counterpart, as NMOS technology was dominant since 1980s by virtue of its higher packing density and higher switching speed. In addition, the impact-ionization rate of electrons is one to two orders of magnitude higher than that of holes, and the barrier height for electron injection (~ 3.2 eV) is also much lower than that for holes (~ 4.7 eV). As a result, more serious oxide damage can be created by hot-electron injection in N-channel devices [31]. Nevertheless, such intensive studies in the early 1980s did not lead to an unanimous agreement on the degradation mechanisms. This may be attributed to the shortage of reliable and sensitive technique to evaluate the interface damage caused by the hot-carrier injection and the lack of understanding on different degradation mechanisms caused by a wide range of stress conditions.

Later on, with the realization of a bias-dependent hot-carrier effect, studies became more focused and they shed more light on the physics and kinetics of the hot-carrier injection induced device degradation. Hofmann *et al.* [42] (1985), Doyle *et*

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

al. [43] (1987), Tsuchiya *et al.* [44] (1987) and Schwerin *et al.* [45] (1987) presented detailed studies on the gate and drain voltage dependency of hot-carrier effect through a series of experiments and simulations, revealing the influence of stress voltage on the nature of hot-carrier induced degradation. According to their studies, hot-hole injection was the dominant degradation mechanism at low gate bias condition ($V_g < V_d/2$), where the positively trapped charge can mask the effect of acceptor-like interface states. As V_g approaches $V_d/2$, hot-electron injection begins to play an important role in neutralizing the positively charged trap centers, significantly enhancing the impact of the acceptor-like interface states on the transistor characteristics. At $V_g \sim V_d/2$, peak substrate current ($I_{b,max}$) appears, denoting a worst-case primary impact ionization and hot-electron generation; hence, this biasing condition is traditionally believed to be the 'worst-case' stress condition for hot-carrier induced devices degradation.

In 1990, an improved understanding of HCE degradation was achieved by Doyle *et al.* [46]. They claimed that in addition to hole trapping and generation of acceptor-like interface states, creation of neutral electron traps was also possible for N-channel device stressed under low gate bias condition. At higher gate bias ($V_g \sim V_d$) where electron injection is dominant, these neutral electron traps are negatively charged, leading to an increased degradation of the N-channel device. This finding points out that the worst-case stress condition for N-MOSFETs can shift from $V_g \sim V_d/2$ to $V_g \sim V_d$, as the transistor channel length is scaled down towards 0.1 µm in more

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

advanced CMOS technology.

In 2001, Li *et al.* verified the above finding by employing a more reliable charge pumping technique. They observed that for an N-channel device with L_{eff} = 0.15 μm, biasing at $V_g \sim V_d$ yielded the largest increment in the charge pumping current, indicating the worst-case stress condition for interface trap generation and therefore the most severe device degradation. They explained that it was because the device degradation was not only controlled by the number of hot carriers available, which is the greatest at $I_{b,max}$, but also by the primary injection location; only energetic carriers bombarding the Si-SiO_2 interface can cause device degradation. Under $I_{b,max}$, although there are a lot of hot carriers near the drain, they are directed away from the interface by the vertical electric field. In contrast, at $V_g \sim V_d$, channel electrons are distributed, on average, closer to the interface than those bias at $I_{b,max}$. As a result, injection of energetic electrons is enhanced, leading to more severe Si-SiO_2 interface damage thus greater device degradation.

In 2003, Ang *et al.* reported an experimental evidence based on modulating the concentration of the high-energy tail of the electron energy distribution function [47] via the application of a reverse-biased V_b during hot-carrier stress. It is revealed that the shift in the worse-case stress condition from $V_g \sim V_d/2$ towards V_d should be mainly caused by the increased injection of the 'high-energy tail' electrons into the gate oxide as V_g approaches $V_{d.}$ Generation of those hot electrons occurs through

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

non-local effect, such as electron-electron scattering, or impact ionization feedback [48].

2.3 Techniques to Suppress Hot-Carrier Effect

The reliability problems caused by HCE in the MOSFET have led to the development of various techniques to suppress them. As hot-carrier effect is due to high channel-field induced carrier injection, one apparent solution to suppress HCI is by modifying the drain structure, where the channel field is the highest, so as to reduce the injection electric field. A LDD (lightly-doped drain) or DDD (double-diffused drain) structure is usually adopted for this purpose [49], whereby the maximum channel field can be reduced by as much as 20% [50]. Since both the substrate current and electron injection depends exponentially on the electric field, the decrease in channel field is reflected as a large reduction in HCE.

Dielectric engineering is another way to suppress HCE. Devices with reoxidized-nitrided-oxide (RNO) were found to be superior to conventional oxide devices [51]. The oxynitride (Si_xN_yO) showed improved stability and reduced degradation under both AC and DC hot-carrier stresses. This is because the additional re-oxidation steps can effectively reduce the number of electron traps introduced in the nitrided gate oxide. Moreover, the Si-N bond is stronger than the Si-O bond, which renders the former more immune to breakage by hot-electron bombardment.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

2.4 Hot-Electron Gate Current Models

Gate current (I_g) is an important parameter to determine the intensity of electron injection as well as the programming efficiency; at the same time, it is also identified as a sensitive measure for carrier-injection related oxide defect generation. Thus, it has been studied intensively and various models have been proposed in literature. However, due to a complex two-dimensional (2D) nature of the phenomenon and the many unknown physical parameters, there is no single close-form analytical expression for the hot-electron injection current, unlike the 1D FN-tunneling case. Those HCI models are mainly qualitative and will be discussed in this section.

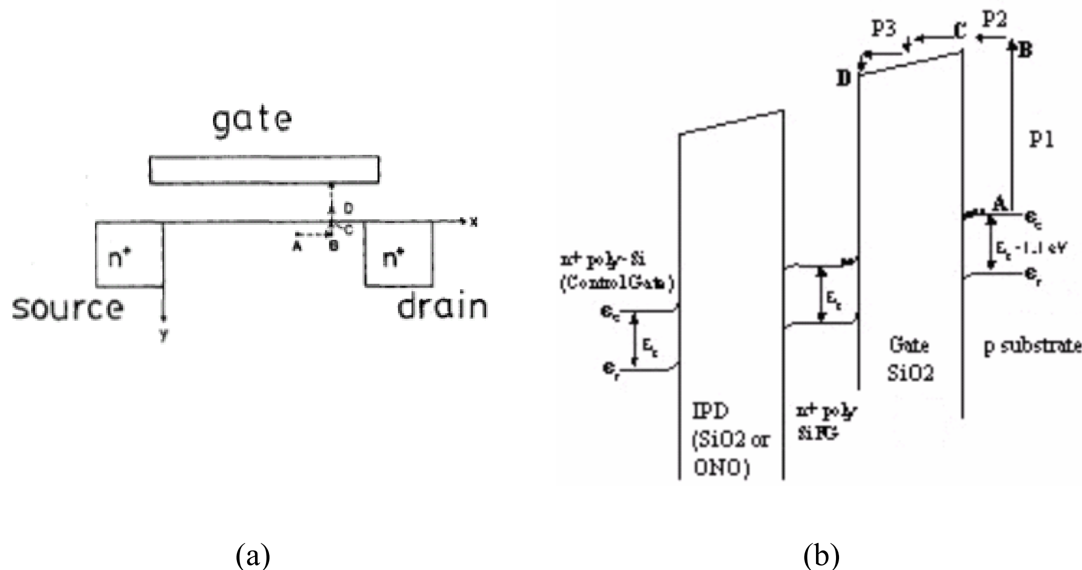
2.4.1 Classical lucky-electron model (LEM)

The idea of applying Lucky-Electron Model (LEM) to simulate hot-electron injection current was originally proposed by Shockley [52]. Verwey *et al.* adopted this concept in their study of substrate hot-electron injection in MOSFETs [53] which was subsequently verified by Ning *et al*. [54]. Hu *et al.* in 1979 [55] and Ko *et al.* in 1981 [56] refined this model and applied it to CHE injection in the N-channel MOSFET.

Conceptually, LEM can be described as follows [57]. In order to surmount the Si/SiO_2 energy barrier and reach the gate electrode, the channel electrons must gain sufficient kinetic energy from the lateral field ($E_{//}$) and possess a vertical momentum directed towards the Si-SiO_2 interface. Fig. 2.9(a) illustrates the three steps of CHE

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

injection by applying the lucky-electron concept. The corresponding energy band diagram was shown in Fig. 2.9(b).



(a) (b)

Fig. 2.9 (a) A cross-sectional view of an N-channel MOSFET showing the three scattering possibilities in the LEM. (b) The energy band diagram of the three-step CHE injection process in the LEM [57][58].

- ***Stage I (from A to B)***

A channel electron gains energy from the channel field $E_{//}$ and becomes 'hot'. If we assume the accelerating electric-field to be constant, the probability that a channel electron can travel a distance *d* without suffering any collision to acquire kinetic energy equal to the Si-SiO_2 barrier height Φ_B can be written as $e^{(-d/\lambda)}$ or $e^{(-\Phi_B/E_{//}\lambda)}$ [34], where λ is the scattering mean-free-path (MFP) of the hot electron. Then, a large vertical momentum due to elastic phonon scattering re-directs the hot electron towards the Si-SiO_2 interface. Assuming isotropic scattering, the overall probability (P_1) for an electron to acquire enough kinetic energy ($\Phi = \Phi_B + \Delta\Phi$) as well as sufficient normal

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

momentum in order to surmount the Si-SiO_2 potential barrier height can be expressed as [59]:

$$P_1 = \int_{\Delta\Phi=0}^{\Delta\Phi=\infty} \frac{\Delta\Phi}{4\Phi_B} e^{-[(\Phi_B+\Delta\Phi)/E_{//}\lambda]} \frac{d(\Delta\Phi)}{E_{//}\lambda} = 0.25 \frac{E_{//}}{\lambda} e^{-(\Phi_B/E_{//}\lambda)} \qquad (2.1)$$

- ***Stage II (from B to C, where C is situated at the interface)***

Once a hot electron is re-directed, it must not suffer any 'energy-robbing' collision so as to retain the energy required to surmount the potential barrier, assuming the electron will lose most of its kinetic energy once it encounters any inelastic collision. The possibility for a hot electron traveling to the Si-SiO_2 interface without suffering any inelastic collision after the re-direction is defined as P_2, which can be weighted by the electron concentration in the inversion layer. Defining *n(y)* as the electron concentration at depth *y* and position *x* in the channel, and assuming strong inversion and gradual channel approximation, P_2 can be approximated as [57]:

$$P_2 = \frac{\int_{y=0}^{y=\infty} n(y) e^{-(y/\lambda)} dy}{\int_{y=0}^{y=\infty} n(y) dy} = 1 - \alpha e^{\alpha} E_1(\alpha) \qquad (2.2)$$

where $\alpha = \frac{6kT}{q\lambda E_{ox}}$. Here, $E_{ox} \approx \frac{V_{gs} - V_{ds}}{t_{ox}}$, T, t_{ox} and E_1 are the absolute temperature, the gate oxide thickness and the exponential integral, respectively.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

- ***Stage III (from C to D)***

The last factor that needs to be considered is scattering in the oxide image-potential well. When a hot electron transits from the Si-SiO_2 interface to the floating gate, it must not suffer any collision in the oxide image potential well located between C and D; once a hot electron arrives at location D, it will be swept towards the gate electrode by the aiding oxide field. The probability associated with this process is P_3 and is defined as the probability of an electron suffering no collision in the oxide image potential well, which can be written as [60]:

$$P_3 = e^{(-y_0/\lambda_{ox})} = e^{(-300/\sqrt{E_{ox}})} \tag{2.3}$$

where $y_0 = \sqrt{\dfrac{q}{16\pi E_{ox}\varepsilon_{ox}}}$, $E_{ox} \approx \dfrac{V_{gs} - V_{ds}}{t_{ox}}$ and $\lambda_{ox} = 3.2\text{nm}$.

In summary, the LEM assumes that an electron can inject into the floating gate only if: *(1)* It can gain enough energy in the large lateral field without undergoing a significant collision and can be re-directed by an elastic phonon scattering towards the Si-SiO_2 interface (P_1); *(2)* it can reach the Si-SiO_2 interface and have enough energy to surmount the barrier (P_2); *(3)* it is not scattered by the oxide image potential well when it transits from the interface to floating gate (P_3). In view of all the above processes, the overall probability for gate injection is the product of each individual probability factor. Therefore, the gate current can be expressed as [57]:

$$I_g = \int_0^{L_{eff}} \frac{I_{ds} P_1 P_2 P_3}{\lambda_r} dx \approx I_{ds} \frac{\Delta L}{\lambda_r} [P_1 P_2 P_3]_{\max} \tag{2.4}$$

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

where L_{eff} is the effective channel length of the floating-gate transistor, λ_r is the re-direction scattering MFP (~ 92 nm) and ΔL is the length of the drain side region where most significant CHE injection occurs.

The central theme of LEM is that the potential profile in the channel is the basic driving force for hot electron emission, so the potential difference over the channel (i.e. V_{ds}) must be at least 3 V [55] for CHE injection to take place. If V_{ds} is less than 3 V, there should be negligible hot-electron injection, regardless of how short the channel length or how shallow the junction depth is.

2.4.2 Lattice-temperature assisted hot-carrier injection model

Although the classical LEM predicts a complete absence of hot-electron induced gate current and interface-state generation when V_{ds} is below 3.7 V, many researchers have reported the occurrence of hot-electron induced degradation at drain biases well below this voltage [61][62][63]. The simplest mechanism capable to provide excess energy to channel carriers is the net phonon absorption from lattice, which can generate an electron-energy-distribution function featuring a thermal tail with slope equal to the lattice temperature (T_L) for $E > qV_{ds}$ [64].

Eitan *et al*. [65] suggested that in the case of small potential drop along the channel (i.e. small V_{ds}), the initial energy distribution of electrons within the lattice in thermal

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

equilibrium should be taken into account, besides the energy gain from electric field. When the potential drop in the drain depletion region is less than the impact ionization threshold, channel hot-electron injection will exhibit an exponential dependence on the initial thermal distribution. This initial thermal distribution can result in a positive temperature dependence of impact ionization at low drain voltages. Su *et al*. [66] reported a thermally-assisted impact ionization mechanism which was believed to be activated by lattice temperature. It can further enhance hot-electron injection at low drain bias condition even when the drain bias is in the sub-bandgap ($V_{ds} < 1.2$ eV) regime.

2.4.3 Electron-electron scattering (EES) model and the refined effective electron-temperature (EET) model

Electron-electron scattering (EES) has been proposed to be another possible energy-gain mechanism which provides additional energy to the channel electrons for sub-3V electron injection [67,68,69,70,71,72]. The role of EES in heating the electrons above energy qV_{ds} was mainly studied by means of simulation in the literature. The simulation results tend to predict a faster decay of the electron-energy distribution without considering the EES mechanism [73]. This is because electron-electron interaction near the drain edge can transfer excess energy to some electrons, and it is believed to be the dominant energy-gain mechanism responsible for the 'high-energy tail' in the electron energy (E_e) distribution under low-voltage condition. Hence, CHE injection can still occur in the sub-3V regime. EES can also

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

lead to a 'spatial retardation' of the maximum hot-electron injection point with respect to the peak of the channel electric field [74], causing interface state generation to take place after the maximum lateral field point. In fact, considering EES to be simply a means of thermalizing the electron-energy-distribution function (EEDF) at carrier energy above the applied drain voltage would lead one to conclude the continued validity of the Lucky-Electron Model.

However, several authors have come to the conclusion that the EES mechanism is too weak to completely thermalize the EEDF in a high electric-field region. Instead, an effective electron-temperature (EET) model should be employed as proposed by Hess and Sah [75]. This model is considered an extension of the LEM for a non-thermal EEDF. It is assumed that the EEDF is Maxwellian with an effective temperature T_e that can be determined for any given field strength. Based on this argument, the electron mobility and concentration as well as the eventual device current can be obtained. Takeda *et al.* [76] applied this model to calculate gate current as the thermionic emission of heated electrons over the interface barrier. In their work, electrons at the high electric field region near the drain-end are considered as heated electron-gas with 2D characteristics, and may be expressed as:

$$I_g^{\,e} \propto I_s \exp(-\frac{\Phi_B}{kT_e}) \tag{2.5}$$

where I_s, Φ_B and T_e are source current, barrier height for the injection of electrons into the gate oxide and electron effective temperature, respectively. In contrast to the

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

empirical Lucky-Electron Model, the gate current was numerically modeled as the thermionic emission from heated electron gas over the Si-SiO_2 energy barrier, so the gate current density is expressed as

$$J_g = qn_s(\frac{kT_e}{2\pi m^*})^{1/2} \exp(\frac{-\Phi_B}{kT_e}) \quad (2.6)$$

where m^* is the effective mass of electron. Surface electron density (n_s) and electric field (E) were calculated by a 2D analysis program. T_e was obtained by assuming the displaced Maxwellian distribution of electron gas as expressed in equation (2.7) below, where E_x is x-component of the electric field, v_s is saturated electron velocity parallel to the surface and τ_ε is energy relaxation time.

$$T_e(x) = \frac{2q}{5k}\int_0^\infty E_x(x-u)\exp(\frac{-3u}{5\tau_\varepsilon v_s})du \quad (2.7)$$

Therefore, I_g can be obtained as

$$I_g = W\int_0^{\Delta L} J_g dy \quad (W\text{: channel width}) \quad (2.8)$$

2.4.4 Impact-ionization feedback (IIF) model

As technology develops, typical voltages for memory programming are reduced along with device dimension. Although CHE injection has become the most common programming scheme for Flash memory, the high gate and drain voltages needed to generate sufficient hot-electron injection current are problematic for deep sub-micrometer memory operation. This has led to a stringent demand for a

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

low-voltage hot-electron injection scheme which can achieve high programming efficiency while consume relatively low electrical power.

As mentioned earlier, a novel hot-electron injection mechanism, called channel-initiated secondary electron (CHISEL) injection, is recently reported to be a better alternative for low-power memory applications (V_{ds} < 2.5 V) due to its better tradeoff between programming efficiency and device reliability [77,78,79,80]. It is activated upon a reverse body bias (i.e. negative V_b for n-MOSFETs), which creates energetic tertiary electrons in the bulk region, via impact-ionization feedback (IIF) [78] – [80] (Fig. 2.10), and provides additional energy for subsequent electron injection.

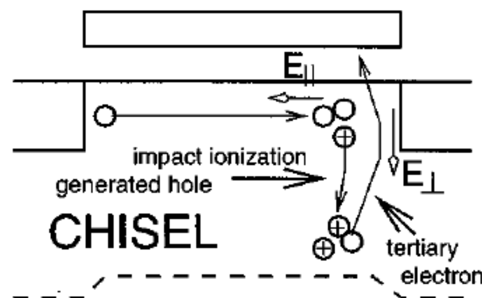


Fig. 2.10 Schematic representation of CHISEL injection in an n-MOSFET by means of impact-ionization feedback of hot holes [80] ($E_{//}$ and $E_{\perp}$ denote the lateral field along the channel and vertical field at the drain end, respectively.)

To induce the IIF process, an N-channel floating-gate transistor has to be operated at moderate gate/drain biasing condition firstly, and a negative voltage is applied at the substrate contact. The lateral field ($E_{//}$) heats up channel electrons and some energetic electrons gain sufficient energy to cause primary impact ionization (PII) in the drain depletion region, in which a large number of hot electron-hole pairs (EHPs)

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

are created. Most of the PII induced electrons are collected at the drain terminal and contribute to the drain current, while a small fraction of them are injected towards the tunnel-oxide and are trapped in the floating gate. This could result in damage to the gate oxide and interface. On the other hand, the induced hot holes are driven towards the bulk contact by a large potential difference established by the positively biased drain and the negatively biased substrate, constituting the substrate current (I_{sub}). Some holes which become extremely energetic due to the vertical heating field ($E_{\perp}$) can induce a secondary impact ionization (SII) in the bulk region, which creates tertiary electron-hole pairs. Tertiary holes will turn out to be part of the I_{sub} and leave through the substrate terminal, while tertiary electrons can flow back to the Si-SiO_2 interface and inject towards the floating-gate under the influence of a high $E_{\perp}$. This feedback process continues with tertiary electron ionization, leading to a multiplicative production of EHPs which enhances the programming current (I_g). Since tertiary electrons in CHISEL injection are created by hot holes arising from PII in the channel, their injection is regarded as 'channel initiated' and can be considered as a 'feedback' of the PII process.

It has been shown by experiments and simulation [81] [82] that tertiary electrons possess higher energies than those deflected electrons in the CHE injection mode, due to additional heating by the reverse body bias. Moreover, as the injection of tertiary electrons tends to be more distributed in the channel rather than localized at the drain-end, interface states created by CHISEL at the drain depletion region are not as

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

severe as that in the CHE injection. Consequently, better cell reliability in terms of data retention and cycling endurance at comparable programming efficiency can be achieved [82] [83] using the CHISEL programming scheme. Nevertheless, the wider damage region may result in larger channel roughness adjacent to the Si-SiO_2 interface, which degrades the carrier effective mobility and hence the drain current more severely.

Although the IIF-induced injection is initiated by channel electrons, carriers heated by the channel field themselves do not directly contribute to the measured I_g. Therefore, the Lucky-Electron Model which correlates the increase in electron injection efficiency purely to the increment of impact ionization for a fixed V_{gd} is not directly applicable here. Nevertheless, an increment in V_d will indeed enhance the primary impact ionization rate and generate more secondary EHPs, which subsequently leads to a stronger secondary impact ionization, and creating more tertiary electrons for CHISEL injection.

The impact ionization feedback model for CHISEL injection was first suggested by Bude *et al.* [77] [78], with the expression for hot-electron gate current given as

$$I_g = I_{ds} \cdot M_1 \cdot M_2 \cdot P_{inj} \quad (2.9)$$

where M_1 and M_2 are current multiplication factors of the impact ionization process at drain-end and substrate, respectively. P_{inj} is the probability that a tertiary electron will be accelerated to the oxide interface and inject into the gate. This expression is rather

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

general and is not able to reveal physical parameters which are directly related to CHISEL injection. In order to relate the CHISEL gate current directly to channel heating process, Esseni and Selmi suggested another expression [83][84][85]:

$$I_{gr} = C_b \cdot I_{br} = T_b \cdot (M-1) \cdot I_{br} \tag{2.10}$$

where I_{gr} is defined as I_g/I_d and is considered as the normalized gate current or gate injection efficiency, while I_{br} is defined as I_{sub}/I_d and is regarded as the normalized substrate current or the channel heating efficiency. C_b is a product of injection probability T_b and multiplication factor M, and is a strong function of the drain voltage and the vertical field in the proximity of the drain junction. Comparatively, the model for classical CHE injection can be written as:

$$I_{gr} = C \cdot I_{br}^{\alpha(\mathrm{E_{OX}})} \tag{2.11}$$

where $\alpha(E_{ox}) = \Phi_B(E_{ox}) / \Phi_{II} \sim 2.5 - 3.5$, which is the ratio of Si-SiO_2 barrier height (~ 3.2 eV) to threshold energy for impact ionization (~ 1.12 eV). E_{ox} denotes oxide field at the point of maximum hot-electron injection.

According to equation (2.10) and (2.11), experimental signatures of CHISEL and CHE manifest themselves in the distinct values of the power-law exponent n (i.e. the characteristic slope of the bi-logarithm plot for I_{gr} versus I_{br}). A unitary slope (i.e. $n \sim 1$) indicates a CHISEL-dominated electron injection process, whereas $n \sim 2.5 - 3.5$ denotes a CHE-dominated injection process. This experimental feature is as illustrated

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

in Fig. 2.11.

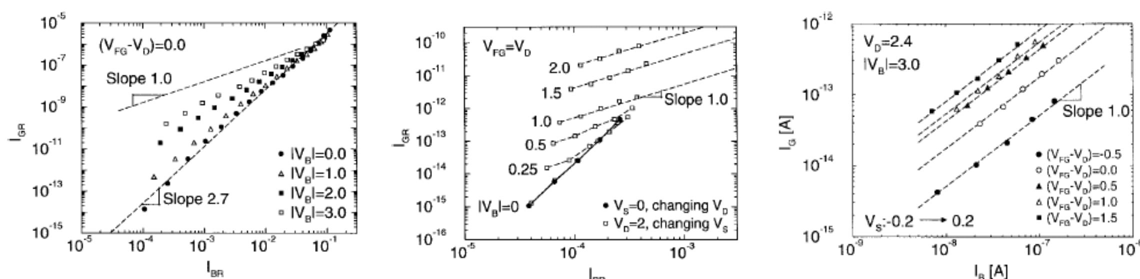


Fig. 2.11 Characteristic plots of I_{gr} versus I_{br} for different V_b. A large slope exists at $V_b = 0$ implying this biasing condition is dominated by CHE injection. As V_b becomes more negative, the slope gradually approaches 1, indicating the CHISEL injection becomes dominant at reversed-V_b condition [85].

The injection efficiency of the IIF-induced tertiary electrons is very sensitive to device design parameters. A thinner gate oxide, a shallower source/drain junction, and an optimized channel engineering scheme with heavy channel implantation dose can all enhance the significance of tertiary electron injection in a memory cell. Among all those parameters, channel dopant profiling is the most important factor to deep submicrometer Flash memory cells. This is because an abrupt channel doping profile, which is often realized by heavy halo/pocket implantation and heavy V_t-adjustment dose, is very common in aggressively scaled Flash memory cells in order to alleviate short-channel effect due to the use of thick tunnel-oxide for good data retention and endurance performance. It is also anticipated that a higher substrate doping concentration and a higher dose of punch-through prevention halos, as device feature size shrinks and supply voltage reduces, are likely to further enhance the IIF-induced tertiary-electron injection even at low or zero $|V_b|$ [86].

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

2.5 Summary

Down-scaling of MOS transistors to the deep sub-micrometer regime can improve the performance and packing density of non-volatile memory, but it also aggravates reliability problems. Hot-electron injection induced oxide defects is one of the most important reliability issues to be considered, as the injection of electrons acts as a 'double-sword' for memory performance. That is, on one hand, common programming schemes rely on hot-electron injection mechanisms (such as CHE and CHISEL injections); but on the other hand, severe oxide defect generation, and subsequent device degradation can be induced because of bombardment by energetic electrons. In addition, in modern floating-gate memory cell whose feature size is aggressively scaled down, an unscalable tunnel oxide due to a stringent requirement on data retention entails the use of heavy channel doping concentration for suppressing short-channel effect. This heavy channel doping scheme will also increase the susceptibility of memory cell to hot-electron effect unintentionally.

In this chapter, various hot-electron injection mechanisms which are commonly observed in MOS transistors are introduced. Several hot-electron injection mechanisms, such as CHE injection, DAHC injection, SHE injection and SGHE injection, are first presented. Among them, only CHE and CHISEL injections are used for Flash memory programming, in view of their better trade-off between the programming efficiency and potential reliability concerns. Next, common hot-carrier effect, including hot-hole substrate current, hot-electron gate current, and hot-electron

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

induced oxide defects and interface traps are briefly discussed. The worst-case hot-carrier stress condition for N-channel MOSFETs is reported to shift from $V_g \sim V_d/2$ to $V_g \sim V_d$, as transistor channel length scaled down to 0.1-μm in advanced CMOS technology. Several techniques that can suppress hot-electron effect, such as implementing LDD or DDD for drain engineering or adopting nitrided tunnel oxide in memory fabrication, are also discussed.

Furthermore, common hot-electron gate current models proposed by using different degradation-monitoring parameters and computer simulation are reviewed. The Lucky-Electron Model (LEM) forms a conceptual basis for hot-carrier lifetime prediction of MOS device. It suggests the potential profile in the channel is the basic driving force for electron injection, and the maximum energy that an electron can obtain is limited to the potential difference over the channel (i.e. V_{ds}). Thus, when V_{ds} falls below critical Si-SiO_2 barrier height (~ 3.2 eV), the hot-electron gate current should drop drastically, and virtually no hot electron emission can be present regardless of how the channel length or junction depth is reduced. As it cannot explain the presence of HCI in sub-3 V condition (i.e. when $V_{ds} < 3$ V), many researchers have looked for alternative mechanisms by which the channel electrons can attain excess energy and contribute to the electron gate injection even when the channel potential difference is less than the Si-SiO_2 barrier height.

The first alternative model proposed is the electron-electron scattering (EES)

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

model. From simulation result, it shows that EES can increase the high-energy tail of the electron-energy-distribution function and contribute to hot-electron injection at sub-3 V. This model was later refined to be the effective-electron-temperature (EET) model (also known as the ‘quasi-thermal equilibrium’ model [87]), which assumes the channel electron gas is quasi-equilibrium with the electric field at all times, and that the energy distribution of electrons can be characterized by an ‘effective electron temperature’ depending on the drift field. Based on this model, hot-electron effect will always be a potential problem as long as channel electrons can gain sufficient energy to initiate primary impact ionization at the drain end.

Recently, a new model called impact ionization feedback (IIF) is found to be able to provide the extra energy needed for sub-3 V hot-electron injection. This mechanism is first observed in channel-initiated secondary electron (CHISEL) injection, which is activated upon the application of a reverse substrate bias. In this IIF model, the feedback heating of hot holes can lead to secondary impact ionization in the bulk, whereby tertiary carriers are generated. These tertiary electrons will subsequently accelerate and inject to the gate, yielding an increased electron gate current. The experimental signature of the vertical-IIF induced CHISEL injection is introduced and compared with respect to the traditional CHE injection. The former is found to be prominent for reverse V_b, and its characteristic slope in a log (I_{gr}) versus log (I_{br}) plot approaches 1. On the contrary, the latter dominates electron injection for V_b = 0 V and the characteristic slope is often at about 2.5 to 3.5.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Among the abovementioned injection models, the IIF model deserves the most careful investigation, in view of its ability to generate highly energetic tertiary electrons for low-power memory programming, which is an important feature for modern mobile and portable electronic applications. In addition, a tight relationship between the IIF-induced injection and device design parameters has been observed. A thinner tunnel oxide, a shallower source/drain junction depth and a more abrupt channel doping profile can all enhance the IIF-induced electron injection. Since an aggressive channel engineering scheme is often needed in Flash memory technology development to ensure the channel scalability without compromise on performance as explained earlier, there is a need to examine how channel doping profile can influence the IIF and the associated tertiary electron injection in deep-submicrometer Flash memory cells in the case even when no substrate bias is deliberately applied.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Chapter 3: Test Devices and Experimental Method

3.1 Equipment Set-up and Device under Test (DUT)

Experiments in this work were conducted with several characterization equipment interfaced to a computer terminal via the IEEE-488 bus. A block diagram for the current-voltage (*I-V*) measurement setup is illustrated in Fig. 3.1.

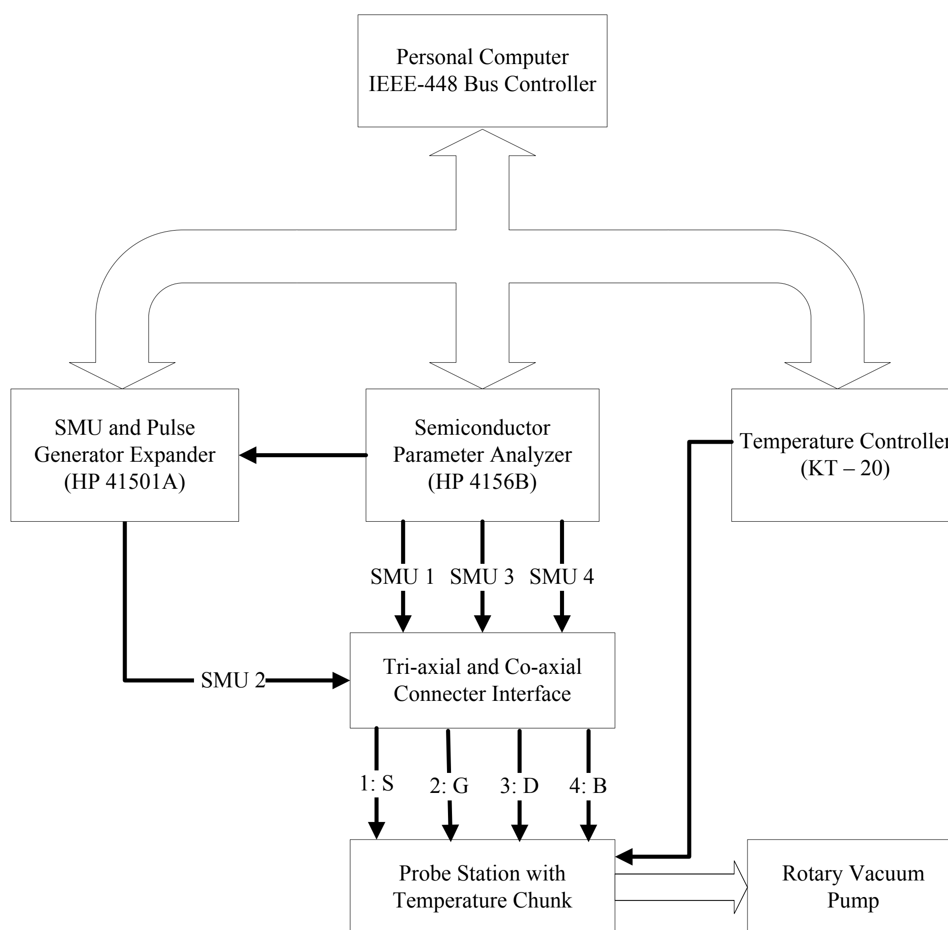


Fig. 3.1 Block diagram of the experimental setup for common *I-V* measurement. (S, G, D, B respectively denotes the source, gate, drain and substrate pads of the test cell.)

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

A Hewlett-Packard Semiconductor Parameter Analyzer (HP 4156B SPA) with a resolution of 1 fA is employed in the experiments. It can supply the prescribed biasing condition to and collect the corresponding electrical feedback from the DUT. Basically, input electrical signals generated by the analyzer are sent to the designated pads of the DUT through direct connection between the device and the analyzer, via good quality tri-axial and co-axial cables to minimize leakage and noise. The electrical feedback of DUT is then returned to the analyzer, where the results are to be stored and/or displayed.

The front-view of the semi-automated probe station used in this project is shown in Fig. 3.2(a). Probers of this station are properly shielded from external interference by using a grounded Faraday cage. Sharp probe needles are used in to minimize the contact resistance. In addition, all measurements are carried out in the dark (with all microchamber shields closed) to reduce any further external noise.

(a) (b)

Fig. 3.2 (a) Front view of the Suss MicroTec probe station used in this project with closed prober shield. (b) The test wafer is carefully placed on the thermal chuck of the probe station with tight suction provided by the vacuum pump.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

During measurement, a test wafer is mounted on the thermal chuck inside the microchamber of the probe station, as shown in Fig. 3.2(b). Typical test wafers employed in this project are fabricated by commercial 0.14-µm and 0.18-µm twin-well CMOS Flash technology. The test structure is one-transistor (1T) N-channel Flash memory cell. In this kind of cell, the floating-gate is 'shorted' to the control gate to allow a direct measurement of the gate current generated by hot-carrier injection process. The drawn channel length-to-width ratio (*L/W*) of a typical test cell is 0.176 µm/10 µm, and the tunnel oxide is about 8 nm.

Comparing to test devices fabricated by 0.18-µm CMOS technology ('0.18-µm devices' hereinafter), devices based on 0.14-µm technology (i.e. '0.14-µm devices') possess heavier doping concentration in the channel region. This is because additional source/drain extension and p-type halo implantations have been adopted in the latter to suppress severe short-channel effect, such as threshold voltage roll-off and drain-induced barrier lowing (DIBL), which are more prominent in scaled cells.

Among the 0.14-µm devices, there are two types of cells with certain process variations available. As tabulated in Table 3.1, one major difference in the fabrication steps is that for type A devices a higher V_t-adjustment implantation is involved comparing to that for type B devices. This can lead to a great difference in the distribution of p-type dopants (mainly Boron) in the channel region of these two types of cells, especially in the vertical direction, as V_t-adjustment is implanted perpendicular to the inversion channel. That would in turn lead to different electron injection behaviors in type A and type B cells (hereinafter called 'cell A' and 'cell B',

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

respectively). Thus, one main task of this project is to investigate the dependence of hot-electron injection and device degradation on the vertical channel doping profile for Flash memory cells fabricated by commercial CMOS technology.

Table 3.1 Main process variations involved in two different types of test devices fabricated by the commercial 0.14-µm CMOS technology.

Wafer Type	Description
A	• Higher V_t-adjustment implantation • Halo implantation • Oxynitrided tunnel-oxide (Nitrided SiO_2) • TiN-cap / Co salicidation
B	• Lower V_t-adjustment implantation • Halo implantation • Extended degas and dry process • Oxynitrided tunnel-oxide (Nitrided SiO_2) • Ti rich TiN-cap / Co salicidation

3.2 Experimental Method

Measurement of basic electrical and hot-electron injection characteristics forms an integral part of a number of semiconductor memory fabrication processes. It is especially important in device channel optimization, which can greatly influence the level of gate injection, the programming efficiency and the defects generation in a memory cell. This is because the device channel doping profile is tightly related to the hot-electron gate injection. A precise knowledge on hot-electron injection and its

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

dependence on channel doping profile is thus crucial for accurate circuit simulation and device lifetime estimation.

In this project, classical electrical characterization is first conducted to compare the fundamental properties of the 0.18- and 0.14-µm test cells. The 0.14-µm devices shows higher electron injection rate and are thus chosen for further hot-carrier injection related study. Next, a direct *I-V* measurement [88] is applied to extract the channel doping profile for two different 0.14-µm cell, that is, cell A and cell B. The result clearly reveals differences in the vertical channel dopant distribution for the two types of cells.

In the second phase of this project, a systematic investigation on the impact ionization feedback (IIF) process of hot holes in N-channel Flash memory cell is carried out. The IIF process can take place along a vertical (from drain to substrate) and/or along a lateral (from drain to source) trajectory. When the effect is prominent enough, the former can essentially lead to CHISEL injection, while the latter can result in a non-classical hot-electron gate current component. In our experiments, we have directly observed the experimental signature of the non-classical hot-electron component at V_b = 0 V, and proposed a lateral-IIF phenomenological model for the first time. Several supplementary experiments are performed subsequently and the results are all shown to be consistent with the proposed model. A strong dependence of this non-classical hot-electron component on channel doping profile is also observed by comparing the electron injection results obtained in cell A and cell B test devices.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

In the last phase of this project, we have performed a comparative study on hot-electron stress induced device degradation for cells featuring different channel doping concentration and vertical profiling, taking source bias V_S as the parameter. By applying a DC hot-electron stress technique, degradation of device parameters with stress time as well as the dependence of degradation on stress bias condition is obtained. Next, to analyze the spatial distribution of the stress induced oxide damage, a simple *I-V* measurement [89] is employed, instead of the charge-pumping (CP) method [90] [91] [92]. This is because the CP technique is only applicable and accurate to analyze the interface and oxide traps of a transistor if the DUT is well isolated (i.e. stand-alone). As all test cell employed in this project are not isolated from the adjacent structures in their respective cluster, that is, their substrate, gate and source are shared with other cells (with different dimension) in the same module, conventional CP technique cannot be used directly in this work.

3.3 Summary

In this chapter, equipment setup for common *I-V* measurement is presented firstly, followed by an introduction of typical test devices. A brief discussion on experiment techniques and arrangements are given in the second part of this chapter.

Commercial 1-T memory cells with a shorted CG-FG structure are employed as test devices in this project. Cells fabricated by the 0.14-μm CMOS technology are found to be more suitable for hot-carrier study than cells fabricated by the 0.18-μm technology. In 0.14-μm devices, two types of cells (i.e. cell A and B) featuring

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

different channel doping profiles are available. Our investigation on hot-electron injection and related device degradation is focused on these two kinds of cells, through which the influence of channel doping profile on carrier injection and associated defects generation can be identified.

All major measurements in this work are conducted with extreme care to mitigate possible noises due to probing, lighting and transient, so as to ensure accurate and reliable experimental results. Crucial experimental observations which are unique in our work and can lead to important conclusion are repeated on fresh cells at different locations of the test wafer to ensure the reproducibility. Moreover, some of the important experimental results are verified by another measuring technique or by applying different stress conditions on test cells with diverse channel dimensions.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Chapter 4: Fundamental Device Characterization

4.1 Introduction

Before embarking on an in-depth study of the hot-electron injection and its effect, basic electrical characterization on both 0.18-µm and 0.14-µm test cells is performed. It includes the output characteristics (I_d versus V_d), transfer characteristics (I_d versus V_g), and related electrical parameters extraction, such as threshold voltage (V_{th}), transconductance (G_m), subthreshold swing (S_t) and on-/off-state current (I_{on}/I_{off}). As these basic characteristics are similar for all semiconductor devices, only typical results of two types of 0.14-µm test devices, i.e. cell A and cell B featuring different channel doping profiles, are presented here. In addition, a simple *I-V* measurement to extract approximately the channel doping profile of the two types of test cells is also presented. Based on the results, the role of vertical channel profile on the hot-electron injection characteristics is demonstrated.

4.2 Basic electrical characteristics

Fig. 4.1 shows the linear and saturation transfer characteristics of a test cell B with channel dimension (*W/L*) = 10/0.176 µm/µm. I_d is the drain current; V_d and V_g are the drain and gate bias, respectively. In Fig. 4.1(a), linear V_{th} is extracted as 0.75 V, according to the 'maximum-G_m' method, where G_m is defined as $\partial I_d / \partial V_g$. Fig. 4.1(b) plots the saturation transfer characteristics and it shows the extraction of I_{on}, I_{off} and S_t.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Here, I_{on} refers to the drain current when $V_g = V_d = V_{dd}$, while I_{off} is the current when $V_g = 0$ V and $V_d = V_{dd}$. S_t (V/decade) is the subthreshold swing, which is defined as the reciprocal of the sub-threshold slope of the saturation characteristic curve.

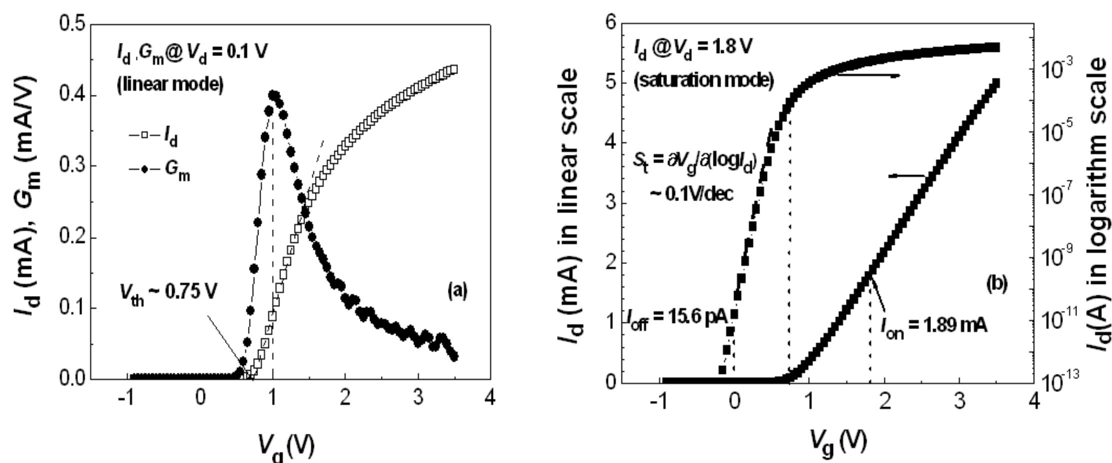


Fig. 4.1 (a) Extracting threshold voltage (V_{th}) in the linear mode (when V_{ds} ~ 0.1 V) by 'maximum-G_m' method. (b) Extracting the on-/off- state current and the sub-threshold swing in the saturation mode (when $V_{ds} = V_{dd} = 1.8$ V).

4.2.1 Comparison of vertical channel doping profiles in cell A and cell B

As mentioned in chapter 3, two types of 0.14-µm N-channel test devices (cell A and cell B) are available in this project, which possess different channel doping profile due to the variations in their V_{th}-adjust implantation and following thermal steps. Since hot-electron injection behavior in a semiconductor device is tightly related to the channel doping profile, it is essential for us to first investigate the channel doping profile of these two types of 0.14-µm test cells.

Fig. 4.2 depicts the output and transfer characteristics of the two types of cells. As expected, heavy V_{th}-adjust implantation in cell A has resulted in a higher channel

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

doping concentration, which is manifested by the higher linear V_{th} and lower saturation I_d than those in cell B.

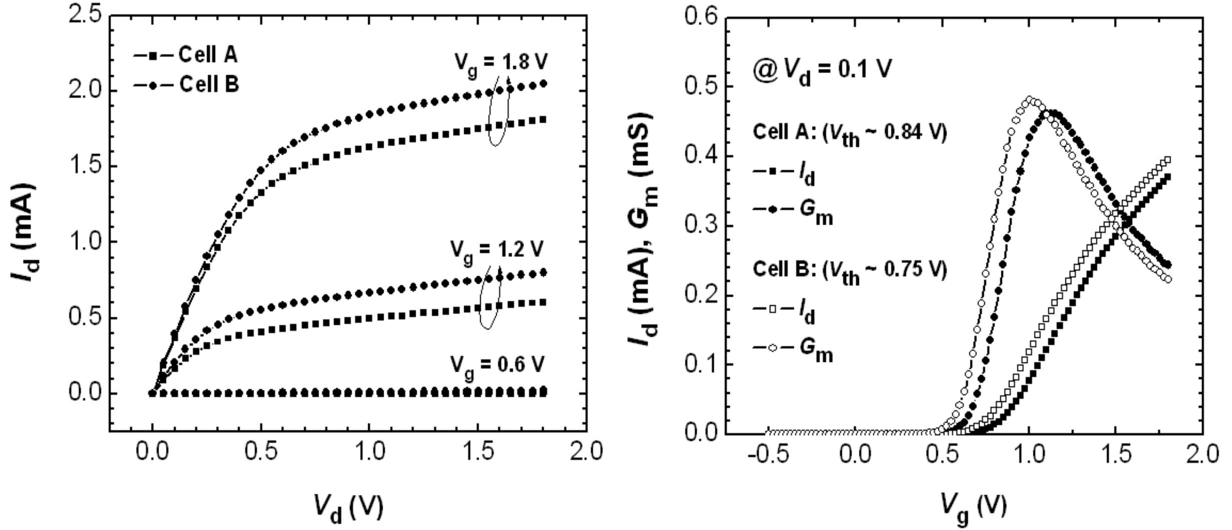


Fig. 4.2 Output and transfer characteristics of test cell A and B with drawn W/L = 10/0.176 μm/μm. Cell A possesses a higher V_{th} than cell B as expected.

Next, a simple and non-destructive technique that can quickly compare the vertical channel doping profile is applied. In this technique, the extent of body-effect, or in other words, the influence of body bias (V_{sb}) on threshold voltage (V_{th}) is measured, since the gradient ($\partial V_{th} / \partial V_{sb}$) can be correlated to the average effective channel doping concentration (N_{eff}), according to [93]:

$$\frac{\partial V_{th}}{\partial V_{sb}} \approx \frac{1}{C_{ox}} \sqrt{\frac{\varepsilon_{Si} q N_{eff}}{2(2\psi_B + V_{sb})}} \tag{4. 1}$$

Fig. 4.3 shows a gradual increment of V_{th} as V_b (i.e. V_{sb}) varies from a small positive value (0.5 V) to a large negative voltage (-5 V). V_b = 0.5 V is set to be the positive limit to avoid any significant junction leakage due to the forward-bias of two p-n$^+$ diodes (substrate/source and substrate/drain). As shown, cell A exhibits higher V_{th} as well as larger

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

gradient ($\partial V_{th} / \partial V_{sb}$) than cell B for all V_b, implying that not only a heavier channel doping concentration but also a much steeper doping profile along the vertical direction is present in cell A than that in cell B.

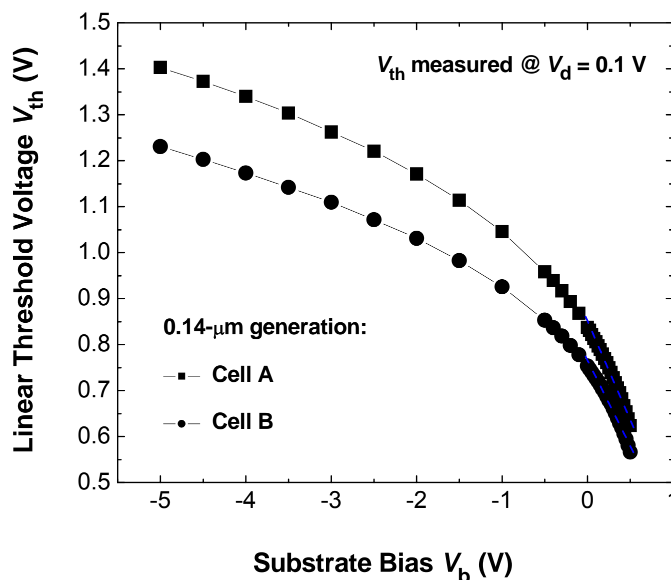


Fig. 4.3 Variation of the linear threshold voltage V_{th} (measured at V_d = 0.1 V) due to the change of substrate bias V_b in two types of 0.14-µm test devices, suggesting a larger channel doping concentration present in cell A.

The above technique offers a fast but qualitative approach to compare the effective channel doping concentration in different devices. In order to conduct a quantitative comparison, that is, to extract approximately the effective channel doping concentration at a particular distance from the Si-SiO_2 interface, more detailed analysis is required. One technique is developed as shown below. The linear threshold voltage can be expressed as:

$$V_{th} = V_{FB} + 2|\phi_p| + \frac{1}{C_{ox}} \int_0^{x_d} qN_a(x)dx \qquad (4.2)$$

where ϕ_p is assumed constant for dopant distribution which varies slightly under the channel. Taking the first-order differentiation of equation (4.2), we have:

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

$$dV_{th} = \frac{q}{C_{ox}} \cdot N_a(x_d) \cdot dx_d \tag{4. 3}$$

As an increment in body bias (dV_{sb}) leads to a differential increment in the width of the depletion region (dx_d), the total charge in the depletion region increases by an amount of qN_adx_d, which is also equal to the differential increment dV_{sb} multiplied by the capacitance of the depletion region (ε_s/x_d). Therefore, we obtain:

$$dV_{sb} = \frac{-qx_d}{\varepsilon_s} \cdot N_a(x_d) \cdot dx_d \tag{4. 4}$$

Based on above equations, the depletion region depth (x_d) and the doping concentration at a particular depletion width ($N_a(x_d)$) are related to the increment of the gradient dV_{sb}/dV_{th}. They are expressed as [94]:

$$x_d = -\frac{\varepsilon_s}{C_{ox}} \cdot \frac{dV_{sb}}{dV_{th}} \tag{4. 5}$$

$$N_a(x_d) = -(\frac{C_{ox}^2}{q\varepsilon_s}) \cdot (\frac{d^2V_{sb}}{dV_{th}^2}) \tag{4. 6}$$

The above two equations form the basis to determine the vertical distribution of the p-type dopant profile in the channel region where the edge of the depletion region is nearly parallel to the Si-SiO_2 interface, i.e. $V_{ds} << |\phi_p|$. Effective dopant concentration $N_a(x_d)$ at a distance x_d from the Si-SiO_2 interface can thus be extracted approximately. As shown in Fig. 4.4, cell A is found to have a heavier and a more abrupt channel doping than cell B in the proximity of the Si-SiO_2 interface. This observation agrees well with the qualitative conclusion. It should be taken note that in above technique there is a pre-assumption that the change in V_{th} with V_{sb} in device sub-threshold region is almost the same as that in the strong inversion region. This pre-assumption is

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

proven to be valid for typical test devices employed in this project. As illustrated in the inset of Fig. 4.4, the 1st- and 2nd- order derivatives of V_{th} with respect to V_{sb} for V_{th} exacted in both methods are almost equal, i.e. $\partial V_{th(max.\ gm)}/\partial V_b \approx \partial V_{th(subthreshold)}/\partial V_b$ and $\partial^2 V_{th(max.\ gm)}/\partial V_b^2 \approx \partial^2 V_{th(subthreshold)}/\partial V_b^2$, for the same V_{sb} bias.

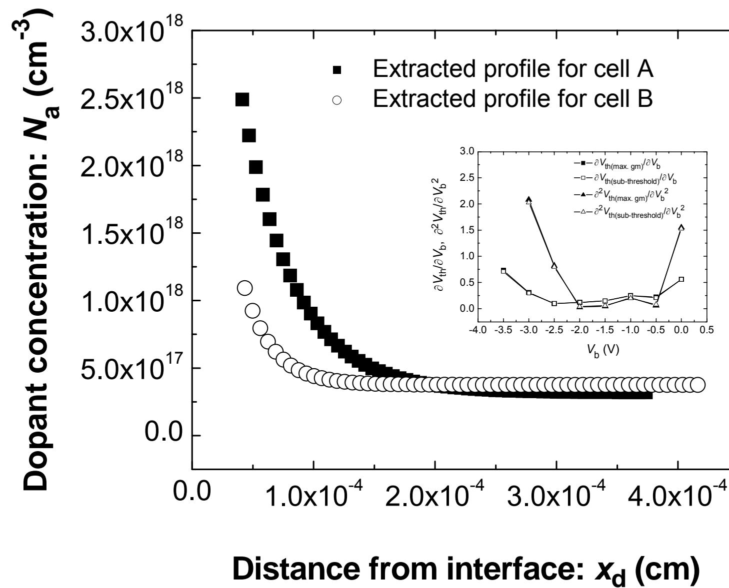


Fig. 4.4 Extracted dopant concentration as a function of the vertical distance x_d from the Si-SiO_2 interface in the two different sample devices.

4.2.2 Comparison of short channel effect (SCE) in cell A and cell B

Previous comparison only confirms that cell A has a larger amount of doping concentration and a steeper vertical doping profile than that of cell B in the proximity of the Si-SiO_2 interface; however, no direct information on the two-dimensional distribution of the channel dopants is available. In order to probe the channel doping profile along the horizontal direction (i.e. from source to drain), an investigation on short channel effect (SCE) in cell A and B is conducted. The parameters examined include the V_{th} roll-off and Drain-Induced Barrier Lowing (DIBL).

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Fig. 4.5(a) depicts the V_{th} roll-off phenomenon in cell A and cell B, as the drawn channel length is shrunk from 0.4 to 0.16 µm. The rate of V_{th} decrement with respect to the channel length scaling is illustrated in Fig. 4.5(b). It is observed that $\Delta V_{th}/\Delta L_g$ decays exponentially with ΔL_g and the roll-off rate is almost the same in both cells.

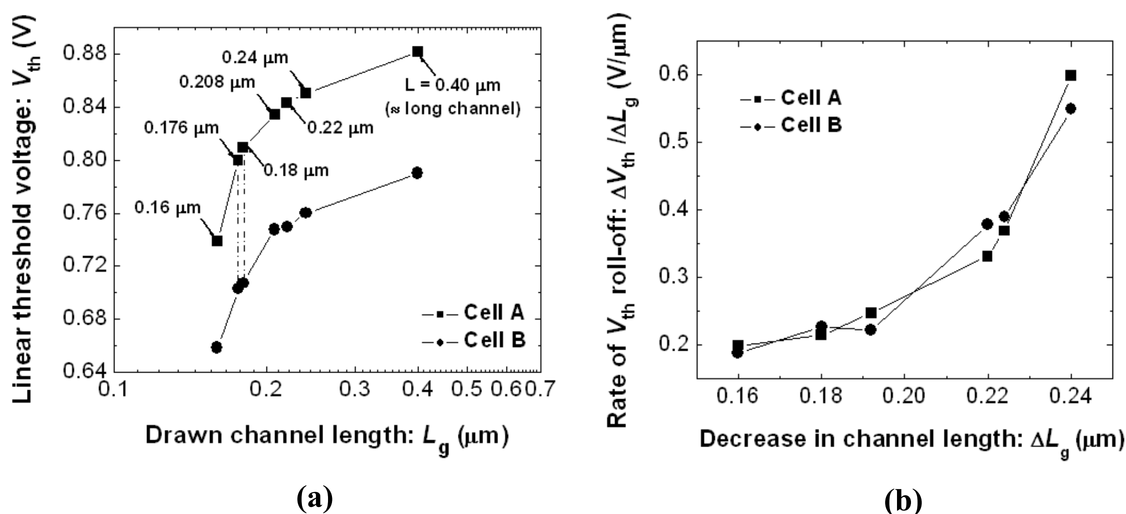


Fig. 4.5 Reduction of linear threshold voltage (V_{th}) as the drawn channel length (L_g) decreases from 0.4 µm to 0. 16 µm for cells A and B.

The monotonic decrease of V_{th} with L_g shrinkage is because a significant amount of depletion charges are provided by the space charges in the source and drain depletion regions; as a result, the gate voltage required for the onset of channel inversion (i.e. V_{th}) is reduced as channel length is scaled down. Since a similar rate of V_{th} reduction is observed for cell A and cell B (Fig. 4.5(b)), it may be concluded that the p-type dopant profile along the channel horizontal direction is almost the same for the two types of cells.

In modern short-channel FETs, drain and source fields can penetrate deeply into the middle of the channel, lowering the potential barrier between the source and drain. This leads to a substantial increase in subthreshold current; in other words, the

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

threshold voltage becomes lower than that of the long-channel counterpart. When a high V_d is applied, the injection barrier height is lowered even more, resulting in further reduction of V_{th}. The effect that a high V_d can bring down the injection potential barrier at the source side thereby reducing V_{th} in short-channel devices is therefore named as Drain-Induced Barrier Lowing. By comparing the significance of DIBL effect in the two different Flash memory cells, information on their relative channel doping profile along the horizontal direction can be deduced.

Fig. 4.6 plots the DIBL characteristics of type A and type B cells for different L_g. Here, DIBL is measured as the shift in V_g for a given subthreshold I_d when V_d is increased from 0.05 V to 1.8 V. Apparently, both cells exhibit a equal amount of barrier lowing at the source/inversion-layer junction when V_d increases significantly. This again confirms the similarity in p-type dopants distribution along the lateral direction in the channel region for type A and B cells.

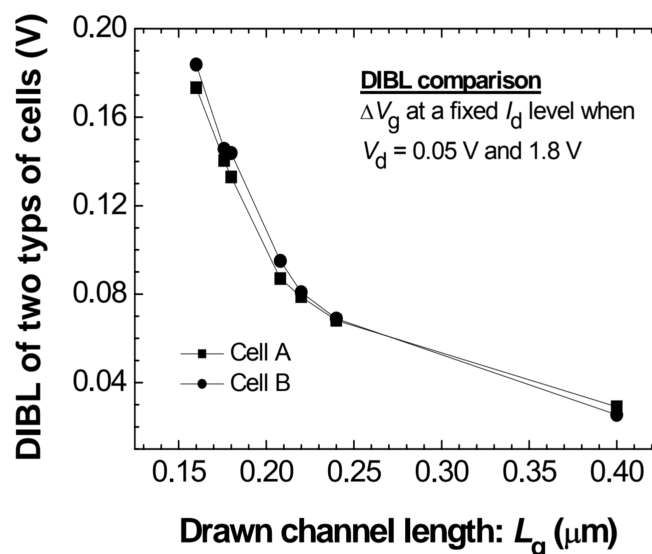


Fig. 4.6 Drain-induced barrier lowing effect in cells A and B at different drawn channel length as V_d changes from 0.05 V to 1.8 V.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

4.2.3 Summary of channel doping profile comparison

Previous comparison of short-channel effect in the two types of 0.14-μm cells strongly suggests that the major abruptness/difference in the channel doping profile in these cells is along the 'vertical' or bulk-to-interface direction, rather than in the 'horizontal' or source-to-drain direction.

This is because any major variation in the lateral doping profile should lead to a distinct V_{th} roll-off or DIBL behavior for different transistors [95] [96]. Since no appreciable difference is observed, no significant disparity is revealed in the horizontal distribution of channel dopants for the two types of 0.14-μm cells. Nevertheless, strong evidence shown in body-effect measurement, as demonstrated by the larger $\partial V_{sb} / \partial V_{th}$ gradient, has confirmed that cell A does possess a heavier and more abrupt vertical doping profile in the channel region. The above experimental results clearly indicate that the key difference in the channel doping profile of cell A and cell B lies in the vertical direction.

A more abrupt vertical doping profile is able to enhance the local built-in vertical electric field, and the variation could result in more efficient hot-electron injection. Therefore, through the above study, any differences in the hot-carrier injection characteristics to be shown subsequently in this work may be directly related to the different vertical channel profiles.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

4.3 Hot-Carrier Generation and Injection Characteristics

Substrate current I_{sub} (or I_b) is usually used to monitor hot-electron effect in the N-channel MOSFET. This is because it consists of the hot-hole current originated by the primary impact ionization process. However, I_{sub} often has adverse influences on the operation of substrate-bias generator and the subthreshold characteristic of neighboring devices. Furthermore, it also induces ohmic drops which can forward-bias the source-substrate junction and lead to parasitic bipolar phenomena such as snap-back and latch-up [97]. This section represents the typical channel hot-carrier generation and injection characteristics of test cells employed in this project. As the general characteristics are similar in both types of 0.14-µm cells, only cell A results are presented here for demonstration.

Fig. 4.7 shows the hot-carrier generation characteristics in the CHE injection dominated regime (i.e. at V_b = 0 V), over a wide range of comparable gate (V_g) and drain (V_d) biases. As it shows, all I_{sub} curves exhibit a conventional bell shape, denoting that the strength of the impact ionization first increases then diminishes with the growing V_g. The initial increment of I_{sub} may be attributed to the increase of hot electron population in the channel (i.e. inversion layer) as V_g increases. When V_g reaches ~ $V_d/2$, I_{sub} is at its peak value, indicating the conventional 'worst-case' condition where most impact ionization takes place for a given drain bias. After that, I_{sub} decreases with further increase in V_g, as a result of the reduction in maximum lateral field (E_{max}) in the drain pinch-off region (or effective ionization length (l_d)) where most primary impact ionization occurs; because $E_{\max} = (V_d - V_{dsat}) / l_d$ and V_{dsat} of short-channel device will increase with V_g according to [98]:

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

$$V_{dsat}(short-channel) = \frac{v_{sat}}{\mu_{if}} \cdot L \cdot [(1 + \frac{2\mu_{if}(V_g - V_{th})}{v_{sat}L})^{1/2} - 1] \quad (4.\ 7)$$

where v_{sat} refers to the saturation velocity, μ_{if} denotes the low filed mobility and L is the channel length. Eventually, when both V_d and V_g are large enough, a plateau of high I_{sub} appears. This may be attributed to the enhanced self-heating effect at high drain and gate biases that increases the high energy tail of the hot-electron distribution function.

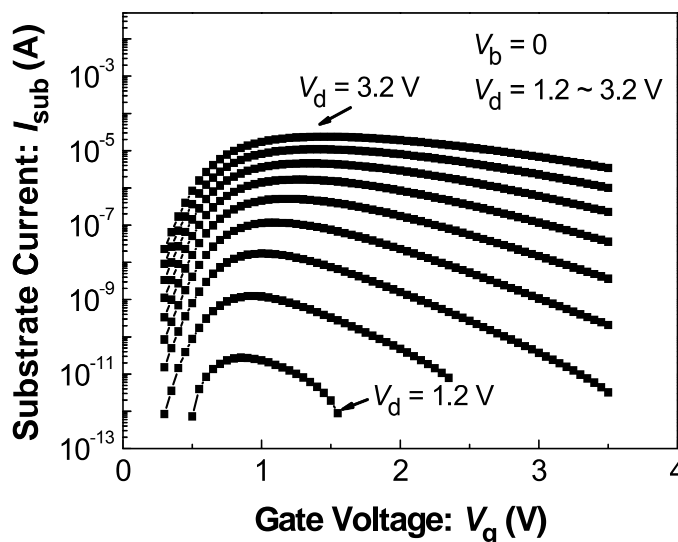


Fig. 4.7 Hot carrier characteristics of a cell A device, taking drain bias (V_d) as the parameter for $V_b = 0$ V.

Fig. 4.8 illustrates the influence of back bias (reverse-V_b) on hot electron generation when V_d = 1.2, 2.2 and 3.2 V for a type A cell. At a more negative V_b, the I_{sub} characteristic is shifted to the right, while the peak I_{sub} value remains unchanged. This observation suggests that although the extent of hot carrier generation is insensitive to back bias, a reverse V_b can significantly delay the occurrence of maximum impact ionization, owing to the body effect which increases V_g for channel inversion (i.e. V_{th}).

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

While I_{sub} is often used to monitor the strength of hot carrier generation, the gate current I_g, measured quantitatively at the gate electrode, is usually employed to examine the injection characteristic of hot carriers.

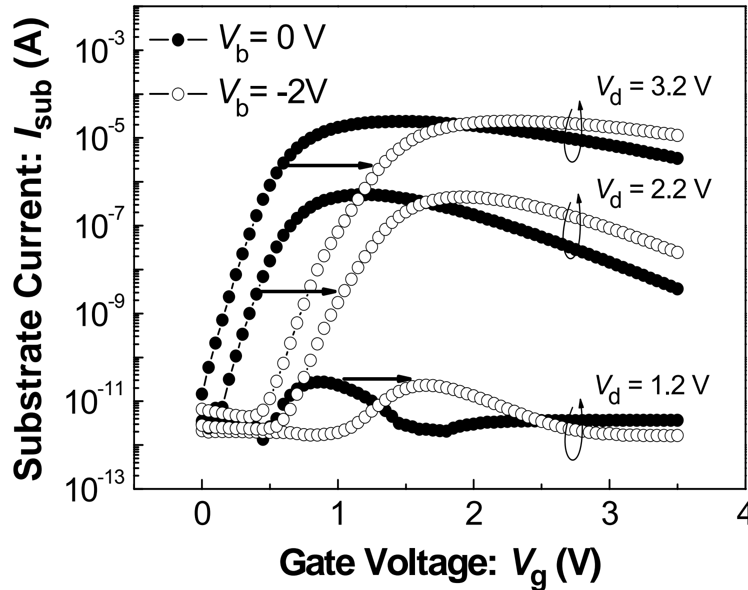


Fig. 4.8 Hot carrier characteristics of a type A cell, taking the drain (V_d) and the substrate (V_b) biases as parameters. (Solid circle: V_b = 0 V; open circle: V_b = -2 V.)

Fig. 4.9 depicts the hot-electron injection characteristics of an N-channel type A cell. Here, V_g is swept from ~ $V_d/2$ to $\geq V_d$ as V_b is changed from zero to negative bias. As shown by the trend of I_g, the injection of electrons first increases then decreases with V_g. This is a combined result from two possible effects: *1)* when $V_g \geq V_d$, the reduction of E_{max} with V_g inhibits the high energy tail of the hot-electron distribution; *2)* the accumulation of negative charges as trapped in the gate oxide and Si-SiO_2 interface can gradually suppress the further injection of hot electrons and thus yields a decreasing I_g. However, at $V_b < 0$, an increase of I_g is observed. As to be discussed in next section, this is expected because a reverse body bias can enhance the gate

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

injection of energetic tertiary electrons as modeled in the CHISEL injection mode.

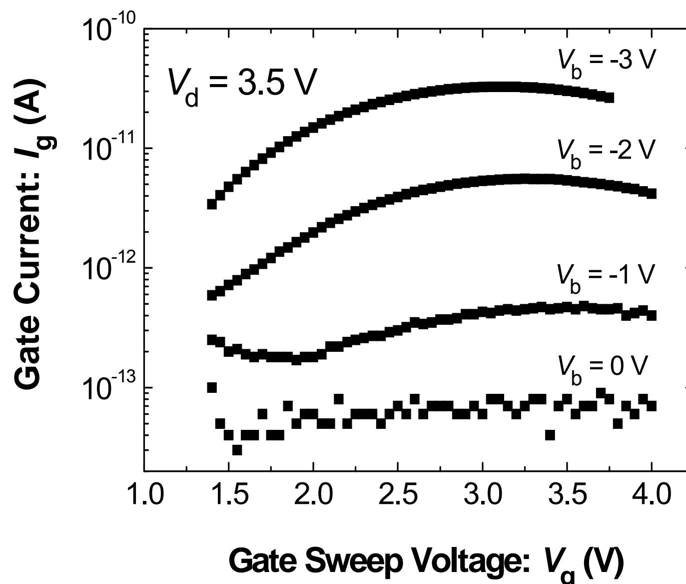


Fig. 4.9 Electron injection current (I_g) versus gate sweep voltage (V_g) characteristics for a typical N-channel test cell with drawn channel dimension (W/L) = 10/0.176 μm/μm and tunnel-oxide thickness (t_{ox}) = 8.1 nm.

4.4 Investigation on CHISEL injection

After a general study on hot-electron generation and injection behaviors, the impact-ionization feedback (IIF) induced tertiary-electron injection now deserves to be investigated in more detail. Channel-initiated secondary electron (CHISEL) injection is a good example of such injection mechanism. As introduced earlier, it is activated upon the application of a reverse body bias ($V_b < 0$ for N-channel devices), through which hot holes flowing towards the substrate contact are heated up. As a consequence, secondary impact ionization (SII) occurs, resulting in the generation of energetic tertiary electrons.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

4.4.1 Experimental signatures of CHE and IIF-induced gate currents

The distinct values of exponent n in the CHE and CHISEL models are identified as their experimental signatures. In the CHE model, hot-electron injection efficiency ($I_{g\ r}^{e}$) exhibits a power-law dependence on the primary impact ionization (or channel heating) efficiency (I_{br}). The relationship is [84]:

$$I_{g\ r}^{e} \propto I_{br}^{\ n} \qquad (4.\ 8)$$

where $I_{g\ r}^{e} = I_{g}^{e}/I_{d}$ and $I_{br} = I_{b}/I_{d}$. I_{g}^{e} refers to the hot-electron gate current, I_{b} (or I_{sub}) and I_{d} are the substrate and the drain current, respectively. In the classical CHE model, the exponent n is linked to the ratio ϕ_{ox}/ϕ_{ii}, where ϕ_{ii} (~ 1.12 eV for Si) is the threshold energy for impact ionization, ϕ_{ox} is the Si-SiO_2 energy barrier (~ 3.2 eV) that channel electrons have to overcome in order to inject into the floating-gate. Thus, a value of 3 is commonly observed for the exponent n in CHE injection, although it may increase for a device fabricated by older technology with longer channel length.

On the other hand, in an IIF assisted electron injection mode, such as CHISEL, the generation and injection of tertiary electrons in low gate and drain bias regime are enhanced via the feedback heating of holes. Hence, in the $I_{g\ r}^{e}$ versus I_{br} logarithm plot, the characteristic exponent n approximately equals to 1, that is [85]:

$$I_{g\ r}^{e} \propto I_{br} \qquad (4.\ 9)$$

It should be emphasized that the CHISEL injection is induced by a vertical IIF mechanism, as it is the vertical feedback heating of holes (from drain to the bulk region) that induces the secondary impact ionization and generates CHISEL in the bulk.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Fig. 4.10 shows the $I_{g\ r}^{e}$ versus I_{br} characteristics of a typical type A cell, where the substrate bias V_b is taken as the parameter. In this experiment, both the gate (V_g) and the drain (V_d) voltages are increased with the same step-size, maintaining a constant vertical injection field (E_{ox}) by always keeping $V_g = V_d$. For $V_b = 0$, the exponent n is about 2.5, which implies that the carrier heating by lateral channel field is the main driving force for I_g^e, as predicted by the conventional Lucky-Electron Model [57, 61]. For $V_b < 0$, however, n reduces gradually and approaches 1, indicating a change in the main driving force for I_g^e from the lateral channel field to the feedback heating of the holes [78, 84, 85], according to the IIF-induced injection mechanism.

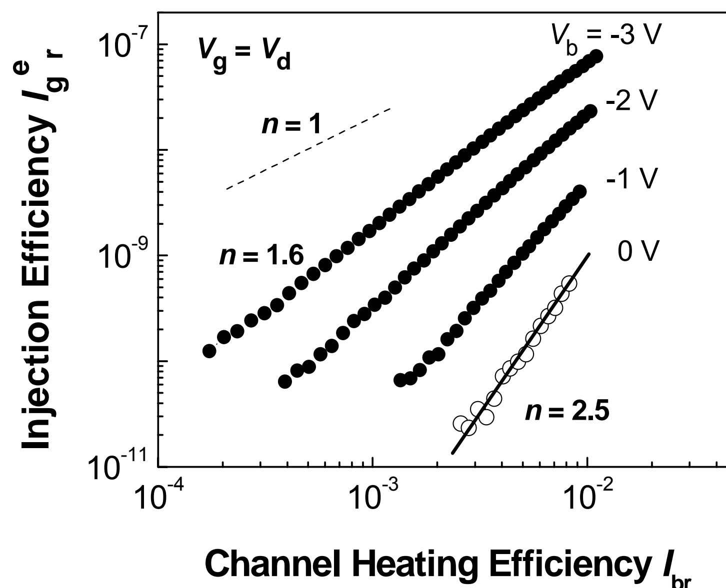


Fig. 4.10 Hot-electron injection efficiency $I_{g\ r}^{e}$ versus primary impact ionization efficiency I_{br} of a typical 0.14-μm generation type A cell. Device dimension (W/L) = 10/0.176 μm/μm, tunnel-oxide thickness t_{ox} = 8.1 nm.

However, we noticed that the expected unitary slope does not appear for the IIF-induced injection, even when V_b = -3 V. This is because the coefficient C_B in the CHISEL model ($I_{gr} = C_B \cdot I_{br}^{\ n}$) can change with the drain bias V_d [85] even when E_{ox} is fixed by the constant potential difference (i.e. $V_g - V_d = 0$). Hence, the experimental

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

signature of CHISEL is covered by this effect and a high exponent appears. In order to eliminate this effect, Esseni and co-workers proposed a new measurement method, in which only V_s was varied in a small range while both V_g and V_d were fixed [85]. By adopting this method, the vertical electric field is essentially kept no change and the unitary slope for reverse-V_b condition shows up. In this project, we have applied this technique to study the two types of 0.14-µm test cells, through which a non-classical hot electron effect at zero body-bias condition is observed in cell A, as to be discussed in the next chapter.

4.4.2 Impact of channel engineering

It is known that the efficiency of hot-electron injection in a memory cell can be influenced by a number of design parameters, such as the tunnel-oxide thickness, the engineering of source/drain and extension regions, as well as channel engineering (i.e. V_t-adjustment, halo and pocket dopant implantations, etc.).

This section presents the investigation details of the impact of channel engineering on the IIF-induced electron injection. It is achieved by first comparing the tertiary electron injection in two generations of memory cells, fabricated by the commercial 0.18-µm and 0.14-µm CMOS technology. For the former generation, a lower V_t-adjust implantation dose was used and no halo or pocket dopants were involved in the fabrication. In contrast, to suppress severe short-channel effect that normally arises as cell dimension shrinks, the latter generation employs a higher V_t-adjust implantation dose as well as p-type halo dopants for N-type device fabrication. Therefore, cells

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

fabricated by 0.14-μm technology possess a heavier channel doping concentration than that of 0.18-μm technology. Hence, by comparing the hot-electron injection behavior in these two generations of devices, we can determine the influence of the channel doping on the IIF-induced tertiary electron injection.

Fig. 4.11 depicts the $I_{g\ r}^{e}$ versus I_{br} characteristics of a test cell fabricated by 0.18-μm CMOS technology. As shown explicitly, there is no significant enhancement of the gate injection current till a highly negative body bias (V_b = -3 V) is applied. This is because the IIF-induced electron injection is heavily dependent on the channel doping profile, especially the halo implantation [86] [99]. As 0.18-μm generation involves low V_t-adjust implantation and no halo implant, the injection of tertiary electrons is suppressed in this generation of devices.

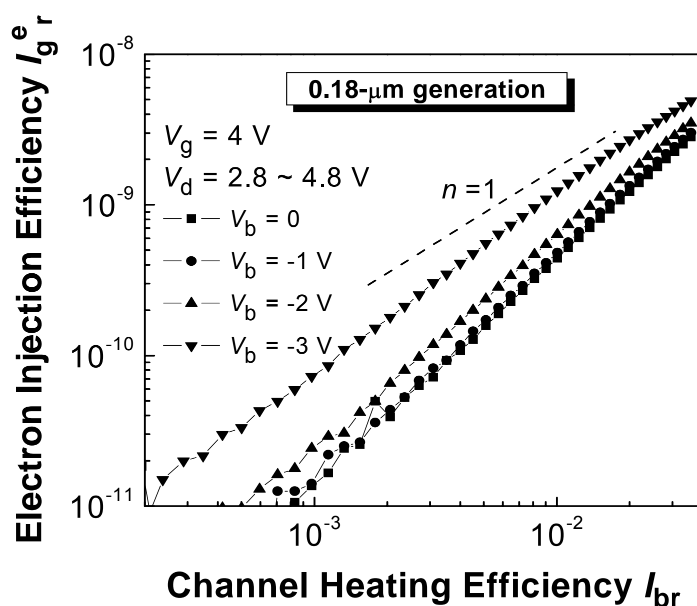


Fig. 4.11 $I_{g\ r}^{e}$ versus I_{br} characteristics (with parameter V_b) for an N-channel memory cell fabricated by commercial 0.18-μm technology. The gate bias V_g is fixed at 4 V, while the drain bias V_d is varied from 2.8 to 4.8 V. The drawn channel dimension (W/L) = 10/0.176 μm/μm and the tunnel-oxide thickness t_{ox} = 8.1 nm.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

For comparison, a similar experiment was conducted on a 0.14-μm generation type A cell with the same drawn channel dimension. Fig. 4.12 shows a substantial increment of the gate current due to the IIF-induced injection of tertiary electrons throughout the whole negative V_b bias range (from -1 to -3 V). This large enhancement in gate injection current may be attributed to the higher channel dopant concentration in 0.14-μm cells, as compared to the 0.18-μm counterparts.

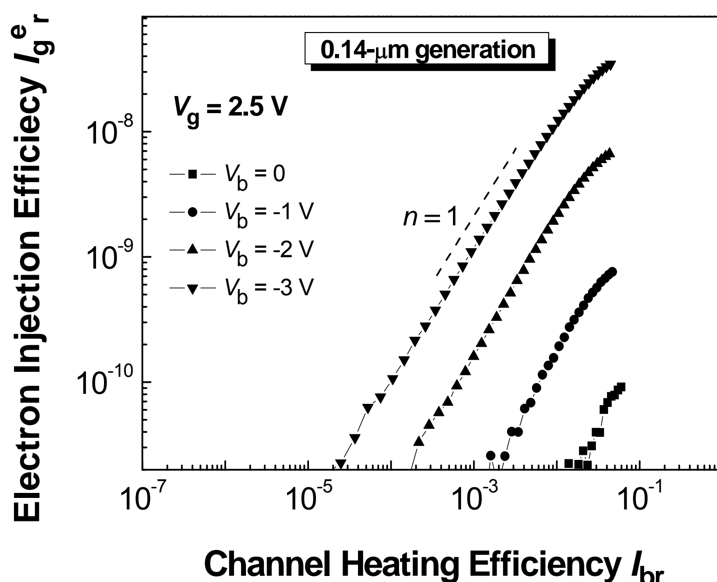


Fig. 4.12 $I_g{}^e{}_r$ versus I_{br} characteristics (with parameter V_b) for an N-channel type A cell fabricated by commercial 0.14-μm technology. V_g is fixed at 2.5 V, while V_d is varied from 1.2 to 4.5 V (small $I_g{}^e{}_r$ noise at low V_d range is not shown here). The drawn channel dimension (W/L) = 10/0.176 μm/μm and the tunnel-oxide thickness t_{ox} = 8.1 nm.

To further examine the impact of channel engineering on tertiary electron injection for the same device technology, CHISEL injection characteristics in two types of cells fabricated by the 0.14-μm technology are compared in Fig. 4.13. Cell B exhibits weaker electron injection efficiency for a given I_{br}, and its $I_g{}^e{}_r$ is not as sensitive to back bias as compared to that for cell A. From previous study, this may be attributed

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

to the more abrupt channel doping profile in cell A than that in cell B, especially along the vertical direction.

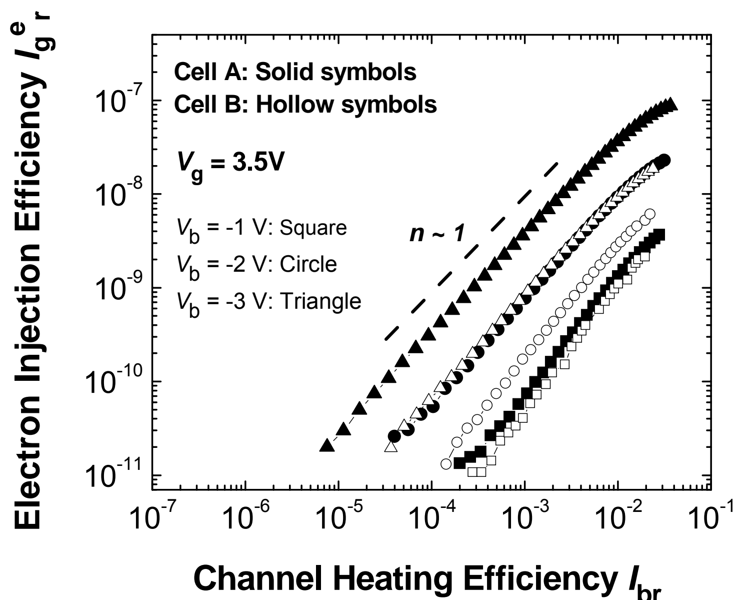


Fig. 4.13 Comparison of the IIF-induced tertiary electron injection in two types of 0.14-μm devices (test cell A and cell B). V_g is fixed at 3.5 V, while V_d is varied from 1.2 to 4.5 V (small $I_g^e{}_r$ noise at low V_d range is not shown here). The drawn channel dimension (W/L) for both cells = 10/0.176 μm/μm and the tunnel-oxide thickness t_{ox} = 8.1 nm.

Fig. 4.14(a) and (b) depict another set of experimental results showing the enhancement of electron injection in the two types of 0.14-μm cells, when the gate and drain biases are swept together to maintain a constant and positive vertical field in the channel region by keeping $V_g - V_d$ = 0.5 V. As shown in the figures, although the net electron injection efficiency at low $|V_b|$ is comparable for these two types of cells (cell B even shows a bit larger CHE injection at V_b = 0 V), however, the increase of $I_g^e{}_r$ with $|V_b|$ is more significant in cell A.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

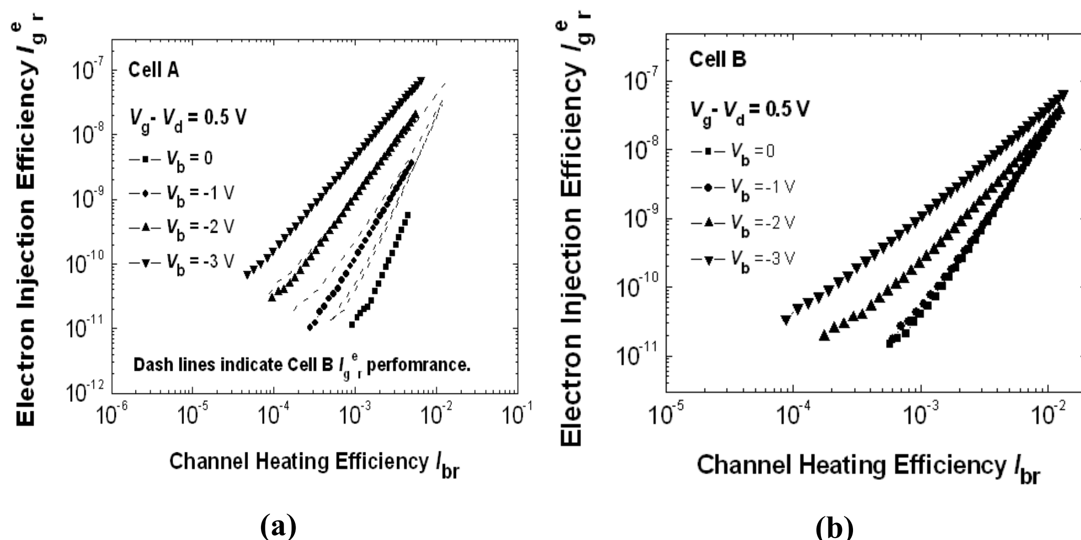


Fig. 4.14 Another experimental result showing the more abrupt and heavier channel doping in cell A can enhance the significance of IIF and promote tertiary electron injection. V_d is varied from 1.2 to 4.5 V (small $I_g{}^e{}_r$ noise at low V_d range is not shown here). The drawn channel dimension (W/L) for both cells = 10/0.176 μm/μm and the tunnel-oxide thickness t_{ox} = 8.1 nm.

It is interesting to observe that when V_g and V_d are increased together while keeping a constant difference between each other (i.e. $V_g - V_d$ = constant), the injection power-low exponent n does not approach 1 even when V_b is at its highest negative value (i.e. -3 V in Fig. 4.10 and Fig. 4.14). This is because for short-channel devices, the increase in V_g (i.e. ΔV_g) > ΔV_{dsat} according to equation (4.7) [98]. For $\Delta V_g = \Delta V_d$, $\Delta V_d > \Delta V_{dsat}$ and the maximum channel field E_{max} (= $(V_d - V_{dsat})/l_d$) thus increases with the positive (V_d- V_{sat}); on the other hand, the vertical injection field E_{ox} (= $(V_g - V_d)/t_{ox}$) remains unchanged as $\Delta V_g = \Delta V_d$. As a result, although a reverse V_b can enhance the IIF-induced hot-electron injection at low I_{br}, the CHE injection is simultaneously enhanced by the increasing E_{max} over the whole I_{br} regime. Therefore, tertiary-electron injection, as well as its unitary exponent signature, is masked by the significant CHE component. It is also worthy to note that the injection efficiency ($I_g{}^e{}_r$) for different V_b tends to converge to one single point at high biases (Fig. 4.14(b)). This

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

is essentially ascribed to the constant E_{ox}, which fixes the maximum reachable $I_g{}^e{}_r$ and sets same limit for hot-electron injection. Nevertheless, when only V_d is varied while V_g is unchanged (i.e. E_{max} is largely increased while E_{ox} is reduced), the experimental signature of tertiary-electron injection starts to appear at $V_b = 0$ V (Fig. 4.12 and Fig. 4.13). This may be explained by the enhanced feedback heating of hot holes at higher E_{max}, which increases the population of tertiary electrons and the associated electron injection, as compared to the suppressed CHE injection at reduced E_{ox} condition.

Moreover, in Fig. 4.13, cell A shows a more prominent $I_g{}^e{}_r$ increment with V_b as compared to cell B, indicating that the built-in vertical field due to the abrupt channel doping can indeed assist the injection of tertiary electrons. Based on above experimental results, we can thus conclude that for deep submicrometer N-channel memory cells, the p-type dopant concentration and its profile along the vertical direction in the channel region do have a significant influence on the IIF-induced tertiary electron injection. Similar to the impact of horizontal profile on classical hot-electron injection, a heavier and more abrupt vertical doping scheme can result in a more prominent injection of tertiary electrons.

4.4.3 Comparison of FN, CHE and CHISEL injections

As introduced earlier, CHISEL injection was considered to be a good candidate for low power programming of modern Flash memory cells, in view of the enhanced injection of energetic tertiary electrons. In this project, several experiments were conducted to compare the injection efficiency and characteristics of CHISEL mechanism with two other common electron injection schemes, namely,

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Fowler-Nordheim (FN) tunneling and CHE injection. Fig. 4.15 depicts the hot-electron gate current (I_g^e) originated by the three injection mechanisms in two generations of cells. Here, FN-tunneling, CHE injection and CHISEL injection are achieved by applying comparable biasing levels at the gate and drain terminals. As our main focus is on the IIF-induced tertiary-electron injection, which is more effect in low voltage regime, a general guideline is to set the bias at around 3 ~ 4.5 V for 0.18-μm and 0.14-μm generations.

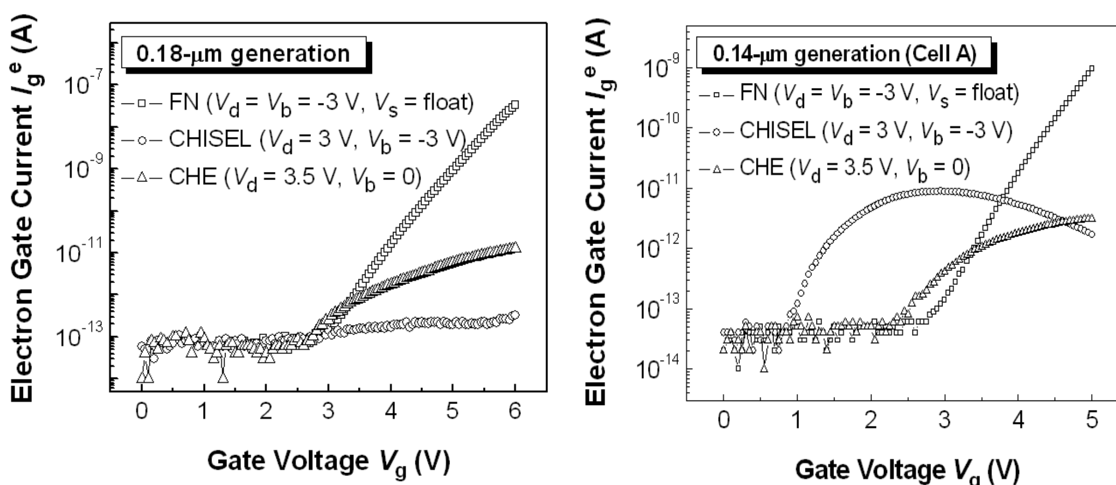


Fig. 4.15 Programming current (I_g^e) measured in three different electron injection schemes for N-channel transistors fabricated by 0.18-μm and 0.14-μm technology (cell A). The drawn channel dimension (*W/L*) for both cells = 10/0.176 μm/μm and the tunnel-oxide thickness t_{ox} = 8.1 nm.

For cells fabricated by the 0.18-μm CMOS technology, FN-tunneling yields the largest I_g^e at high gate bias level, followed by CHE injection, whereas CHISEL injection mode generates the lowest I_g^e over the whole gate bias range. In contrast, for cell A fabricated by 0.14-μm technology, a relatively high gate injection current is attained in CHISEL mode even when the V_g is low. A peak I_g^e occurs when V_g ~ 2.5 – 3 V. On the other hand, FN-tunneling only becomes important for large V_g (~ 4 V) to generate comparable I_g^e as given by peak I_g^e in CHISEL mode. This difference may

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

not seem to be significant in high-voltage driven memory arrays; however, for low-voltage programming or power-saving purposes, the capability of IIF-induced injection, that is, contributing comparable programming current at less power and lower voltages, is very important.

Fig. 4.16 shows the electron gate current (I_g^e) characteristics of a type B cell which possesses a more graded channel doping profile than cell A. As expected, cell B exhibits a much lower CHISEL current than cell A in low V_g regime.

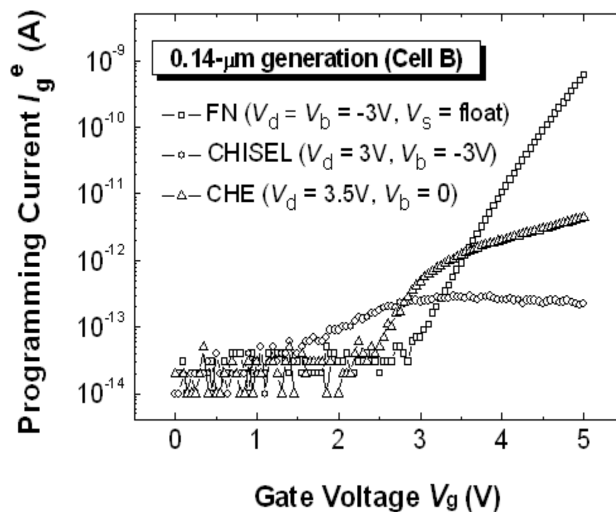


Fig. 4.16 Programming current (I_g^e) measured in three different electron injection schemes for N-channel transistors fabricated by 0.14-μm technology (cell B). The drawn channel dimension (W/L) = 10/0.176 μm/μm and the tunnel-oxide thickness t_{ox} = 8.1 nm.

The above experimental results are consistent with those shown in Fig. 4.13 and Fig. 4.14, that is, a higher IIF-induced tertiary-electron injection is present in cells with more abrupt channel engineering (higher V_t-adjust and halo implantations). Fig. 4.15 and Fig. 4.16 also confirm that CHISEL injection can be used for low-power programming in deep-submicron devices with abrupt vertical channel doping profile, as it can yield a significant amount of hot-electron gate current even at low V_g biases.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

4.5 Summary

This chapter presents the electrical characteristics of 0.18-µm and 0.14-µm (type A and B) N-channel memory cells used in our project. Typical output/transfer characteristics were illustrated first, followed by a comparison of the channel doping profile of the test memory cells. By applying either qualitative or quantitative characterization technique, cell A was shown to possess a heavier and more abrupt channel doping profile than cell B. However, as these cells show no distinct difference in the SCE performance (both V_t-roll off and DIBL) which is an indicator of the lateral distribution of channel dopants, we can conclude that the lateral channel doping profiles of both cells are comparable. This finding provides an excellent opportunity to distinguish the effect of the vertical channel doping profile on hot-electron injection characteristics.

Typical channel hot-electron injection characteristics were presented next, where I_{sub} and I_g were adopted as the electrical parameters to monitor the impact ionization and electron injection efficiency, respectively. Experimental characteristics of CHISEL injection, which is induced by a vertical-IIF mechanism, were compared to those of the traditional CHE injection. It is shown that the two injection modes may be clearly distinguished based on the values of the exponent n in the empirical power-law injection model $I_g{}^e{}_r \propto I_{br}{}^n$.

By comparing the hot-electron injection efficiency ($I_g{}^e{}_r$) among different test cells (of 0.18-µm and of 0.14-µm generations), a heavier and more abrupt channel doping

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

profile (as in 0.14-µm cell type A) formed by a heavy V_t-adjust implantation and halo doping scheme was shown to be able to yield a much more significant injection of tertiary electrons in deep submicrometer memory cells. This clearly indicates the significant influence of the vertical channel doping profile on hot-electron injection in modern Flash memory cells.

Last but not least, electron programming current generated by three biasing configurations (FN-tunneling, CHE injection and CHISEL injection) in different test cells were demonstrated. The results reconfirmed that an abrupt channel doping profile can lead to significant tertiary-electron injection. Thus, CHISEL injection can be employed for low-power programming or power-saving purposes in such memory cells, where appreciable amount of electron current can be generated even at low voltage biases.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Chapter 5: Non-Classical Hot-Electron Gate Current Injection

5.1 Observation of Non-Classical Gate Current Component

As presented in last chapter, the CHISEL injection of feedback-heating induced tertiary electrons is much more efficient than the CHE injection of primary channel electrons, especially in the low gate bias regime. In view of its growing importance in low-power memory operation for modern portable systems and relatively few studies available on this topic, it is necessary for us to investigate further the mechanisms involved in the impact-ionization feedback (IIF) process, which can have great influence on cell programming and device degradation.

Fig. 5.1 depicts a set of I_{gr}^{e} versus I_{br} characteristics for a 0.14-μm type A cell. In this measurement, the gate bias V_g was fixed while drain voltage V_d was varied. This is different from Fig. 4.14, where V_g and V_d were swept together with the same step-size in order to keep a constant vertical injection field ($E_{inj} \approx (V_g - V_d) / t_{ox}$) across the oxide. For $V_d < V_g$, a positive vertical field is presented in the channel pinch-off region; it gradually reduces as V_d increases, and finally reaches zero when $V_g = V_d$. With any further increment in V_d, the vertical electric field near the drain end becomes negative, a condition where the field is against the injection of hot electrons.

As shown in Fig. 5.1, for $V_b < 0$ V, the characteristic exponent n for the empirical

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

power-law model $I_{g\ r}^{\ e} \propto I_{br}^{\ n}$ is close to 1, consistent with previous work [84, 85] on CHISEL injection attributing the increased efficiency ($I_{g\ r}^{\ e}$) and reduced exponent (n) to the feedback heating of holes by the reverse substrate bias. However, at $V_g < V_d$, where the vertical oxide field at the drain end (E_{inj}) is opposing hot-electron injection, the efficiency $I_{g\ r}^{\ e}$ continues increasing. This is different from conventional understanding based on Lucky-Electron Model, in which a positive electric field is one of the main driving forces for hot-electron gate injection. Hence, the injection should be suppressed for a non-positive field. The result suggests that some other kind of energy gaining mechanism is responsible for the substantial hot-electron injection at a repulsive E_{inj}.

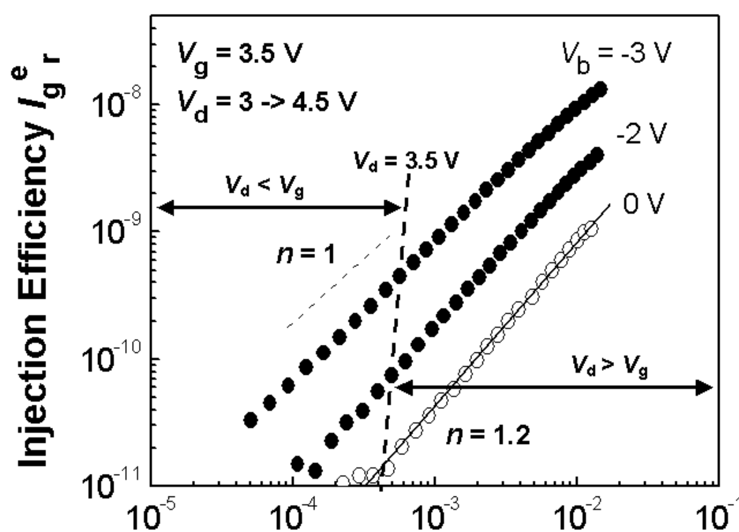


Fig. 5.1 $I_{g\ r}^{\ e}$ versus I_{br} characteristics for Cell A (14-μm generation) when V_g was fixed while V_d was increased from 3 to 4.5 V. Despite $V_g < V_d$, the hot-electron gate current continues increasing with V_d. Drawn channel dimension of the test cell (W/L) = 10/0.176 μm/μm, and tunnel-oxide thickness t_{ox} = 8.1 nm.

In addition, another interesting observation is that in the CHE dominated regime (i.e. at $V_g \sim V_d$ and V_b = 0 V), the characteristic exponent n is about 1.2, much lower than that reported in literature, which is about 2.5 ~ 3.5 as determined by the ratio ϕ_{ox}/ϕ_{ii}. This near-unity exponent at V_b = 0 V implies that the injection current $I_g^{\ e}$ does not originate from

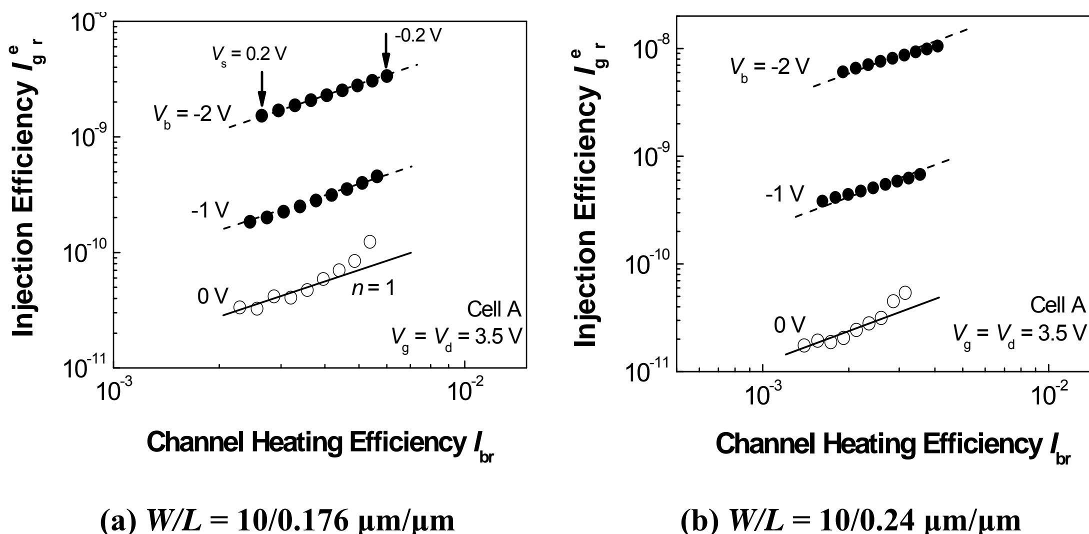
ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

primary channel hot electrons; instead, the injection mechanism should be similar to the vertical IIF (VIIF) mechanism in CHISEL injection [78, 84, 85]. But as no reverse V_b was applied here, the unitary injection feature may not be simply ascribed to the conventional VIIF mechanism. An alternative IIF process is believed to be responsible for this non-classical hot-electron effect in the traditional CHE injection dominated regime.

Fig. 5.2 depicts another set of experimental evidence obtained in test cell A, showing a non-classical hot-electron effect under conventional CHE biasing condition. In this experiment, V_g and V_d were fixed at 3.5 V, while the source voltage V_s was swept in a small range (i.e. -0.2 to 0.2 V). Varying V_s alters the lateral channel field slightly and thus the heating efficiency (I_{br}), but not the vertical field [84]. Therefore, the vertical electric field distribution at the drain end is virtually unchanged and any observed change in $I_{g}{}^{e}{}_{r}$ can be regarded as modulated by the lateral channel field ($E_{//}$) only.

In Fig. 5.2(a), the exponent $n \sim 1$ for $V_b < 0$ V, implying that the hot-electron gate current arises from the VIIF mechanism [85]. For $V_b = 0$ V, one would expect $n \sim 2.5 - 3.5$ as predicted by the LEM [61]; however, here n is still close to 1 at $V_b = 0$ V, although it tends to increase at larger I_{br}. Similar results were attained for type A cell with a different channel length (Fig. 5.2(b)), implying that the non-classical hot-electron gate injection component at $V_b = 0$ V is related to an inherent physical mechanism. Furthermore, the data points begin to deviate from the $n = 1$ extrapolation line as V_s becomes more negative (i.e. larger V_{ds}), indicating a transition to a possibly different hot-electron injection mechanism from low to high I_{br} regime.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library



(a) W/L = 10/0.176 μm/μm **(b) W/L = 10/0.24 μm/μm**

Fig. 5.2 $I_g^e{}_r$ versus I_{br} characteristics are modulated by varying V_s from 0.2 to -0.2 V. Evident non-classical hot-electron injection is shown at V_b = 0 V. The test cells are type A cells fabricated by commercial 0.14-μm CMOS technology.

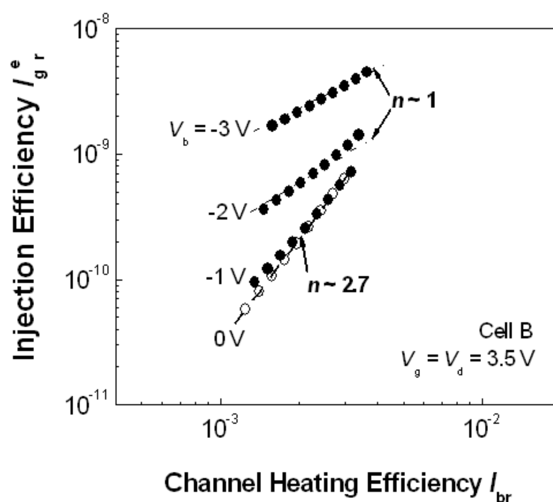


Fig. 5.3 $I_g^e{}_r$ versus I_{br} characteristics of type B cell as V_s is varied from 0.2 to -0.2 V. Typical CHISEL injection signature appears at high reverse V_b bias (-2 and -3 V); however, no non-classical HCI component at V_b = 0 V.

Similar experiment was also carried out on type B cell. Fig. 5.3 shows the experimental result for a cell with drawn channel dimension W/L = 10/0.176 μm/μm. Unlike cell A, a high exponent value (n ~ 2.7) is presented in cell B at V_b = 0 V for the

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

empirical power-law relationship, indicating that the hot-electron injection behavior follows the prediction by LEM and the non-classical injection component is suppressed. Moreover, the conventional CHE injection is so significant that it even covers the effect of CHISEL injection at V_b = -1 V, and the unitary exponent only starts to appear at V_b = -2 V when CHISEL injection becomes prominent. The suppression of CHISEL and non-classical electron injection in cell B may be attributed to its lower and more graded channel doping as compared to that in cell A. Therefore, we have focused our study of the non-classical component on type A cell as it is more sensitive to the non-classical hot-electron injection.

5.2 Experimental Feature of the Non-Classical Hot-Electron Effect in Conventional CHE Injection Mode

To further investigate the non-classical hot-electron effect, the range for source voltage (V_s) variation was increased from $\pm$0.2 to $\pm$0.5 V. The results of hot-electron gate injection efficiency ($I_g{}^e{}_r$) at different gate (V_g) and drain (V_d) biases as V_s changes are shown in Fig. 5.4. Here, a more negative V_s corresponds to a higher $V_{ds,}$ which in turn yields a larger I_{br}.

At V_b = 0 V, two distinct values of exponent n can be observed in Fig. 5.4. In the high I_{br} regime (when V_s is towards the negative limit), the gate injection efficiency increases rapidly with the channel heating and the exponent n is about 2.7. This is expected since a more negative source bias can lead to a larger drain-source potential difference and thus a greater channel heating efficiency. Therefore, injection of

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

channel hot electrons is prevailing, giving rise to a high value of n as predicted by the classical LEM model. On the contrary, when V_s switches from negative to positive and increases in the positive direction, channel hot-electron injection induced by the electric field is strongly suppressed because V_{ds} decreases and approaches the Si-SiO_2 barrier height (~ 3.2 eV). The injection efficiency ($I_g^e{}_r$) is observed to be correlated to I_{br} by a linear relationship, i.e. exponent $n \sim 1$, which unambiguously denotes the presence of a non-classical hot electron effect at low I_{br} regime, under a conventional CHE injection dominated biasing condition.

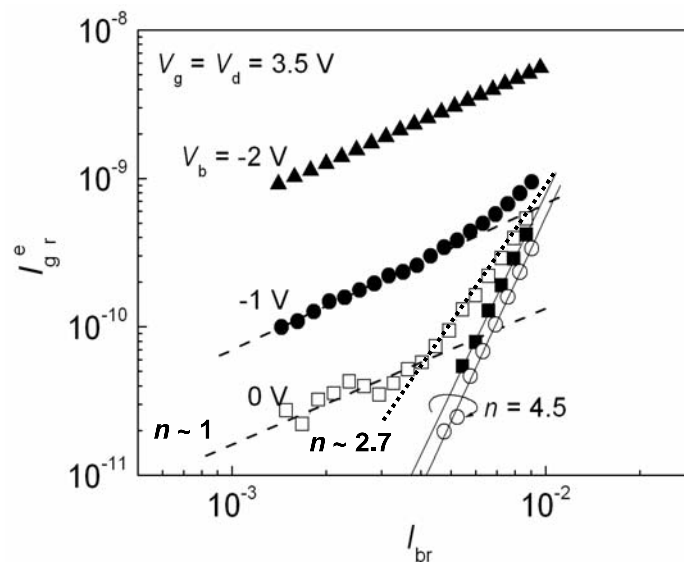


Fig. 5.4 Modulating the injection efficiency ($I_g^e{}_r$) by a wider V_s variation (from 0.5 to –0.5 V) on a 0.14-μm type A cell (drawn channel dimension W/L = 10/0.176 μm/μm). Dash lines (with $n \sim 1$) are the best fit of low I_{br} data points and correspond to the $I_g^e{}_r$ component contributed by the non-classical hot-electron effect. Subtracting this non-classical $I_g^e{}_r$ component from the total $I_g^e{}_r$ in high I_{br} regime can yield the characteristics of high-field electron injection solely (open circles: for V_b = –1 V, solid squares: for V_b = 0 V).

The above result clearly indicates that there are two different hot-electron injection mechanisms in the test cell at V_b = 0 V. By extrapolating $I_g^e{}_r$ measured in low I_{br} to high I_{br} regime (as dashed lines in Fig. 5.4) and subtracting this component from the

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

overall $I_{g\ r}^{e}$ obtained in the high I_{br} regime, the respective contribution of the two mechanisms to the total high field $I_{g\ r}^{e}$ may be clearly delineated. The part of $I_{g\ r}^{e}$ which results from high-field injection is denoted by filled squares and open circles for $V_b = 0$ V and –1 V, respectively, where a value of ~ 4.5 for n is extracted for $I_{g\ r}^{e}$. Theoretically speaking, if I_{br} can be extended further while keeping the vertical injection field unchanged, exponent n of total $I_{g\ r}^{e}$ at high I_{br} regime can approach 4.5 ~ 5 as well since the high-field electron injection will become predominant.

The result also suggests that the empirical exponent n (2.5 ~ 3.5) reported in the past for hot-electron injection at $V_b = 0$ V may be a consequence of the superposition of two different hot-electron injection components, which are:

(1) The non-classical hot-electron effect (with a low exponent n ~1); and

(2) The high-field hot-electron injection (with a higher exponent around 4 ~ 5).

However, the former effect is usually masked by the latter one as the non-classical electron injection possesses a low exponent value and is only noticeable by V_s variation at low field with fixed V_g and V_d.

5.3 A Phenomenological Model based on Impact-Ionization Feedback Mechanism

In view of the similar experimental signature between the non-classical hot-electron injection and the CHISEL injection, an IIF process analogous to that in CHISEL injection is believed to be the governing mechanism for the former. Moreover, as this non-classical injection is modulated by V_s instead of by V_b, the feedback heating of holes is deduced to

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

be along a lateral path. Consequently, a phenomenological model based on a lateral impact-ionization feedback (LIIF) mechanism was proposed. This alternative IIF mechanism, together with the conventional CHE and CHISEL injection modes, are illustrated in Fig. 5.5.

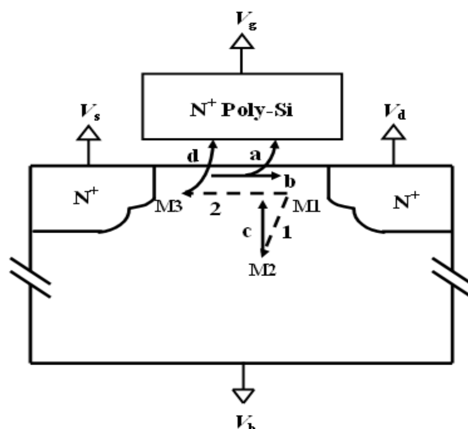


Fig. 5.5 Cross-sectional view of an N-channel MOSFET under different hot-electron injection mechanisms.
'*a*': Classical CHE injection. '*b*': Electron-flow to drain terminal. '*1*': Vertical heating of hot holes generated in the primary impact ionization (PII) process. '*c*': CHISEL injection induced by a vertical impact-ionization feedback (LIIF) mechanism. '*2*': Lateral heating of hot holes. '*d*': Non-classical tertiary-electron injection induced by the sub-surface lateral impact-ionization feedback (LIIF). M_1, M_2 and M_3 are the multiplication factors for PII, VIIF and LIIF, respectively. ('- ►' indicates electron-flow, while '⟶' denotes hole-flow.)

In Fig. 5.5, arrow 'a' refers to the part of channel electrons which are deflected to the gate in the CHE injection mode. The rest of the electrons continue traveling along the channel and are finally collected as the device current I_d, indicated by arrow 'b'. Some electrons are energetic enough to cause primary impact ionization (PII) near the drain end with a multiplication factor M_1. The secondary hot holes generated in this process, in general, are able to travel along two possible trajectories.

Dashed arrow '1' indicates a vertical trajectory due to the potential difference

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

between the drain and the substrate. It is dominant when a reverse V_b is applied. The vertical electric field accelerates holes as they are flowing from the positively-biased drain to the negatively-biased substrate terminal, during which the secondary impact ionization (SII) with a multiplication factor M_2 takes place. Tertiary electrons are then generated and can inject towards the gate through the vertical feedback path ('c'). This process is known as the CHISEL injection of tertiary electrons for $V_b < 0$ V.

Alternatively, the hot holes generated in PII can also drift along a sub-surface lateral trajectory, denoted by dashed arrow '2' in Fig. 5.5, since the potential at source side is often much smaller than that at the drain side under most biasing conditions. During the lateral drifting, a SII process with multiplication factor M_3 will occur in the mid-channel region towards the source side. As the attractive oxide field $E_\perp \approx (V_g - V_{channel}(x))/t_{ox}$ becomes larger when carriers move further away from the drain end, where x is the distance of the carrier injection point to the drain end, the injection probability of tertiary electrons in the sub-surface path is enhanced and becomes prominent. Thus, the non-classical component of the LIIF-induced tertiary electron injection ('d') appears.

Fig. 5.6 depicts a schematic diagram of the drain region for an N-channel device when LIIF injection occurs at $V_g \geq V_d$. Electrons accelerated along the channel *(solid arrow)* can induce PII at the drain end (event *(1)*). A unitary exponent n in low I_{br} regime when V_s varies strongly suggests that a lateral sub-surface path is present in the channel for the feedback heating of holes. This feedback heating (*open arrow*) can result in SII (event *(2)*). Tertiary electrons are produced at a distance in the order of the mean-free-path of holes.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Subsequently, those electrons are injected towards the Si-SiO_2 interface by the attractive vertical field inside the channel region. In 0.14-µm type A cells, as the vertical channel doping profile is more abrupt than that in type B cells, a large local built-in electric field exists, which further enhances the injection of tertiary electrons.

Fig. 5.7 represents another schematic diagram of the drain region for an N-channel device when LIIF injection occurs at $V_g < V_d$, corresponding to a repulsive electric field for hot-electron injection at drain side. Under this condition, the primary electron path at the drain side deviates deeper into the substrate (*solid arrow*), reducing the probability of CHE injection; however, this negative field favors the transport of holes towards the Si-SiO_2 interface. At the same time, the holes tend to drift towards the source, forming a sub-surface hole flow (*open arrow*). The feedback heating of holes results in SII (event *(2')*) during which tertiary carriers are generated along the mid-channel region towards the source side. Again, the attractive vertical field in this region would assist the injection of tertiary electrons, thus the hot-electron gate current I_g^e increases proportionally with the channel heating efficiency (I_{br}) even when $V_d > V_g$.

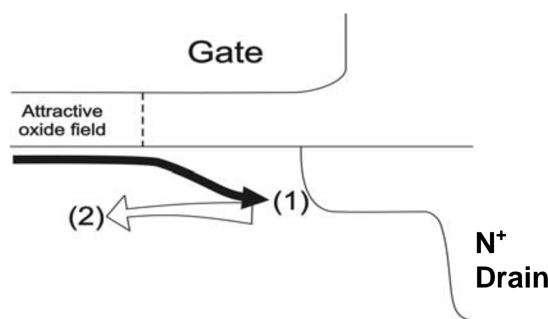


Fig. 5.6 Schematic diagram of the drain region illustrating the primary electron path (*solid arrow*) and the lateral feedback heating of holes (*open arrow*) at biasing condition $V_g \geq V_d$. Event *(1)*: primary impact ionization at drain side. Event *(2)*: secondary impact ionization in the channel.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

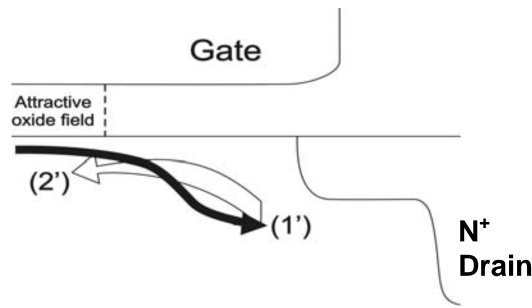


Fig. 5.7 Schematic diagram of the drain region at biasing condition $V_g < V_d$. The corresponding primary electron path is deeper and the feedback hole path is shallower as compared to the condition when $V_g \geq V_d$.

5.4 Summary

In the study of the IIF-induced hot-electron injection in 0.14-µm type A cells, a non-classical hot-electron effect was directly observed in the conventional CHE injection dominated regime (i.e. $V_g \sim V_d$ and $V_b = 0$ V). It exhibits the same experimental signature ($n \sim 1$) as that of the CHISEL injection. Thus, it is believed to be due to the injection of tertiary electrons originated by a kind of IIF mechanism, which is similar to that in CHISEL injection but requires no application of reverse substrate bias. In addition, varying source bias can alter the lateral channel field, which can effectively modulate the injection efficiency of the non-classical component. Therefore, a phenomenological model based on a concept of lateral impact-ionization feedback (LIIF) mechanism is proposed to explain the non-classical hot-electron injection at $V_b = 0$ V.

In this model, some of the hot holes generated in primary impact ionization drifts laterally from the drain to the source side via a sub-surface trajectory. In this lateral path, holes are further heated up, which induces the secondary impact ionization and

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

creates tertiary electrons in the channel region. Subsequent injection of tertiary electrons towards the floating-gate is assisted by an attractive vertical field across the tunnel-oxide, which may be further enhanced by a high local built-in electric field due to an abrupt channel doping profile. Therefore, electron injection efficiency at low V_{ds} is enhanced and the experimental signature of IIF is observed even at zero body bias. As the source bias V_s is varied negatively (i.e. towards larger V_{ds}), a transition from the $n \sim 1$ to the $n \sim 2.5$ regimes occurs, as shown in Fig. 5.4. It is well-known that at high V_{ds} condition the conventional CHE injection becomes dominant, which yields an $n \sim 2.5 – 3.5$ for the empirical power-law relationship. A hypothesis is thus established in which the exponent $n \sim 2.5 – 3.5$ shown at high V_{ds} may be a result of the superposition of two hot-electron injection mechanisms; one is the lateral-IIF injection, and the other is the conventional CHE injection which is strongly suppressed as V_{ds} falls below3 V.

The direct observation of the LIIF-induced non-classical hot-electron effect is very important for accurate modeling and simulation of device hot-carrier lifetime. This is because the lifetime test is often carried out under accelerated stress conditions, where the stress field can be very high in order to reduce the test cycle-time. By simply extrapolating the accelerated test results acquired in the high field to low field working condition, we may over-estimate the actual lifespan of the test cell at normal condition, in view of a transition to a different (other than CHE) hot-electron injection regime. Consequently, to understand how the non-classical tertiary hot-electron injection would affect lifetime estimation, a comprehensive study has to be carried out as shown in the next chapter.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Chapter 6: Influence of Non-Classical Injection on Device Degradation

6.1 Introduction

One of the main concerns in deep-submicrometer Flash memory technology is device degradation induced by hot-carrier injection (HCI) into the tunnel oxide. This is especially important for future scaling of the memory cell, as the programming voltage is rarely scaled or is scaled at a much slower rate than the device channel length due to performance consideration. As a consequence, a large electric field exists in the channel region. On one hand, this high injection field can enhance programming efficiency; on the other hand, it also leads to hot-electron injection that could result in severe oxide damage (including both oxide and Si-SiO_2 interface traps). The degradation of oxide and Si-SiO_2 interface integrity will limit the lifetime of memory cells as many important transistor parameters, such as threshold voltage, drive current and transconductance, are largely dependent on the quality of the tunnel oxide and the Si-SiO_2 interface.

Many hot-carrier lifetime models have been proposed in the literature. Most of them describe empirically the time dependence of device degradation in accelerated stress modes, where degradation is often monitored by changes in electrical parameters, such as a shift in threshold voltage (ΔV_{th}), the reduction in drain current

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

(ΔI_d) and decrease in transconductance (ΔG_m), and can be generally expressed as:

$$\Delta D(t) = D_0 t^n \quad (6.1)$$

where $\Delta D(t)$ is the absolute parameter shift during aging, D_0 refers to the specific degradation parameter and is a strong function of V_d, n is the exponent which depends on V_g [100][101]. For long-time stress, ΔD tends to saturate, possibly because of the negative feedback of the trapped charge on the electric field at the injection point; however, when degradation saturates, ΔD often exceeds the maximum acceptable limit and equation (6.1) is sufficient for lifetime evaluation in most cases.

Derivation of the conventional lifetime model starts with the assumption that both I_b and I_g are proportional to the number of carriers with an energy level above the effective threshold (Φ_{IIeff} for impact ionization and Φ_{Beff} for overcoming the Si-SiO_2 barrier height). Based on the effective-temperature model (discussed in chapter 2), the following relationship can be obtained for gate and substrate currents [81]:

$$\frac{I_b}{I_s} = \exp(-\frac{\Phi_{IIeff}}{k_B T_e}) \quad (6.2) \quad \text{and} \quad \frac{I_g}{I_s} = C(E_{ox})\exp(-\frac{\Phi_{Beff}}{k_B T_e}) \quad (6.3)$$

where I_s is channel current , T_e is effective temperature, $C(E_{ox})$ is a function of oxide field that accounts for the back scattering in the image potential well. Hence,

$$\frac{I_g}{I_s} = C(E_{ox})(\frac{I_b}{I_s})^{(\frac{\Phi_{Beff}}{\Phi_{IIeff}})} \quad (6.4)$$

It predicts a linear relationship between I_b/I_s and I_g/I_s in a log-log plot. This correlation

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

between I_b and I_g was suggested in [102] and has been well verified over a broad range of technologies and bias conditions. In fact, we have also observed this linear relationship in our work, although the exponent is reduced significantly as a result of the non-classical hot-electron effect in cells with abrupt channel doping.

Assuming the interface traps are generated by only electrons with energy level larger than Φ_{ITG} (threshold for interface-state generation), we can then express D_0 and ΔD based on equations (6.1) – (6.4). After eliminating T_e in the above correlations, the hot-carrier lifetime τ_L to reach maximum allowable degradation can be expressed as:

$$\tau_L = \frac{C_\tau}{I_s} \cdot \left(\frac{I_b}{I_s}\right)^{\left(\frac{\Phi_{ITG}}{\Phi_{IIeff}}\right)} \quad \text{or} \quad \tau_L = \frac{C_\tau}{I_d} \cdot \left(\frac{I_b}{I_d}\right)^{(-m)} \tag{6. 5}$$

where C_τ is an empirical constant. The source current I_s is approximately equal to the drain current I_d assuming a low avalanche multiplication condition. According to (6.5), plotting I_b/I_s against $\tau_L I_d$ (i.e. expressing the device degradation as a function of stress time) yields a linear relationship in a log-log graph [103]. A slope of 2.9 is often observed for stress at $V_{gs} \sim V_{ds}/2$, which agrees well with the ratio $|-\Phi_{ITG}/\Phi_{IIeff}|$. Nevertheless, the slope can go up to ~ 5.5 for stress at low V_{gs} and high I_b/I_d. The larger slope may be due to the generation of fast interface traps by hole injection [100] and parasitic bipolar action [104].

In our work, a transition of the dominant hot-electron injection mechanism, from high-field CHE to low-field non-classical hot-electron injection, has been observed as V_{ds} decreases. The latter is attributed to tertiary electron injection induced by the lateral

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

feedback heating of holes, which occurs more homogeneously across the channel as compared to the localized CHE injection. As will be shown, the resultant oxide damage induced by the non-classical hot-electron injection is more spatially distributed in the cell [105] [106]. Therefore, simply extrapolating the HCI data from high field to evaluate the cell lifetime at normal operating field may result in a risk of over-estimation, as this technique ignores the fact that more spatially distributed oxide/interface traps can be introduced by the non-classical tertiary injection in the low-field regime.

6.2 Experimental

In order to study the influence of non-classical tertiary-electron injection on device degradation and related lifetime estimation, hot-carrier stress experiments were performed on two types of N-channel cells by applying constant gate and drain stress voltages and varying the source bias. This is to make sure that the vertical oxide field at the drain end remains unchanged by fixing V_g and V_d while the maximum lateral field in the drain pinch-off region is modulated by varying V_s. Test cells are fabricated by commercial 0.14-μm twin-well CMOS Flash technology. As discussed earlier, type A cell possesses a more abrupt vertical channel doping profile than cell B due to a higher V_t-adjust implantation dose. The tunnel-oxide thickness of both types of cell is about 8.1 nm. Source/drain junctions were achieved by a double-diffusion method (first implantation of arsenic with dose 1.7×10^{15} cm^{-2} and second implantation of phosphorous with dose 6×10^{13} cm^{-2}). The n^+ MDD regions were formed by arsenic

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

implantation (5 x 10^{13} cm^{-2}); the junction depth is ~0.05 µm. All sample cells have a 10-µm drawn channel width while the drawn channel length is either 0.176 µm or 0.24 µm. The purpose of employing test cells with different channel dimension is to verify the repeatability of the non-classical hot-electron effect. The experimental results obtained are to be presented and discussed in following sections.

6.3 Results and Discussion

The spatial distribution of hot-electron injection induced oxide damage may be evaluated by comparing the fractional degradation of the linear drain current ($\Delta I_d/I_{do}$) measured at $V_g \sim V_{tho}$ (0.7 to 0.8 V) and $V_g >> V_{tho}$, as proposed by Ang *et al.* [89,107]. For the latter biasing case, V_g is chosen to be ~ 3 V to meet the $V_g >> V_{tho}$ requirement. Here, V_{tho} refers to the threshold voltage of a fresh device, I_{do} denotes the linear drain current of the fresh device measured at a specific V_g, and ΔI_d is the absolute change in the measured drain current at that specific V_g. Superscripts 's' and 'm' are used to distinguish *stress* voltage and *measurement* bias in subsequent result discussion for easier reference and clearer explanation.

At a low gate voltage V_g^m (i.e. when $V_g^m \sim V_{tho}$), the channel resistance $R_c >>$ source/drain series resistance R_{sd}. The influence of ΔR_{sd} (due to negative oxide and interface trapped charge) on I_d is small and the degradation of linear I_d mainly depends on ΔR_c. Due to the stress-induced interface states (ΔN_{it}) and electron trapping (ΔN_{ot}) in the

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

channel region, both concentration and mobility of channel electrons decrease, leading to an increase in R_c and a decrease in linear I_d. In contrary, at a high V_g^m, $R_c << R_{sd,}$ it is the increase of R_{sd}, due to ΔN_{it} generated in the gate-drain overlap region, that accounts for the linear I_d degradation, as reported in earlier works [108] [109]. This means that the effect of ΔN_{it}, induced by electron injection at different channel locations (in the gate-drain overlap or spacer-oxide region), can be investigated separately by studying the behavior of linear I_d degradation measured at different V_g^m [107]. In other words, linear I_d degradation measured at $V_g^m \sim V_{tho}$ probes the hot-electron injection induced damage in the channel, while the degradation measured at $V_g^m >> V_{tho}$ probes oxide damage localized at the drain end [89][107]. Besides linear I_d, transconductance (G_m) can also be used to monitor device deterioration; however, as this parameter is influenced by both R_c and R_{sd} in a similar manner and extent, it cannot be used to differentiate the hot-electron induced damage in the channel from that in the spacer region. As a consequence, in this work we only the characteristics of linear I_d to monitor the degradation of memory cells.

Fig. 6.1 shows the percentage of linear I_d reduction ($\Delta I_d/I_{do} \times 100\%$) measured at $V_g^m \sim V_{tho}$ (left figure) and at $V_g^m = 3$ V (right figure) for a type A cell. Same measurements have been also carried out for a type B cell but the results are not presented here as they are similar to those of cell A. In both experiments, stress voltages V_g^s and V_d^s are kept constant at 3.5 V to establish a CHE injection condition (at $V_b^s = 0$ V) for N-channel cells with channel length ~ 0.176 μm. At $V_g^s \sim V_d^s$, hot-electron injection dominates, electron trapping and interface trap generation coexist [18]. Device lifetime limited by hot-electron

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

injection can be extracted from Fig. 6.1. Here, lifetime (τ) is defined as the hot-carrier stress time when 10% degradation of I_{do} occurs for a specific V_g^m. Based on this definition, hot-carrier lifetime of cell A and cell B were determined, and the result was plotted against the channel heating efficiency (I_{br}) as shown later.

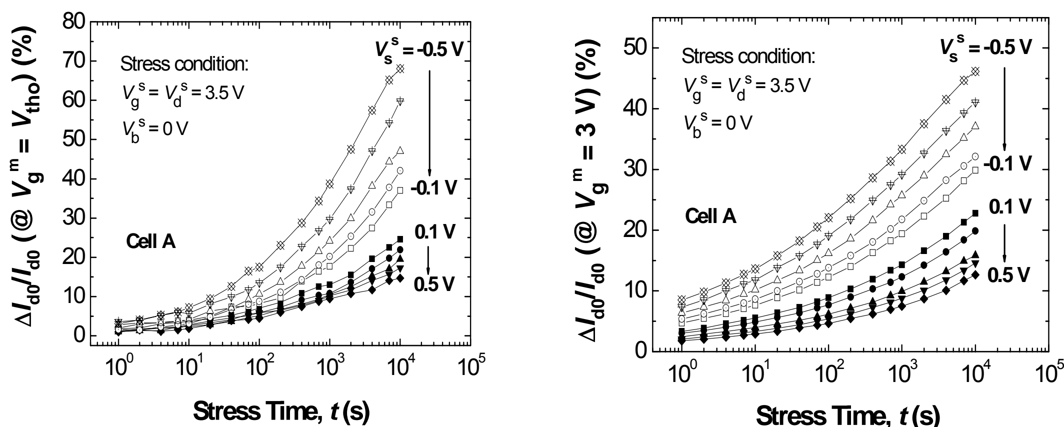


Fig. 6.1 Degradation of the linear drain current I_{do} (when $V_d^m = 0.1$ V) versus the hot-electron stress time for cell A. Left figure depicts $\Delta I_d/I_{do}$ at $V_g^m \sim V_{tho}$, indicating a condition where the oxide damage is extended into channel region. Right figure depicts $\Delta I_d/I_{do}$ at $V_g^m = 3$V, which is mainly due to the damage induced at the drain side.

Fig. 6.2 depicts the classical evolution of the hot-electron induced oxide damage from the drain-end into the channel region for both types of test cell, revealed by a cross-over in the $\Delta I_d/I_{do}$ characteristics measured at different V_g^m as stress time increases. Here, the source stress voltage $V_s^s = -0.3$ V, corresponding to an effective $V_{ds}^s = 3.8$ V.

Fig. 6.3 shows that the lateral spread of the interface states generation in cell B reduces significantly when the stress V_s^s becomes more positive (as indicated by solid squares), suggesting most of the damage tends to be confined in the drain-end as V_{ds}^s decreases. We can thus conclude that the generation of oxide damage in cell B is mainly ascribed

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

to the conventional CHE injection, which is dominant at high field and creates oxide/interface traps at the drain depletion region. Therefore, when the channel field is reduced (by a positive increase in V_s^s), the spread of oxide damage in cell B drops drastically.

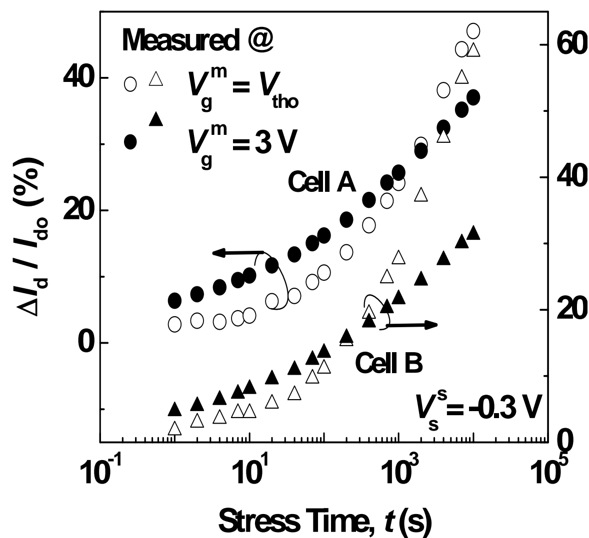


Fig. 6.2 Degradation of the linear drain current I_{do} (when $V_d^m = 0.1$ V) versus the hot-electron stress time, presenting a classical evolution of the main hot-electron damage region from the drain-end towards the middle-channel region.

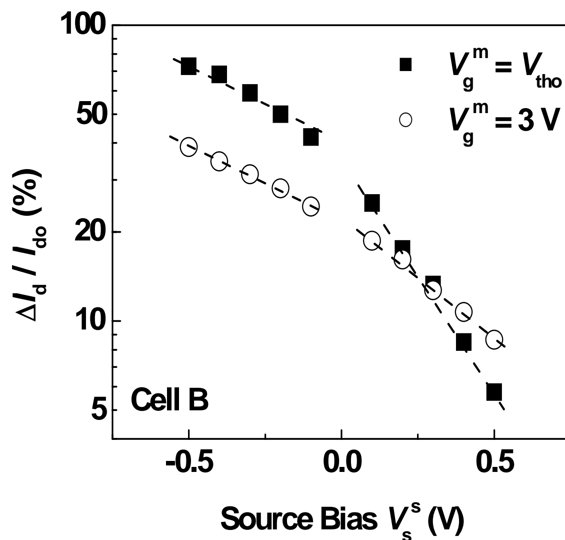


Fig. 6.3 The lateral spread of hot-electron damage region in cell B decreases drastically as V_{ds}^s reduces for a positive increase in V_s^s.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

However, for cell A, there is still a substantial spread of the damage region into the channel even when V_{ds}^{s} is reduced due to the positive increase of V_{s}^{s}. The result of cell A is shown by the solid circles in Fig. 6.4.

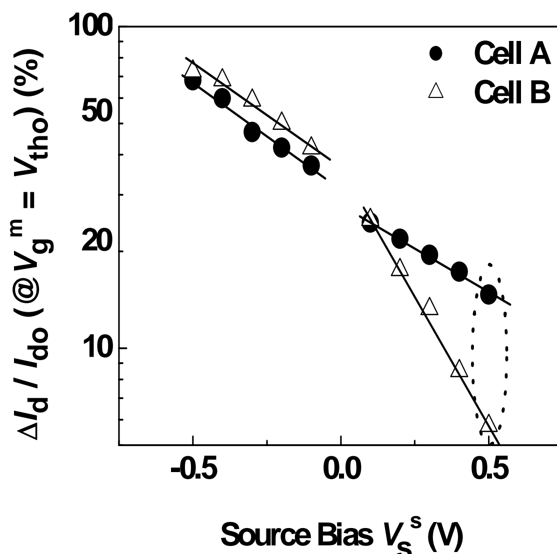


Fig. 6.4 A substantial lateral spread of the hot-electron damage region into device channel is observed for cell A even when V_{ds}^{s} is decreased.

Moreover, it is noticed that the increment of spatial distribution of oxide damage in cell A can not be well sensed by linear I_d degradation measured at high V_g^m. As depicted in Fig. 6.5, the degradation of linear I_d at a high V_g^m (= 3 V) for both test cell A and B yields similar characteristic over the entire range of V_s^s, leading to an incorrect impression that the hot-electron effect in cell A completely conforms to the LEM, just as in the case of cell B. However, Fig. 6.3 and Fig. 6.4 clearly indicate that cell A exhibits a different defect/damage distribution characteristic compared to cell B. The above result actually suggests that degradation of linear I_d at a high V_g^m is not able to reflect the actual spatial extent of hot-electron induced damage in cell A, which

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

arises from non-classical tertiary-electron injection.

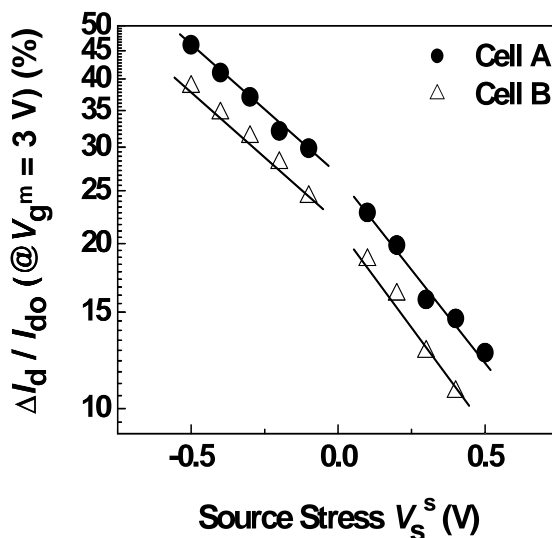


Fig. 6.5 At a much higher gate overdrive, the effective hot-electron damage region in cell A appears to be reduced and the I_{do} degradation characteristic follows that of cell B.

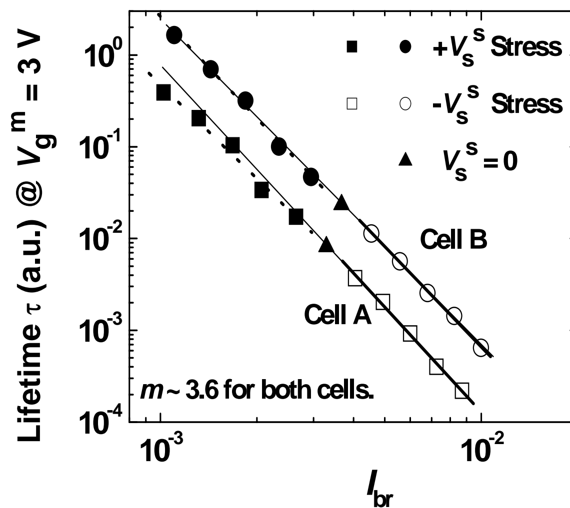


Fig. 6.6 Hot-electron lifetime characteristic when the degradation metric ($\Delta I_d/I_{do}$ x 100% = 10%) is measured at V_g^m = 3 V. Damage in the channel region is not probed and thus the lifetime characteristic of cell A appears similar to that of cell B. The thick-solid line is the best-fit for the open symbols, while the dotted line is the best-fit for the filled symbols. The thin-solid line is an extrapolation of the thick-solid line, showing almost the same slope (m ~3.6) over the whole I_{br} range under high gate-drive condition.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Fig. 6.6 shows the classical power-law dependence of the hot-electron lifetime τ on channel heating efficiency I_{br} (=I_b/I_d), that is, $\tau \propto I_{br}^{m}$ according to equation (6.5). As evident in Fig. 6.6, the exponent *m* for both cells is almost the same and is approximately equal to 3.6. This result is not unexpected since the degradation metric measured at a high gate bias (V_g^m) is mainly determined by the oxide damage at drain side. Therefore, consistent with Fig. 6.5, no difference was observed in the lifetime characteristic for both cells.

Fig. 6.7 depicts cell lifetime characteristics when the degradation metric (($\Delta I_d/I_{do}$ x 100%) is measured at $V_g^m \sim V_{tho}$. As mentioned earlier, this parameter more sensitively probes the spread of oxide damage into the channel region. As shown in Fig. 6.7(a), a unique exponent m = 2.53 is obtained for cell B over the whole stress range. This conforms to the LEM model and suggests that CHE injection is the main driving force for cell B degradation. In contrast, Fig. 6.7(b) reveals a different lifetime characteristic of cell A. The power-law exponent (i.e. *m*) switches from 2.97 to 1.43 as V_s^s increases positively, indicating that a significant amount of oxide damage are created and it extends into the channel region despite the reduction of lateral V_{ds}^s. This may be attributed to the heavier and more abrupt channel doping profile of cell A, which enhances the non-classical tertiary-electron injection at low lateral field (or V_{ds}^s). As a result, significant oxide/interface traps can still be induced in the channel region, in spite of the reduction of CHE-induced damage at reduced V_{ds}^s. The above explains satisfactorily the increased cell A degradation at low field regime.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

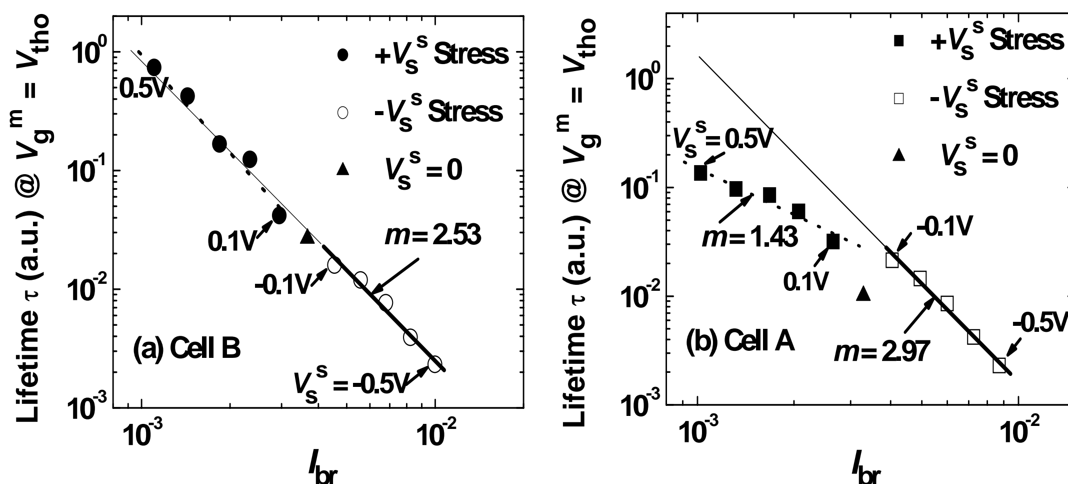


Fig. 6.7 Lifetime characteristic of cell A and cell B when the degradation metric ($\Delta I_d/I_{do}$ x 100%= 10%) is measured at V_g^m ~ V_{tho} to probe the oxide damage spread in the channel region. The thick-solid (dotted) line is the best-fit for the open (filled) symbols. The thin-solid line is the extrapolation of the thick-solid line.

Next, the influence of V_g^m selection on cell lifetime extraction was also studied. Experiments similar to that depicted in Fig. 6.1 were performed to measure linear I_d degradation at two other gate biases, that is, when V_g^m = 1 V and 1.3 V.

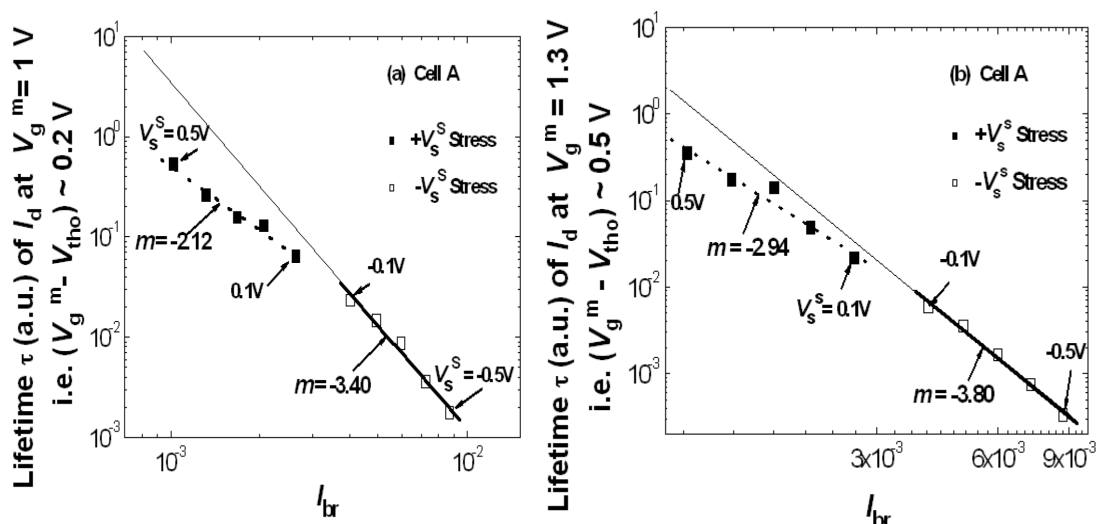


Fig. 6.8 Lifetime characteristic of cell A extracted at (a) V_g^m = 1 V and (b) V_g^m = 1.3 V. Degradation monitored at higher V_g^m mainly probes the oxide damage due to CHE injection. The thick-solid (dotted) line is the best-fit for the open (filled) symbols. The thin-solid line is the extrapolation of the thick-solid line.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Fig. 6.8(a) represents the lifetime characteristic of cell A when the degradation metric ($\Delta I_d/I_{do}$) is measured at a lower gate drive (i.e. V_g^m = 1V), whereas Fig. 6.8(b) shows the result measured at a higher gate drive (i.e. V_g^m = 1.3V). In the former case, a larger difference between the power-law exponents (m) in the positive and negative V_s^s stress regimes is observed; in addition, the magnitude of m in both regimes is also smaller. At higher V_g^m, m increases for both regimes and the difference diminishes; eventually, a unique m appears for the entire range of V_s^s stress (as shown in Fig. 6.6). This again confirms that linear I_d degradation measured at high V_g^m tends to probe the oxide damage at channel drain end, as that localized damage is mainly contributed by CHE injection which is invariant by fixing V_g^s and V_d^s, a very close m is obtained for lifetime characteristics of different V_{ds}^s (or I_{br}) conditions.

Fig. 6.9 depicts the lifetime characteristics of type A and B cells, when the drawn channel dimension (W/L) of both cells equals to 10 μm/0.24 μm. This is to verify the reproducibility of low field non-classical hot-electron effect on cells possessing different channel dimension. Fig. 6.9(a) represents the lifetime estimation based on the parametric degradation measured at $V_g^m \sim V_{tho}$. The result is shown to be consistent with that in Fig. 6.7. For type A cell, the actual cell lifetime directly measured in low-field regime is shorter than that extrapolated from the high-field stress data, as shown by the difference between the thin-solid line and the dotted line, regardless of channel dimension. However, for type B cell, the high-field extrapolation works fine for the lifetime prediction at operating (low field) condition.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

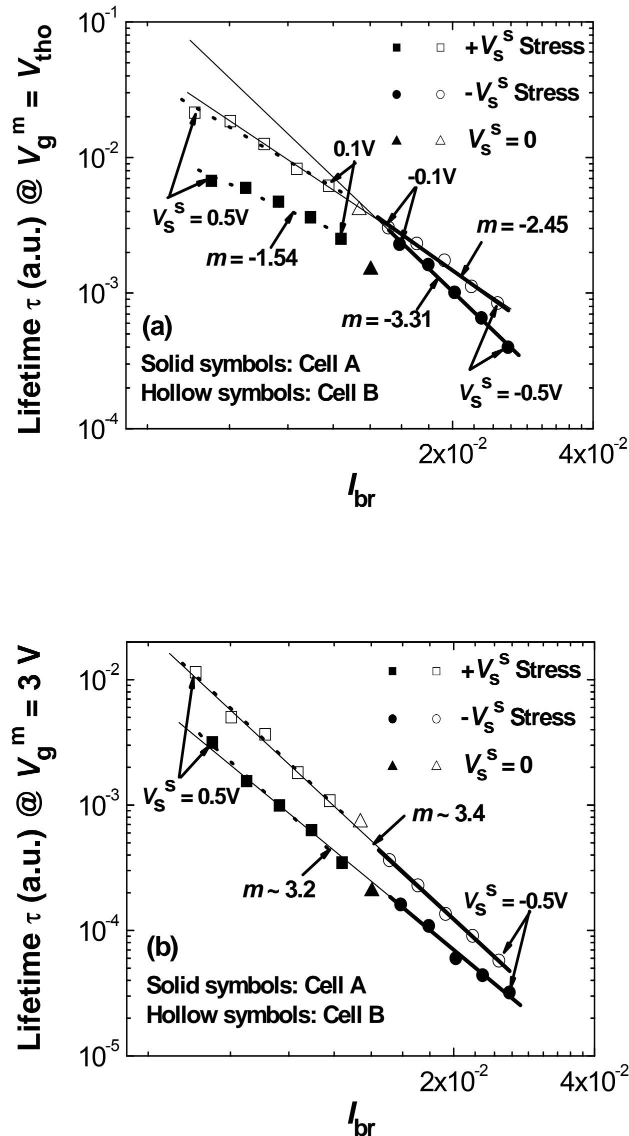


Fig. 6.9 Extraction of lifetime characteristics for type A and type B cells (both with W/L = 10 μm /0.24 μm) at (a) $V_g^m \sim V_{tho}$ and (b) V_g^m = 3 V. Similar trend as for cells with channel dimension W/L = 10 μm / 0.176 μm is also observed. The thick-solid (dotted) lines are the best-fit for data measured in high (low) I_{br} region, while the thin-solid lines are the extrapolation of the corresponding thick-solid lines.

The above result clearly points out that extrapolating the stress lifetime attained under high field (i.e. high I_{br}) condition to low field for hot-carrier lifetime prediction could lead to an over-estimation of transistor reliability, especially for scaled Flash

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

cell with abrupt channel doping profile. This is because the simple extrapolation neglects the fact that the oxide damage can extend into the channel region at a low lateral field condition due to the sub-surface LIIF induced tertiary electron injection. In addition, this non-local electron injection is more significant in memory device like cell A, in which the vertical abruptness of the channel doping profile can enhance the local built-in field and favors tertiary electron injection. In contrast, for cell B as its channel doping profile is relatively gradual, the non-classical electron injection is effectively suppressed by the dominant CHE injection. Therefore, the oxide damage and interface traps does not spread considerably into the channel region of cell B, and the lifetime extrapolation from high-field experimental data yield a similar result as that measured directly in the low-field (hollow symbols in Fig. 6.9(a)).

Fig. 6.9(b) illustrates the lifetime characteristic of cell A and cell B monitored by the linear I_d degradation at a higher gate drive (i.e. $V_g^m = 3V$). Similar to Fig. 6.6, both cells show unique exponents across the entire range of V_s^s variation. This shows that the oxide damage at drain end, which is mainly induced by the CHE injection component, always decreases exponentially with the reduction in lateral field.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

6.4 Summary

Device degradation induced by hot-electron stress and the influence of non-classical tertiary electron injection on cell lifetime prediction were investigated on two types of commercial 0.14-um Flash cells (i.e. cell A and B). The study was carried out by examining the stress-time dependent degradation of linear I_d at different gate drives for both cells. The location of main oxide damage, at either the drain side or along the channel region, can be clearly delineated by comparing the percentage of linear I_d reduction ($\Delta I_d/I_{do}$ x 100%) monitored at $V_g^m \sim V_{tho}$ and $V_g^m >> V_{tho}$, respectively.

First, a classical evolution of the hot-electron induced oxide damage from the drain-end into the channel region for both types of test cells was revealed by a cross-over in the $\Delta I_d/I_{do}$ versus stress time characteristics measured at different V_g^m.

Next, it was observed that in type A cells, the non-classical tertiary injection mechanism was more significant than that in type B cells at a decreasing channel field condition; the oxide damage induced by hot-electron injection can still be prominent in cell A, regardless of channel dimension (*W/L*). Furthermore, the non-local nature of non-classical tertiary injection was shown to be able to yield a substantial spread of oxide damage into the channel region even at a reduced V_{ds}. This is because tertiary electrons which are responsible for this non-classical effect are created in a sub-surface path due to the lateral feedback heating of secondary hot holes. As a

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

consequence, tertiary electron injection and its induced oxide damage are more spatially distributed along the channel region.

Last but not least, it was proven by experimental evidence that a simple extrapolation of high-field HC stress data for evaluating device lifetime at low operating field may lead to an over-estimation of the lifetime result. This is because it neglects the fact that a Flash cell with abrupt vertical channel profile could suffer more severe and distributed hot-electron induced damage under low-field condition due to the non-classical tertiary-electron injection which is non-local in nature. Thus, the hot-carrier lifetime does not follow the prediction of the conventional CHE model, which predicts a consistent inverse power-law dependency of hot-carrier lifetime on channel heating efficiency. The power-law exponent m does not remain unchanged across the whole biasing condition due to the non-classical hot carrier injection mechanism.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Chapter 7: Conclusion and Recommendations for Future Work

7.1 Conclusion

A comprehensive study of the impact-ionization-feedback (IIF) induced tertiary electron injection and its influence on the degradation of deep submicrometer floating-gate memory cell has been carried out in this work.

Initially, our focus is on the study of CHISEL injection, which is activated upon the application of a reverse body bias (i.e. $-V_b$ for N-channel device). It was reported to be an efficient and effective programming scheme for low-power memory applications, due to the injection of energetic tertiary electrons generated by the impact-ionization-feedback (VIIF) mechanism. In this project, we have verified the conclusion on our sample cells fabricated by commercial 0.18- and 0.14-µm twin-well CMOS technology, and between two types of 0.14-µm generation test cells (e.g. A and B) featuring different channel doping profiles. Electron gate currents generated by three different injection schemes (i.e. Fowler-Nordheim tunneling, CHE and CHISEL injection) under similar biasing voltages were compared. CHISEL injection indeed yields the highest programming current in low gate voltage range for optimized test cell (e.g. cell A) while FN-tunneling only becomes important when the injection field is large enough. This confirms that CHISEL is a good candidate for low-voltage

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

memory programming. It is also shown that CHSEL injection exhibits a unitary exponent for the power-law relationship $I_{g\ r}^{\ e} \propto I_{br}^{\ n}$ (i.e. $n \sim 1$), while the classical high-field channel hot-electron injection shows a higher $n \sim 2.5 - 3.5$. Therefore, the unitary injection exponent in the log-log plot of $I_{g\ r}^{\ e}$ versus I_{br} characteristic can be considered as the experimental signature for the IIF-induced tertiary electron injection, which reveals a tight linear relationship between the channel heating ($I_{br} = I_b/I_d$) and the tertiary-electron gate injection ($I_{g\ r}^{\ e} = I_g^{\ e}/I_d$). Last but not least, the dependence of tertiary hot-electron injection on effective channel doping concentration has been investigated. A significant CHISEL injection current was observed in scaled cell with heavier and more abrupt channel doping profile, whereby local built-in electric field due to the abruptness in channel doping profile is shown to be able to assist the injection of tertiary hot electrons.

In addition to verifying the characteristics of VIIF governed CHISEL injection, a non-classical hot-electron injection component, which can be attributed to the lateral impact-ionization-feedback (LIIF) heating of secondary holes, was reported for the first time. Unlike CHISEL injection which is activated upon a reverse substrate bias, this non-classical hot-electron injection is observed in the conventional CHE biasing regime (i.e. $V_g \sim V_d$ and $V_b = 0$ V). However, it also exhibits a unitary injection slope, i.e. $I_{g\ r}^{\ e} \propto I_{br}$, which is the experimental signature of CHISEL injection. This observation aroused our interest to investigate the possible mechanism that governs the non-classical hot-electron effect, the influence of this mechanism on device

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

degradation and lifetime prediction, as well as its implications on device modeling, memory scaling and further applications; the dependence of this non-classical effect on channel doping profile is also examined accordingly.

First, similar experimental signatures of the non-classical hot-electron effect and the CHISEL injection suggest that the two processes arise from analogous hot-carrier generation and injection mechanism. Since the former does not depend on body bias and its efficiency can be readily modulated by a slight variation of source voltage under low channel field, a phenomenological model, based on the concept of sub-surface lateral impact-ionization feedback (LIIF) heating of holes, has been proposed as its origin. Essentially, the LIIF model claims that besides the conventional vertical feedback heating of hot holes, which can finally yield CHISEL, a sub-surface heating trajectory is also possible whereby hot holes drift from the drain towards the source side under the influence of the channel electric field. This lateral feedback heating process can also induce secondary impact ionization (SII) and create tertiary electrons in the channel for subsequent gate injection. Since both CHISEL and LIIF-induced non-classical injection originate from tertiary electrons, which are resulted from the SII of holes, the two mechanisms possess similar experimental characteristics.

In theory, the LIIF-induced non-classical component should always co-exist with CHE injection at V_b = 0 V; however, in practical experiment results, the former is

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

often masked by the more pronounced CHE component due to its larger power-law exponent in the latter case. Nevertheless, if the scaled device is doped with a heavy and abrupt channel doping profile, especially in the vertical direction as in cell A of this project, the non-classical component will distinctly show up under low-field condition.

Next, by monitoring the device degradation characteristics, it is evident that the non-local nature of the non-classical hot-electron injection at low electric field would lead to a substantial spread of oxide damage into the channel region. The interface states created in the channel and at the drain side can be probed by measuring the linear current degradation at $V_g^m \sim V_{tho}$ and $V_g^m >> V_{tho}$, respectively, where V_{tho} refers to the initial threshold voltage of a fresh cell. It is observed that although the lateral field is reduced (via a positive increment in source bias V_s^s), a significant amount of oxide damage can still be generated by the prominent LIIF-induced tertiary electron injection, in spite of the large reduction of CHE induced damage as V_{ds}^s is reduced.

The spread of oxide damage into the channel region at low lateral field implies possible over-estimation of device lifetime when extrapolating the test data attained under a highly accelerated stress condition to a normal work condition, based on the classical CHE model. By comparing the lifetime of two types of 0.14-μm cells (lifetime is defined as the stress time when 10% degradation of linear drain current occurs at a specific V_g^m), we can see that there is over an order of magnitude over-estimation of lifetime for cell

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

A when extrapolation is based on linear I_d degradation data measured at $V_g^m \sim V_{tho}$. This is because cell A is more prone to the LIIF-induced non-classical injection which induces a greater spread of oxide damage at low channel field. In contrast, if device degradation at $V_g^m >> V_{tho}$ is taken to be the metric, both cell A and cell B show a degradation characteristic that conforms to the conventional Lucky-Electron Model. This is because this degradation metric can only examine oxide damage created at the drain-end which is not affected by the non-classical tertiary injection.

In the last part of this work, we verified the reproducibility of the LIIF-induced non-classical effect and its influence on device degradation. Similar hot-electron stress test was performed on type A and type B cells of different channel lengths. The result was shown to be consistent with previous conclusion, regardless of device dimension. This confirms an inherent dependence of the non-classical hot-electron injection on channel doping profile and its impact on oxide damage generation. It is also noticed that the higher gate drive at which the device degradation is measured, the smaller deviation between the extrapolated and the actual measured low-field lifetime. This is because at higher gate drive, the measurement has a stronger tendency to monitor the damage induced by traditional CHE injection. As CHE injection is comparable in both types of cell, the difference between extrapolated and actual lifetime diminishes.

In summary, the work presented here has contributed towards understanding the hot-electron effect; particularly, the characteristics and influence of the

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

impact-ionization-feedback induced tertiary-electron injection for deep submicrometer floating-gate memory, in the following areas:

- Initial experiments conducted on our sample cells verified the experimental signatures of the CHISEL and CHE injections. CHISEL is confirmed to be a good candidate for low-power memory applications through a comparison of gate current generated by three different common programming schemes. These results add values to the understanding of hot-electron injection for programming state-of-the-art Flash memory cell.

- Next, the dominance of a non-classical hot-electron effect under conventional CHE biasing condition, due to aggressive channel profile engineering in scaled Flash memory cells, is observed directly and reported for the first time. It exhibits similar experimental signature as IIF-induced tertiary-electron injection and can be modulated by varying the channel field at low V_{ds} for $V_b = 0$ V.

- Furthermore, by virtue of the tight relationship between this non-classical injection component and channel heating efficiency, a phenomenological model based on the concept of lateral impact-ionization feedback of secondary hot holes is proposed to be the responsible mechanism for the non-classical injection component, and is verified by experimental means.

- The non-local nature of the non-classical hot-electron injection and a substantial

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

spread of oxide damage into channel region due to this effect at low field is validated. This work highlights an intricate issue that may arise during hot-carrier reliability assessment. An over-estimation of device lifetime may occur when simply extrapolating the data acquired in high field stress test while neglecting the non-classical effect under low-field condition.

7.2 Recommendations for Future Work

The study of hot-carrier injection and its influence on memory programming and reliability has received wide spread attention over the past three decades. Although the basic degradation mechanisms are not expected to change significantly, some reliability issues have nowadays become more outstanding as technology scaling approaches the physical limit and electron injection behavior varies depending on the changes in process condition and device structure. One good example is that the adoption of heavy and abrupt channel doping profile in an aggressively scaled memory cell would enhance the non-classical hot-electron effect in the low electric field regime and impose adverse effect on device lifetime. It is therefore important for the 'well-established' models to be continuously improved in order to incorporate the less predictable impact of technology advancement on the reliability of integrated circuits. In particular, the following areas may be of future interest:

- In view of the possible dominance of the LIIF-induced non-classical

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

hot-electron injection in the low field condition as well as its influence on device degradation (or hot-electron lifetime estimation), it is recommended to include this non-local injection phenomenon into the HCI model of a device simulator, especially when modeling the hot-carrier effect in a scaled memory cell with abrupt channel doping profile, so as to achieve more accurate and reliable simulation results.

- In future generation of memory cells which require to be programmed at a much lower supply voltage, electrons injected by the non-classical hot-electron mechanism may be considered for low power programming. This is because energetic tertiary electrons are involved in this mechanism, similar to CHISEL injection but without the need of a substrate bias.

- The non-local nature of this non-classical hot-electron injection can result in a substantial spread of oxide damage into the channel region even at reduced V_{ds}. This suppressed scalability of the oxide damage region poses a potential challenge to the scaling of high-κ dielectric based memory cell which is targeted for single-cell multi-bit application by employing the separate source-/drain-side hot-electron injection technique. Consequently, when one is designing a novel cell structure for reliable memory application, this limiting factor has to be taken into account.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Publication List

[1] Y. Zhang and D. S. Ang, "Non-classical hot-electron mechanism and its implications on the reliability and scalability of the high-κ dielectric N-MOS Flash memory cell," *Microelectronic Engineering*, vol. 84, Issues 9-10, pp. 1929 – 1933, September – October 2007.

[2] Y. Zhang, D. S. Ang, H. P. Kuan and K. T. Tan, "Non-classical hot-electron gate current in the deep submicrometer N-MOS Flash memory cell," *13th International Symposium on the Physical and Failure Analysis of Integrated Circuits (IPFA)*, pp. 75 – 79, July 2006.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Bibliography

[1] R. R. Schaller, “Moore's law: past, present and future,” *Spectrum IEEE*, vol. 34, pp. 52 – 59, 1997.

[2] D. Kahng and S. M. Sze, “A floating gate and its application to memory devices,” *Bell Syst. Tech. J.*, vol. 46, p. 1288, 1967.

[3] Jitu J. Makwana, Dr. Dieter K. Schroder. (2005). A Nonvolatile Memory Overview. [Online]. Available: http://aplawrence.com/Makwana/nonvolmem.html

[4] H. Iizuka, F. Masuoka, T. Sato, and M. Ishikawa, “Electrically alterable avalanche-injection type MOS read-only memory with stacked-gate structures,” *IEEE Trans. Electron Devices*, ED-23, 379, 1976.

[5] F. Masuoka, M. Asano, H. Iwahashi, and T. Komuro, “A new flash EEPROM cell using triple poly-Si technology,” *IEEE IEDM Tech. Dig*., p. 464, 1984.

[6] Y. Tarui, Y. Hayashi, and K. Nagai, “Electrically reprogrammable non-volatile semiconductor memory,” *IEEE J. Sol.St. Circ.,* vol. SC-7, p. 369, 1972.

[7] D. Guterman, *et al.*, “An electrically alterable nonvolatile memory cell using a floating gate structure*,” IEEE Trans. Electron Devices*, p. 576, 1979.

[8] J. Yeargain and K. Kuo, “A high density floating gate EEPROM cell,” *IEEE IEDM Tech. Dig.,* p. 24, 1981.

[9] M. Lenzlinger and E. H. Snow, “Fowler-Nordheim tunneling in thermally grown SiO2,” *J. Appl, Phys.,* vol. 40, p. 278, 1969.

[10] A. Wu, T. Chan, P. Ko and C. Hu, “A novel high-speed, 5-V programming EPROM structure with source-side injection,” *IEEE IEDM Tech. Dig.,* p. 584, 1986.

[11] S. S. Chung, “Low voltage/power and high speed Flash memory technology for high performance and reliability,” *The 3rd Workshop and IEEE EDS Mini-colloquium on Nanometer CMOS Technology*, 2003, Singapore.

[12] P. E. Cottrell, R. R. Troutman, and T. H. Ning, “Hot electron emission in n-channel IFGET’s,” *IEEE J. Sol. St. Circ.,* vol. SC- 14, p. 442, 1979.

[13] S. Schmitt-Landsiedel, and G. Dorda, “Interface states in MOSFETs due to hot-electron injection determined by the charge pumping technique,” *Electron. Lett.,* vol. 17, pp. 761 – 763, 1981.

[14] E. Takeda, A. Shimizu, and T. Hagiwara, “Role of hot-hole injection in hot-carrier effects and the small degraded channel region in MOSFETs,” *IEEE Electron Device Lett.,* vol. 4, pp. 329 – 331, 1983

[15] L. Su, *et al*., “A high-performance 0.08µm CMOS,” IEEE VLSI Symp., pp. 12-13, 1996.

[16] K. Takeuchi, *et al*., “0.15 µm CMOS with high reliability and performance*,” IEDM Tech. Dig.,* 1993.

[17] E. Li, E. Rosenbaum, J. Tao and P. Fang, “Projecting lifetime of deep submicron MOSFETs,” *IEEE Trans. Electron Devices*, vol. 48, No. 4, pp. 671 - 678, 2001.

[18] P. Cappelletti, C. Golla, P. Olivo., E. Zanoni, Flash Memories. Boston: Kluwer Academic Publishers, 1999.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[19] Yoshikawa K., *et al.,* "A Flash EEPROM cell scaling including tunnel oxide limitations," *Proc. European Solid State Device Res. Conf.*, p. P/2, 1990.

[20] T. H. Ning, C. M. Osburn, and H. N. Yu, "Effect of electron trapping on IGFET characteristics," *J. Electron. Mater.,* vol. 6, pp. 65 – 76, 1977.

[21] E. Takeda, Y. Nakagome, H. Kume, and S. Asai, "New hot-carrier injection and device degradation in submicron MOSFETs," *Proc. Inst. Electr. Eng.*, vol. 130, Part I, pp. 144 – 150, 1983.

[22] E. Takeda, H. Kume, T. Toyabe, and S. Asai,"Submicron MOSFET structure for minimizing hot-carrier generation*," IEEE Trans. Electron Devices*, vol. 27, pp. 1359 – 1367, 1980.

[23] B. Marchand, *et al.*, "Generation of hot carriers by secondary impact ionization in deep submicron devices: model and light emission characterization," *Proceed. Of 38th Annual 2000 IEEE International Reliability Physics Symposium,* pp. 93 – 97, 2000.

[24] T. H. Ning, and H. N. Yu, "Optically induced injection of hot electrons into silicon dioxide*," J. Appl. Phys*., vol. 45, pp. 5372 – 5378, 1974.

[25] J. D. Bude, M. R. Pinto, and R. K. Smith, "Monte carlo simulation of the CHISEL flash memory cell," *IEEE Trans. Electron Devices*, vol. 24, pp. 357 – 359, 2003.

[26] F. Driussi, D. Esseni, and L. Selmi, "On the electrical monitor for device degradation in the CHISEL stress regime*," IEEE Electron Device Lett.,* vol. 8, p. 237, 1987.

[27] C. Hu, "Hot-carrier effects in Advanced MOS Device Physics", VLSI Electronics Microstructure Science, vol. 18, ed. by N. G. Einspruch and G. S. Gildeblat, p. 119, Academic Press, San Diego, 1989.

[28] C. Hu, S. C. Tam, F. –C. Hsu, P. –K. Ko, T. –Y. Chan, and K. W. Terrill, "Hot-electron-induced MOSFET degradation: model, monitor, and improvement," *IEEE Trans. Electron Devices*, vol. ED-32, p. 375, 1985.

[29] S. Ogura, *et al.*, "Design and characteristics of the lightly-doped drain-source (LDD) insulated gate filed effect transistor, *IEEE Trans. Electron Devices*, vol. 29, pp. 611 – 618, 1982.

[30] R. Izawa, T. Kure, and E. Takeda, "Impact of the gate-drain overlapped device (GOLD) for deep submicrometer VLSI," *IEEE Trans. Electron Devices*, vol. 35, pp. 2088 – 2093, 1988.

[31] J. Y. Chen, "CMOS Devices and Technology for VLSI", *Prentice Hall,* p. 183, 1990.

[32] E. Takeda, C. Y. Yang, and A. M. –Hamada, *Hot-Carrier Effects in MOS Devices*. Academic Press, San Diego, 1996.

[33] P. E. Cottrell, R. R. Troutman and T. H. Ning, "Hot-electron emission in N-channel IGFETs,*" IEEE Trans. Electron Devices*, vol. 26, pp. 520 – 533, 1979.

[34] T. H. Ning, "Hot-carrier emission currents in N-channel IGFETs," *Tech. Dig. – Int. Electron Devices Meet*., pp. 144 – 147, 1997.

[35] D. Esseni, L. Selmi, R. Bez, E. Sangiorgi and B. Ricco, "Bias and temperature dependence of gate and substrate currents in n-MOSFETs at low drain voltage," *IEDM*, pp. 307 – 310, 1994.

[36] Y. Nakagome, E. Takeda, H. Kume, and S. Asai, "New observation of hot-carrier injection phenomena," *Jpn. J. Appl. Phys*., vol. 22, Suppl. 1, pp. 99 – 102, 1983.

[37] 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.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[38] Y. Nakagome, E. Takeda, H. Kume and S. Asai, "New oabervation of hot-carrier injection phenomena," *Jan. J. Appl. Phys.,* vol. 22, Suppl. 1, pp. 99 – 102, 1983.

[39] S. Tam, C. Hsu, P. K. Ko, C. Hu, and R. S. Muller, "Hot-electron induced excess carriers in MOSFET's," *IEEE Electron Device Lett*., vol. 3, p. 376, 1982.

[40] I. –C. Chen, C. W. Teng, "A quantitative physical model for the band-to-band tunneling-induced substrate hot electron injection in MOS devices," *IEEE Trans. Electron Devices*, vol. 39, pp.1646 - 1651, 1992.

[41] A. Roy, R. Kazerounian, A. Kablanian and B. Eitan, "Substrate injection induced program disturb: A new reliability consideration for Flash-EPROM arrays," *Proc. Int. Reliability Physics Symp*. p. 68, 1992.

[42] K. R. Hofmann, C. Werner, W. Weber, and G. Dorda, "Hot-electron and hole emission effects in short n-channel MOSFET's," *IEEE Trans. Electron Devices*, vol. 32, p.691, 1985.

[43] B. S. Doyle, M. Bourcerie, J. C. Marchetaux, and A. Boudou, "Dynamic channel hot-carrier degradation of NMOS transistors by ehnaced electron-hole injection into the oxide*," IEEE Electron Device Lett.,* vol. 8, p. 237, 1987.

[44] T. Tsuchiya, "Trapped-electron and generated interface-trap effects in hot-electron-induced MOSFET degradation," *IEEE Trans. Electron Devices*, vol. 34, p. 2291, 1987.

[45] A. Schwerin, W. Hansch, and W. Weber, "The relationshfip between oxide chage and device degradation: A comprehensive study of n- and p-channel MOSFET's," *IEEE Trans. Electron Devices*, vol. 34, p. 2493, 1987.

[46] B. S. Doyle, *et al*., "The generation and characterization of electron and hole traps created by hole injection during low gate voltage hot-carrier stressing of n-MOS transistors," *IEEE Trans. Electron Devices,* vol. 37, p. 1869, 1990.

[47] D. S. Ang, T. W. H. Phua, H. Liao, and C. H. Ling, "High-energy tail electrons as the mechanism for the worst-case hot-carrier stress degradation of the deep submicrometer N-MOSFET," *IEEE Electron Devices Lett.,* vol. 24, pp. 469 – 471, 2003.

[48] D. S. Ang, H. Liao, T. W. H. Phua, and C. H. Ling, "Evidence for a composite interface state generation mode in the CHE-stressed deep-submicrometer n-MOSFET," *IEEE Electron Devices Lett.,* vol. 51, pp. 2246 – 2248, 2004.

[49] F. –C. Hsu, and K. –Y. Chiu, "Evaluation of LDD MOSFETs based on hot-electron induced degradation," *IEEE Electron Devices Lett.,* vol. 4, pp. 71 – 74, 1984.

[50] F. –C. Hsu, and H. R. Grinolds, "Structure-enhanced MOSFET degradation due to hot-electron injection," *IEEE Electron Devices Lett.,* vol.5, pp. 71 – 74, 1984.

[51] T. Hori, H. Iwasaki, "Improved hot-carrier immunity in submicrometer MOSFETs with reoxidized nitrided oxides prepared by rapid thermal processing*," IEEE Electron Devices Lett.*, vol. 10, pp. 64 – 66, 1989.

[52] W. Shockley, "Problems related to p-n junctions in silicon," *Solid-State Electronics*, vol.2, pp.35-67, 1961.

[53] J. F. Verwey, *et al.*, " Mean free path of hot electrons at the surface of boron-doped silicon,*" J. Appl. Phys.,* vol. 46, pp. 2612 – 2619, 1975.

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

[55] C. Hu, "Lucky electron model of hot electron emission," *IEEE IEDM Tech. Dig.,* p. 22, 1979.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[56] P. K. Ko, R. Muller, and C. Hu, "A unified model for hot electron currents in MOSFET's," *IEEE IEDM Tech. Dig.,* p. 600, 1981.

[57] S. Tam, P. K. Ko and C. Hu, "Lucky-electron model of channel hot-electron injection in MOSFET's," *IEEE Trans. Electron Devices,* vol. ED-31, pp. 1116 - 1125, 1984.

[58] R. Bez., E. Camerlenghi, A. Modelli and A. Viscounti, "Introduction to Flash memory," *Proc. IEEE*, vol. 91, pp. 489 – 501, 2003.

[59] C. N. Berglund and R. J. Power, "Photoinjection into SiO2, electron scattering in the image force potential well," *J. Appl. Phys.,* vol. 42, pp. 573 – 579, 1971.

[60] D. R. Young, "Electron current injected into SiO2 from p-type Si depletion regions," *J. Appl. Phys.*, vol. 47, pp. 2098 – 2102, 1976.

[61] J. E. Chung, M.C. Jeng, J.E. Moon, P.K. KO, and C.M. Hu, "Low-Voltage Hot-Electron Currents and Degradation in Deep-Submicrometer MOSFET's," *IEEE Trans. Electron Devices*, vol. 37, No. 7, pp. 1651-1656, 1990.

[62] T. Mizuno, *et al*., "Hot-Hot-Carrier Effects in 0.1 mm Gate Length CMOS Devices," *IEDM Tech. Dig.*, pp. 305-308, 1992.

[63] E. Stewart, G. L. Rosa and F. J. Guarin, "Role of e-e scattering in the enhancement of channel hot carrier degradation of deep sub-micron NMOSFETs at high Vgs condition," *Int. Reliability Phy. Symp.,* pp. 399 – 405, 2001.

[64] A. Abramo, C. Fiegna and F. Venturi, "Hot carrier effects in short MOSFETs at low applied voltages," *IEDM Tech. Dig.*, p. 301, 1995.

[65] B. Eitan, D. Frohman-Bentchkowsky, and J. Shappir, "Impact ionization at very low voltages in silicon," *J. Appl. Phys.,* vol. 53, pp. 1244 – 1247, 1982.

[66] P. Su, K. Goto, T. Sugii and C. M. Hu, "A thermal activation view of low voltage impact ionization in MOSFETs," *IEEE Electron Device Lett.,* vol. 23, pp. 350 – 352, 2002.

[67] P. A. Childs and C.C. Leung, "New mechanism of hot carrier generation in very short channel MOSFETs," *Electronics Lett.,* vol. 31, pp. 139-141, 1995.

[68] P. A. Childs and C.C. Leung, "A one-dimensional solution of the Boltzmann transport equation including electron-electron interactions," *J. Appl. Phys.,* vol. 79, pp. 222- 227, 1996.

[69] M.Y. Chang, D.W. Dyke, C.C. Leung, and P.A. Childs, "High energy electron-electron interactions in silicon and their effect on hot carrier energy distributions," *J. Appl. Phys.,* vol. 82, pp. 2974-2979, 1997.

[70] A. Ghetti, J. Bude, and C.T. Liu, "Monte Carlo Simulation of Hot-Carrier Degradation in Scaled MOS Transistors for VLSl Technology," *IEDM Tech. Dig.*, pp. 893-896, 1998.

[71] E. Sangiorgi, "Historical perspective and recent developments of hot-carrier generation modeling for device analysis," Int. *Conf. Simulation of Semi. Processes and Devices Tech. Dig.,* pp. 5-8, 1997.

[72] J.J. Ellis-Monaghan, R.B. Hulfachor, K.W. Kim, and M.A. Littlejohn, "Ensemble Monte Carlo Study of Interface-State Generation in Low-Voltage MOS Devices," *IEEE Trans. Electron Devices*, vol. 43, pp. 1123-1132, 1996.

[73] A. Ghetti, L. Selmi, R. Bez and E. Sangiorgi, "Monte carlo simulation of low voltage hot carrier effects in non-volatile memory cells," *IEDM*, pp. 379 – 382, 1996.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[74] R. B. Hulfachor, *et al.,* "Spatial retardation of carrier heating in scaled 0.1-um n-MOSFET's using Monte carlo simulations," *IEEE Trans. Electron Devices*, vol. 43, pp. 661 - 663, 1996.

[75] K. Hess and C. T. Sah, "Hot carriers in silicon surface inversion layers," *J. Appl. Phys.*, vol. 45, pp.1254 –1257, 1974.

[76] E. Takeda, H. Kume, T. Toyabe, and S. Asai, "Submicrometer MOSFET structure for minimizing hot-carrier generation," *J. Solid-State Circuits,* vol. SC-17, pp. 241 – 248, 1982.

[77] J. D. Bude, "EEPROM/Flash sub 3.0 V drain-source bias hot carrier writing," *IEDM*, pp. 989 – 991, 1995.

[78] J. D. Bude, "Gate current by impact ionization feedback in sub-micron MOSFET technologies," *Proc. Symp. on VLSI Technology,* pp. 101 – 102, 1995.

[79] J. D. Bude, M. R. Pinto and P. K. Smith, "Monte carlo simulation of the CHISEL Flash memory cell," *IEEE Trans. Electron Devices*, vol. 47, pp. 1873 - 1881, 2000.

[80] D. Esseni, L. Selmi, A. Ghetti, and E. Sangiorgi, "Injection efficiency of CHISEL gate currents in short MOS devices: physical mechanisms, device implications, and sensitivity to technological parameters," *IEEE Trans. Electron Devices*, vol. 47, pp. 2194 - 2200, 2000.

[81] S. Mahapatra, S. Shukuri and J. Bude, "CHISEL Flash EEPROM – part I: performance and scaling," *IEEE Trans. Electron Devices*, vol. 49, pp. 1296 – 1301, 2002.

[82] F. Driussi, D. Esseni and L. Selmi, "Performance, degradation monitors, and reliability of the CHISEL injection regime," *IEEE Trans. Device and Materials Reliability,* vol. 4, pp. 327 - 334, 2004.

[83] S. Mahapatra, S. Shukuri and J. Bude, "CHISEL Flash EEPROM – part II: reliability," *IEEE Trans. Electron Devices*, vol. 49, pp. 1302 - 1307, 2002.

[84] D. Esseni and L. Selmi, "Experimental signature and physical mechanisms of substrate enhanced gate current in MOS devices," *IEDM*, pp. 579 – 582, 1998.

[85] D. Esseni and L. Selmi, "A better understanding of substrate enhanced gate current in VLSI MOSFET's and Flash cells – part I: phenomenological aspects," *IEEE Trans. Electron Devices*, vol. 46, pp. 369 - 375, 1999.

[86] D. Esseni and L. Selmi, "A better understanding of substrate enhanced gate current in VLSI MOSFET's and Flash Cells - Part II: Physical Analysis," *IEEE Trans. Electron Devices*, vol. 46, pp. 369 - 375, 1999.

[87] S. Tam, F. Hsu, C. M. Hu, R. S. Muller and P. K. Ko, "Hot-electron currents in very short channel MOSFET's," *IEEE Electron Device Lett.,* vol. EDL-4, pp. 249 – 251, 1983.

[88] D. W. Feldbaumer, D. K. Schroder, "MOSFET doping profiling," *IEEE Trans. Electron Devices*, vol. 38, pp. 135 - 140, 1991.

[89] D. S. Ang and C. H. Ling, "A unified model for the self-limiting hot-carrier degradation in LDD n-MOSFET's," *IEEE Trans. Electron Devices*, vol. 45, pp. 149-159, 1998.

[90] G. Groeseneken, H. E. Maes, "Basics and applications of charge pumping in submicron MOSFET's," *Proc. 21st International Conference on Microelectronics*, vol. 2, pp. 581 - 589, 1997.

[91] G. Groeseneken, H. E. Maes, N. Beltran and R. F. De Keersmaecker, "A reliable approach to charge-pumping measurements in MOS transistors," *IEEE Trans. Electron Devices*, vol. ED-31, pp. 42 – 53, 1984.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[92] S. Mahapatra, C. D. Parikh, V. Ramgopal Rao, C. R. Viswanathan and J. Vasi, "A comprehensive study of hot-carrier induced interface and oxide trap distributions in MOSFET's using a novel charge pumping technique," *IEEE Trans. Electron Devices,* vol. 47, pp. 171-177, 2000.

[93] Y. Taur, T. K. Ning, *Fundamentals of Modern VLSI Devices*, Cambridge Univerisity Press, 1998.

[94] D. W. Feldbaumer, D. K. Schroder, "MOSFET doping profiling," *IEEE Trans. Electron Devices*, vol. 38, pp. 135-140, 1991.

[95] J. Tanaka, T. Toyabe, S. Ihara, S. Kimura, H. Noda, K. Itoh, "Simulation of sub-0.1µm MOSFETs with completely suppressed short-channel effect," *IEEE Electron Device Lett.,* vol. 14, pp. 396-399, 1993.

[96] X. D. Chen, Q. C. Ouyang, G. Wang, S. K. Banerjee, "Improved hot-carrier and short-channel performance in vertical nMOSFETs with graded channel doping," *IEEE Trans. Electron Dev*, vol.49, pp. 1962-1968, 2002.

[97] K. Sakui, S. S. Wong and B. A. Wooley, "The effects of impact ionization on the operation of neighboring devices and circuits," *IEEE Trans. Electron Devices,* vol. 41, p. 1603, 1994.

[98] B. L. Anderson and R. L. Anderson, *Fundamentals of Semiconductor Devices,* McGraw-Hill Education (Asia), New York, 2005.

[99] N. R. Mohapatra, D. R. Nair, S. Mahapatra, V. R. Rao and S. Shukuri, "The impact of channel engineering on the performance reliability and scaling of CHISEL NOR Flash EEPROMs," *European Solid-State Device Research 2003 (ESSDERC'03), 33rd Conf.,* pp. 541 – 544, 2003.

[100] P. Heremans, R. Bellens, G. Groeseneken and H. Maes, "Consistent model for the hot-carrier degradation in n-channel and p-channel MOSFETs," *IEEE Trans. Electron Devices,* vol. 12, p. 2194, 1988.

[101]W. Wong and A. Icel, "A comprehensive methodology and model for the characterization of hot-carrier induced MOS device degradation (Invited)," *Proceedings of the 1995 First IEEE International Caracas Conference on Devices, Circuits and Systems*, pp. 183-187, 1995.

[102] B. Eitan and D. Frohman-Bentchkowsky, "Hot-electron injection into the oxide in n-channel MOS devices," *IEEE Trans. Electron Devices,* vol. 28, p. 328, 1981.

[103] C. Y. Chang, S. M. Sze, *ULSI Devices*. John Wiley & Sons Inc., 2000.

[104] G. Krieger, P. P. Cuevas and M. N. Misheloff, "The effect of impact ionization induced bipolar action on n-channel hot-electron degradation," *IEEE Electron Device Lett.,* vol. 9, p. 26, 1988.

[105] F. Driussi, D. Esseni, L. Selmi, F. Piazza, "Damage generation and location in n- and p-MOSFETs biased in the substrate-enhanced gate current regime," *IEEE Trans. Electron Devices*, vol. 49, pp. 787-794, 2002.

[106] D. S. Ang, H. Liao, C. H. Ling, "Nonlocal hot-electron injection as the mechanism for the predominant source-side gate-oxide degradation in CHE-stressed deep submicrometer n-MOSFETs," *IEEE Electron Device Lett.,* vol. 25, pp. 417-419, 2004.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[107] D. S. Ang and C. H. Ling, "A new assessment of the self-limiting hot-carrier degradation in LDD nMOSFET's by Charge Pumping Measurement," *IEEE Electron Device Lett.,* vol. 18, pp. 299-301, 1997.

[108] V. H. Chan and J. E. Chung, "Two-stage hot-carrier degradation and its impact on submicrometer LDD NMOSFET lifetime prediction," *IEEE Trans. Electron Devices*, vol. 42, p. 957, 1995.

[109] F. C. Hsu and H. R. Grinolds, "Structure-enhanced MOSFET degradation due to hot-electron injection," *IEEE Electron Device Lett.,* vol.EDL-5, p. 71, 1984.