國立臺灣大學電機資訊學院電子工程學研究所
博士論文

Graduate Institute of Electronics Engineering
College of Electrical Engineering and Computer Science
National Taiwan University
Doctoral Dissertation

具超厚薄氧化層之鋁/二氧化矽/p 型矽金氧半穿隧結構研究

Study on Al/SiO_2/Si(p) Metal-Oxide-Semiconductor Tunnel Structures with Ultra-High-Low Oxides

林冠文
Kuan-Wun Lin

指導教授：胡振國 博士
Advisor: Jenn-Gwo Hwu, Ph.D.

中華民國 111 年 6 月
June, 2022

doi:10.6342/NTU202304241

國立臺灣大學博士學位論文
口試委員會審定書

具超厚薄氧化層之鋁/二氧化矽/p 型矽金氧半穿隧結構研究

Study on Al/SiO_2/Si(p) Metal-Oxide-Semiconductor Tunnel Structures with Ultra-High-Low Oxides

本論文係林冠文君 (D07943010) 在國立臺灣大學電子工程學研究所完成之博士學位論文，於民國 111 年 6 月 27 日承下列考試委員審查通過及口試及格，特此證明

口試委員：

胡振國
(指導教授)

王永和

陳敏璋

連振炘

林浩雄

系主任、所長

doi:10.6342/NTU202304241

doi:10.6342/NTU202304241

Study on Al/SiO$_2$/Si(p) Metal-Oxide-Semiconductor Tunnel Structures with Ultra-High-Low Oxides

By

Kuan-Wun Lin

DISSERTATION

Submitted in partial fulfillment of the requirement
for the degree of Doctor of Philosophy in Electronics Engineering
at National Taiwan University
Taipei, Taiwan, R.O.C.

June 2022

Approved by: Jenn-Gwo Hwu

Chenhsin Lien

Yeong-Her Wang   Miin-Jang Chen   Hao Hsiung Lin

Advised by: Jenn-Gwo Hwu

Approved by Director:

doi:10.6342/NTU202304241

誌謝

我能順利完成博士論文，首先必須感謝我的指導教授胡振國老師。從大三進入 C–V Lab 做專題以來，和老師的每次個人 meeting 總是醍醐灌頂。我仰慕老師深厚的知識底蘊與煥發的正能量，更感念老師在研究方向上給我極大的自由發揮空間。我也十分欽佩老師對於教職的耐心、熱忱與執著。特別感謝老師鼓勵並支持我在畢業前出國訪問，讓原本有點畏縮的我，毅然下了人生中最正確的決定。感謝六位口試委員曾俊元教授、王永和教授、蘇彬教授、陳敏璋教授、連振炘教授、林浩雄教授，在疫情籠罩下排除萬難，共同出席令人永生難忘的線上口試，也針對我的研究內容給出許多寶貴的指教和建議，讓我的博士論文更臻完善。

謝謝大學摯友陳威成，你的熱情感染了我一同跨入半導體領域。我們還一起在宿舍開過讀書會，這是一切旅程的起點。你一定能很快取得史丹佛的博士學位。由衷感謝辛苦的實驗室學長楊昶豐老大，總是親自帶我們設計實驗、操作製程、跑模擬、分析數據，並把半導體知識和行政工作傳承給我，讓我從懵懂的小專題生蛻變成獨當一面的研究戰力和掌櫃；我將繼續傳承你的奉獻精神。感謝帶過我製程的楊明翰學長、詹前修學長、黎鎬均學長、張博凱學長，每次與你們的知識交流和閒話家常，都讓我的大學生活更加繽紛。感謝江育誠學長、鄭捷方學姐、劉宇恒學長、陳昱諠學姐、以及同屆的夥伴陳柏均、黑暗料理大師江子豪、許庭昊、楊詠竣，你們的上進心是推動我持續研究不放棄的引擎，你們對實驗室的小心維護是我研究順利的功臣，與你們在 118 的每次飯局都是我的靈感來源和精神食糧，使我的博班生涯沒有一刻感到失落與徬徨。感謝簽博的學弟陳冠竹，你是和我共事最久的，替我分擔不少行政工作，一起處

理麻煩狀況，也最常陪我討論研究和聊天，讓我收穫滿滿。感謝學弟妹林建宇、黃琛云、陳人豪把 C–V Lab 的研究氣氛推高到極致。感謝學弟黃崧瑋總嚴謹地和我討論機制，常能指出我論述中的盲點，同時熱衷召集各項活動幫大家實踐 work-life balance。感謝學弟妹林津丞、林俊諭、林彥瑜、高綺憶、陳舜啓、林郁芹、沈祐德、林軒毅、廖威騏、林俊億、龔泰銘、羅雅云、王瑞賢、與後繼的學弟妹們，讓實驗室在疫情的低氣壓下依然生氣蓬勃，尤其和我同梯進台積電實習，下班總一起吃飯、出遊、逛百貨公司、探險的威騏、軒毅、祐德、郁芹、舜啓。祝福 C–V Lab 的每一份子，學業、事業都一帆風順。

感謝國科會補助我到 UCSD 訪問八個月，期間我有幸很快步入研究正軌，更不忘用心體驗異國生活，深入學習在地民情，這對我的人生觀有巨大而正向的衝擊。謝謝治學嚴謹卻像朋友般親切的 Yuan Taur 教授首肯我的到訪，從您身上我學到紮實的數學方法和物理基礎，也知曉美國學界、業界、生活上的酸甜苦辣。尤其感謝您主動提議彈性的研究進度，並提醒我要多出門走走別只顧著做研究，讓我得以騰出許多時間深度探索這個美麗的濱海城市。謝謝教授的末代研究生蘇美樺學姐帶我認識環境，也常和我分享美國業界近況或陪我逛超市。謝謝在機場被我搭訕而相識，最後讓我搭便車來 UCSD 並帶我初識校園的 Ryan Wang，你是我在美國的第一個朋友，更是我的恩人。謝謝不同科系的千里馬同梯 — 時常牽線飯局、酒局和旅遊團的賴怡如，和總是開車載我們出遊、還陪我去墨西哥吃 taco 的張凱捷。兩位將比我待得更久，所以還請把握時間享受南加州的沙灘和暖陽。

感謝所有摯友的陪伴和鼓勵。研究之餘與你們同樂，讓本該單調的研究生活增添不少色彩。

最後，由衷感謝父母一路以來的支持。我們偶有觀點分歧的時候，但你們總願意聆聽並尊重我的想法。你們時常提點投資理財、身體健康、心靈富足的重要，我都銘記在心並予以實踐。沒有你們的默默支持，我將無法完成這趟博士之旅！

林冠文

2024.01.05 美國加州聖地牙哥

doi:10.6342/NTU202304241

摘要

本論文詳細研究具超厚薄氧化層之鋁/二氧化矽/p 型矽金氧半穿隧元件 (或稱超厚薄元件，即氧化層在特定區域爲超薄，其餘爲超厚) 的電特性、靜電學特性和其中的載子傳輸機制。首先，我們對眞實元件進行量測與定性研究，並以 TCAD 模擬輔助觀察。吾人發現元件施加反偏壓時，儘管載子穿隧只在薄氧區發生，穿隧電流卻正比於閘極面積，而非薄氧區面積。換言之，金氧半電容氧化層中的微小局部薄化區域 (只佔閘極面積的百萬分之幾) 便可導致巨大的漏流，顯示了氧化層品質控制的重要。我們將此現象歸因於一種新發現的、反轉電荷的水平耦合機制，能將厚氧區中的電子耗盡，從而影響其靜電學特性，影響範圍深遠。爲定量探討這些發現，吾人首次提出一個適用於較簡單平面結構 — 金氧半穿隧二極體 — 的靜電學解析模型。本模型顯示了此平面結構的靜電學特性在某個臨界偏壓前後截然不同，而吾人也推導出此臨界偏壓對氧化層厚度的閉形近似式。模型與 TCAD 模擬結果高度吻合 (誤差小於 2Å)，並成功重現金氧半穿隧二極體的實驗特性。隨後，本模型被推廣至超厚薄元件中，尤其著重厚氧區中電子準費米能階的建模，而成功預測前述穿隧電流正比於閘極面積的現象，以及前述耦合機制的作用距離超過毫米等級。這些模型有助我們從物理和直觀的面向加深對金氧半穿隧二極體與超厚薄元件的瞭解。最後，吾人探討以超厚薄元件作爲溫度與環境光感測器之應用。在低偏壓 (不高於 0.3 伏特) 之下，與平面的金氧半穿隧二極體相比，超厚薄元件的溫度響應度與光電流均提升超過百倍，顯示其具有作爲低電壓下感測器應用的潛力。

關鍵字： 解析模型；深空乏；靜電學；反轉電荷；金氧半穿隧二極體；金氧半電容；氧化層軟崩潰；氧化層穿隧；感測器。

Abstract

The electrical characteristics, electrostatics, and carrier transport mechanisms in Al/SiO_2/Si(p) metal-oxide-semiconductor (MOS) tunnel structures with ultra-high-low oxides (the "ultra-high-low devices", where the oxide layer is ultrathin at specific locations and ultrathick otherwise) have been comprehensively studied in this dissertation. First, a qualitative study has been conducted on characterizing experimental devices with an aid of TCAD simulation. Under reverse bias, device tunnel currents were found proportional to the gate area rather than the low-region area, even though carrier tunneling only occurs at the low region. In other words, a tiny local oxide thinning spot (accounting only millionths of the gate area) in an MOS capacitor can lead to a giant gate leakage, affirming the importance of oxide quality control. This phenomenon was ascribed to a newly-discovered lateral electron coupling mechanism, which is capable of depleting electrons in the high region and affecting the electrostatics thereof over a wide distance. To study these findings quantitatively, an analytical electrostatics model for the degenerate planar structures, metal-oxide-semiconductor tunnel diodes (MIS TDs), under reverse bias was established the first time. The model reveals a critical gate voltage that demarcates the electrostatics into two dissimilar regimes. A closed-form approximation for the critical voltage-oxide thickness relation was then derived. The model highly agrees with TCAD simulation results (with $< 2\,\text{Å}$ discrepancy) and man-

doi:10.6342/NTU202304241

ages to reproduce the experimental MIS TD characteristics. Subsequently, the model was generalized into ultra-high-low devices with an emphasis on modeling the electron quasi Fermi level in the high region. It successfully predicts the aforementioned gate area proportionality of the tunnel current, as well as a lateral coupling distance beyond the millimeter scale. The models may help us understand MIS TDs and ultra-high-low structures in a more physical and intuitive aspect. Finally, temperature and ambient light sensor applications for the ultra-high-low devices were investigated. Compared to a planar MIS TD, the thermal responsivity and photocurrent in an ultra-high-low device can be increased $>$ 100-fold under low applied bias (no greater than $0.3\,\mathrm{V}$), making them a competitive candidate for low-voltage sensor applications.

Keywords: Analytical model; Deep depletion; Electrostatics; Inversion charge; MIS tunnel diode; MOS capacitor; Oxide soft breakdown; Oxide tunneling; Sensors.

Epitoma

Proprietates electricae electrostaticaeque ac mechanismi fluminum latorum oneris electrici in structuris tunnelativis metalli-oxidi-semiconductri (Al/SiO_2/Si(p)) quae et regiones laminarum oxidi percrassi et pertenuis possident (de quibus praedicatur quod “apparatus” vel “apparatus percrassi-pertenues”) sunt funditus attacta in hac dissertatione. Imprimis, studium qualitativum characterificando apparatuum experimentalium adducebatur per auxilium simulatoris (TCAD). Penes tensionem (electricam) reversam, in apparatibus compertum est quod fluxiones (electricae) tunnelantes sunt proportionales areis electrodorum, sed non eis regionum tenuium, tametsi modo hae regiones a latoribus oneris tunnelari possunt. Id est, parvula macula localis de oxido tenuato (quae $\sim 10^{-6} \times$ tam magna est quam area electrodi) in MOS capacitore praegrandem fluxionem fugientem adducere quit, quod importantiam dispensationis qualitatis oxidi affirmat. Hoc phaenomenon mechanismo copulationis electronum lateralis attributum est, qui primum inventus erat et electrones ex regione crassa deplere atque electrostaticam ea longam distantiam afficere potest. Ut his repertis quantitative studeremus, primus modulus analyticalis structurae planae degeneratae, “diodo” tunnelativo metalli-insulatri-semiconductri (MIS TD) penes tensionem reversam statutus est. Is electrodo tensionem criticam revelat, quae electrostaticam diodorum in duo regimina dissimilaria dividit. Expressio tensioni criticae per oxidi crassitudinem, approximata at algebraica,

doi:10.6342/NTU202304241

deinde derivata est. Modulus proventibus simulationum magnopere congruit (discrepantia $< 2\,\text{Å}$) ac proprietates diodorum experimentalium recreare pervenit. Postea, hic modulus generalificatus est pro apparatu percrassopertenui, ubi quasienergia electronum Fermii modulanda est in regione crassa. Is superdictam proportionalitem (inter fluxiones et areas electrodorum) prospere praeloquitur etiamque distantiam copulationis lateralis supra millimetrum ominatur. Hi moduli diodos tunnelativos apparatusque in aspectu physicaliore et intuitiviore intellegi adiuvent. Tandem, applicationes apparatuum ut sensores temperaturae et lucis ambientis investigabantur. Parva tensione ($\leq 0.3\,\mathrm{V}$) applicata, responsivitas thermalis et fluxio luce in apparatu $> 100\,\times$ augescere queunt prae diodo plana, quod pro applicationibus super parva tensione apparatum candidatum sensoris faciat.

Descriptores: Modulus analyticalis; Depletio profunda; Electrostatica; Onus electricum inversionis; Diodus tunnelativus metalli-insulatri-semiconductri; Capacitor metalli-oxidi-semiconductri; Ruptura mitis oxidi; Tunnelatio trans oxidum; Sensores.

Table of Contents

doi:10.6342/NTU202304241

doi:10.6342/NTU202304241

List of Figures

doi:10.6342/NTU202304241

doi:10.6342/NTU202304241

EXCURSUS

doi:10.6342/NTU202304241

doi:10.6342/NTU202304241

doi:10.6342/NTU202304241

List of Tables

EXCURSUS

doi:10.6342/NTU202304241

Nomenclature

Acronyms

ANO	Anodic oxidation
BOE	Buffered oxide etchant
CBE	Conduction band edge
CEST	Classical MOS electrostatics theory
CI	Classical inversion
DD	Deep depletion
DDE	Deep depletion effect
DT	Direct tunneling
FFE	Fringing field effect
FNT	Fowler-Nordheim tunneling
GOI	Gate oxide integrity
LOT	Local oxide thinning
MIS	Metal-insulator-semiconductor
MOS	Metal-oxide-semiconductor
PR	Photoresist
QC	Quantum confinement
QFL	Quasi Fermi level
RTP	Rapid thermal process
S/D	Source/drain
SBH	Schottky barrier height
SBHME	Schottky barrier height modulation effect
SP	The surface potential-based compact model

TD	Tunnel diode
TEM	Transmission electron microscopy
UHL	Ultra-high-low

List of Fundamental Physical Constants

$\hbar$	Reduced Planck constant
k_B	Boltzmann constant
m_0	Electron mass
q	Elementary charge
ϵ_0	Vacuum permittivity

List of Symbols

A^*	Effective Richardson constant for holes
A_G	Gate area
A_H	High-region area
A_L	Low-region area
a	Gate side length of square-shaped experimental devices
B	A field constant in electron tunneling probability equations (See eq. 3.6)
b	Low-region side length of square-shaped experimental devices
C_Acc	Accumulation capacitance
C_acc	Accumulation capacitance per unit area
C_G	Gate capacitance
C_HF	High-frequency gate capacitance
C_LF	Low-frequency gate capacitance
C_ox	Oxide capacitance per unit area
C_oxH	Oxide capacitance per unit area in the high region

doi:10.6342/NTU202304241

C_{oxL}	Oxide capacitance per unit area in the low region
D_{it}	Interface trap density
D_n	Diffusion coefficient for electrons
$\mathscr{D}_{\mathrm{ox}}$	Characteristically-normalized oxide thickness (See eq. 3.38)
d_{ox}	Oxide thickness
d_{H}	High oxide thickness
d_{L}	Low oxide thickness
d_{L}^{*}	Characteristic low oxide thickness
d_{ox}^{*}	Characteristic oxide thickness
E	Energy
E_{a}	Activation energy
E_{C}	Conduction band energy
E_{Fm}	Fermi level in metal
$E_{\mathrm{F}n}$	Electron quasi Fermi level in semiconductor
$E_{\mathrm{F}p}$	Hole quasi Fermi level in semiconductor
E_{V}	Valence band energy
E_{g}	Semiconductor band gap
$\mathscr{E}$, $\mathbf{E}$	Electric field
$\mathscr{E}_{\mathrm{ox}}$	Oxide electric field
$\mathscr{E}_r$	Radial electric field
e_{V}	Illuminance
f	Frequency
G	Net generation rate
$\mathscr{G}$	Conductance
I_{G}	Gate current
I_n	Electron current
I_{gn}	Electron generation current
$I_{\mathrm{t}n}$	Electron tunnel current
I_{ph}	Photocurrent
I_{sat}	Reverse saturation current of MOS tunnel structures
J_n, $\mathbf{J}_n$	Electron current density

doi:10.6342/NTU202304241

$J_{\mathrm{d}n}$	Electron diffusion current density
$J_{\mathrm{g}n}$	Electron generation current density
$J_{\mathrm{t}n}$	Electron tunnel current density
$J_{\mathrm{t}p}$	Hole tunnel current density
L_n	Electron diffusion length
m^*	Longitudinal effective mass for electrons in silicon
m_{ox}	Oxide tunneling effective mass for electrons across SiO_2
N_{A}	Doping concentration of acceptors
n	Electron concentration
n_{p0}	Electron concentration in bulk p-type semiconductor
n_{S}	Surface electron concentration
n_{i}	Intrinsic carrier concentration
P_{t}	Electron tunneling probability
P_{tC}	Critical electron tunneling probability
$P_{\mathrm{t}p}$	Hole tunneling probability
p	Hole concentration
p_{p0}	Hole concentration in bulk p-type semiconductor
p_{S}	Surface hole concentration
Q_{d}	Depletion charge per unit area
Q_{dH}	Depletion charge per unit area in the high region
Q_{eff}	Oxide effective charge per unit area
Q_{I}	Inversion charge
Q_{IH}	Inversion charge in the high region
Q_{i}	Inversion charge per unit area
Q_{iH}	Inversion charge per unit area in the high region
Q_{S}	Surface charge per unit area
Q_{SH}	Surface charge per unit area in the high region
Q_{SL}	Surface charge per unit area in the low region

doi:10.6342/NTU202304241

r	Radius
r_{G}	Gate radius
r_{L}	Low-region radius
S	Supply function
$\hat{S}$	(See eq. 3.29)
T	Temperature
t	Time
u	An auxiliary quantity related to $\Delta\phi_n$ (See eq. A.11)
V_{C}	Critical voltage
V_{FB}	Flat-band voltage
V_{G}	Gate voltage
V_{ox}	Oxide voltage drop
W	Depletion width
W_0	Depletion width at zero applied bias (See eq. 3.28)
W_{B}	Depletion width at the high-low boundary
W_{H}	Depletion width in high region
W_{L}	Depletion width in low region
W_{g0}	Generation-depletion width offset
y	Depth across the structure, starting from the semiconductor surface
α	An auxiliary material constant (See eq. A.19)
$\hat{\alpha}$	(See eq. 3.26)
β	An auxiliary quantity (See eq. A.25)
β_{H}	β in high region
γ	Body effect coefficient
γ_{H}	γ in high region
$\Delta\phi_p$	Quasi Fermi level splitting for holes, with respect to the bulk value
$\Delta\phi_n$	Quasi Fermi level splitting for electrons, with respect to the bulk value

$\Delta\phi_{n\mathrm{L}}$	$\Delta\phi_n$ value in the low region
$\Delta\phi_{n\mathrm{S}}$	$\Delta\phi_n$ value at semiconductor surface
ϵ_{ox}	Permittivity of SiO_2 ($= 3.9\,\epsilon_0$)
ϵ_{s}	Permittivity of Si ($= 11.9\,\epsilon_0$)
η_{C}	Coupling efficiency
κ	An auxiliary material constant (See eq. A.9)
Λ	Coupling length
λ_{p}	Extrinsic Debye length in p-type semiconductors
μ_n	Electron mobility
ρ	Normalized radius, by Λ
ρ_{G}	Normalized gate radius ($= r_{\mathrm{G}}/\Lambda$)
ρ_{L}	Normalized low-region radius ($= r_{\mathrm{L}}/\Lambda$)
σ_κ	An auxiliary quantity related to κ (See eq. A.48)
τ_0	$\equiv \frac{1}{2}(\tau_{n0} + \tau_{p0})$
τ_{n0}	Excess electron lifetime
τ_{p0}	Excess hole lifetime
ϕ_{Bp}, ϕ^*_{Bp}	A hole Schottky barrier. (See eq. 1.2)
Φ_{b}	Conduction band discontinuity at the oxide-semiconductor interface
ϕ_{CF0}	Potential difference from semiconductor bulk Fermi level to semiconductor bulk conduction band energy
ϕ_{F}	Semiconductor bulk potential
ϕ_{m}	Metal work function
ϕ_{t}	Thermal voltage. ($\equiv k_{\mathrm{B}}T/q$)
χ_{ox}	Oxide electron affinity
χ_{s}	Semiconductor electron affinity
ψ	Band bending
ψ_{S}	Surface band bending
ψ_{S0}	Surface band bending at zero applied bias
ψ_{SH}	Surface band bending in high region
$\psi^{\wedge}_{\mathrm{SH}}$	Maximum surface band bending in high region (see eq. 4.5)

ψ_{SL}	Surface band bending in low region

List of Special Functions

erf	Error function
erfi	Imaginary error function
$\mathscr{F}_j$	Complete Fermi-Dirac integral, for index j
I_ν	Modified Bessel function of the first kind, of order ν
$\mathscr{I}_\kappa$	An auxiliary function (See eq. A.46)
K_ν	Modified Bessel function of the second kind, of order ν
Li_s	Polylogarithm, of order s
$\Gamma_{\alpha,\kappa}$	An auxiliary function (See eq. A.28b)
$\Phi_{\alpha,\kappa}$	An auxiliary function (See eq. A.28a)

1

Introduction

1.1 Motivation

Gate oxide integrity (GOI) has been one of the leading issues arising from the very first metal-insulator-semiconductor field-effect transistors (MOSFETs) [1, 2] to state-of-the-art ultra-large-scale integrated (ULSI) circuits [3–5]. Defects in gate oxide and oxide wear out give rise to local conductive paths across the oxide, rendering gate leakage. This can pose negative impacts on multiple performance factors such as noise margin, power consumption, and speed in CMOS logic circuits, even leading to logic failure [6]. In nonvolatile memory devices (e.g., Flash), degraded GOI is responsible for retention loss [5]. It is therefore of great importance to elucidate

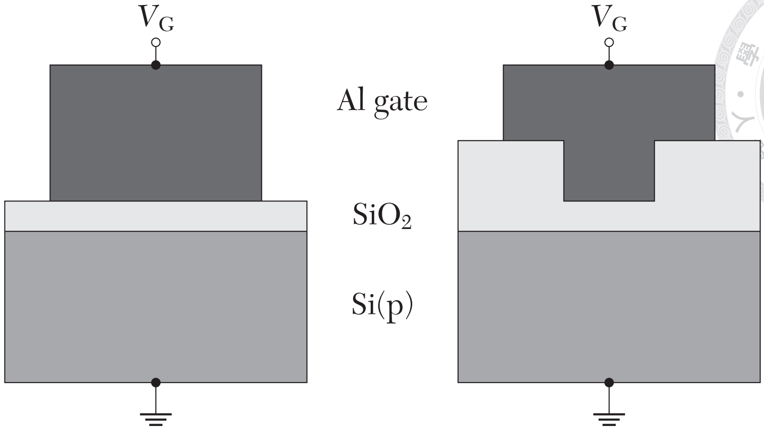
doi:10.6342/NTU202304241

the influence of local conductive spots in the gate oxide on the electrical characteristics of metal-oxide-semiconductor (MOS) capacitors.

Non-catastrophic oxide wear out, such as oxide soft breakdown [7, 8], can be modeled as introduction of local oxide thinning (LOT) spots into the oxide layer [9, 10], while physical LOT spots, either inadvertently introduced during the fabrication process, or emerged post-stressing, have been discovered under the microscope and reported [8, 11–14]. The scope of discussion may not be limited to SiO_2 insulating layers. For high-k dielectric stacks, bulk defects in the high-k films can form local conductive paths [15], where the LOT model may also apply. Unfortunately, besides the increase in gate leakage current, the spots' influence in other aspects, especially on the device electrostatics, is yet to be addressed. As such, characterizing and modeling MOS capacitors with artificial (well-controlled) LOT spots may help expound the influence.

On the other hand, metal-insulator-semiconductor (MIS) tunnel diodes (TDs), namely MOS capacitors with an ultrathin, extremely leaky oxide layer (**Figure 1–1 (a)**), manipulates the leakage-induced electrostatics change and finds multiple sensor applications with highly responsive current readouts (**Section 1.2**). While planar MIS TDs as shown in **Figure 1–1 (a)** have been relatively well-studied and theorized, with the company of thick oxide under the gate (*abbrv.* high-low structures, **Figure 1–1 (b)**), the electrical characteristics and carrier transport mechanisms remain elusive. Research on the high-low structures is not of mere interest, but also of practical concerns. The high-low structure can be inadvertently introduced into expectedly planar MIS TDs and related devices; e.g., when scaling down a charge-coupled MIS tunnel transistor [16, 17] and extending the contact pads on the isolation oxide for the ease of probing

doi:10.6342/NTU202304241



(a) Planar **(b) High-low**

Figure 1–1. Schematic cross sections of **(a)** a planar Al/SiO_2/Si(p) MIS tunnel diode, and **(b)** a high-low Al/SiO_2/Si(p) MOS tunnel structure.

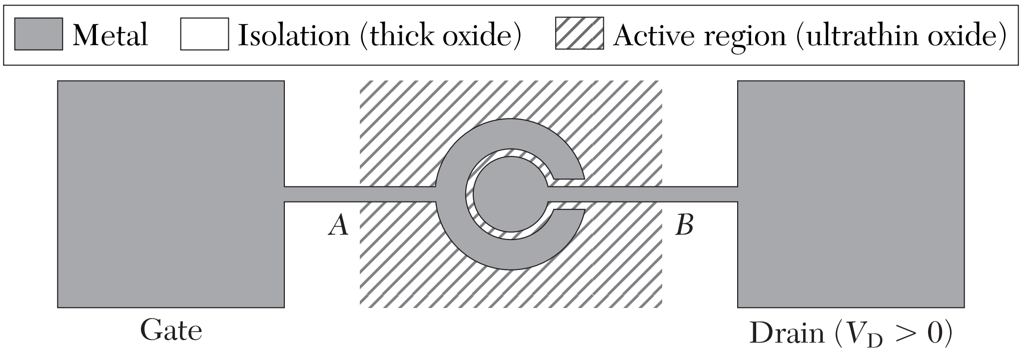


Figure 1–2. Layout of a scaled charge-coupled MIS tunnel transistor with extended pads on the isolation oxide, for the ease of probing. High-low structures were inadvertently introduced at *A* and *B*.

(**Figure 1–2**). With limited knowledge about the high-low structure, it is questionable whether the pads have truly negligible effects on the device characteristics.

Remarkably, MOS capacitors with LOT spots are essentially high-low structures. Hence, in this work, we are aiming to study the high-low structures by electrical characterizations, TCAD simulations, electrostatics theory development, and modeling. For simplicity, every device is associated with only one LOT spot, which is photolithographi-

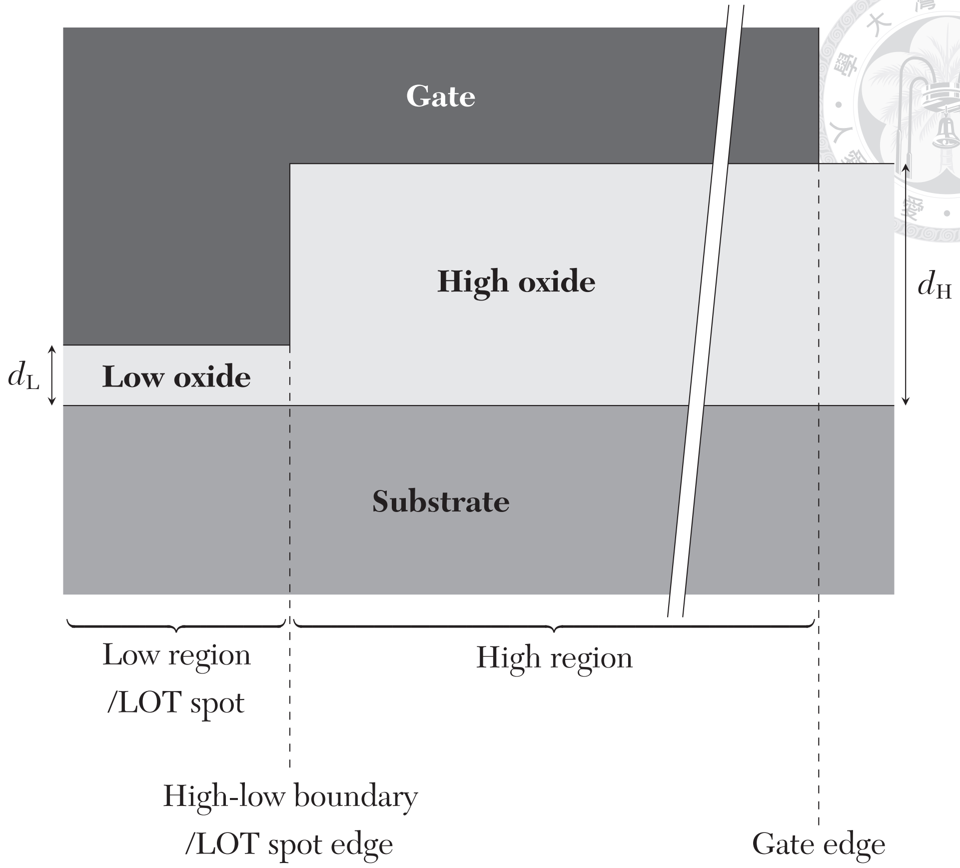


Figure 1–3. Schematic cross section of a high-low device, and terminologies for the structure.

cally defined at the device center. Terminologies for the device structure are defined in **Figure 1–3** . Specifically speaking, the combination of ultrathick high oxide ($\geq 100\,\text{Å}$) and ultrathin low oxide ($\leq 30\,\text{Å}$), composing the *ultra-high-low device*, is the scope of this work. The ultrathick high oxide forbids carrier tunneling, replicating normal gate oxide outside the spot in MOS capacitors, while the ultrathin oxide facilitates substantial tunneling, mimicking the spot.

Moreover, the electrostatics theory for planar MIS TDs will be revisited in this work, and an analytical electrostatics model for planar MIS TDs will be proposed and generalized to that for high-low MOS tunnel structures. By the end of this work, we will also seek possible applications for the high-low devices.

1.2 Current-Voltage Characteristics and Electrostatics for Planar Al/SiO$_2$/Si(p) MIS Tunnel Diodes

Planar MIS TDs (**Figure 1–1 (a)**) are essentially MOS capacitors with ultrathin oxide and, as a result, significant gate current. With structural similarities to Schottky diodes [18, 19], MIS TDs are found to exhibit diode-like current-voltage characteristics [20]. Furthermore, with the current being very sensitive to ambient conditions, MIS TDs find their place in sensor applications, including but not limited to temperature sensors [21], strain sensors [22], and photodetectors [23].

Without otherwise specified, the material system is Al/SiO$_2$/Si(p) with a doping concentration $N_\mathrm{A} = 10^{16}\,\mathrm{cm}^{-3}$ for the single crystal Si substrate. Assuming negligible oxide charge (Q_eff) and interface trap density (D_it), the gate voltage can be expressed as

$$\begin{aligned} V_\mathrm{G} &= V_\mathrm{FB} + V_\mathrm{ox} + \psi_\mathrm{S} \\ &= V_\mathrm{FB} - \frac{Q_\mathrm{S}}{C_\mathrm{ox}} + \psi_\mathrm{S}, \end{aligned} \tag{1.1}$$

where $V_\mathrm{FB} = \phi_\mathrm{ms} = \phi_\mathrm{m} - (\chi_\mathrm{S} + E_\mathrm{g}/2q + \phi_\mathrm{F}) = -0.9\,\mathrm{V}$ is the flat-band voltage for the material system (ϕ_ms: metal-semiconductor work function difference, ϕ_m: metal work function, χ_S: semiconductor electron affinity, E_g: semiconductor band gap, q: elementary charge, $\phi_\mathrm{F} = \phi_\mathrm{t}\ln(N_\mathrm{A}/n_\mathrm{i})$: semiconductor bulk potential, $\phi_\mathrm{t} = k_\mathrm{B}T/q$: thermal voltage, k_B: Boltzmann constant, T: temperature, N_A: acceptor doping concentration in the semiconductor, n_i: intrinsic carrier concentration for the semiconductor), V_ox is the oxide voltage drop, ψ_S is the semiconductor band bending, $C_\mathrm{ox} = \epsilon_\mathrm{ox}/d_\mathrm{ox}$ is the oxide

doi:10.6342/NTU202304241

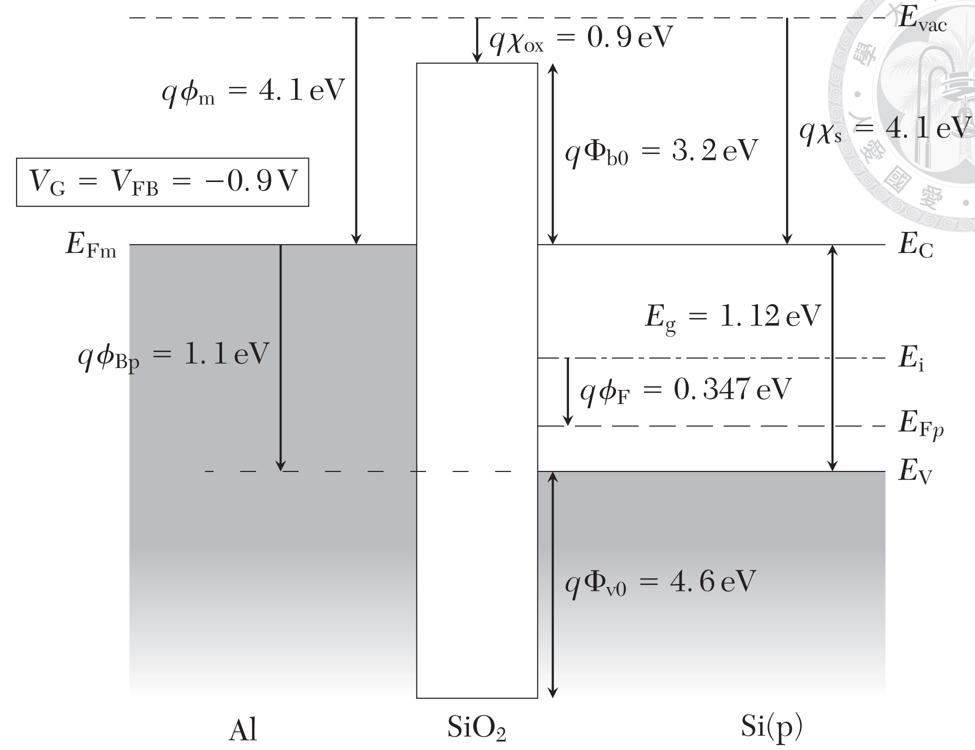
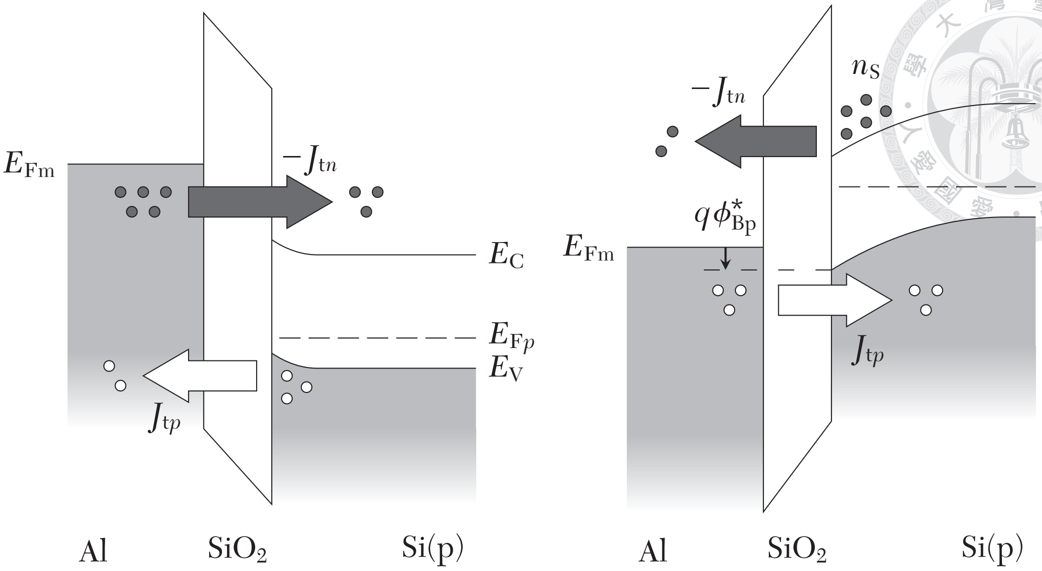


Figure 1–4. Band diagram (not to scale) for an Al/$\mathrm{SiO_2}$/Si(p) MOS capacitor with doping concentration $N_{\mathrm{A}} = 10^{16}\,\mathrm{cm}^{-3}$, biased under flat-band condition $V_{\mathrm{G}} = V_{\mathrm{FB}} = -0.9\,\mathrm{V}$. ($\phi_{\mathrm{m}}$: metal work function, χ_{ox} and χ_{S}: oxide and semiconductor electron affinities, E_{g}: semiconductor band gap, Φ_{b0} and Φ_{v0}: conduction and valence band discontinuities, ϕ_{F}: semiconductor bulk potential, q: elementary charge.)

capacitance (ϵ_{ox}: oxide permittivity, d_{ox}: oxide thickness), and Q_{S} is the semiconductor surface charge. This equation describes how the gate voltage is distributed across the oxide (as V_{ox}) and semiconductor (as ψ_{S}).

Figure 1–4 depicts the band diagram for such MOS capacitor (and also for MIS TDs) biased under the flat-band condition $V_{\mathrm{G}} = V_{\mathrm{FB}}$. The Fermi level in metal (E_{Fm}) roughly aligns with the conduction band of Si, and there is a hole Schottky barrier ($q\phi_{\mathrm{Bp}}$) seen from the metal side. **Figure 1–5 (a, b)** depict the band diagrams where $V_{\mathrm{G}} < V_{\mathrm{FB}}$ and $V_{\mathrm{G}} > 0 > V_{\mathrm{FB}}$, respectively. These bias conditions are referred to as the accumulation and inversion regimes in MOS electrostatics theories, respectively [24].



(a) Accumulation ($V_{\mathrm{G}} < V_{\mathrm{FB}}$) **(b)** Inversion, $V_{\mathrm{G}} > 0$

Figure 1–5. Band diagrams and tunnel current components of an MIS(p) tunnel diode biased under **(a)** accumulation regime ($V_{\mathrm{G}} < V_{\mathrm{FB}}$), and **(b)** inversion regime with $V_{\mathrm{G}} > 0$.

Tunneling current of both carrier types are present under either bias condition. Under accumulation, accumulated holes tunnel from the insulator/semiconductor (I/S) surface towards the metal, and electrons tunnel from the metal towards the Si conduction band. Contrarily, under inversion, it is the inversion charge (electrons) that tunnel towards the metal, and holes that tunnel towards the Si valence band. As shown in **Figure 1–5 (b)**, such tunneling holes under positive gate bias will encounter a Schottky barrier ($q\phi^{*}_{\mathrm{Bp}}$) at the I/S surface, given geometrically by

$$\begin{aligned}\phi^{*}_{\mathrm{Bp}} &= \left(\chi_{\mathrm{s}} - \phi_{\mathrm{m}} + \frac{E_{\mathrm{g}}}{q}\right) - V_{\mathrm{ox}} \\ &= \phi_{\mathrm{Bp}} - V_{\mathrm{ox}}. \end{aligned} \tag{1.2}$$

Figure 1–6 presents the $I - V$ characteristics for a collection of MIS TDs with dif-

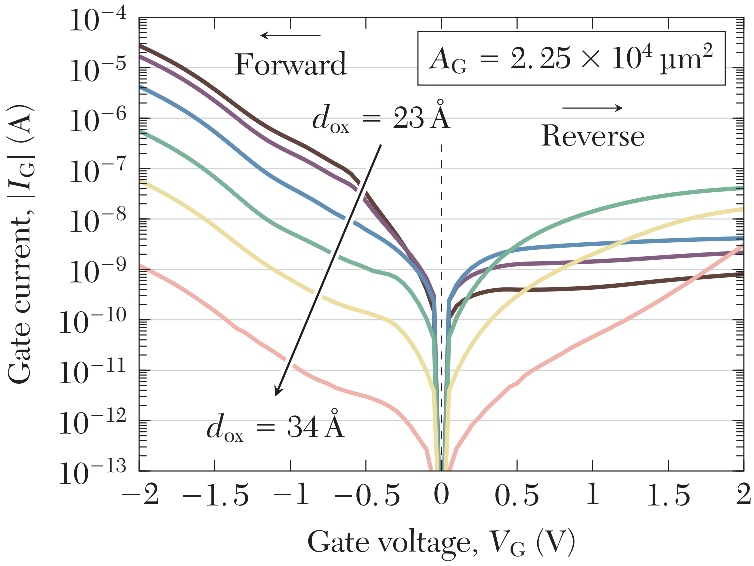


Figure 1–6. Current-voltage characteristics for a collection of Al/SiO$_2$/Si(p) MIS tunnel diodes with oxide thicknesses ranging from 23 Å to 34 Å.

ferent oxide thicknesses, ranging from 23 Å to 34 Å. The diode-like nature of the $I - V$ curves is apparent. Under forward bias ($V_G < 0$, inherited from what we call "forward bias" in Al/Si(p) Schottky diodes), the tunnel current decreases with increasing d_{ox}. This can be attributed to curtailed tunnel probabilities for both carrier types. Under reverse bias ($V_G > 0$), the current saturates like Schottky diodes [20, 25]. Nonetheless, the saturation current anomalously increases with d_{ox} [17]. The increase of "the voltage after which current saturation takes place" (defined as the critical voltage V_C) with increasing d_{ox} [26] is also remarkable. Devices with the thickest oxides in this collection demonstrate off-scale critical voltages, and what remains in the plot is only their pre-saturation behavior. The anomaly only seems to occur post-saturation.

This phenomenon is known as the hole effective Schottky barrier height modulation effect (SBHME) in MIS TDs [27–29]. We may consider two MIS TDs with slightly different d_{ox} but biased under identical V_G, with the band diagrams depicted

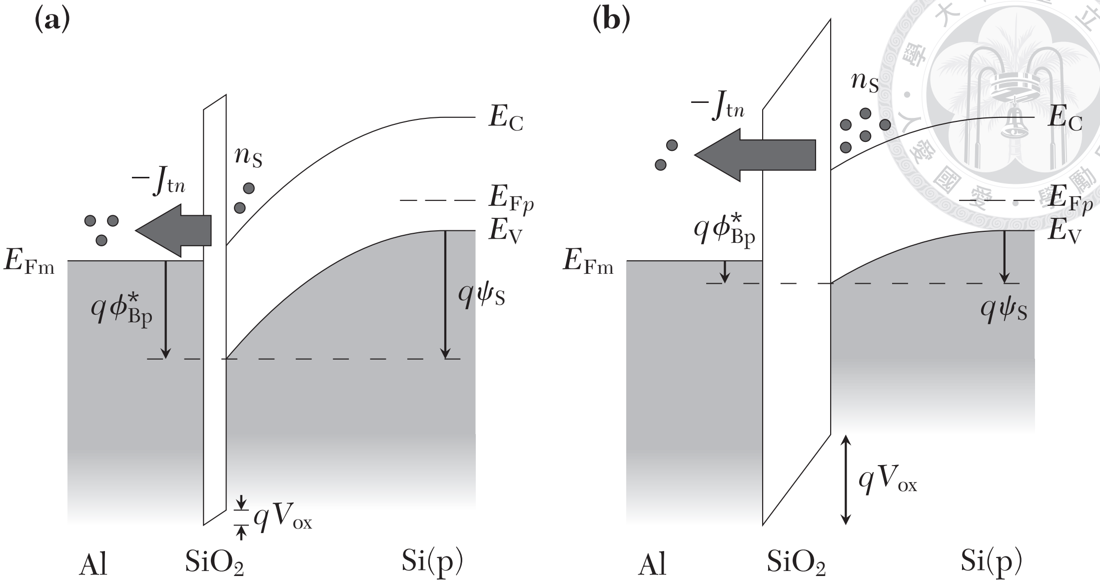


Figure 1–7. Band diagrams of MIS(p) tunnel diodes biased under identical positive V_G. **(a)** With ultrathin oxide, surface electron concentration (n_S) is low, oxide voltage drop is low, semiconductor band bending is high, and the hole Schottky barrier ($q\phi^*_{Bp}$) is high. **(b)** with slightly thicker oxide, conversely.

in **Figure 1–7**. The device with thicker d_{ox} is associated with higher V_{ox} and lower ψ_S, partly due to lower C_{ox} in (1.1), but more importantly, also due to a higher saturation value of surface electron concentration (n_S). With very thin d_{ox}, inversion charge can barely be contained, rendering early saturation of n_S to a low value at a relatively low V_G. Conversely, with thicker d_{ox}, n_S can saturate to a higher value at a higher V_G, yielding also a higher saturation value of V_{ox}, and ultimately a lower ϕ^*_{Bp} according to (1.2).

The hole tunnel current, J_{tp}, can be modeled with a combined tunnel-and-thermionic-emission equation [27, 30]:

$$(V_G \gg \phi_t) \qquad J_{tp} = A^* T^2 P_{tp} \exp(-\phi^*_{Bp}/\phi_t), \qquad (1.3)$$

where A^* is the effective Richardson constant for holes [31], and P_{tp} is the tunneling probability for holes. From the exponential dependence to ϕ^*_{Bp} in this current model (that may overwhelm the $P_{tp} - d_{\mathrm{ox}}$ dependence), devices with thicker d_{ox} and thus lower ϕ^*_{Bp} will present higher J_{tp}, resolving the anomaly. As n_{S} is highly involved in this model with J_{tp} still being the dominant current component, we may classify Al/SiO$_2$/Si(p) MIS TDs as *minority-controlled majority devices.* The high sensitivity of J_{tp} to n_{S} is the essence of MIS TDs' sensor applications. What's more, transistor applications have been demonstrated by gating an MIS TD by another MOS capacitance and manipulating n_{S} by charge coupling [16, 17]. Recently, a more detailed and realistic model for J_{tp} has been established [32].

Diverged from leakage-free MOS capacitors where ψ_{S} is pinned above threshold [24], in MIS TDs, the pinning of V_{ox} prompts any further increase in V_{G} to fall on ψ_{S}, known as the deep depletion (DD) effect (DDE) [33, 34]. **Figure 1–8** demonstrates the TCAD simulation results of n_{S}, V_{ox}, and depletion width (W) as functions of V_{G} in an planar MIS TD with $d_{\mathrm{ox}} = 20\,\text{Å}$. For comparison purpose, simulation results with the electron tunneling model switched off (pretending if there were zero tunneling probability) are also presented. The "tunnel off" results are precisely described by the classical MOS electrostatics model. Contrariwise, the realistic "tunnel on" results suggest n_{S} and V_{ox} saturation, along with a continually-growing W. It is noteworthy that the discrepancies arise primarily under the inversion regime, affirming the necessity of a reliable electrostatics model for MIS TDs under reverse bias. A comprehensive analytical model for this will be developed in **Chapter 3**.

doi:10.6342/NTU202304241

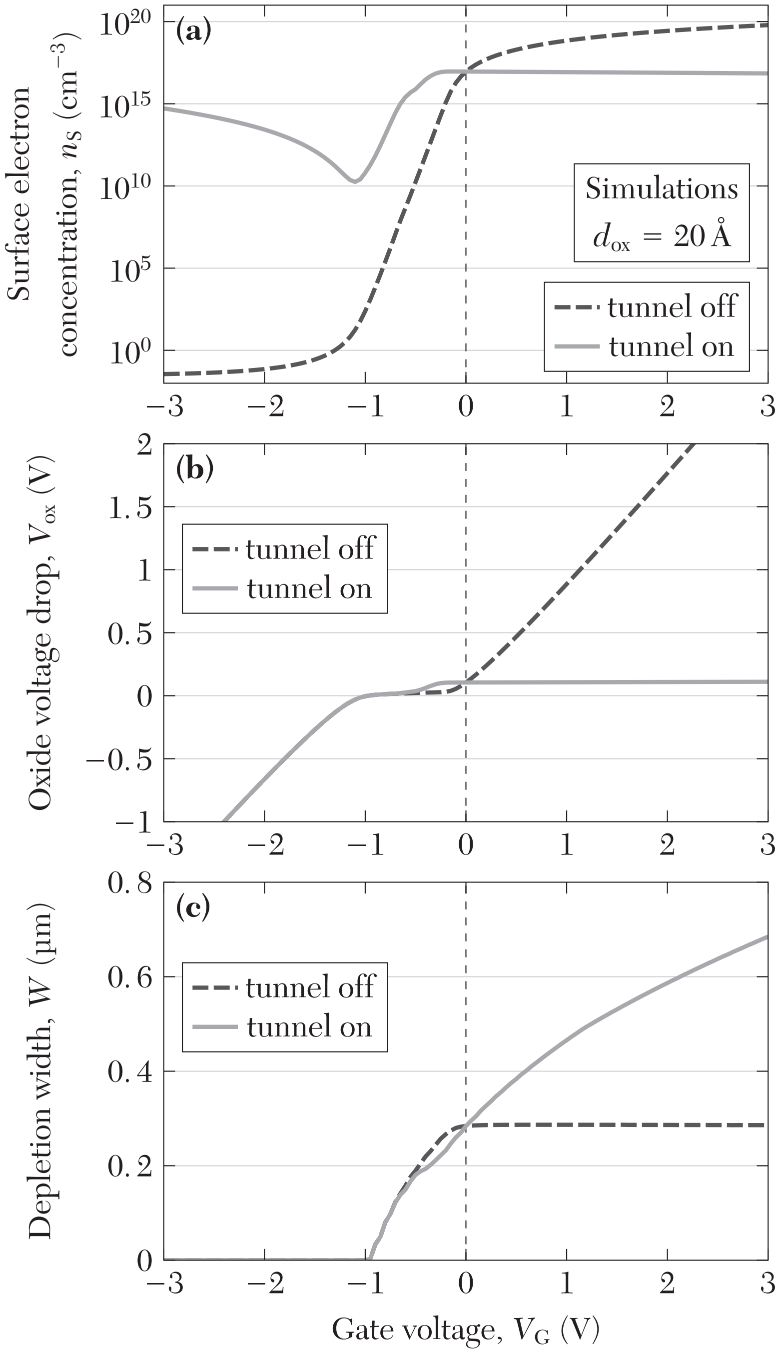


Figure 1–8. Simulated **(a)** surface electron concentration, **(b)** oxide voltage drop, and **(c)** depletion width vs. V_G plots for a planar Al/SiO$_2$/Si(p) MIS TD with d_{ox} = 20 Å, and with the electron tunneling model switched on and off.

1.3 High-Low and Ultra-High-Low MOS Tunnel Structures

High-Low MOS tunnel structures are encountered in defective/worn MOS structures and scaled MIS TDs, as mentioned in **Section 1.1**, and also in MOS capacitors with high surface roughness [35]. The thin oxide thickness d_L must be ultrathin for the devices' being *tunnel* structures. Every tunnel device with non-planar oxide may be categorized as high-low. However, devices with only one to a few low regions scattered under the gate area would be the most practical configurations. We will therefore concentrate on devices with a single low region at the gate center, not touching the gate edge (**Figure 1–3**). Depending on the high oxide thickness d_H, a further but crude classification may be useful:

1. The *demi-high-low* devices, with very subtle difference between d_H and d_L (up to a few angstroms).

2. The *moderate high-low* devices, with thicker d_H (up to a few nanometers) while tunneling is still possible.

3. The *ultra-high-low* devices, with ultrathick d_H ($\geq 10\,\mathrm{nm}$) that blocks any tunneling.

Several MIS-TD-like applications have been demonstrated with the demi-high-low devices [36, 37], the ultra-high-low devices [38], and even with MIS TDs with intentional soft breakdown [39], while the physical mechanisms are not entirely known. It is one of our main objectives to unroll the phenomena and mechanisms in ultra-high-low devices, by qualitative and quantitative means (**Chapter 2** and **Chapter 4**) in this

doi:10.6342/NTU202304241

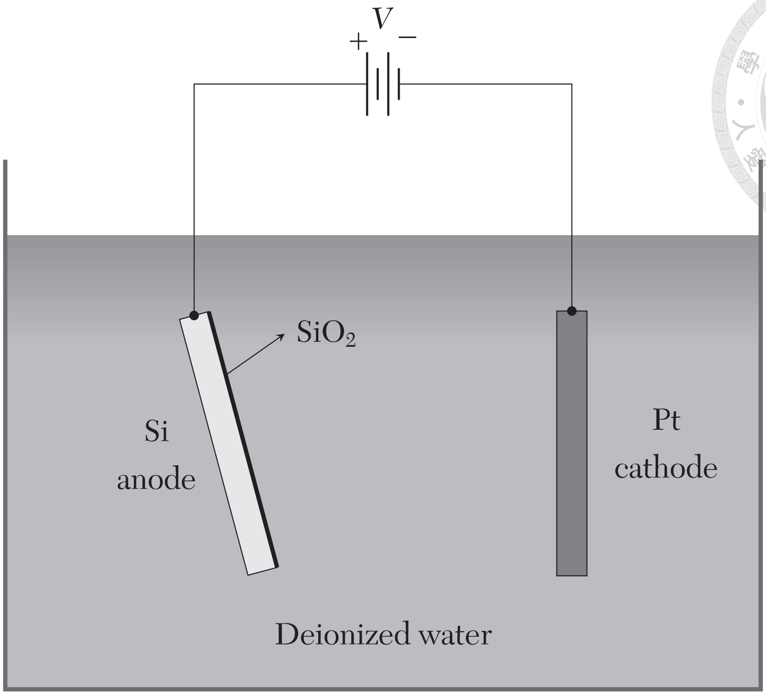


Figure 1–9. Apparatus for Si wafer anodic oxidation.

work.

1.4 On Fabricating the High-Low Oxide Structures

The high-low oxide profiles in this work and other [36–38] were fabricated with in-situ oxidation of the silicon substrate, with an aid of photolithography to define the high-low patterns. Anodic oxidation (ANO) was adopted as the oxidation method to produce high-quality SiO$_2$ layers [40, 41]. The apparatus for Si wafer anodic oxidation is shown in **Figure 1–9** . The silicon wafer and an inert electrode (Pt) is submerged in a water tank filled with room-temperature deionized water. The front side of the wafer faces towards the inert electrode with a separation of $1 - 3\,\text{cm}$. A DC power source ($15 - 30\,\text{V}$) is applied across the two electrodes, with the wafer serving as the anode, where oxidation reaction would take place. The anode-side half reaction is given by

$$\mathrm{Si} + 4\,\mathrm{OH}^- \longrightarrow \mathrm{SiO_2} + 4\,\mathrm{e}^- + 2\,\mathrm{H_2O}. \tag{1.4}$$

The wafer may be tilted (off-parallel to the cathode) or not — Tilted wafers are subject to graded electric field at the surface, rendering an oxide thickness gradient. A range of oxide thicknesses on a single wafer is therefore achievable. On the other hand, parallelly-aligned electrodes results in very uniform d_{ox}. Rapid thermal process (RTP) must be conducted post ANO to assure oxide quality.

High-low oxide structures can be achieved by a combination of two ANO runs and one photolithography step. Two possible pathways are shown in **Figure 1–10**: the *thin-first* process, by growing a thin planar oxide layer first, masking the low region with photoresist, and conducting a second ANO to thicken the exposed oxide, and the *thin-last* process, by growing the high oxide first, oxide etching at the low region, and oxidizing the exposed surface to form the low oxide. In the thin-first process, d_{L} is fixed right after the first ANO. This process is therefore optimal for precise d_{L} requirements. It is also ideal for fabricating demi-high-low structures. However, for growing thicker d_{H}, the second ANO will be so long ($>$ 30 minutes) that it prompts integrity issues in the photoresist, such as peeling off or pattern deformation. Stress could also build up at the high-low boundary if d_{H} is too thick, like the LOCOS process [42]. Hence, the thin-first process is not suitable for growing ultra-high-low structures. On the other hand, the thin-last process, while capable of growing ultra-high-low structures with relatively low stress and better maintaining silicon surface flatness, suffers from difficulty controlling d_{L}. This may be attributed to reinforced electric field at the exposed silicon

doi:10.6342/NTU202304241

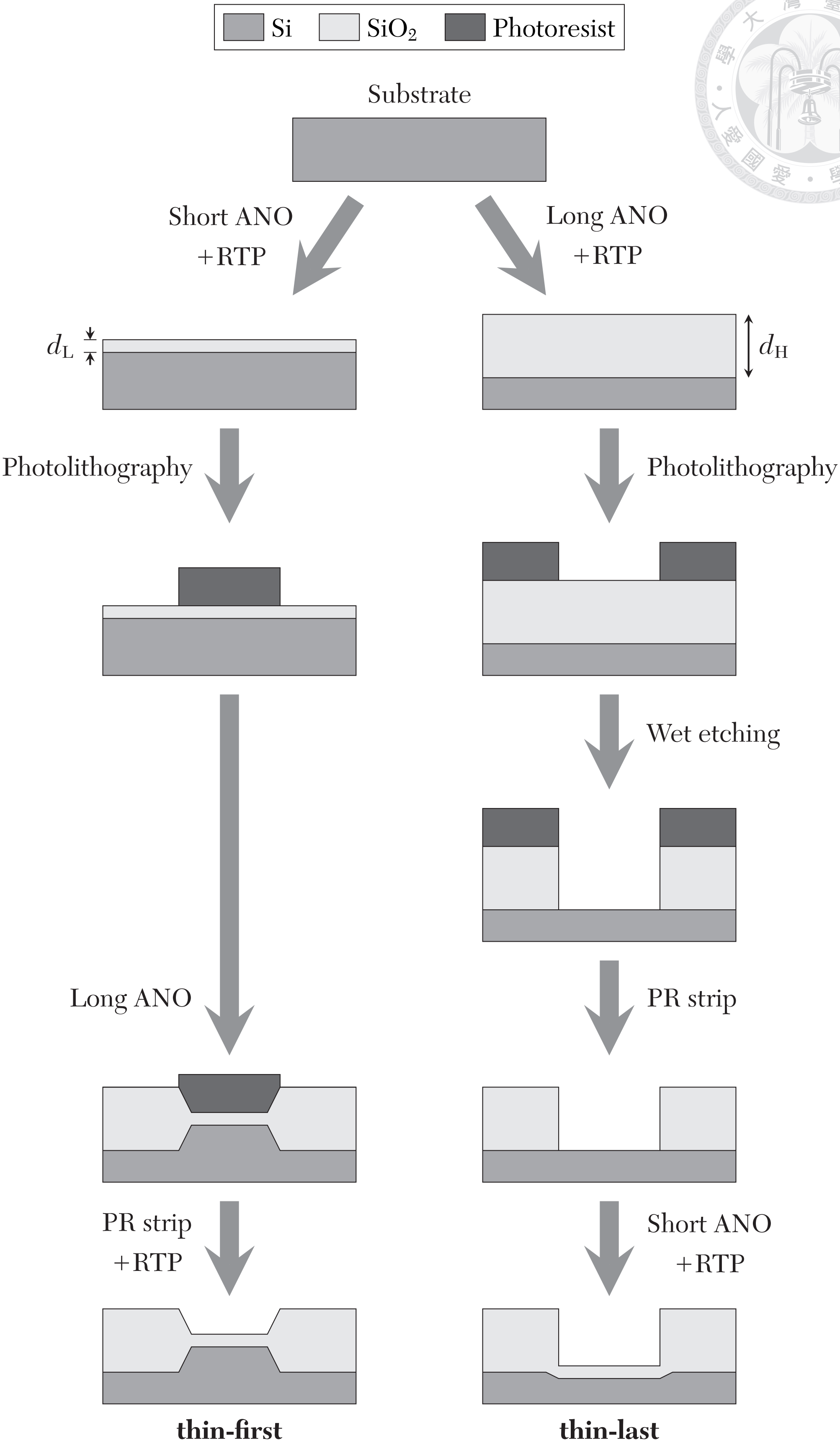


Figure 1–10. Process flow for manufacturing high-low oxide structures by the thin-first and thin-last processes.

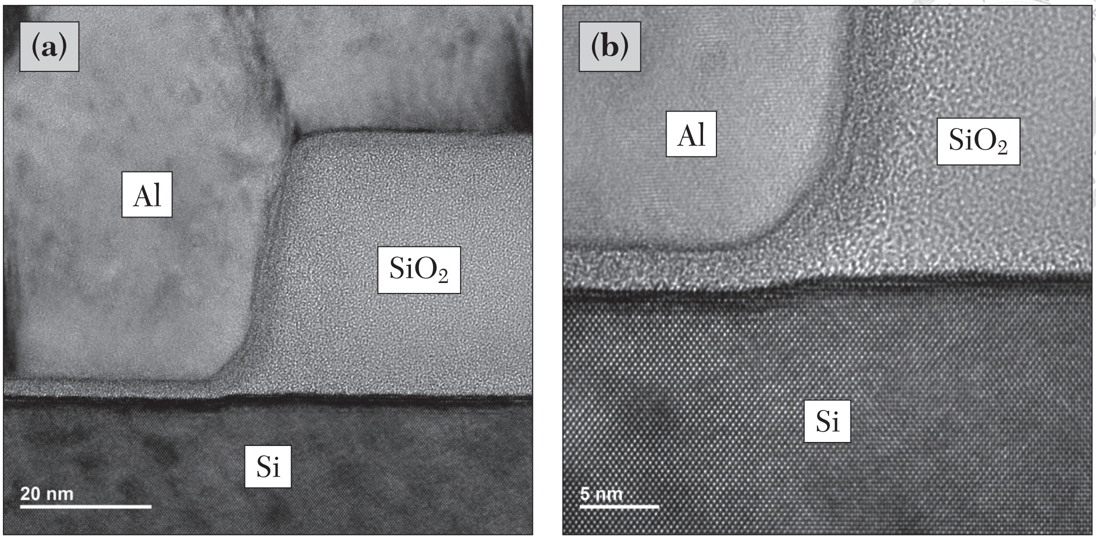


Figure 1–11. **(a)** Cross-sectional TEM image of an ultra-high-low device fabricated by the thin-last process, at the high-low boundary. Physical oxide thicknesses are $d_{\mathrm{L}} = 31\,\text{Å}$ and $d_{\mathrm{H}} = 400\,\text{Å}$. **(b)** High magnification.

surface (i.e., low regions) that constitutes only a small portion of the total wafer area during the second ANO run. Replacing the second ANO with other field-free oxidation techniques such as thermal oxidation or rapid thermal oxidation (RTO) may be a viable option for better control on d_{L}. Yet, for fabricating only a small number of test devices, several trial-and-error runs on the second ANO for obtaining the desired d_{L} may still be acceptable. Thus, in this work, the thin-last process was adopted to fabricate ultra-high-low devices. **Figure 1–11 (a, b)** show the cross-sectional transmission electron microscope (TEM) images of an ultra-high-low device fabricated by the thin-last process, at the high-low boundary. Wet etching is conducted with buffered oxide etchant (BOE, HF:NH_4F). The first ANO lasted 8 hours at an electric field of 15 V/cm and yielded $d_{\mathrm{H}} = 400\,\text{Å}$. The second ANO lasted 3 minutes at an electric field of 7.5 V/cm and yielded $d_{\mathrm{L}} = 31\,\text{Å}$. Wet etching is responsible for the tapered wall and the smooth high-low boundary. Nonetheless, superior uniformities of d_{H} and d_{L} were

doi:10.6342/NTU202304241

observed in the images.

1.5 Dissertation Organization

In **Chapter 2**, experimental details for fabricating the ultra-high-low devices are presented. Electrical characteristics, electrostatics, and carrier transport phenomena for the devices are studied qualitatively with an aid of TCAD simulations. For quantitative electrostatics analysis, an comprehensive, analytical electrostatics model for planar MIS TDs under reverse bias is established and studied in **Chapter 3**, and extended to an ultra-high-low electrostatics model in **Chapter 4**. Last but not least, sensor applications for ultra-high-low MOS(p) tunnel devices with suitable d_{L} and the presence of SBHME are discussed and compared to planar MIS TDs in **Chapter 5**. The dissertation is concluded in **Chapter 6** with suggestions for future works.

Some detailed mathematical derivations can be found in the appendix.

doi:10.6342/NTU202304241

2

Qualitative Study on Electrical Characteristics, Electrostatics, and Carrier Transport Phenomena in Ultra-High-Low MOS(p) Tunnel Structures

doi:10.6342/NTU202304241

2.1 Background

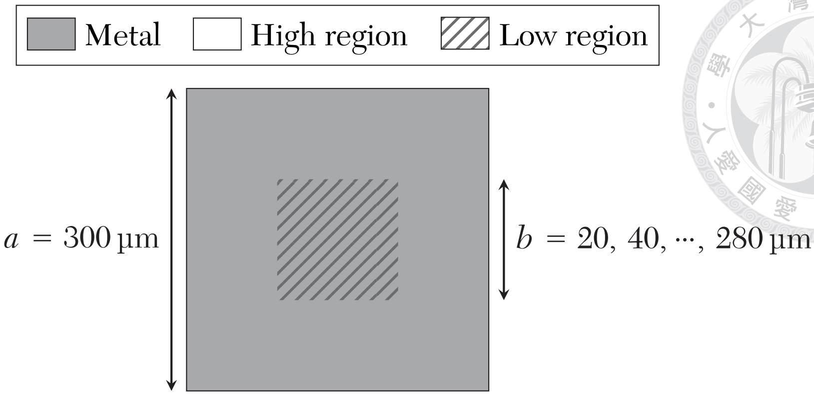
Ultra-high-low (UHL) MOS tunnel structures may or may not exhibit average traits of planar MOS capacitors or MIS TDs alone, despite structurally being a blend of the two. Tunneling-induced deep depletion phenomena in MIS TDs render very different electrostatics from the classical MOS electrostatics theory (**Section 1.2** and **Figure 1–8**) under the inversion regime. Ideal MOS capacitors has no carrier transport in steady state, but DC tunneling is prominent in MIS TDs. There are also very dissimilar $I - V$ and $C - V$ characteristics between the two categories of planar devices [34]. All in all, planar MOS capacitors and MIS TDs can essentially be treated as one-dimensional devices (except at the gate edge) and may come with negligible lateral carrier movement. However, the high and low regions are laterally connected in UHL devices, and therefore lateral carrier coupling may effectuate unforeseen outcomes on device electrostatics, or even on their $I - V$ and $C - V$ characteristics, which are to be qualitatively studied in this chapter by characterizing physical, photolithographically-defined UHL devices, and also by TCAD simulation.

2.2 Experimental Details

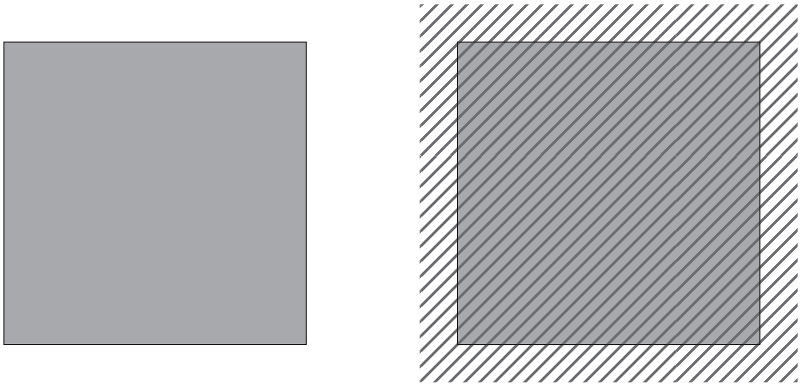
A 3-inch boron-doped (100)-oriented single-crystal p-type Si wafer with a resistivity of $1 - 10\,\Omega \cdot \mathrm{cm}$ (corresponding to $N_\mathrm{A} = 10^{16}\,\mathrm{cm}^{-3}$) and a thickness of $370\,\mu\mathrm{m}$ was utilized as the substrate. After performing a standard Radio Cooperation of America (RCA) clean [43] for particle and impurity removal, as well as a dilute hydrofluoric acid (HF) dip for stripping the native oxide, the thin-last process (**Section 1.4**, **Figure 1–10**) was conducted to obtain the desired high-low oxide profile on the wafer surface. The first ANO lasted 8 hours under a 30 V DC voltage across the electrodes, and the second ANO lasted 2.5 minutes under a 15 V DC voltage. Electrode spacing was $\sim$ 2 cm. RTP was performed, at 950 °C for 15 s in a 20-torr N_2 ambient, right after each ANO. Oxide wet etching was conducted in BOE for 80 seconds at room temperature. Following the completeness of the high-low oxide structure, a layer of Al metal (200 nm thick) was thermally evaporated onto the oxide surface. It was then patterned into metal gates with photolithography and wet etching. The fabrication process was concluded by cleaning the back side of the wafer with BOE and evaporating another layer of Al metal to serve as the back contact.

Device $I - V$ and $C - V$ characteristics were measured with an Agilent® B1500A Semiconductor Device Analyzer. TCAD simulations were conducted with SILVACO® ATLAS™. The measured C_{ox} value for the high-only device suggests $d_{\mathrm{H}} = 320$ Å, while the low V_{C} and early current saturation behavior (**Section 1.2**) in the low-only device suggests $d_{\mathrm{L}} \leq 25$ Å.

Figure 2–1 depicts the top view of the experimental devices. Metal gates and low regions are all in square shapes. Dimensions a and b denote the side lengths of the metal gate and the low region, respectively, in a UHL device. The low region is centered in the metal gate for each UHL device. For every device in this experimental set, $a = 300$ µm. "UHL-b" (b in µm) is the shorthand notation for the device with low region of side length b. For comparison purposes, planar devices equipped only with the high oxide (high-only, ≡ MOS capacitor) or the low oxide (low-only, ≡ MIS TD) were co-fabricated. In the UHL devices, the *low area proportion* (low-to-gate area



(a) UHL-*b* device



(b) High-only device **(c)** Low-only device

Figure 2–1. Top view of the experimental devices. **(a)** UHL-b (b in micrometers; e.g., UHL-20) are ultra-high-low devices with the low regions being $b \times b$ square shapes. **(b)** High-only and **(c)** low-only reference devices are co-fabricated planar devices for comparison purposes. All metal gates are in $a \times a = 300\,\mu\text{m} \times 300\,\mu\text{m}$ square shapes.

ratio) is given by

$$A_\text{L}/A_\text{G} = \frac{b^2}{a^2}, \tag{2.1}$$

where $A_\text{G} = a^2$ and $A_\text{L} = b^2$ denotes the gate area and low-region area, respectively.

We may also denote the high-region area by $A_\text{H} \equiv A_\text{G} - A_\text{L} = a^2 - b^2$.

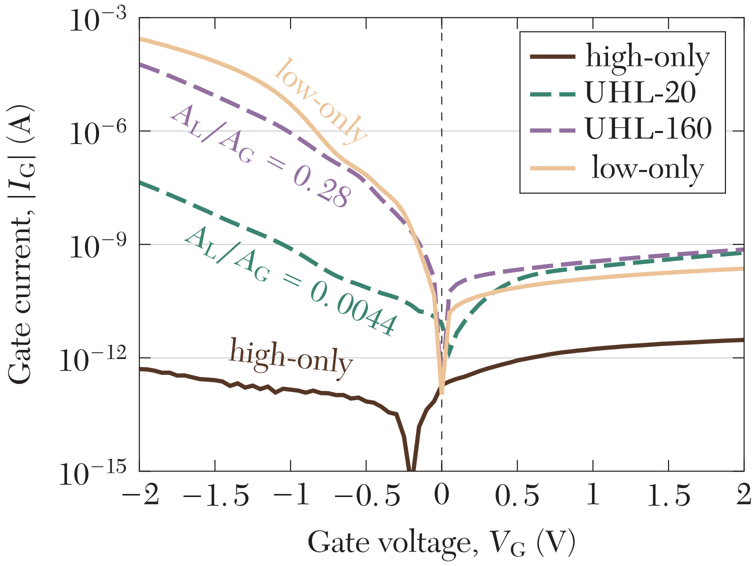


Figure 2–2. Current-voltage characteristics for selected UHL devices and the reference devices. The characteristics are independent of voltage ramp rate (except for the high-only device with apparent displacement current.)

2.3 Results and Discussion

2.3.1 Experimental Current-Voltage and Capacitance-Voltage Characteristics

Figure 2–2 shows the $I - V$ characteristics for selected UHL devices (UHL-20 and UHL-160) along with the high-only and low-only reference devices. The low-only device demonstrates typical $I - V$ characteristics for an MIS TD (**Figure 1–6**) with relatively thin d_L ($\leq 25\,\text{Å}$) such that the reverse ($V_G > 0$) current saturates to a relatively low magnitude at a small V_C and hole current is negligible due to unpronounced SBHME. The high-only device presents little current, indicating negligible carrier tunneling with such d_H. On the other hand, in the UHL devices, tunneling current becomes prominent once a low region is introduced. While the forward-bias ($V_G < 0$) currents show proportionality to A_L for each UHL and the thin-only device, under reverse bias, their saturation currents become comparable. Even for the UHL-20 device with $A_{\mathrm{L}}/A_{\mathrm{G}}$ being as low as 4.4×10^{-3}, its leakage current under the inversion regime is as significant as that of the planar MIS TD.

Figure 2–3 (a) shows the device $C - V$ characteristics under $f = 10\,\mathrm{kHz}$, and **Figure 2–3 (b)** plots the extracted values vs. $A_{\mathrm{L}}/A_{\mathrm{G}}$ under the accumulation regime (at $V_{\mathrm{G}} = -1.7\,\mathrm{V}$). The near-ideal flat-band voltages ($V_{\mathrm{FB}} \approx -0.9\,\mathrm{V}$, $|\Delta V_{\mathrm{FB}}| \leq 0.2\,\mathrm{V}$) for all samples in **Figure 2–3 (a)** suggest negligible effective oxide charge ($Q_{\mathrm{eff}}/q \leq 1.5 \times 10^{11}\,\mathrm{cm}^{-2}$). The high-only device demonstrates the lowest capacitance under the accumulation regime, and conversely for the low-only device. Moreover, in **Figure 2–3 (b)**, the accumulation capacitance for each device shows linearity to $A_{\mathrm{L}}/A_{\mathrm{G}}$. That is, provided the high-only and low-only accumulation capacitances (per unit area) C_{accH} and C_{accL}, the C_{Acc} for any UHL device can be predicted by the equivalent circuit in the inset figure:

$$\begin{aligned} C_{\mathrm{Acc}} &= C_{\mathrm{accL}} A_{\mathrm{L}} + C_{\mathrm{accH}} A_{\mathrm{H}} \\ &= A_{\mathrm{G}} C_{\mathrm{accH}} + A_{\mathrm{G}}(C_{\mathrm{accL}} - C_{\mathrm{accH}}) \cdot (A_{\mathrm{L}}/A_{\mathrm{G}}), \end{aligned} \tag{2.2}$$

bringing in the linearity to $A_{\mathrm{L}}/A_{\mathrm{G}}$. Notwithstanding, in **Figure 2–4** showing the magnified $C - V$ characteristics under the inversion regime, such linearity is no longer present. In **Figure 2–4 (a)** showing the low-frequency capacitance values (C_{LF}), the high-only device's capacitance restores to a higher value under the inversion regime due to the emergence of inversion charge [24]. However, the low only device loses its inversion charge from tunneling and demonstrates no capacitance restoration. Meanwhile, in **Figure 2–4 (b)** showing the high-frequency capacitance values (C_{HF}), the

doi:10.6342/NTU202304241

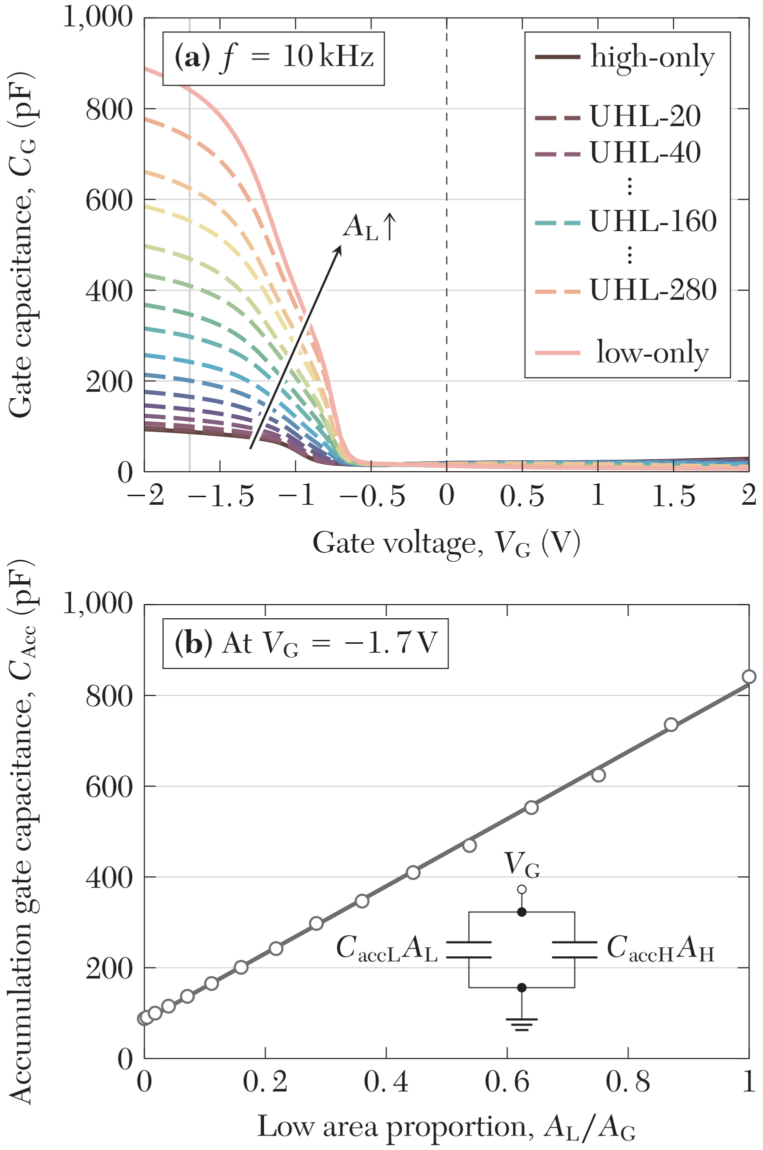


Figure 2–3. **(a)** Capacitance-voltage characteristics for all UHL devices (dashed) and the reference devices (solid) under 10 kHz. **(b)** Scatter plot of accumulation capacitance vs. (A_L/A_G) exhibits linearity. The capacitance values were extracted at $V_G = -1.7\,\text{V}$ from **(a)**. Inset shows the equivalent circuit under accumulation.

inversion capacitance is near constant in the high-only device due to pinned ψ_S [18], but the low-only device undergoes DD; i.e., W increases with gate voltage, and C_{HF} decreases. For the UHL devices, it is noteworthy that both their $C_{LF}-V_G$ and $C_{HF}-V_G$ characteristics are analogous to the low-only device regardless of A_L/A_G. Even in the

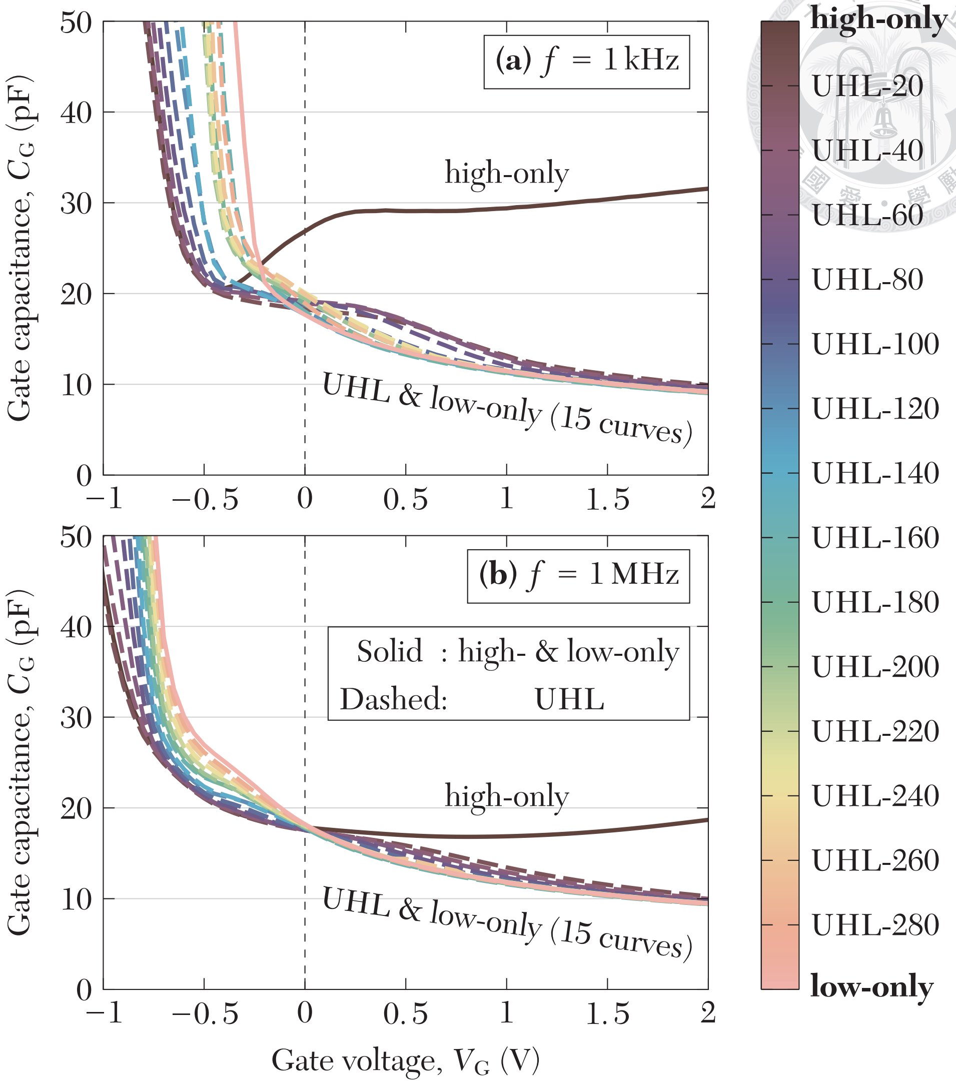


Figure 2–4. **(a)** Low-frequency (1 kHz) and **(b)** high-frequency (1 MHz) $C - V$ characteristics for all UHL devices (dashed) and the reference devices, featuring the inversion regime.

UHL-20 device with A_H accounting 99.6 % the total gate area, the $C - V$ characteristics under the inversion regime shows no similarities to the high-only device. That is, in the UHL devices, the high and low regions are apparently *charge-coupled* under the inversion regime: Not only the inversion charge under the low oxide must be lost, but also under the high oxide, to prompt C_{HF} deep depletion and impaired C_{LF} restoration throughout the UHL device. One concludes that a local oxide thinning region can have

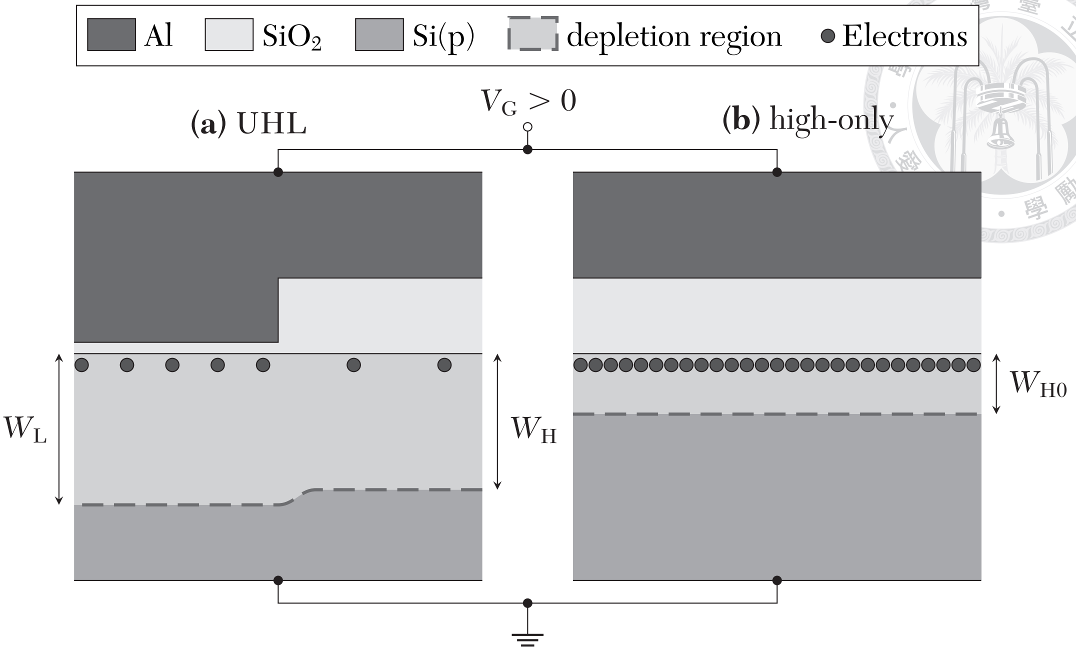


Figure 2–5. Schematic cross-sectional view of **(a)** a UHL and **(b)** the high-only device, illustrating the surface electrons and depletion widths.

very significant impact on the device $C-V$ characteristics and electrostatics, even with a small A_L/A_G ratio.

To further inspect the electrostatics at the high region of the UHL devices, its depletion width may be extracted from the $C_{HF}-V$ curves. **Figure 2–5** depicts the cross-sectional view of a UHL and the high-only device, illustrating the surface electrons (inversion charge) and depletion regions, and also featuring a lowered electron concentration and consequent DD in the UHL device. W_{L0} and W_{H0} denotes the depletion widths in the low-only device (not shown in this figure) and the high-only device, respectively, under a certain positive V_G. W_L and W_H denotes the depletion widths (assumed to have little dependence on lateral location) in a UHL device at the low and high regions, respectively. The lack of surface inversion charge is responsible for the fact that $W_H > W_{H0}$.

doi:10.6342/NTU202304241

Under the depletion regime as well as DD, the gate capacitance in the low-only device is given by [24]

$$\text{(Low-only)} \qquad \begin{aligned} C_{\mathrm{GL0}} &\approx A_{\mathrm{G}}\left(\frac{1}{C_{\mathrm{oxL}}}+\frac{1}{C_{\mathrm{dL}}}\right) \\ &= A_{\mathrm{G}}\left(\frac{1}{C_{\mathrm{oxL}}}+\frac{W_{\mathrm{L0}}}{\epsilon_{\mathrm{s}}}\right)^{-1}, \end{aligned} \tag{2.3}$$

where $C_{\mathrm{dL}} = \epsilon_{\mathrm{s}}/W_{\mathrm{L}}$ is the substrate depletion capacitance per unit area, and $C_{\mathrm{oxL}} = \epsilon_{\mathrm{ox}}/d_{\mathrm{L}}$ is the oxide capacitance per unit area at the low region. Likewise, assume deep depletion to take place in both the high and low regions in the UHL device, the gate capacitance would be

$$\text{(UHL)} \qquad C_{\mathrm{G}} \approx A_{\mathrm{L}}\left(\frac{1}{C_{\mathrm{oxL}}}+\frac{W_{\mathrm{L}}}{\epsilon_{\mathrm{s}}}\right)^{-1} + A_{\mathrm{H}}\left(\frac{1}{C_{\mathrm{oxH}}}+\frac{W_{\mathrm{H}}}{\epsilon_{\mathrm{s}}}\right)^{-1}, \tag{2.4}$$

where $C_{\mathrm{oxH}} = \epsilon_{\mathrm{ox}}/d_{\mathrm{H}}$ is the oxide capacitance per unit area at the high region. By assuming similar electrostatics between the UHL low region and the thin-only device ($W_{\mathrm{L}} \approx W_{\mathrm{L0}}$), W_{H} is solved from (2.4) as

$$W_{\mathrm{H}} \approx \epsilon_{\mathrm{s}}\left(\frac{A_{\mathrm{H}}}{C_{\mathrm{G}} - (A_{\mathrm{L}}/A_{\mathrm{G}})C_{\mathrm{GL0}}} - \frac{1}{C_{\mathrm{oxH}}}\right). \tag{2.5}$$

Figure 2–6 depicts the calculated $W_{\mathrm{H}} - V_{\mathrm{G}}$ curves from **Figure 2–4 (b)** and equation (2.5) for the high-only and UHL devices. While all devices share the same W_{H} at zero bias, as V_{G} continues to grow, there is no deep depletion in the high-only device, but deep depletion is prominent in the UHL devices. In fact, inversion carriers are so efficiently depleted in the UHL devices that all W_{H} present similar magnitudes

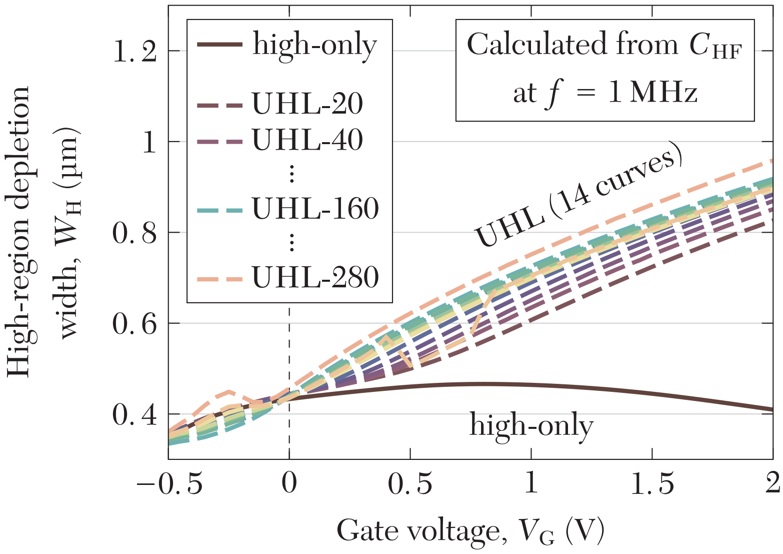


Figure 2–6. Calculated device high-region depletion widths (W_{H}) as functions of V_{G} from **Figure 2–4 (b)** and equation (2.5).

regardless of $A_{\mathrm{L}}/A_{\mathrm{G}}$.

Looking back at the $I-V$ characteristics in **Figure 2–2** , the like magnitudes of saturation currents amongst all devices (except high-only) can be associated with the deep depletion phenomenon in the high region. For sufficiently thin d_{L} such that there is little SBHME, thermal generation current in the space charge region is the dominant current component for the device reverse saturation current. Therefore, the fact that $I_{\mathrm{G}} \propto A_{\mathrm{G}}$ rather than $\propto A_{\mathrm{L}}$ under reverse bias indicates that thermally-generated electrons in the high region must also contribute to the gate current. As a consequence, there is little inversion charge in the high region, leading to very dissimilar $C-V$ characteristics to that of the high-only device, as discussed earlier. On the contrary, electron tunnel current under the accumulation regime is tunneling-rate-limited [44], and I_{G} would exhibit simple proportionality to A_{L} as $V_{\mathrm{G}} < V_{\mathrm{FB}}$. **Figure 2–7** illustrates the different electron transport mechanisms under the two bias regimes in a UHL device. In conclusion, the introduction of a tunneling LOT spot into an MOS capacitor can

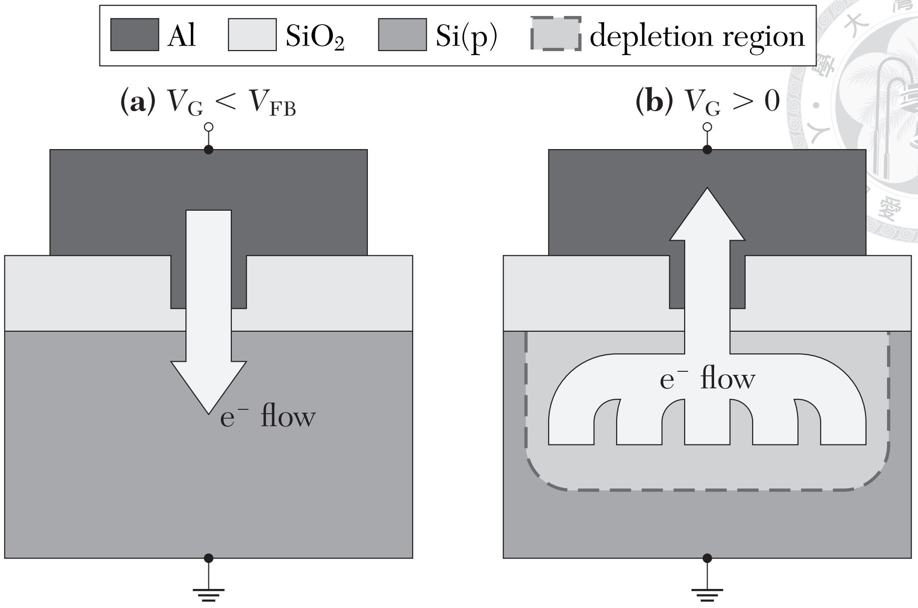


Figure 2–7. Schematic electron flow in a UHL device under **(a)** accumulation regime and **(b)** reverse bias.

extensively impact its leakage current and electrostatics under the inversion regime, making the entire device to behave like an MIS TD.

2.3.2 TCAD Simulation Results

For better understanding on the influence of LOT spots, TCAD simulations were conducted. The simulated three-dimensional structures are designed to be cylindrically symmetric to reduce the azimuthal dimension in cylindrical coordinate systems, and to accomplish two-dimensional cylindrical simulations for better accuracy and performance. **Figure 2–8** depicts the schematic cross-sectional view of the simulated UHL devices across a diameter, and also the definition of the coordinate system. The gate and low regions are in concentric circular shapes with radii r_G and r_L, respectively. Two probing points, O and A, where physical quantities are to be extracted, represent the low and high regions, respectively. O is defined at the device center ($r_O = 0$), and A at midway between the gate edge and the high-low boundary ($r_A = \frac{1}{2}(r_L + r_G)$). In

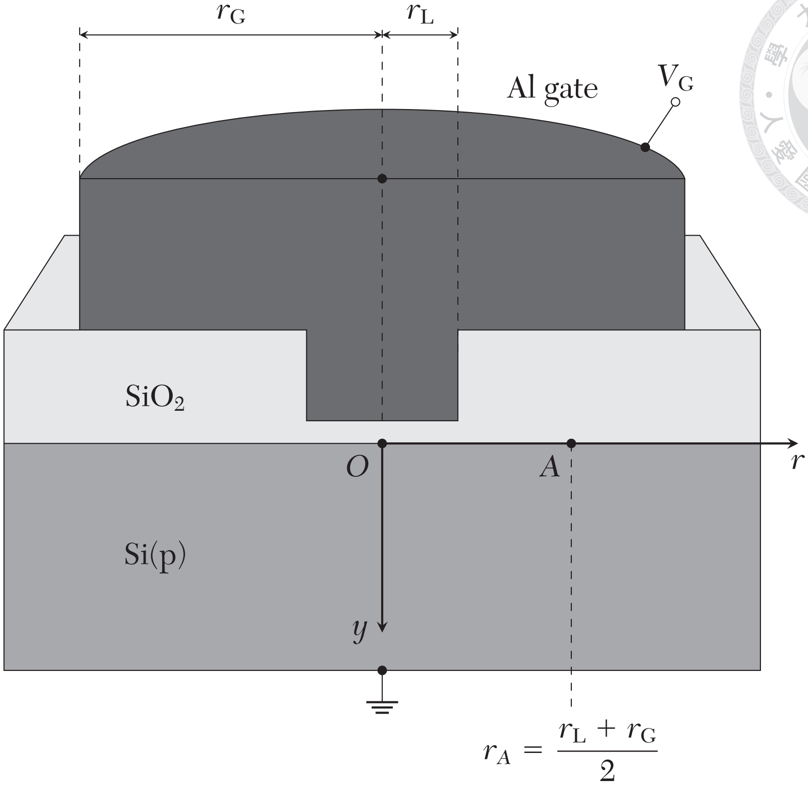


Figure 2–8. Schematic cross-sectional view of the simulated UHL devices across a diameter, and the coordinate system. The simulated devices are cylindrically-symmetric. Two probing points O and A for data extraction are defined in the low and high regions, respectively. (O at the device center, A at midway between the gate edge and the high-low boundary.)

analogy to the experimental devices, high-only and low-only devices were simulated also. r_{G} is fixed at 20 µm for each device. **TABLE 2–I** lists all simulated devices and their dimensions. Without otherwise specified, oxide thicknesses are set identical to the experimental devices: $d_{\mathrm{H}} = 320\,\text{Å}$, and $d_{\mathrm{L}} = 20\,\text{Å}$. **Figure 2–9** demonstrates the $I - V$, quasistatic, and high-frequency $C - V$ characteristics for the simulated devices, reproducing the same traits as the experimental curves.

First, TCAD simulation results under reverse bias reveal the presence of a radial electric field ($\mathscr{E}_r$) in the semiconductor substrate, near the high-low boundary, pointing from the low region towards the high region. **Figure 2–10 (a)** schematically depicts

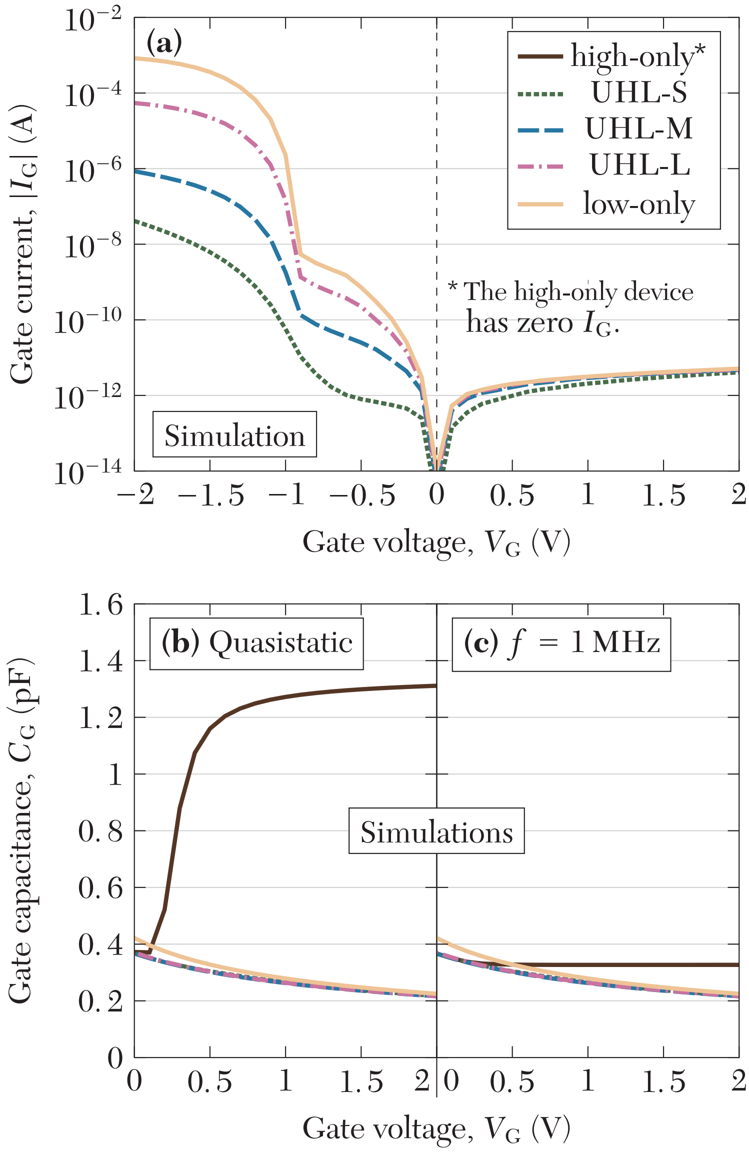


Figure 2–9. Simulated **(a)** $I - V$, **(b)** quasistatic $C - V$, and **(c)** high-frequency $C - V$ characteristics for the high-only, low-only and UHL devices.

TABLE 2–I

List of the simulated devices and their dimensions.

Device	r_G (μm)	r_L (μm)	$A_\mathrm{L}/A_\mathrm{G}$
High-only	20	0	0
UHL-S		0.05	6.25×10^{-6}
UHL-M		0.5	6.25×10^{-4}
UHL-L		5	0.0625
Low-only		∞	1

$\mathscr{E}_r$ in the UHL device about the high-low boundary. The lateral field arises from the surface band bending difference across the boundary and may be responsible for the highly-effective electron depletion underneath the high oxide. Under same V_G, the low region with thinner d_ox is associated with a higher ψ_S (**Section 1.2**), and thus $\mathscr{E}_r$ must point from the low region with higher ψ_S towards the high region with lower ψ_S. As a result, thermally-generated electrons under the high region are drifted towards the low region, constituting the gate current. **Figure 2–10 (b)** shows the simulated $\mathscr{E}_r$ value map about the high-low boundary in the UHL-M device. Noteworthily, $\mathscr{E}_r$ is skewed to the high-region side and is maximized ($\equiv \mathscr{E}_{r(\mathrm{max})}$) near the high-low boundary. **Figure 2–11 (a)** plots $\mathscr{E}_{r(\mathrm{max})}$ vs. V_G in the three simulated UHL devices. $\mathscr{E}_r$ reverts direction below $V_\mathrm{G} < V_\mathrm{FB}$ for each device. In the devices with higher $A_\mathrm{L}/A_\mathrm{G}$ (namely, UHL-M and UHL-L), the monotonic increment of $\mathscr{E}_{r(\mathrm{max})}$ with $V_\mathrm{G} > V_\mathrm{FB}$ implies increasing surface band bending difference (between high and low) as V_G grows, but the increment appears to be minor. However, in the small-$A_\mathrm{L}/A_\mathrm{G}$ device (UHL-S), the lower values of $\mathscr{E}_\mathrm{r}$ implies less discrepancy in the surface band bendings. The values of surface band bending will be analyzed in more details in the upcoming discussion.

Figure 2–12 shows the surface electron concentration (n_S) vs. V_G plot at points

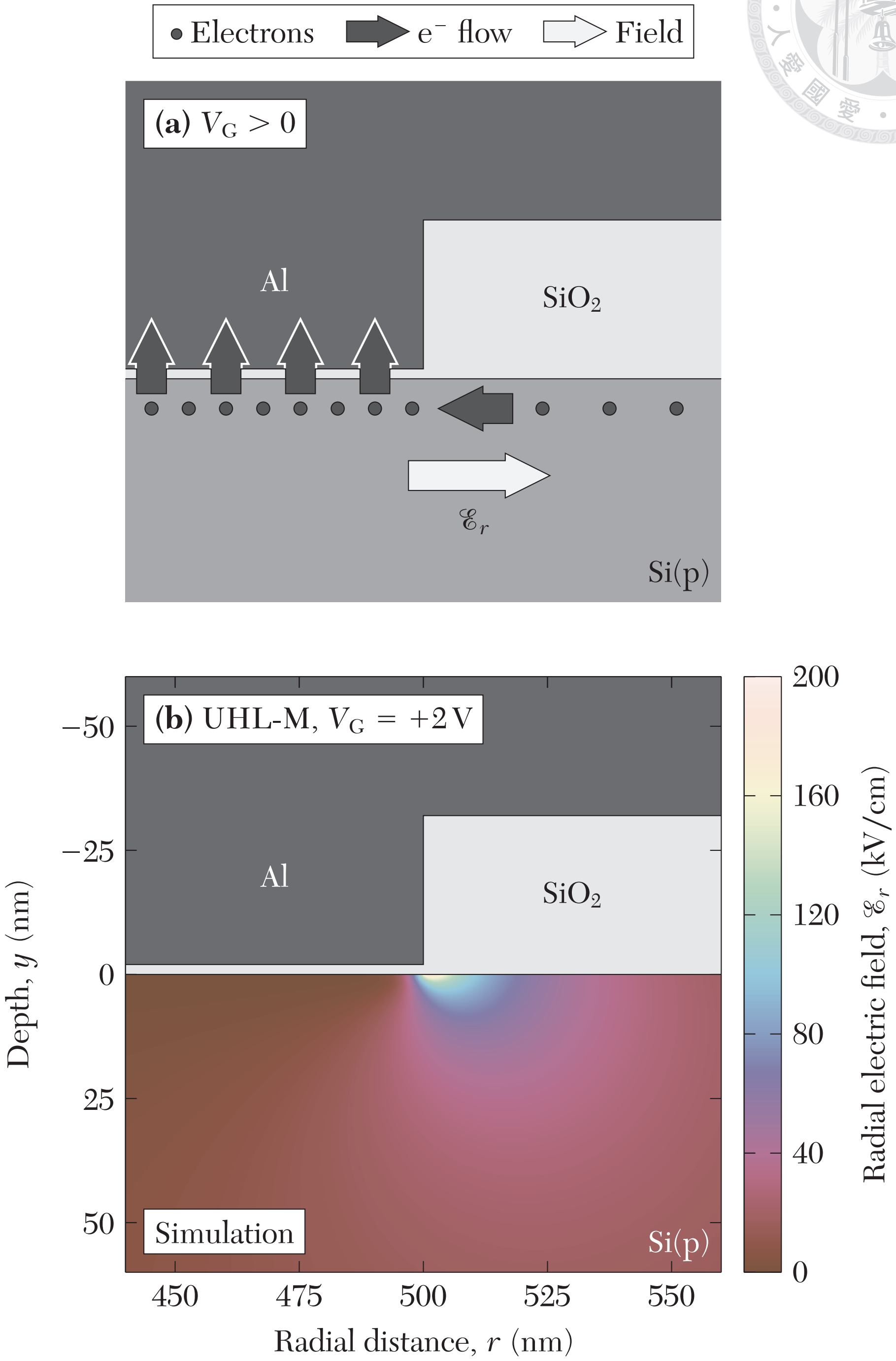


Figure 2–10. **(a)** Schematic cross section of a reverse-biased UHL device about the high-low boundary, showing the electron transport and a radial electric field $\mathscr{E}_r$. **(b)** Simulated $\mathscr{E}_r$ value map about the high-low boundary in the UHL-M device. Gate voltage is +2 V.

doi:10.6342/NTU202304241

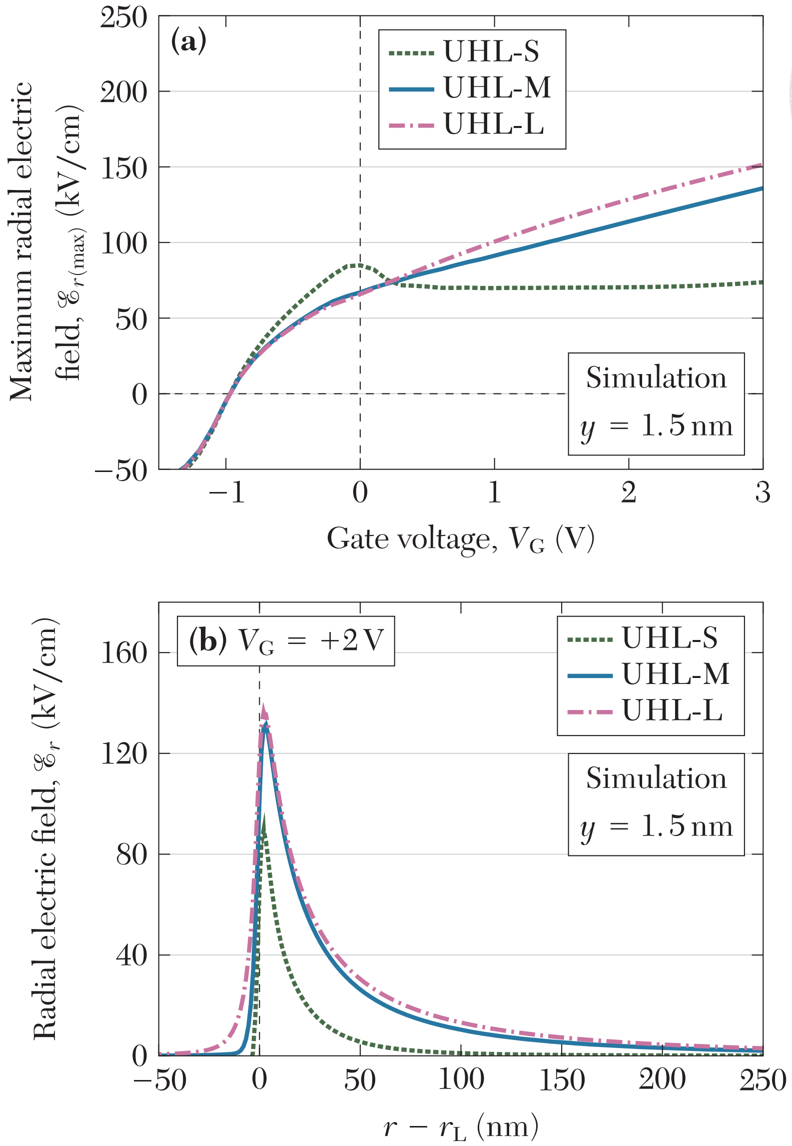


Figure 2–11. **(a)** Plot of maximum $\mathscr{E}_r$ values in the simulated UHL devices about the high-low boundary, as a function of V_G. **(b)** Plot of $\mathscr{E}_r$ profiles along radius in the simulated UHL devices about the high-low boundary, under $V_G = +2\,\mathrm{V}$. Data extracted at $y = 1.5\,\mathrm{nm}$.

O and A in the UHL-M device and the reference devices. In the planar devices, near zero applied bias, the low-only device exhibits higher n_S due to higher C_{ox}. A crossover of $n_S - V_G$ curve with the thick-only device takes place at a higher V_G due to the n_S saturation in the low-only device as a result of tunneling, but also due to the n_S growth in the otherwise non-tunneling device. In the UHL-M device, intriguingly, n_S in the

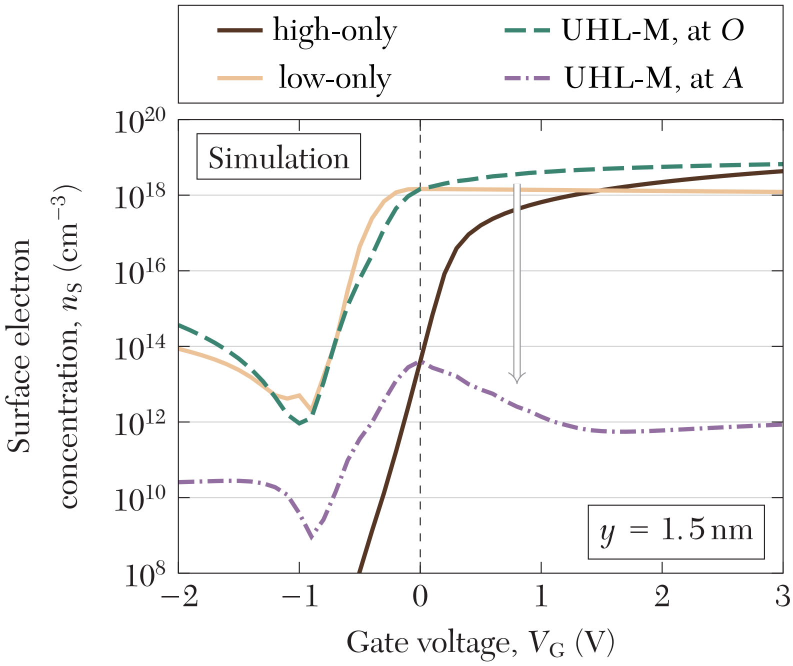


Figure 2–12. Simulated $n_\mathrm{S} - V_\mathrm{G}$ plots in the UHL-M device, at points O and A, and in the reference devices. Data extracted at $y = 1.5\,\mathrm{nm}$.

low region (O) demonstrates comparable magnitude to that in the thin-only device, whereas n_S in the high region (A) undergoes a degradation by 6 decades, compared to that of the high-only device, exemplifying the high effectiveness of inversion carrier depletion by the local oxide thinning spot. **Figure 2–13** plots the radial n_S profiles in the simulated devices under a high positive gate bias (+2 V). n_S appears to be near constant throughout the high region in all UHL devices, except being slightly lower near the high-low boundary, indicating the low region's capability on electron exhaustion over such a wide distance. With smaller A_L (UHL-S), n_S is less affected in the high region. In addition, note the lowered n_S in the neutral region near the gate edge ($r \gtrsim r_\mathrm{G}$) in the thin-only and UHL devices, indicating nonequilibrium in the depletion layer. This neutral-region minority carrier profile is similar to that in the depletion edge of P-N junctions under reverse bias [18, 19] and indicates nonequilibrium (scarcity of

doi:10.6342/NTU202304241

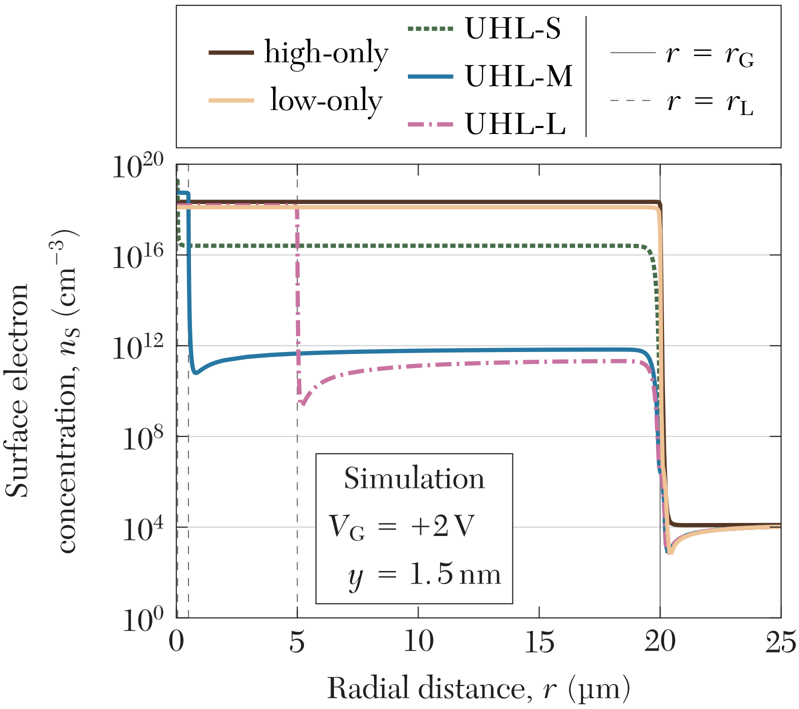


Figure 2–13. Radial n_S profiles in the simulated devices under $V_G = +2\,V$. Data extracted at $y = 1.5\,nm$.

carriers) in the depletion region.

Figure 2–14 shows simulated the ($n_S \cdot p_S$) vs. V_G plot (where p_S is the surface hole concentration) for the UHL-M and reference devices as a judgment of the devices' status away from equilibrium. In the high-only device, $n_S \cdot p_S = n_i^2$, implying equilibrium. Contrarily, scarcity of carriers ($n_S \cdot p_S \ll n_i^2$) is apparent in every other device as long as $V_G > 0$, no matter in the low or high region. This will prompt carrier generation in the entire depletion layer and lead to an A_G-dependent electron current. To make it more clear, **Figure 2–15** demonstrates the simulated cross-sectional electron quasi Fermi level splitting ($\Delta\phi_n$) map in the UHL-M device biased under $V_G = +2\,V$, as a direct clue for nonequilibrium. The high and low regions can be identified from the oxide profile in the figure. Not only the low region, but also the high region, undergo severe EQFL splitting, indicating the entire depletion region's being far below equilib-

doi:10.6342/NTU202304241

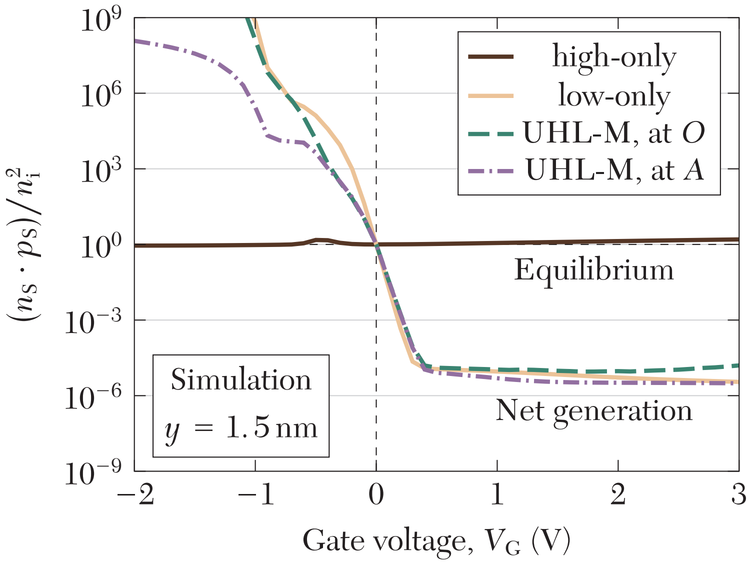


Figure 2–14. Plot of the product of surface electron and hole concentrations, $n_S \cdot p_S$ (normalized by n_i^2), vs. V_G, in the UHL-M device, at points O and A, and in the reference devices. Data extracted at $y = 1.5\,\mathrm{nm}$.

rium. **Figure 2–16** shows the simulated cross-sectional net generation rate (G) map in the UHL-M device biased under $V_G = +2\,\mathrm{V}$. Carrier generation is prominent under the entire gate area. It is noteworthy that the generation rate is not everywhere maximized in the depletion region. In silicon, thermal carrier generation is dominated by the Shockley-Read-Hall (SRH) Process associated with traps in the band gap [45]. For traps being recombination-generation centers (RGC) at the midgap (with trap energy $E_t = E_i$), the net generation rate is given by

$$G = \frac{n_i^2 - np}{\tau_{p0}(n + n_i) + \tau_{n0}(p + n_i)}, \tag{2.6}$$

where τ_{p0} and τ_{n0} are excess hole and electron lifetimes. With scarcity of both types of carriers ($n, p \ll n_i$), G reaches a maximum value $G_{\max} = n_i/(\tau_{n0} + \tau_{p0})$. However, it is evident from the mathematical form of (2.6) that G will be degraded at higher elec-

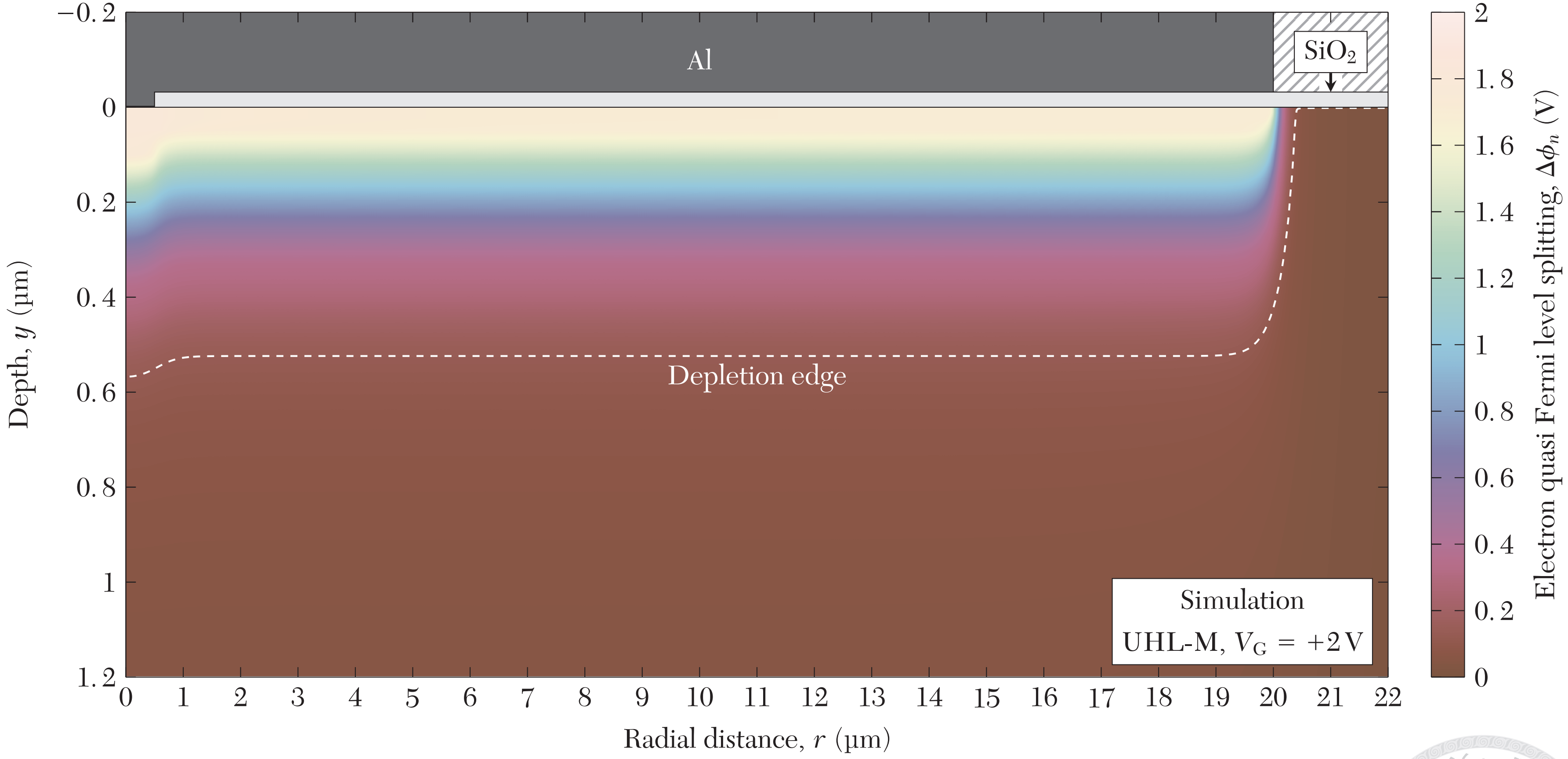


Figure 2–15. Simulated cross-sectional electron quasi Fermi level (EQFL) splitting ($\Delta\phi_n$) map in the UHL-M device with V_G = +2 V. The entire depletion region is far below equilibrium.

doi:10.6342/NTU202304241

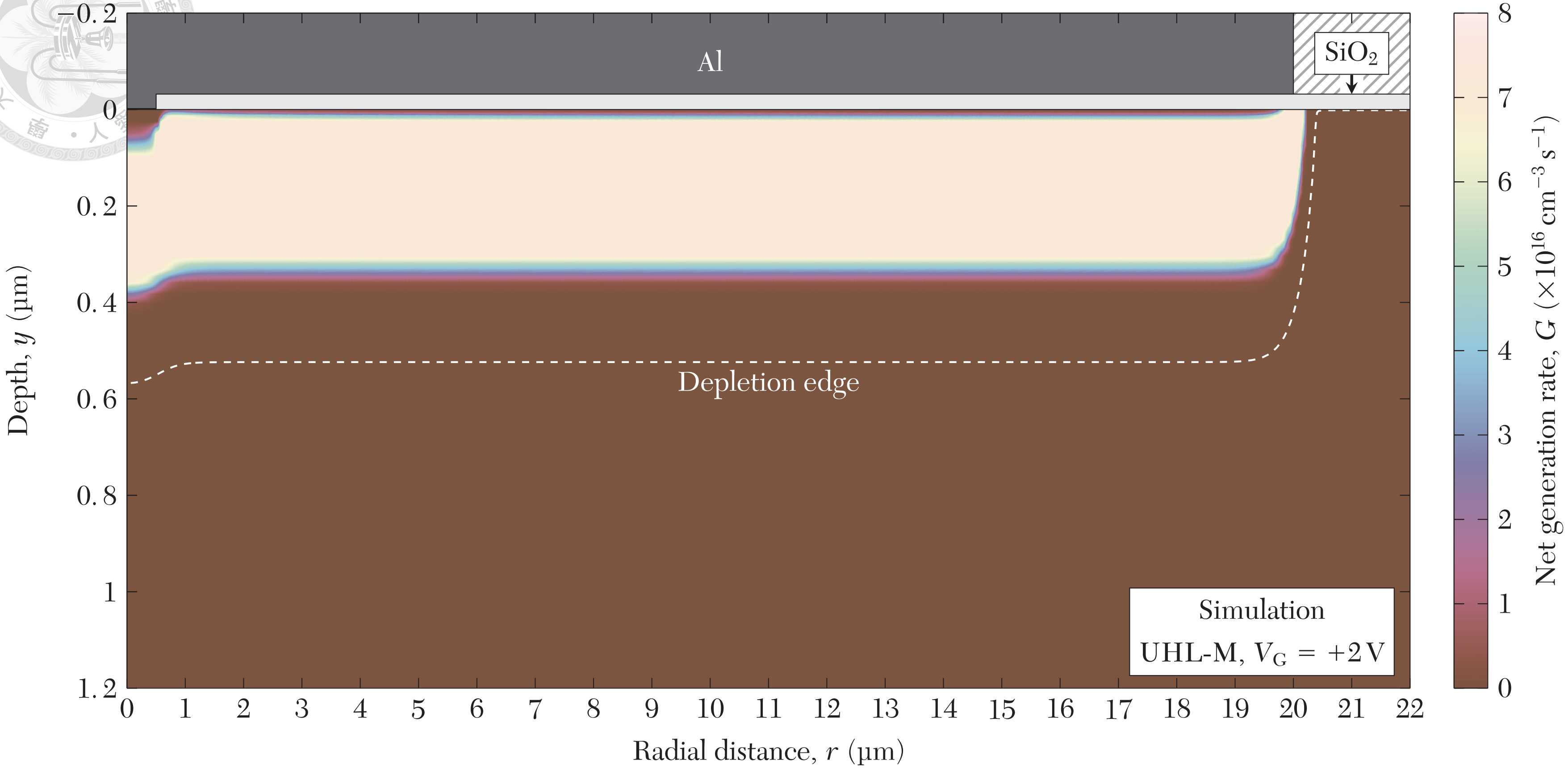


Figure 2–16. Simulated cross-sectional net generation rate (G) map in the UHL-M device with $V_G = +2\,\mathrm{V}$. (The depletion and generation edges differ by a depth $W_{g0} \approx 0.2\,\mu\mathrm{m}$.)

doi:10.6342/NTU202304241

tron or hole concentrations. At the surface, electrons are collected at the low region, lowering G, but only a very small volume proportion of the entire depletion layer is affected. However, near the depletion edge where electrostatic potential is relatively low compared to the surface, hole concentration is high enough to suppress G, in a manner that the generation edge retracts $\approx 0.2\,\mu\mathrm{m}$ away from the depletion edge. This length offset, referred to as the *generation-depletion width offset* (W_{g0}) is a crucial modeling parameter for the device electrostatics, and is to be carefully addressed in **Chapter 3** and **Chapter 4**. Detailed mathematical discussion and modeling on the magnitude of W_{g0} is conducted in **Appendix A**.

With n_{S} exhaustion in the high regions of UHL devices being verified by TCAD simulation results, its influence on the electrostatics under reverse bias is also to be examined. **Figure 2–17 (a)** shows the band diagrams for the UHL-M device (at point A) and the high-only device under a gate bias of $+3\,\mathrm{V}$. Even with identical oxide thickness ($d_{\mathrm{H}} = 320\,\text{Å}$) atop, the band diagram undergoes a dramatic distortion with the presence of the thin region in the device. V_{ox} is considerably lowered, while ψ_{S} drastically increases, in the UHL device. **TABLE 2–II** lists the extracted values of V_{ox}, ψ_{S} and W in the high regions of the simulated devices (high-only and UHL) under this gate bias. In the high-only device, ψ_{S} is pinned nearly at the threshold condition ($\psi_{\mathrm{S}} \approx 2\phi_{\mathrm{F}}$), and V_{ox} is high due to high Q_{S}. Contrariwise, once a low-region is introduced into the device, ψ_{S} rises with V_{G}, and so does W. **Figure 2–17 (b)** demonstrates the $\psi_{\mathrm{S}} - V_{\mathrm{G}}$ plots in the UHL-M device (at points O and A) and the reference devices, in correspondence with **Figure 2–12**. ψ_{S} saturation in the high-only device and DD in the low-only device is clearly observed in the plot. In the UHL-M device, ψ_{S} continues to

doi:10.6342/NTU202304241

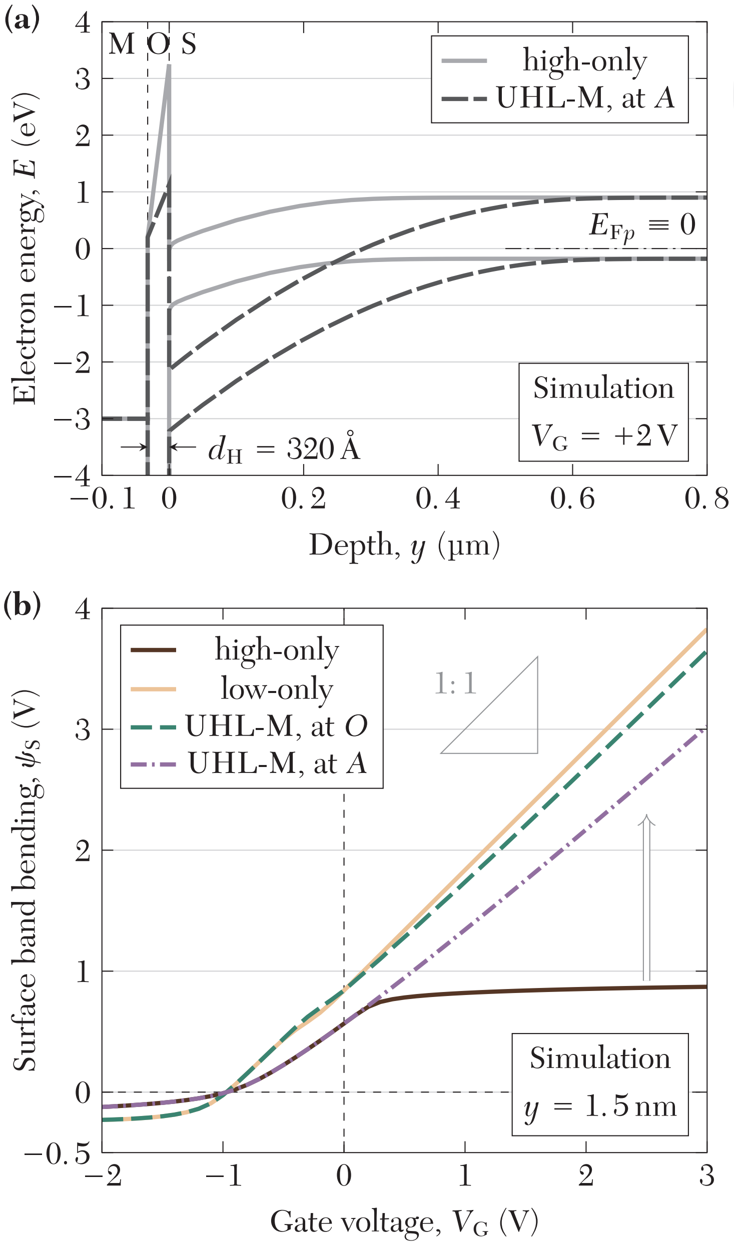


Figure 2–17. **(a)** Simulated high-region band diagrams for the UHL-M device (at the high region, point *A*) and the high-only device under $V_\mathrm{G} = +3\,\mathrm{V}$. **(b)** Simulated $\psi_\mathrm{S} - V_\mathrm{G}$ plots in the UHL-M device, at points *O* and *A*, and in the reference devices. Data extracted at $y = 1.5\,\mathrm{nm}$.

rise with V_G in both the high and low regions. Still, ψ_S at *A* is intrinsically lower than that at *O* under the same reverse bias. This may be attributed to the difference in oxide capacitance, making V_ox higher at *A* even though DD take place in both locations.

TABLE 2–II
Extracted values of oxide voltage drops, surface band bendings, and depletion widths in the high regions of the simulated devices biased under $V_{\mathrm{G}} = +3\,\mathrm{V}$. For each UHL device, extraction location is point A.

Device	V_{ox} (V)	ψ_{S} (V)	W (µm)[†]
High-only	3.054	0.916	0.315
UHL-S	0.935	3.035	0.620
UHL-M	0.931	3.039	0.620
UHL-L	0.931	3.039	0.620

[†] The depth where hole concentration is half of its neutral value, $p = \frac{1}{2}p_{\mathrm{p0}}$. Extracted from simulated hole concentration profiles.

2.3.3 Low Area Proportion and Oxide Thickness Effects

With fixed oxide thicknesses, UHL devices with three different low area proportions (**TABLE 2–I**) have been studied in the previous subsection. In **Figure 2–18**, by varying r_{L} in finer steps in the UHL structure, simulated $n_{\mathrm{S}} - r_{\mathrm{L}}$ and $\psi_{\mathrm{S}} - r_{\mathrm{L}}$ relations at probing points O and A are plotted, under a gate bias of $+3\,\mathrm{V}$. **Figure 2–18 (a)** shows the $n_{\mathrm{S}} - r_{\mathrm{L}}$ plots. As A_{L} goes from a large ($\sim 10^{-1} \cdot A_{\mathrm{G}}$) to medium size, n_{S} barely rises at both points, and DD is still prominent in the high region. However, there seems to be a critical A_{L} ($\sim 10^{-5} \cdot A_{\mathrm{G}}$) below which $n_{\mathrm{S}}(A)$ restores drastically to the high-only value. That is, deep depletion is mitigated below a critical area proportion. **Figure 2–18 (b)** shows the $\psi_{\mathrm{S}} - V_{\mathrm{G}}$ plots. Interestingly, $\psi_{\mathrm{S}}(A)$ appears to be saturated after the critical A_{L}, indicating the presence of a physical limit on $\psi_{\mathrm{S}}(A)$ that would be quantitatively studied in **Chapter 4**. On the other hand, as A_{L} shrinks from its largest value, $\psi_{\mathrm{S}}(O)$ decreases accordingly for all A_{L}, in correspondence with the n_{S} increment in **Figure 2–18 (a)**. This may be attributed to electron crowding at the low region, by the electrons that are exhausted from the high region. Hence, expectedly, as

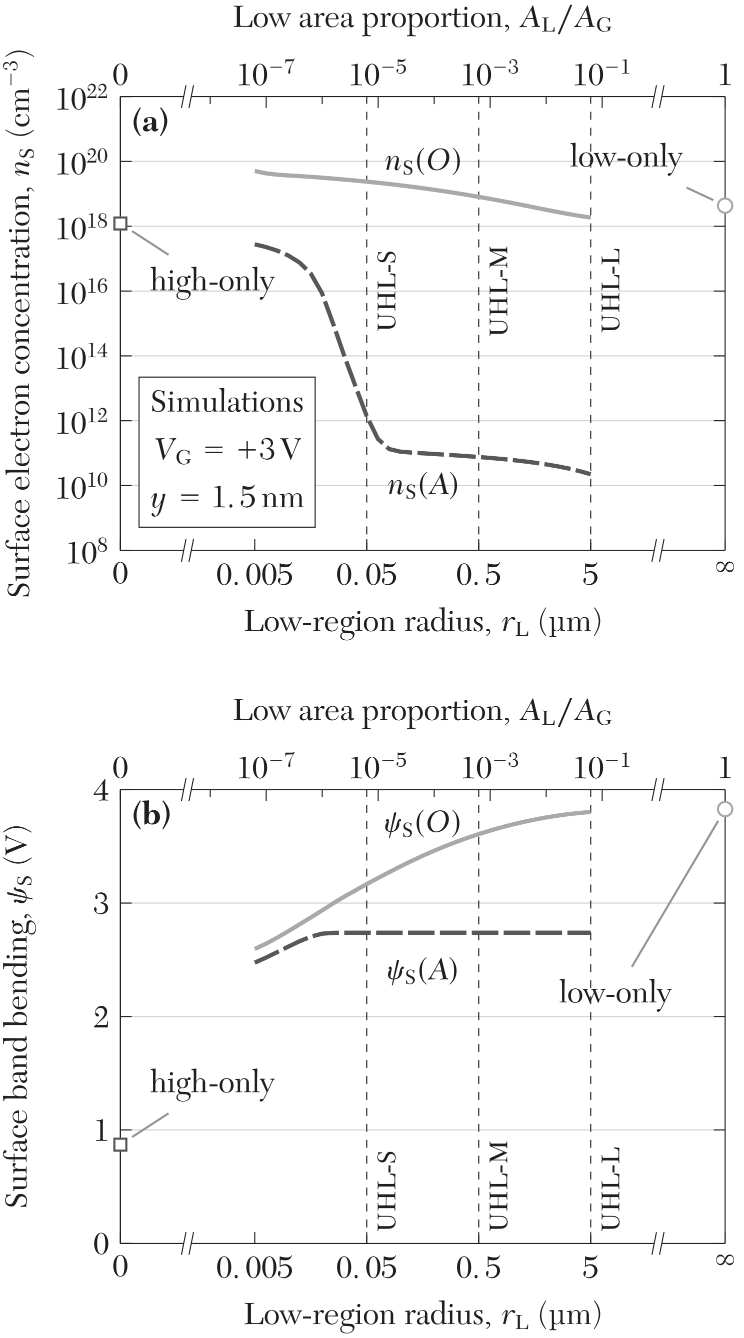


Figure 2–18. Simulated **(a)** n_S and **(b)** ψ_S vs. r_L plots in a collection of reverse-biased UHL devices with different r_L. Data extracted from points O and A at $y = 1.5\,\mathrm{nm}$. Gate bias was +3 V. The (A_L/A_G) axes corresponding to r_L are also shown.

A_L/A_G shrinks, electron crowding becomes more intense, making the low region to be less deep-depleted. Unfortunately, reliability issues are foreseeable in such local oxide

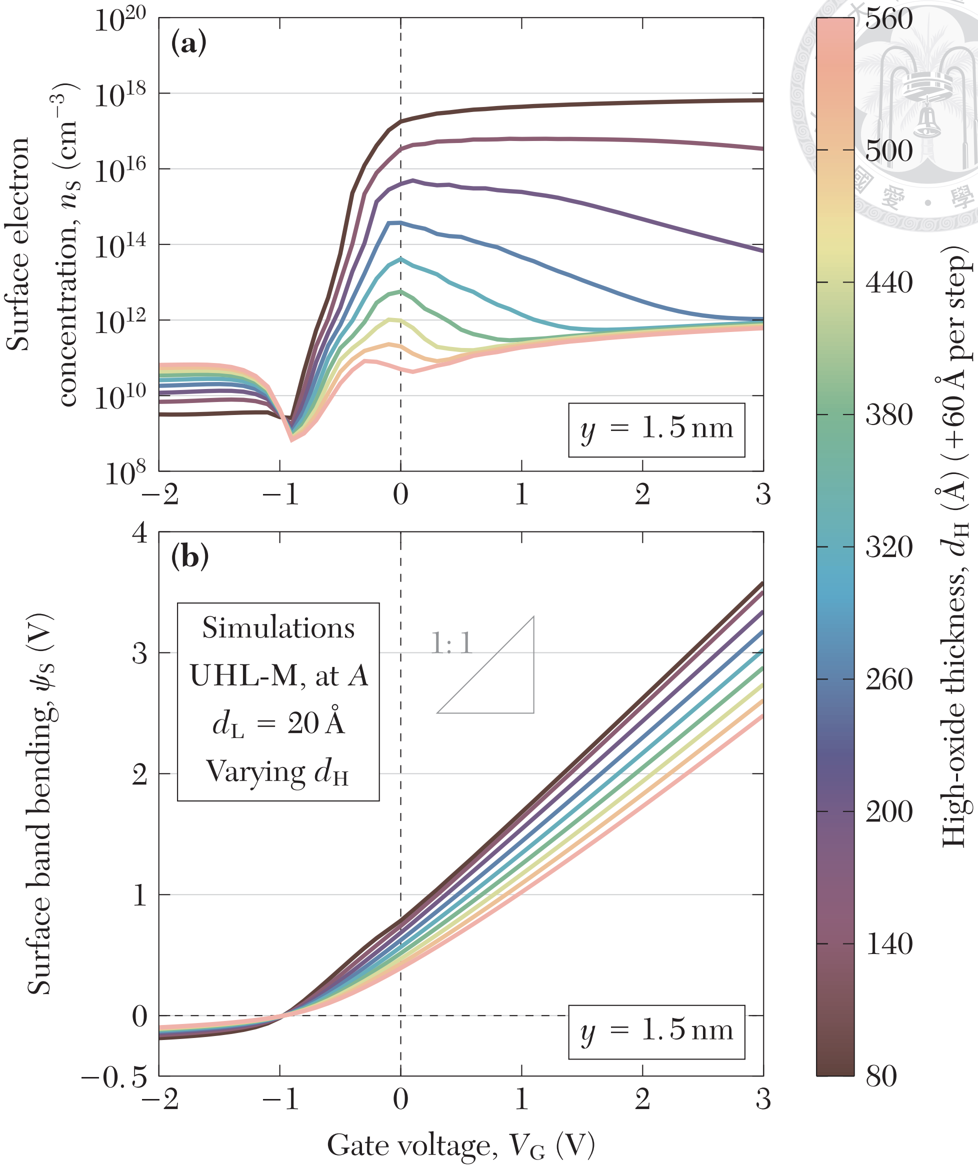


Figure 2–19. Simulated **(a)** $n_S - V_G$ and **(b)** $\psi_S - V_G$ plots at point A in UHL-M devices with $d_L = 20\,Å$ and varying d_H. Data extracted at $y = 1.5\,nm$.

thinning regions with intensified V_{ox} across such a thin d_L. In defective MOS capacitors, this can cause further breakdown events at the already weakened spot, eventually leading to catastrophic outcomes on oxide integrity.

Figure 2–19 (a, b) shows the simulated $n_S - V_G$ and $\psi_S - V_G$ plots, respectively, in devices with the UHL-M dimensions, except for varying d_H. Here d_L is fixed at 20 Å.

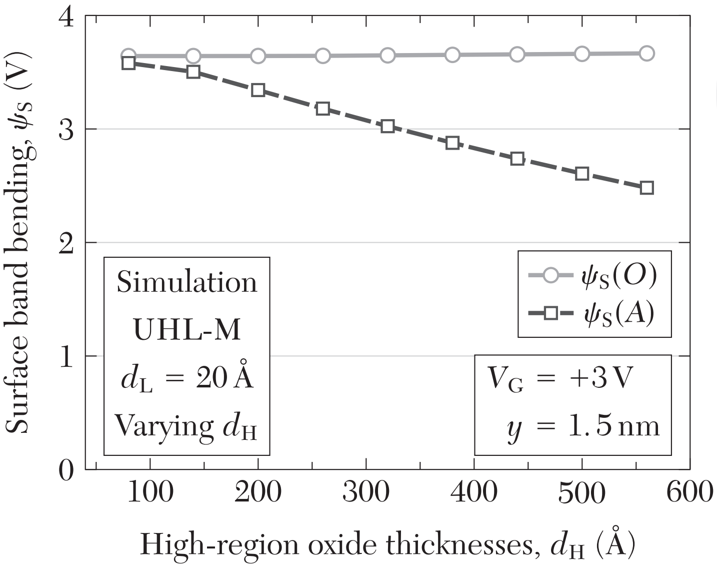


Figure 2–20. Simulated ψ_S vs. d_H plots at points O and A in UHL-M devices with $d_L = 20$ Å and varying d_H, under $V_G = +3$ V. Data extracted at $y = 1.5$ nm.

With the smallest values of d_{ox}, the presence of DD is hard to tell from the $n_S - V_G$ plots. However, The $\psi_S - V_G$ characteristics prove DD to have taken place for every d_H. Therefore, the greater n_S associated with thinner d_H may simply be attributed to higher C_{oxH}, and not to the absence of DD. Still, reduction of n_S can be achieved by increasing d_H. **Figure 2–20** shows, at a fixed V_G of +3 V, the extracted ψ_S values at points O and A from **Figure 2–19 (b)** as functions of V_G. Increasing d_H poses little impact on $\psi_S(O)$ but reduces ψ_{SH}, which may also be attributed to the change in C_{oxH}.

Figure 2–21 demonstrates the simulated $n_S - V_G$ plots at points O and A in devices with the UHL-M dimensions, but with varying d_L. Here d_H is fixed at 320 Å. With thicker d_L, n_S is noticeably restored in the high region. **Figure 2–22** shows the corresponding $\psi_S - V_G$ plots. One may find that DD still occurs, to some extent, in all the devices. Nonetheless, devices with thicker d_L may first undergo a regime where inversion charge accumulates (**Figure 2–21**) and ψ_S is pinned like an ordinary MOS capacitor.

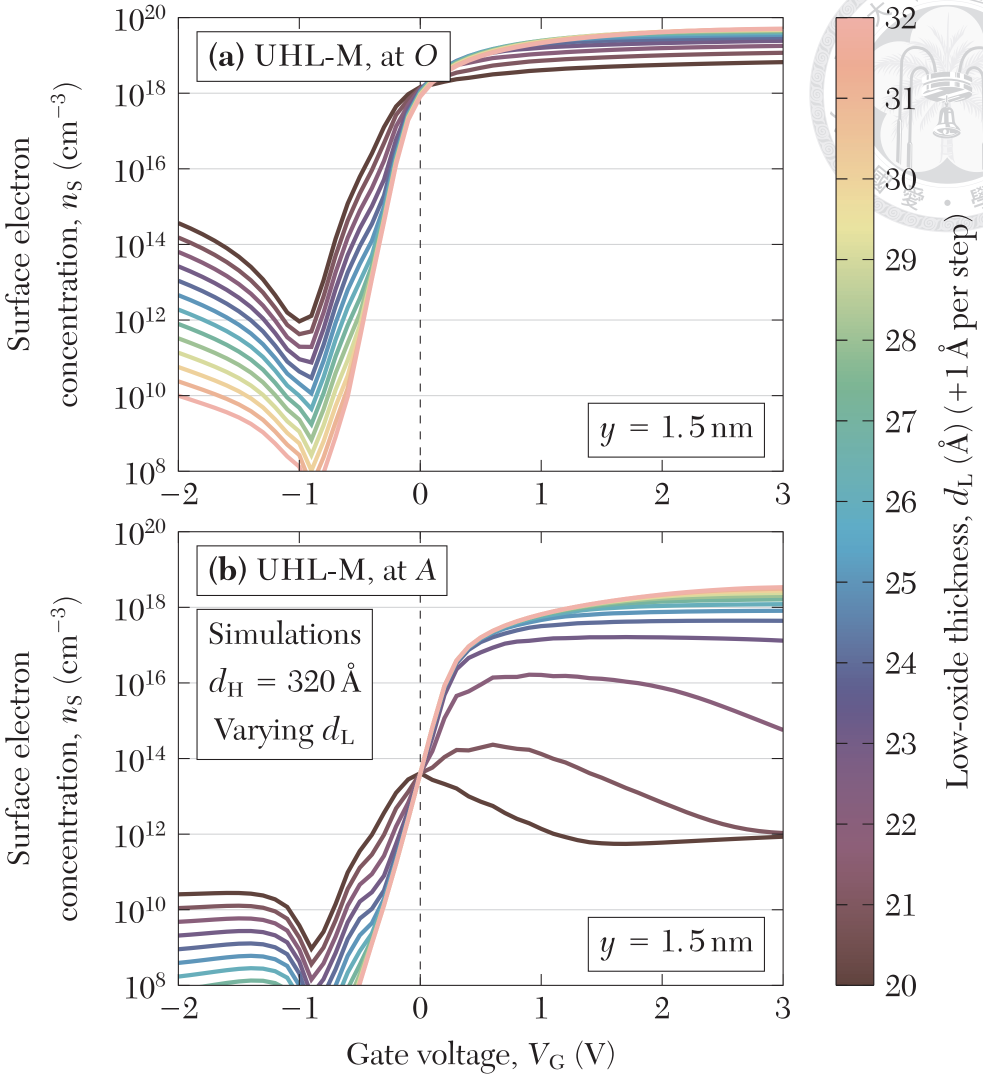


Figure 2–21. Simulated $n_S - V_G$ plots at **(a)** point O and point A, in UHL-M devices with $d_H = 320\,\text{Å}$ and varying d_L. Data extracted at $y = 1.5\,\mathrm{nm}$.

DD only takes over above some critical voltage V_C, and this voltage increases with d_L. In fact, this *two-stage electrostatics model* under the inversion regime is also observed in planar MIS TDs over a wide range of d_{ox}, but with presumably different $V_C - d_{ox}$ relations from the UHL devices. Modeling of this behavior is achieved in **Chapter 3** for planar MIS TDs and in **Chapter 4** for UHL structures. Most importantly, the critical voltages appear to be identical in the high and low regions within a UHL device.

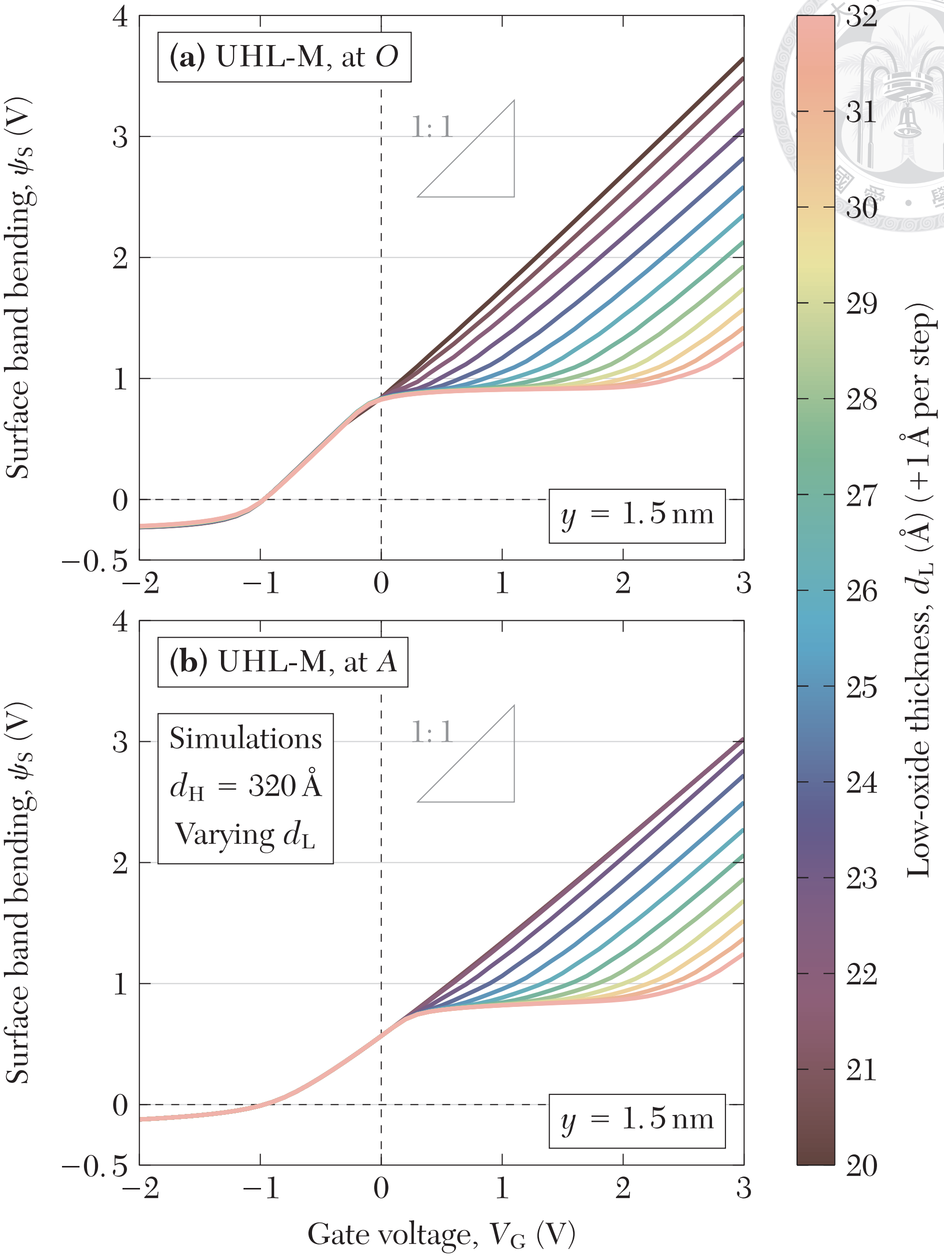


Figure 2–22. Simulated $\psi_\mathrm{S} - V_\mathrm{G}$ plots at **(a)** point O and point A, in UHL-M devices with $d_\mathrm{H} = 320\,\text{Å}$ and varying d_L. Data extracted at $y = 1.5\,\mathrm{nm}$.

In other words, once DD is triggered in the low region, the high region is also driven to DD. **Figure 2–23** shows the $\psi_\mathrm{S} - d_\mathrm{L}$ relations extracted from **Figure 2–22** under $V_\mathrm{G} = +3\,\mathrm{V}$. At smaller d_L, $\psi_\mathrm{S}(A)$ is clamped by the aforementioned physical limit. Otherwise, they share nearly identical magnitudes and both decrease monotonically with

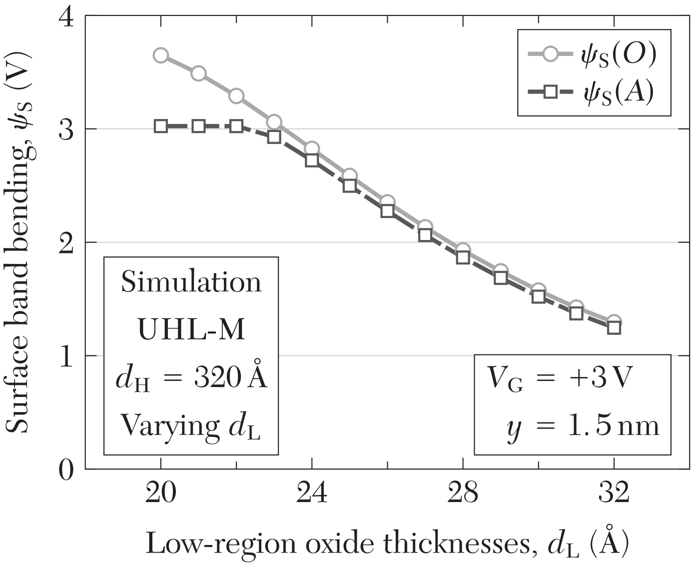


Figure 2–23. Simulated ψ_S vs. d_L plot at points O and A in the UHL-M device with $d_H = 320\,\text{Å}$ and varying d_L, under $V_G = +3\,\text{V}$. Data extracted at $y = 1.5\,\text{nm}$.

d_L.

2.3.4 Prominence of Deep Depletion in Aggressively-Scaled High-Low Devices

As a final remark, **Figure 2–24** shows the simulated band banding maps in aggressively-scaled devices ($r_G = 50\,\text{nm}$) under $V_G = +3\,\text{V}$, with or without the presence of a low region ($r_L = 2\,\text{nm}$). Oxide thicknesses are 80 Å and 20 Å. The high-low device, with $A_L/A_G = 1.6 \times 10^{-3}$, still undergoes severe DD in such a small scale. The electrostatic potential can be up to 2.1 V higher from the high-only device.

2.4 Summary

In this chapter, $I - V$ and $C - V$ measurements on experimental UHL devices with A_L/A_G ranging from 4.4×10^{-3} to 0.87 were performed. Under forward bias, current and capacitance values exhibit proportionality to A_L/A_G. However, regardless of A_L,

doi:10.6342/NTU202304241

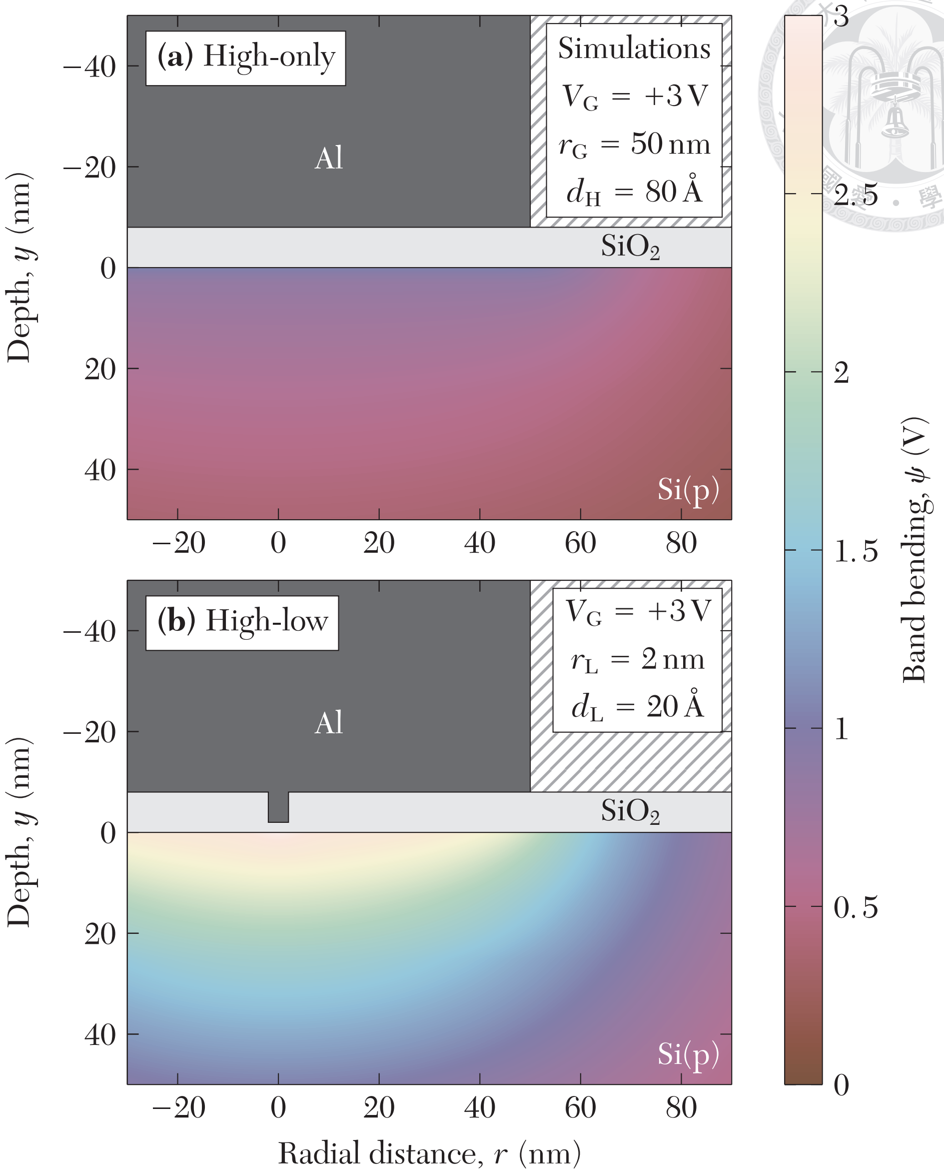


Figure 2–24. Simulated cross-sectional band banding maps in aggressively-scaled cylindrical MOS devices under $V_G = +3\,\mathrm{V}$. **(a)** A high-only device with $r_G = 50\,\mathrm{nm}$ and $d_H = 80\,\text{Å}$. **(b)** A high-low device with the same r_G and d_H, and with $r_L = 2\,\mathrm{nm}$, $d_L = 20\,\text{Å}$. Here $A_L/A_G = 1.6 \times 10^{-3}$.

the reverse saturation currents for all UHL devices are as high as that of the low-only device, and DD phenomena are equally severe among all the devices, except for the high-only device. Both observations were ascribed to inversion charge coupling between the high and low regions, in a manner that DD must be jointly triggered in both regions under reverse bias. In turn, extensive DD prompts carrier generation under the entire gate area, leading to A_G-dependent total electron current. TCAD simulations reveal the presence of a radial electric field near the high-low boundary, which may be responsible for the highly-effective charge coupling. What's more, the exhaustion of inversion charge and the deformation of band diagrams ($V_{ox}\downarrow$, $\psi_S\uparrow$) in the UHL devices were verified by TCAD simulation. Quasi Fermi level splitting and extensive carrier generation under the high region were also directly observed from simulated results. By varying the device dimensions, it is found that device DD is mitigated when A_L/A_G is sufficiently small, and increasing the d_L will shift device V_C towards the positive direction. On the other hand, varying the d_H poses little effects on the electrostatics. However, it was also discovered that electron crowding at the low region of small-A_L devices may lead to high field across the thin oxide and provoke reliability issues. Last but not least, the prominence of DD in an aggressively-scaled high-low device was demonstrated. That is, LOT is still highly influential on device electrostatics even at such a small device size.

doi:10.6342/NTU202304241

doi:10.6342/NTU202304241

3

Electrostatics Theory and Modeling I: Planar MIS(p) Tunnel Diodes Under Reverse Bias

3.1 Background

The high sensitivity of MIS TD reverse saturation current (I_{sat}) on its electrostatics due to SBHME is the basis for its sensor applications (**Section 1.2**). Thus, it is essential to model MIS TD electrostatics under reverse bias for any further quantitative study on MIS TD sensors. Moreover, the planar MIS TD electrostatics model, if established, can serve as the cornerstone for establishing theories in structurally more complex, non-planar devices, such as the UHL devices.

The MOS electrostatics theory is about finding the distribution of V_G over the oxide layer and the semiconductor substrate; that is, about finding V_{ox} and ψ_S under a certain bias (1.1). There is the classical MOS electrostatics theory (CEST) for non-tunneling oxides [24] that defines 3 regimes (accumulation, depletion, inversion) according to the gate bias. CEST has become the foundation of MOSFET current models and, as a result, one of the kernel theories in CMOS ULSI circuit design. However, the collapse of CEST have been demonstrated for ultrathin oxides (**Figure 3–1** ; also see **Section 1.2** and **Figure 1–8**) due to the erroneous quantification of Q_S by not considering the leakage of the inversion charge via gate tunneling. In other words, the electron tunneling current under reverse bias must be carefully addressed and incorporated into the electrostatics theory to construct a self-consistent electrostatics model.

Historically, following the oxide thickness scaling in VLSI circuits [18], several tunneling current models for the inversion charge have been consecutively proposed to examine its impact on power dissipation and circuit functionality and reliability [46]. First, there was the Tsu-Esaki model [47] proposed by Duke (1969) [48] that have been considered highly-accurate. Yet, it involves an integral over energy along the tunneling barrier, making numerical integration inevitable. Next, the Fowler-Nordheim tunneling (FNT) model for triangular barriers [49, 50] and its direct-tunneling variant for trapezoidal barriers [51] are in closed forms for sufficiently large gate bias, but are

doi:10.6342/NTU202304241

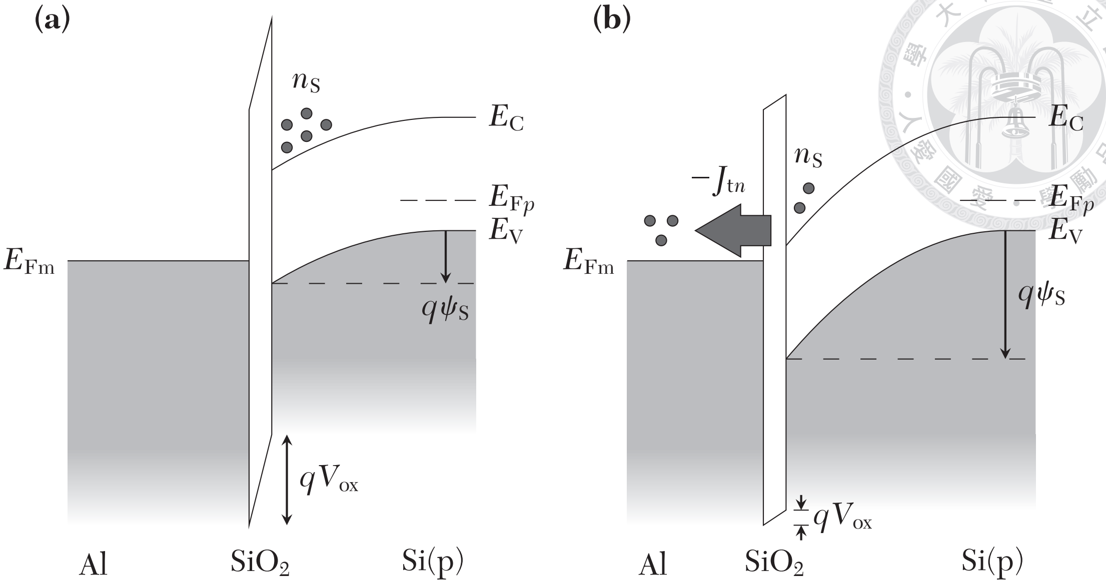


Figure 3–1. **(a)** MIS TD electrostatics under reverse bias described by the CEST, not accounting gate leakage effects, yielding nonphysical results. **(b)** Electrostatics by accounting gate tunnel current (J_{tn}), resulting in QFL splitting, smaller V_{ox}, and greater ψ_S, compared to **(a)**.

not compatible with small V_G. Last but not least, the surface potential-based compact model (*SP* model) [52] is a modified Tsu-Esaki model in which, by assuming equal energy for the incident electrons, the integral in the Tsu-Esaki model is eliminated without compromising model accuracy. With its closed form and compatibility to a wide bias range, the *SP* model is adopted as the electron tunnel current model in this chapter.

It should be noted that the compactness of the *SP* model relies on a known surface potential. In MOSFETs with source/drain (S/D) doping, with abundant inversion charge supply from the S/D region, the surface potential is well-defined [18, 53]. Notwithstanding, in MIS TDs, inversion charge is mainly supplied by the thermal generation in the space charge layer [24], a relatively slow process. Chances are that the shortage of inversion charge, as a consequence of high tunneling rate, perturbs the surface potential itself from the CEST value, in reverse-biased MIS TDs. This makes the determination of J_{tn} and device electrostatics a pair of coupled problems.

Oxide thickness also has a great influence on device electrostatics for its domination on the tunneling rate. Prior studies suggest that, under appropriate d_{ox}, there exists a critical voltage V_{C} that demarcates the electrostatics into two regimes [34] (like **Figure 2–22** for UHL devices, but in planar devices.) DD takes place only when $V_{\mathrm{G}} > V_{\mathrm{C}}$. Otherwise, the leakage current is so small that CEST appears to be valid to some extent under the pre-V_{C} regime. The dissimilar electrostatics around the critical point can clearly be observed from experimental $C - V$ characteristics [34], as well as by TCAD simulations. Unfortunately, there have been little quantitative studies on the magnitude of V_{C} subject to d_{ox}, not to mention the $\psi_{\mathrm{S}} - V_{\mathrm{C}}$ and $V_{\mathrm{ox}} - V_{\mathrm{C}}$ relations at a certain d_{ox}.

Therefore, in this chapter, we proposed an comprehensive analytical model for quantitatively evaluating the MIS TD electrostatics under a given d_{ox} with no differentials, integrals or mesh-based numerical methods involved. Furthermore, we formulated a closed-form approximation for the $V_{\mathrm{C}} - d_{\mathrm{ox}}$ relation that provides deeper physical interpretations. The model have been validated with experimental and TCAD simulation results.

3.2 The Model and Procedure for Evaluation

Figure 3–2 depicts the band diagram of a reverse-biased MIS TD and the current components, in which $q\Phi_{\mathrm{b}} = q(\chi_{\mathrm{s}} - \chi_{\mathrm{ox}}) = 3.2\,\mathrm{eV}$ is the conduction band edge (CBE) discontinuity at the insulator-semiconductor interface, E_{C} and E_{V} are the conduction and valence band energies, respectively, E_{Fm} is the Fermi level in metal, $E_{\mathrm{F}p}$ and $E_{\mathrm{F}n}$

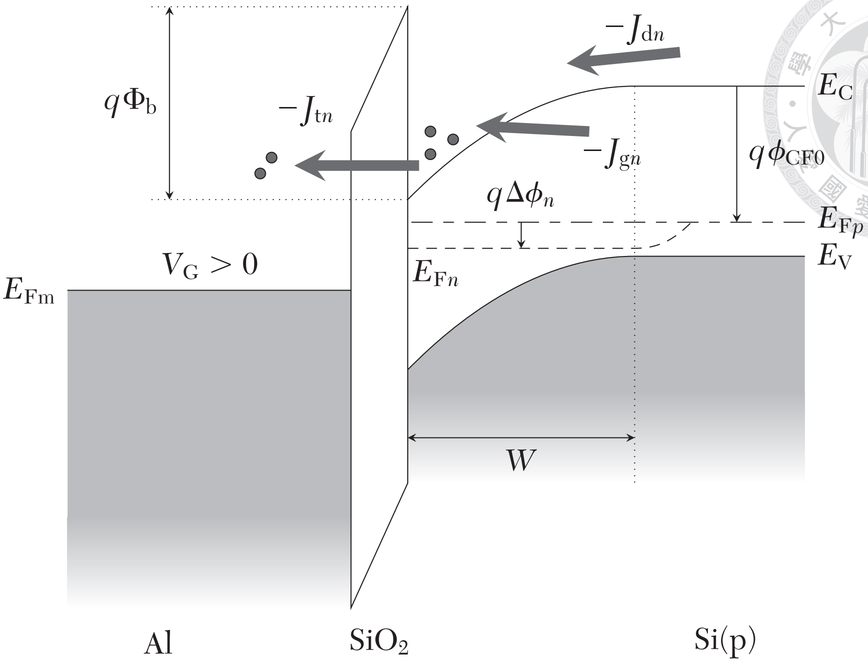


Figure 3–2. Minority carrier current components in a reverse-biased MIS TD.

are the electron and hole quasi Fermi levels (QFLs) in the semiconductor, respectively, and

$$\begin{aligned}\phi_{CF0} &\equiv \frac{E_C(\infty) - E_{Fp}(\infty)}{q} \\ &= \frac{E_g}{2q} + \phi_F \end{aligned} \tag{3.1}$$

is the electrostatic potential difference from semiconductor bulk E_{Fp} to bulk E_C.

The following assumptions have been made in the model:

1. Quantum confinement (QC) effects [54] are neglected.
2. Image charge lowering effect [19] is neglected.
3. A constant E_{Fp} throughout the semiconductor is assumed.
4. The carrier lifetimes are sufficiently long such that the electron QFL is "flat" in the space charge region [55]. Detailed discussions on the "flatness" and the

doi:10.6342/NTU202304241

assumption's validity can be found in **Appendix A**.

5. Doping concentration is moderate ($N_\mathrm{A} \leq 10^{17}\,\mathrm{cm}^{-3}$) such that band-to-band tunneling (BTBT) in the substrate can be neglected.

6. MIS TD electrostatics under reverse bias is governed solely by electrons. This is assumed to hold even if hole tunnel current is present. This assumption is based on the following facts:

    (a) Holes that tunnel from the gate to the substrate, constituting the hole current of the device, will be readily drifted to the neutral region by the semiconductor field, making their influence on the electrostatics negligible.

    (b) Electrons, on the other hand, can accumulate at the semiconductor surface and are capable of modulating the band profile.

    That is, the hole current, even if present, is passively modulated by the electrostatics (e.g., $q\phi^{*}_{\mathrm{Bp}}$) but has no control over it. Electrons have the control.

7. There is no minority carrier diffusion from the back contact [24].

8. There is no oxide charge. ($Q_\mathrm{eff} = 0$)

9. There are no interface states. ($D_\mathrm{it} = 0$)

10. Carrier generation in the depletion region is subject to the Shockley-Read-Hall process.

doi:10.6342/NTU202304241

3.2.1 Electron Current Components

The steady-state solution to the electrostatics must be subject to balanced gate and substrate electron currents:

$$J_{\mathrm{t}n} = J_{\mathrm{d}n} + J_{\mathrm{g}n}. \tag{3.2}$$

The *SP* model, by assigning the lowest possible energy (that at the bottom of the triangular energy well at the insulator-semiconductor interface) to every incident electron in the Tsu-Esaki model, reads [52]

$$J_{\mathrm{t}n} = J_{\mathrm{t}n0} \cdot P_{\mathrm{t}} \cdot S, \tag{3.3}$$

where P_{t} is the tunneling probability across the oxide for incident electrons at the bottom of the well, S is the supply function in the Tsu-Esaki model evaluated at the same energy level, and

$$J_{\mathrm{t}n0} \equiv \frac{qm^{*}k_{\mathrm{B}}^{2}T^{2}}{2\pi^{2}\hbar^{3}} \tag{3.4}$$

is a material constant with a dimension of current density, in which $m^{*} = 0.92m_{0}$ is the longitudinal electron effective mass in silicon (m_{0}: electron mass) [18], and $\hbar$ is the reduced Planck constant. The model involves no integral and conveniently separate the P_{t} and S portions. P_{t} can be formulated by the Wentzel-Kramers-Brillouin (WKB) approximation. For triangular barriers ($V_{\mathrm{ox}} \geq \Phi_{\mathrm{b}}$), or better known as the FNT criterion, P_{t} picks up the form [56]

$$\text{(Triangular)} \qquad P_{\mathrm{t}} = \exp\left(-\frac{B}{\mathscr{E}_{\mathrm{ox}}}\right), \tag{3.5}$$

doi:10.6342/NTU202304241

where $\mathscr{E}_{\mathrm{ox}} = V_{\mathrm{ox}}/d_{\mathrm{ox}}$ is the oxide electric field,

$$B \equiv \frac{4\sqrt{2m_{\mathrm{ox}}}(q\Phi_{\mathrm{b}})^{3/2}}{3\hbar q} \tag{3.6}$$

is a material-system constant with the dimension of electric field, quoted from the FNT model [50], and m_{ox} is the oxide tunneling effective mass, measured as $(0.34 \pm 0.04)m_0$ in the FNT region and $(0.29 \pm 0.02)m_0$ in the DT region. For simplicity, we adopt $m_{\mathrm{ox}} = 0.32m_0$ throughout this work. Likewise, for trapezoidal barriers ($V_{\mathrm{ox}} \leq \Phi_{\mathrm{b}}$) or the DT criterion,

$$\text{(Trapezoidal)} \qquad P_{\mathrm{t}} = \exp\left(-\frac{B\left[1-(1-V_{\mathrm{ox}}/\Phi_{\mathrm{b}})^{3/2}\right]}{\mathscr{E}_{\mathrm{ox}}}\right). \tag{3.7}$$

As a side note, for extremely small V_{ox} such that the barrier is nearly rectangular, P_{t} can be evaluated by taking a limit on (3.7):

$$\begin{aligned} \text{(Rectangular)} \qquad P_{\mathrm{t}} &= \lim_{V_{\mathrm{ox}}\to 0} \exp\left(-\frac{B\left[1-(1-V_{\mathrm{ox}}/\Phi_{\mathrm{b}})^{3/2}\right]}{\mathscr{E}_{\mathrm{ox}}}\right) \\ &= \exp\left(-\frac{3Bd_{\mathrm{ox}}}{2\Phi_{\mathrm{b}}}\right). \end{aligned} \tag{3.8}$$

Moreover, at the transition point from DT to FNT ($V_{\mathrm{ox}} = \Phi_{\mathrm{b}}$), (3.5) and (3.7) yield, in unison,

$$(V_{\mathrm{ox}} = \Phi_{\mathrm{b}}) \qquad P_{\mathrm{t}} = \exp\left(-\frac{Bd_{\mathrm{ox}}}{\Phi_{\mathrm{b}}}\right). \tag{3.9}$$

The supply function is determined by the electron quasi Fermi levels on both sides of the oxide, and is given by [52]

$$\begin{aligned} S &= \ln\left(\frac{1+\exp\left(-\dfrac{E_{\mathrm{CS}}-E_{\mathrm{F}n}}{k_{\mathrm{B}}T}\right)}{1+\exp\left(-\dfrac{E_{\mathrm{CS}}-E_{\mathrm{Fm}}}{k_{\mathrm{B}}T}\right)}\right) \\ &= \ln\left(\frac{1+\exp\left(-\dfrac{\phi_{\mathrm{CF0}}-\psi_{\mathrm{S}}+\Delta\phi_n}{\phi_{\mathrm{t}}}\right)}{1+\exp\left(-\dfrac{\phi_{\mathrm{CF0}}-\psi_{\mathrm{S}}+V_{\mathrm{G}}}{\phi_{\mathrm{t}}}\right)}\right), \end{aligned} \tag{3.10}$$

where $E_{\mathrm{CS}} = E_{\mathrm{C}}(\infty) - q\psi_{\mathrm{S}}$ is E_{C} at the surface.

Meanwhile, the diffusion current at the depletion edge in terms of minority carrier QFL splitting is quoted from P-N junction theories [19], which reads

$$J_{\mathrm{d}n} = \frac{qD_n n_{\mathrm{i}}^2}{L_n N_{\mathrm{A}}}\left(1 - e^{-\Delta\phi_n/\phi_{\mathrm{t}}}\right), \tag{3.11}$$

where D_n is the diffusion coefficient for electrons, and L_n is their diffusion length. On the other hand, the generation current can be obtained by integrating the net generation rate G over the depth:

$$J_{\mathrm{g}n} = q\int_0^W G\,\mathrm{d}y, \tag{3.12}$$

where G is given by the Shockley-Read-Hall model:

$$G = \frac{n_{\mathrm{i}}^2 - np}{\tau_{p0}(n+n_{\mathrm{i}}) + \tau_{n0}(p+n_{\mathrm{i}})}. \tag{from 2.6}$$

A common simplification on G is by letting low carrier concentrations throughout the depletion region [57]:

$$(n,\, p \ll n_\mathrm{i}) \qquad \begin{aligned} G &\approx \frac{n_\mathrm{i}^2 - np}{2n_\mathrm{i}\tau_0} \\ &= \frac{n_\mathrm{i}}{2\tau_0}\left(1 - e^{-\Delta\phi_n/\phi_\mathrm{t}}\right), \end{aligned} \tag{3.13}$$

where (3.12) evaluates to

$$(n,\, p \ll n_\mathrm{i}) \qquad J_{\mathrm{g}n} = \frac{qn_\mathrm{i}W}{2\tau_0}\left(1 - e^{-\Delta\phi_n/\phi_\mathrm{t}}\right), \tag{3.14}$$

which is known as the Sah-Noyce-Shockley model under reverse bias for high $\Delta\phi_n$ ($J_{\mathrm{g}n} \approx qn_\mathrm{i}W/2\tau_0$) [57]. However, in fact, the negligibility of p near the depletion edge is not valid, leading to overestimated G and $J_{\mathrm{g}n}$ from this approximation. Another approximation without the need for letting $p \ll n_\mathrm{i}$ in simplifying G has been proposed in **Appendix A**, which results in

$$(n \ll n_\mathrm{i}) \qquad J_{\mathrm{g}n} = \frac{qn_\mathrm{i}}{2\tau_0}\left(W - W_\mathrm{g0}\right)\left(1 - e^{-\Delta\phi_n/\phi_\mathrm{t}}\right). \tag{from A.53}$$

This differs from (3.14) by a term W_g0, the *generation-depletion width offset*:

$$\begin{aligned} W_\mathrm{g0} &= \sqrt{\frac{\pi}{2}}\left[-\operatorname{Li}_{1/2}\left(-\frac{N_\mathrm{A}\tau_{n0}}{2n_\mathrm{i}\tau_0}\right)\right]\lambda_\mathrm{p} \\ &\approx \left(2\ln\frac{N_\mathrm{A}\tau_{n0}}{2n_\mathrm{i}\tau_0}\right)^{1/2}\lambda_\mathrm{p}, \end{aligned} \tag{from A.54}$$

where λ_p is the extrinsic Debye length in p-type semiconductor, and $\operatorname{Li}_{1/2}$ is the polylogarithm of order 1/2. For $N_\mathrm{A} = 10^{16}\,\mathrm{cm}^{-3}$ and $\tau_{p0} = \tau_{n0}$, W_g0 is evaluated as 208 nm.

Finally, the depletion approximation [24] for W may be useful for evaluating J_{gn}:

$$W = \sqrt{\frac{2\epsilon_s \psi_S}{q N_A}}. \tag{3.15}$$

3.2.2 Potentials and Surface Charge

The distribution of V_G over V_{ox} and ψ_S is subject to

$$\begin{aligned} V_G &= V_{FB} + V_{ox} + \psi_S \\ &= V_{FB} - \frac{Q_S}{C_{ox}} + \psi_S. && \text{(from 1.1)} \end{aligned}$$

For evaluating the surface charge subject to electron QFL splitting, the three-terminal MOS electrostatics model [53] has been quoted. For any positive ψ_S greater than several ϕ_t:

$$(\psi_S \geq 3\phi_t) \qquad Q_S \approx -\sqrt{2q\epsilon_s N_A}\sqrt{\psi_S + \phi_t \exp\left[\frac{\psi_S - (2\phi_F + \Delta\phi_n)}{\phi_t}\right]}. \tag{3.16}$$

3.2.3 Procedure for Evaluation

One of the ultimate goals for establishing the quantitative electrostatics model is to obtain the $\psi_S - V_G$ relation under a given oxide thickness. **Figure 3–3** shows the flowchart that summarizes all the aforementioned variables and their dependencies. In this flowchart, simply put, one finds that J_{dn}, J_{gn} and J_{tn} are sequentially determined from specifying d_{ox}, ψ_S and $\Delta\phi_n$. The current balance equation (3.2) is the only constraint in the system. Therefore, we may input arbitrary values of d_{ox} and ψ_S, combined with an adaptively-chosen $\Delta\phi_n$ that satisfies the constraint, to solve for every other vari-

doi:10.6342/NTU202304241

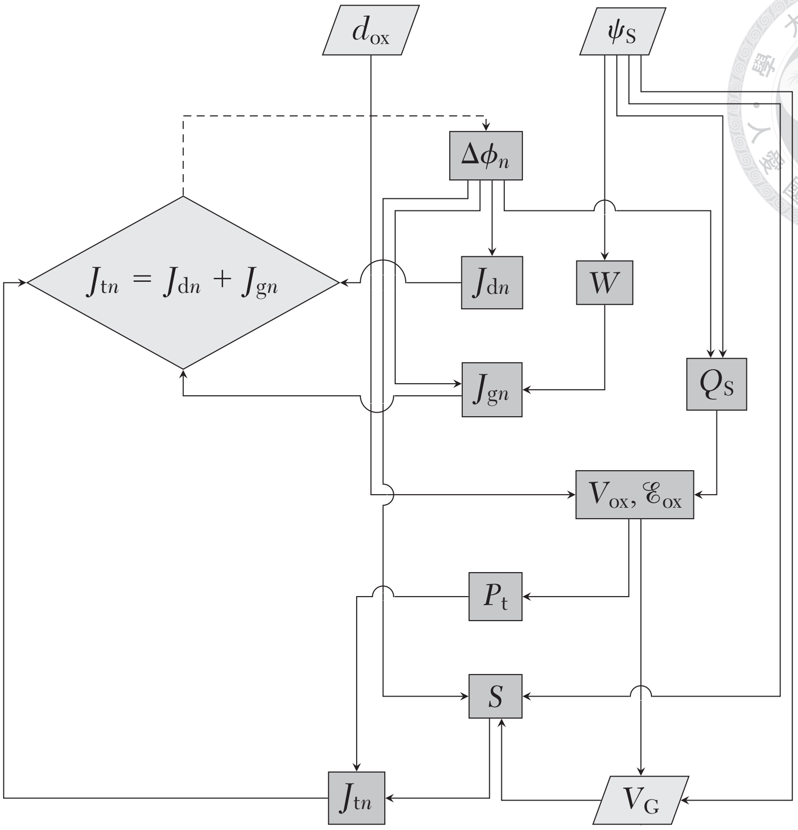


Figure 3–3. Flowchart for all the physical variables involved in the evaluation process and their dependencies.

able, especially for V_G. In other words, via fixing d_ox, the $V_\mathrm{G} - \psi_\mathrm{S}$ is acquired through the evaluation process. Finally, by taking an inverse, the $\psi_\mathrm{S} - V_\mathrm{G}$ relation of our interest is obtained.

TABLE 3–I lists the adopted values of constant physical quantities necessary for the evaluation. **TABLE 3–II** lists the values of secondary constant physical quantities that can be computed from the values given by **TABLE 3–I**. **TABLE 3–III** lists all physical variables and their governing equations.

TABLE 3–I
List of adopted values of constant physical quantities.

Quantity	Unit	Value
D_n	$\mathrm{cm^2/s}$	36
E_g	eV	1.12
L_n	cm	0.03
m^*/m_0	–	0.92 [18]
m_ox/m_0	–	0.32 [58]
N_A	$\mathrm{cm^{-3}}$	1×10^{16}
n_i	$\mathrm{cm^{-3}}$	1.5×10^{10}
T	K	300
V_FB	V	-0.9
W_g0	nm	208 **TABLE A–I**
$\epsilon_\mathrm{ox}/\epsilon_0$	–	3.9
$\epsilon_\mathrm{s}/\epsilon_0$	–	11.9
τ_0	s	1×10^{-7}
Φ_b	V	3.2

TABLE 3–II
List of secondary constant physical quantities that can be computed from the values given by **TABLE 3–I**.

Quantity	Unit	Value	Definition
B	V/cm	2.2120×10^8	(3.6)
$J_{\mathrm{t}n0}$	$\mathrm{A/cm^2}$	9.9504×10^6	(3.4)
ϕ_F	V	0.3473	$= \phi_\mathrm{t} \ln(N_\mathrm{A}/n_\mathrm{i})$
ϕ_t	V	0.0259	$\equiv k_\mathrm{B} T/q$

3.3 Experimental Details

A set of circular MIS TDs, all with a diameter of 170 µm, were freshly fabricated and characterized to serve as the experimental database. A p-type Si wafer with identical parameters to that in **Section 2.2** was utilized as the substrate, and the same cleaning and native oxide stripping processes were conducted. Next, the wafer underwent ANO in

doi:10.6342/NTU202304241

TABLE 3–III
List of physical variables and their governing equations in **Figure 3–3**.

Quantity	Description	Unit	Equation
$\mathscr{E}_{\mathrm{ox}}$	Oxide field	V/cm	$= V_{\mathrm{ox}}/d_{\mathrm{ox}}$
$J_{\mathrm{d}n}$	Electron diffusion current density	A/cm^2	(3.11)
$J_{\mathrm{g}n}$	Electron generation current density	A/cm^2	(A.53)
$J_{\mathrm{t}n}$	Electron tunnel current density	A/cm^2	(3.3)
P_{t}	Tunneling probability	–	(3.5), (3.7)
Q_{S}	Surface charge per unit area	C/cm^2	(3.16)
S	Supply function	–	(3.10)
V_{ox}	Oxide voltage drop	V	$= -Q_{\mathrm{S}}/C_{\mathrm{ox}}$
W	Depletion width	µm	(3.15)
V_{G}	Gate voltage	V	(1.1)

room-temperature deionized water with an electrostatic potential of +15 V with respect to the inert electrode. ANO time was 8 minutes. The wafer was tilted with respect to the inert electrode with a spacing of 1–3 cm. Following the ANO, RTP was performed at 950 °C for 15 s in an 20-torr N_2 ambient. The subsequent metal gate deposition, metal gate patterning, wafer back-side cleaning, and back contact formation processes were identical to those in **Section 2.2**.

The devices were characterized with an Agilent® B1500A Semiconductor Device Analyzer. In addition, one-dimensional TCAD simulations were conducted with SILVACO® ATLAS™. To inspect the influence of QC among the simulated devices, there were simulation runs with the QC model incorporated (QC on) and excluded (QC off) from the TCAD simulator.

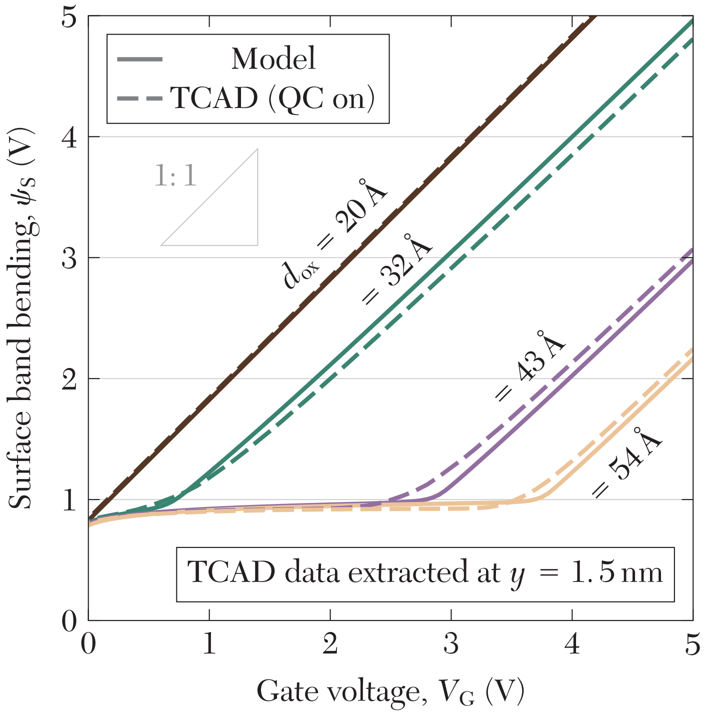


Figure 3–4. Modeled and TCAD-simulated $\psi_{\mathrm{S}} - V_{\mathrm{G}}$ characteristics for MIS TDs with several d_{ox}, showing two-stage electrostatics behavior. TCAD data were extracted at $y = 1.5\,\mathrm{nm}$.

3.4 Results and Discussion

Figure 3–4 shows the $\psi_{\mathrm{S}} - V_{\mathrm{G}}$ curves evaluated from the proposed model, and also the corresponding curves from the TCAD simulation results with QC on, under several oxide thicknesses. While the modeled and simulated curves differ slightly in their shapes and the knee voltage (V_{C}), two-stage electrostatics behavior is apparent and consistent. Also, structures with thicker oxide is associated with a higher V_{C}. For $V_{\mathrm{G}} < V_{\mathrm{C}}$, the near-constant ψ_{S} implies that further increment in V_{G} would drop almost entirely on the oxide as a consequence of (1.1). This regime corresponds to what is known as the inversion regime in the CEST model, and may be coined the *classical inversion (CI) regime* in MIS TDs. On the contrary, the curves exhibit near-unity slopes once

doi:10.6342/NTU202304241

TABLE 3–IV
Nomenclature for the electrostatics regimes in reverse-biased MIS TDs and there criteria.

Regime name	Acronym	Criterion
Classical inversion	CI	$0 \leq V_\mathrm{G} \leq V_\mathrm{C}$
Deep depletion	DD	$V_\mathrm{G} \geq V_\mathrm{C}$

$V_\mathrm{G} > V_\mathrm{C}$, indicating V_ox pinning. This regime above the critical voltage may be referred to as the *deep depletion (DD)* regime. With too small a d_ox ($\lesssim 25\,\mathrm{Å}$), CI would not be visible in the curve. On the other hand, for a very thick d_ox, DD is not likely to take place, or it takes place at an off-the-scale V_C. With an otherwise moderate value of d_ox, the curve transition at V_C appears to be sharp. **TABLE 3–IV** summarizes the electrostatics regimes in reverse-biased MIS TDs and there criteria.

The value of V_C associated with a specific d_ox may be extracted from the $\psi_\mathrm{S} - V_\mathrm{G}$ curve, by finding the intersection of piecewise asymptotes for the CI and DD regimes, respectively. By doing so over a continuous range of d_ox, the $V_\mathrm{C} - d_\mathrm{ox}$ characteristic can be obtained. **Figure 3–5** shows the $V_\mathrm{C} - d_\mathrm{ox}$ curves from the proposed model, from TCAD simulated results with QC on, and from TCAD simulation results with QC off, respectively. A $V_\mathrm{C} - d_\mathrm{ox}$ curve located more to the bottom in this plot (smaller V_C at the same d_ox) indicates the prominence of oxide tunneling over substrate generation, as a smaller V_C can already drive the device into DD given the same d_ox, which can be ascribed to an overall lack of inversion charge, and vice versa. Between the two TCAD-simulated curves, incorporation of QC results in larger V_C. This can be attributed to, as a result of QC, greater effective tunneling lengths due to nonzero inversion layer thickness [59], degrading P_t. The curve for the proposed model in this plot, compared to

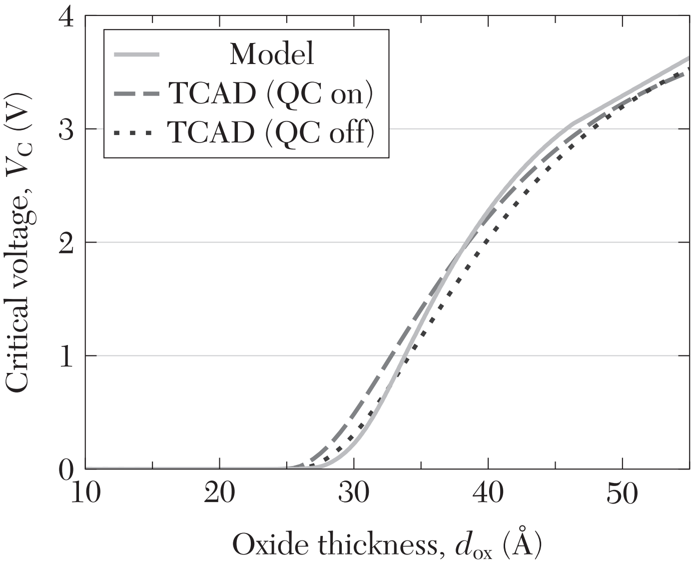


Figure 3–5. Extracted $V_C - d_{ox}$ curves from the proposed model, from TCAD simulated results with QC on, and from TCAD simulation results with QC off, respectively. V_C values were extracted from $\psi_S - V_G$ curves like **Figure 3–4**, by finding the intersection of the piecewise asymptotes for the CI and DD regimes, respectively. TCAD data were extracted at $y = 1.5\,\text{nm}$.

the QC-off TCAD simulation result, also shows slightly higher V_C, especially for thicker oxides. This may be attributed to underestimated Q_S in (3.16) where the electron QFL were assumed flat, but may not be in reality. An underestimated Q_S implies, equivalently, an overestimated $V_G = V_C$ to prompt DD. Nonetheless, the deviations among the curves (< 2 Å horizontally and < 0.3 V vertically) still appear to be very minor.

Figure 3–6 shows the evaluated $(\Delta\phi_n) - V_G$ characteristics. These curves also demonstrate two-stage behaviors in analogy with the $\psi_S - V_G$ relations. Mathematically, as $\Delta\phi_n$ begins to rise when V_G exceeds the critical voltage, the exponential term in the Q_S expression (3.16) saturates, and Q_S stops rising, driving the device into DD. On the other hand, the negligibility of $\Delta\phi_n$ at $V_G < V_C$ reduces the proposed model to CEST, and thus CI behavior is observed. For extremely thin d_{ox}, $\Delta\phi_n \approx V_G$, implying aligned

doi:10.6342/NTU202304241

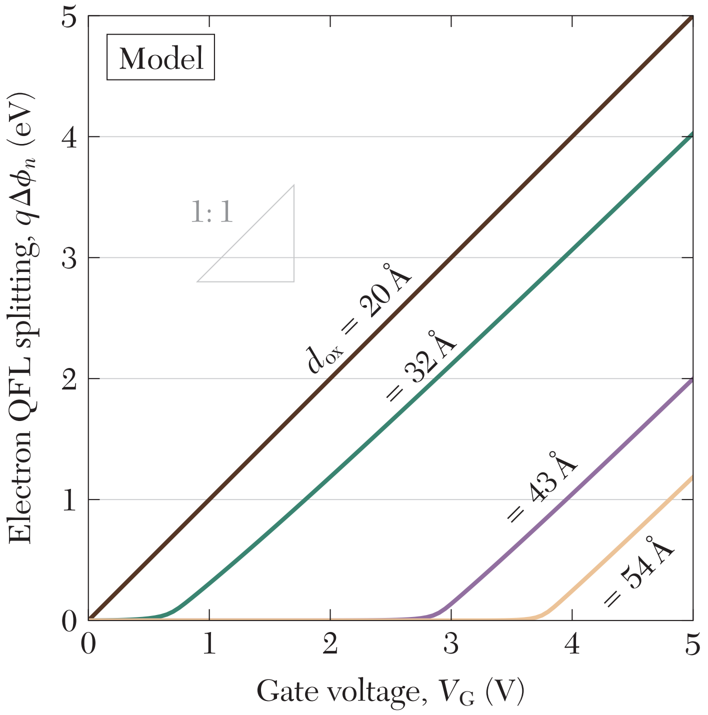


Figure 3–6. Electron QFL splitting vs. V_{G} curves evaluated from the model, for MIS TDs with several d_{ox}.

electron QFLs across the oxide.

The oxide voltage drop is related to Q_{S} by

$$\begin{aligned} V_{\mathrm{ox}} &= -\frac{Q_{\mathrm{S}}}{C_{\mathrm{ox}}} \\ &= \gamma\sqrt{\psi_{\mathrm{S}} + \phi_{\mathrm{t}} \exp\left(\frac{(\psi_{\mathrm{S}} - \Delta\phi_n) - 2\phi_{\mathrm{F}}}{\phi_{\mathrm{t}}}\right)}, \end{aligned} \tag{3.17}$$

where

$$\gamma \equiv \frac{\sqrt{2q\epsilon_{\mathrm{s}}N_{\mathrm{A}}}}{C_{\mathrm{ox}}} \tag{3.18}$$

is the body effect coefficient [53] that increases with oxide thickness. In the scope of our discussion, maximum d_{ox} is 100 Å, which corresponds to a maximum γ of $0.17\,\mathrm{V}^{1/2}$. The small values of γ in the scope implies strong V_{ox} pinning against rising ψ_{S} in the

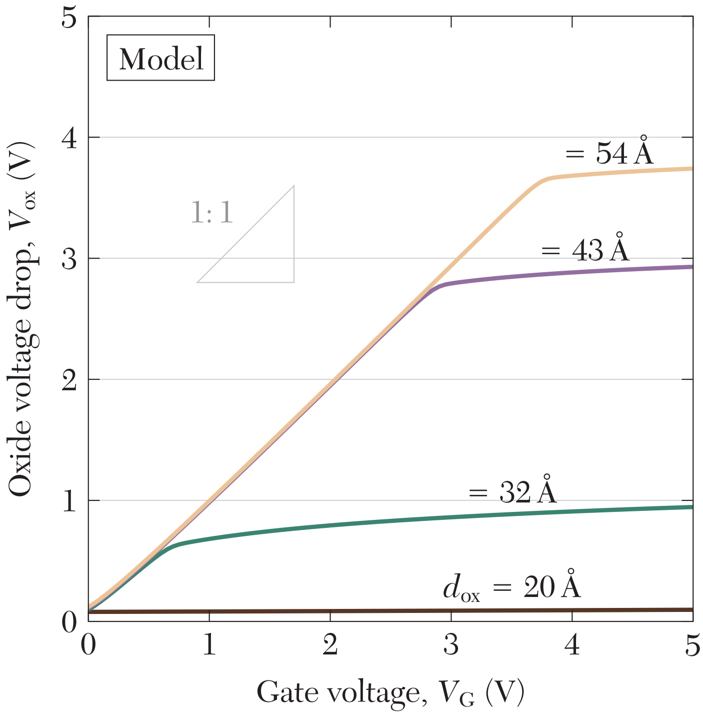


Figure 3–7. $V_{ox} - V_G$ characteristics evaluated from the model, for MIS TDs with several d_{ox}.

DD regime. As a result, any further increment in V_G from V_C would fall almost entirely on the semiconductor substrate, leading to the near-unity slope in the DD regimes of **Figure 3–4**. **Figure 3–7** depicts the $V_{ox} - V_G$ characteristics evaluated from the proposed model that, indeed, shows V_{ox} saturation in the DD regime to a great extent. Contrariwise, in accordance with ψ_S pinning in the CI regime, V_{ox} rises with V_G with near-unity slope before the critical point. That is, devices with higher d_{ox} and thus higher V_C will also possess higher saturation values of V_{ox}. What's more, as CEST (applicable at thermal equilibrium; i.e., $V_G = 0$) indicates negligible V_{ox} at $V_G = 0$ with such thin oxides, a piecewise approximation formula may apply to V_{ox}:

$$V_{ox} \approx \begin{cases} V_G & , V_G \leq V_C; \\ V_C & , V_G \geq V_C. \end{cases} \tag{3.19}$$

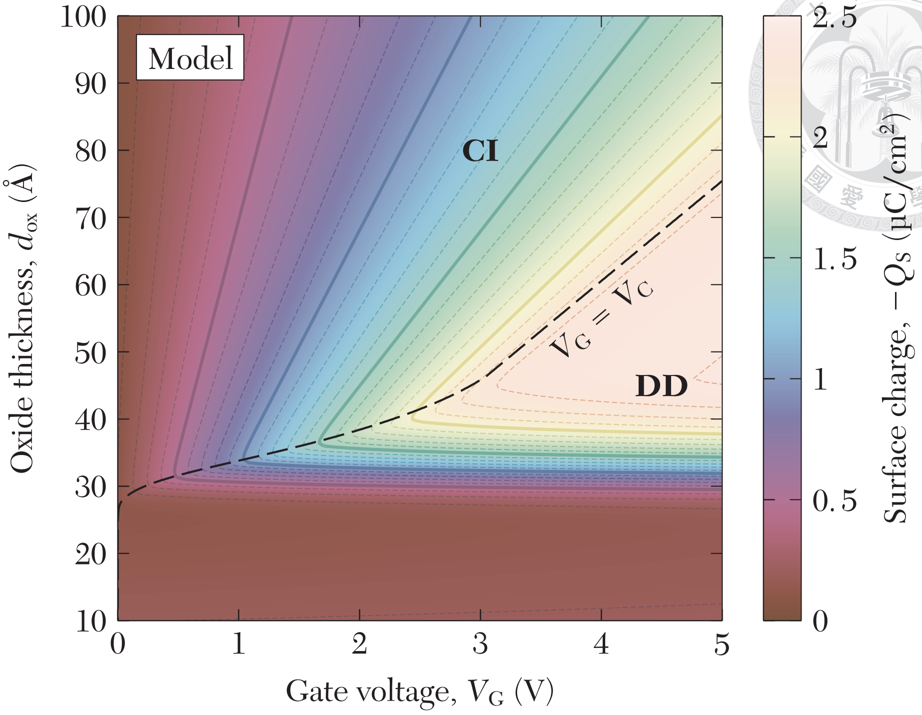


Figure 3–8. Contour plot of Q_S as a function of V_G and d_{ox}, evaluated from the model. The dashed line is the $V_G = V_C$ trace (a replica of that in **Figure 3–5**) that demarcates the electrostatics into two regimes.

Accordingly, as a consequence of (1.1),

$$\psi_S \approx \begin{cases} \psi_{S0} & , V_G \leq V_C; \\ \psi_{S0} + (V_G - V_C) & , V_G \geq V_C; \end{cases} \tag{3.20}$$

where ψ_{S0}, the value of ψ_S at zero applied bias, is given by (1.1) and (3.19):

$$\psi_{S0} \approx -V_{FB}. \tag{3.21}$$

Figure 3–8 shows the contour plot of model-evaluated Q_S as a function of V_G and d_{ox}. The $V_G = V_C\,(d_{ox})$ trace (as a replica of that in **Figure 3–5**) is overlaid in the plot as well. Demarcated by the trace, the dissimilarity of Q_S in the CI and DD regimes is very evident. In the CI regime, at a certain d_{ox}, $|Q_{\mathrm{S}}|$ rises linearly with V_{G}. However, $|Q_{\mathrm{S}}|$ nearly comes to a saturation once $V_{\mathrm{G}} \geq V_{\mathrm{C}}$. This is analogous to the $V_{\mathrm{ox}} - V_{\mathrm{G}}$ characteristics shown in **Figure 3–7**. The sparser contours in the CI regime associated with thicker d_{ox} can be attributed to their lower C_{ox} values. As $V_{\mathrm{ox}} - V_{\mathrm{G}}$ curves possess near-unity slopes in the CI regime, the slopes are $\approx C_{\mathrm{ox}}$ in the $|Q_{\mathrm{S}}| - V_{\mathrm{G}}$ curves, and hence the thinner the oxide, the steeper the curve. After all, the $|Q_{\mathrm{S}}| - (V_{\mathrm{G}},\, d_{\mathrm{ox}})$ characteristics are in accordance with the CEST in the CI regime. On the other hand, the saturation values for $|Q_{\mathrm{S}}|$ in the DD regime is primarily determined by V_{C}; that is, the horizontal margin between $V_{\mathrm{G}} = 0$ and the CI-DD demarcation in this plot, or the margin for V_{ox} to rise (3.19). For ultrathin d_{ox} ($< 30\,\text{Å}$), V_{C} is so small that $|Q_{\mathrm{S}}|$ saturates at a very low value. As the oxide goes thicker, under a high positive V_{G}, $|Q_{\mathrm{S}}|$ surges due to the drastic mitigation in oxide tunneling rate and the resulting V_{C} rise.

Figure 3–9 shows the $Q_{\mathrm{S}} - d_{\mathrm{ox}}$ plots at some fixed values of V_{G}; namely, along some vertical cutlines across **Figure 3–8**. At a positive gate bias, with increasing d_{ox}, the surge of $|Q_{\mathrm{S}}|$ at the demarcation, as well as its slow decline afterwards due to decreased C_{ox}, is clearly observed. The monotonic decline in the $V_{\mathrm{G}} = 0$ curve is also noteworthy. This is ascribed to the fact that CEST applies at thermal equilibrium (i.e., $V_{\mathrm{G}} = 0$), even for ultrathin oxides, due to the absence of all current components.

The surface charge (3.16) can be split into the depletion and inversion charge components:

$$Q_{\mathrm{S}} = Q_{\mathrm{d}} + Q_{\mathrm{i}}, \tag{3.22}$$

doi:10.6342/NTU202304241

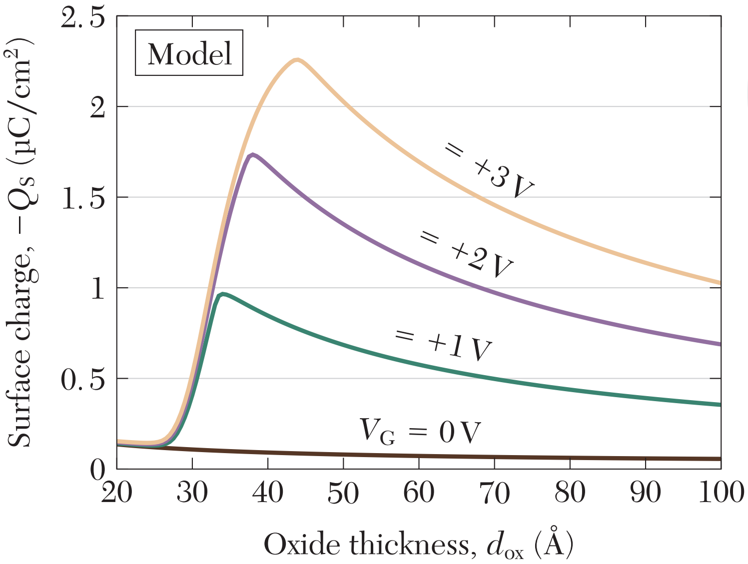


Figure 3–9. $Q_S - d_{ox}$ plots extracted from **Figure 3–8**, at some fixed values of V_G.

in which

$$Q_d = -\sqrt{2q\epsilon_s N_A \psi_S}$$

$$= -qN_A W \tag{3.23a}$$

is the depletion charge per unit area, and

$$Q_i = -\sqrt{2q\epsilon_s N_A}\left(\sqrt{\psi_S + \phi_t \exp\left(\frac{(\psi_S - \Delta\phi_n) - 2\phi_F}{\phi_t}\right)} - \sqrt{\psi_S}\right) \tag{3.23b}$$

is the inversion charge per unit area. **Figure 3–10** shows the contour plot of model-evaluated Q_d as a function of V_G and d_{ox}. Intriguingly, as opposed to $|Q_S|$, at a certain d_{ox}, $|Q_d|$ is constant in the CI regime but rises in the DD regime, in accordance with the behavior of ψ_S. **Figure 3–11** shows the contour plot of model-evaluated Q_i as a function of V_G and d_{ox}. Due to the saturated $|Q_S|$ but rising $|Q_d|$ in the DD regime, $|Q_i|$

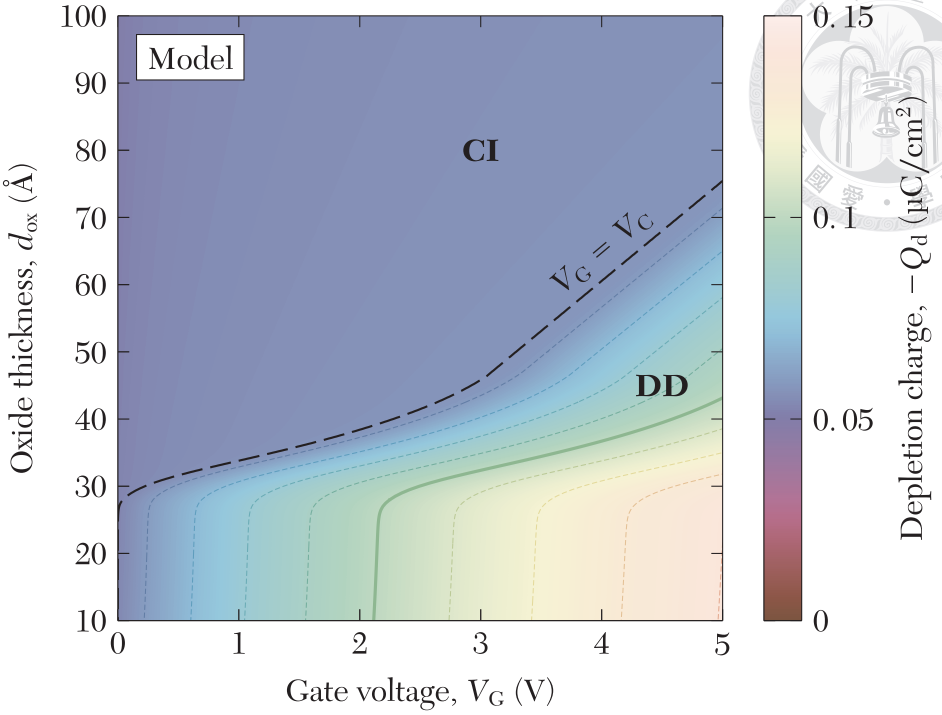


Figure 3–10. Contour plot of Q_d as a function of V_G and d_{ox}, evaluated from the model. The dashed line is the $V_G = V_C$ trace.

can decline with increasing V_G. This is especially evident for $d_{ox} < 30$ Å. Conversely, in the CI regime, $|Q_i|$ rises monotonically with V_G.

Figure 3–12 shows the contour plot of normalized low-frequency capacitance, $C_{LF}/A_G C_{ox}$, as a function of V_G and d_{ox}, evaluated by [24]

$$\frac{C_{LF}}{A_G} = -\frac{dQ_S}{dV_G}. \tag{3.24}$$

In the CI regime, C_{LF}/A_G is comparable to C_{ox}, as predicted by the CEST. However, $|Q_S|$ saturation with V_G in the DD regime yields low C_{LF}. For structures associated with ultrathin oxide, the absence of CI regime results in $C_{LF}/A_G \lll C_{ox}$ throughout the entire reverse bias region. **Figure 3–13** shows the contour plot of normalized high-frequency capacitance, $C_{HF}/A_G C_{ox}$, as a function of V_G and d_{ox}, evaluated by [24]

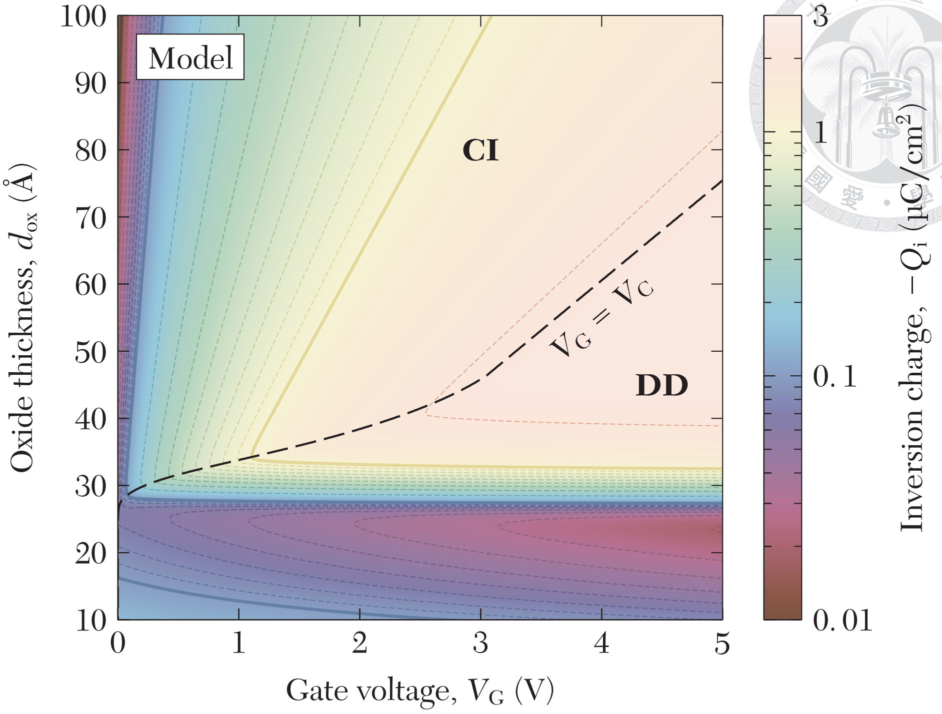


Figure 3–11. Contour plot of Q_i as a function of V_G and d_{ox}, evaluated from the model. The dashed line is the $V_G = V_C$ trace.

$$\frac{C_{HF}}{A_G} = \left(\frac{1}{C_{ox}} + \frac{\epsilon_s}{W}\right)^{-1}. \tag{3.25}$$

For higher d_{ox}, the higher $C_{HF}/A_G C_{ox}$ values in the CI regime can be attributed to a smaller C_{ox} but a fixed, universal W above the threshold, due to ψ_S pinning. Still, the C_{HF} pinning against V_G in the CI regime is in correspondence with the CEST. As $V_G > V_C$, DD drives C_{HF} further lower due to the expansion of W. **Figure 3–14 (a)** shows the experimental $(C_{HF}/A_G) - V_G$ curves for MIS TDs with a collection of d_{ox}. Correspondingly, **Figure 3–14 (b)** shows the modeled curves with like V_C values. In the experimental curves, the two-stage behavior is observed, as reported in prior research [34]. The two-stage behavior in $C_{HF} - V_G$, replicating the modeled characteristics, can be the strongest experimental support to the proposed electrostatics model.

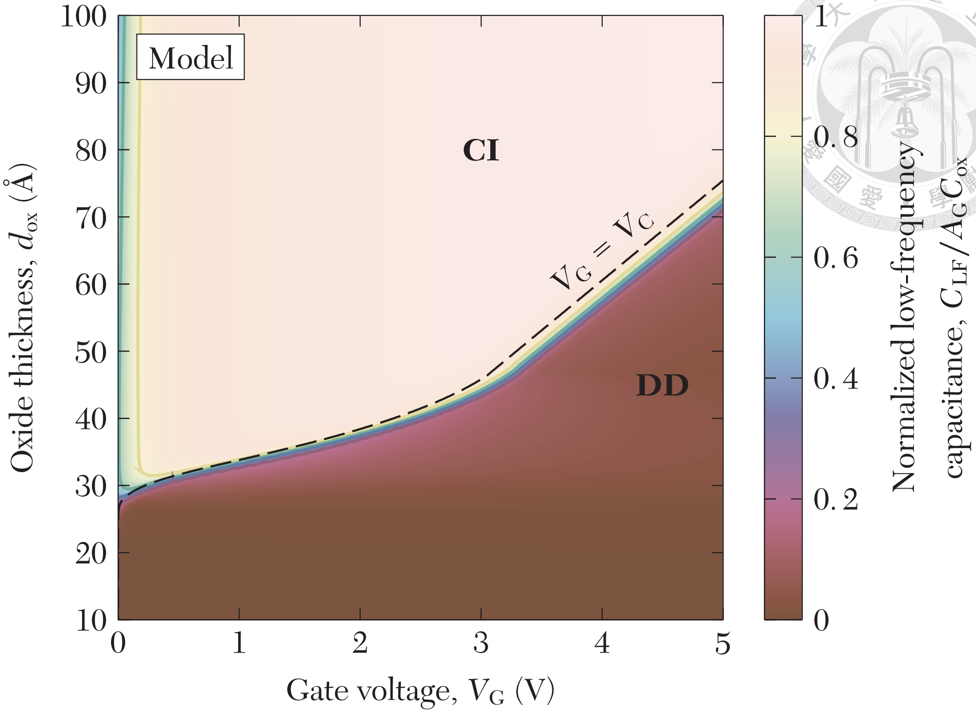


Figure 3–12. Contour plot of the normalized low-frequency capacitance, $C_{LF}/A_G C_{ox}$, as a function of V_G and d_{ox}, evaluated from the model. The dashed line is the $V_G = V_C$ trace.

However, to achieve the same V_C for each curve in **Figure 3–14 (a)**, the corresponding curve in **Figure 3–14 (b)** must be associated with a d_{ox} that is ~ 3 Å thicker than the experimental thickness. In other words, the model is prone to a systematic ~ 3 Å overestimation on d_{ox}. This can be attributed to QC (see **Figure 3–5**), and also to D_{it} in the experimental samples but neglected in the model. D_{it} can act as recombination-generation centers [19] and contribute to J_{gn}, shifting the $V_C - d_{ox}$ more to the left. In addition, they can also perturb the electrostatics [24]. Further corrections on the model are mandatory for curve-fitting purposes with experimental $V_C - d_{ox}$ relations.

Figure 3–15 shows the modeled $J_{tn} - d_{ox}$ curves at several fixed values of positive V_G. It should be emphasized that J_{tn} is not the device total current ($J_{tn} + J_{tp}$). In the plot, J_{tn} also demonstrates a two-stage behavior (DD for lower d_{ox}, CI for higher ones.)

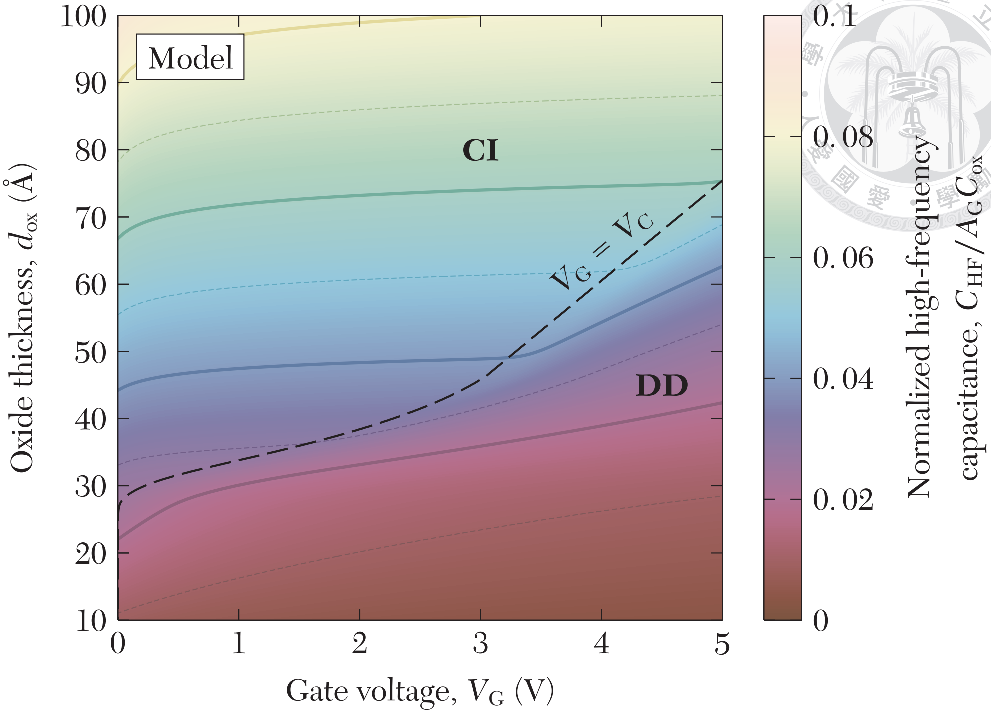


Figure 3–13. Contour plot of the normalized high-frequency capacitance, $C_{HF}/A_G C_{ox}$, as a function of V_G and d_{ox}, evaluated from the model. The dashed line is the $V_G = V_C$ trace.

The falling portions of the curves (CI) can be attributed to exponentially decreasing P_t as d_{ox} increases, where the current is limited by the oxide tunneling process itself, whereas the curve plateaux before the critical points can be ascribed to the limit in electron supply rate from the substrate (J_{gn} and J_{dn}). All in all, it is the shortage of electron supply for ultrathin d_{ox} that pins V_{ox} and drives the device into DD. For this reason, the CI regime may alternatively be referred to as the *tunneling-limited regime*, and the DD regime, as the *supply-limited regime*.

Figure 3–16 addresses the modeled $V_C - d_{ox}$ characteristics if we were to utilize the Sah-Noyce-Shockley model (3.14), in place of the W_{g0}-corrected model (A.53), in calculating J_{gn}. By not considering W_{g0} in the former model, the resulting higher J_{gn} shifts the $V_C - d_{ox}$ curve more to the left. This is equivalent to a reduction in V_C at a

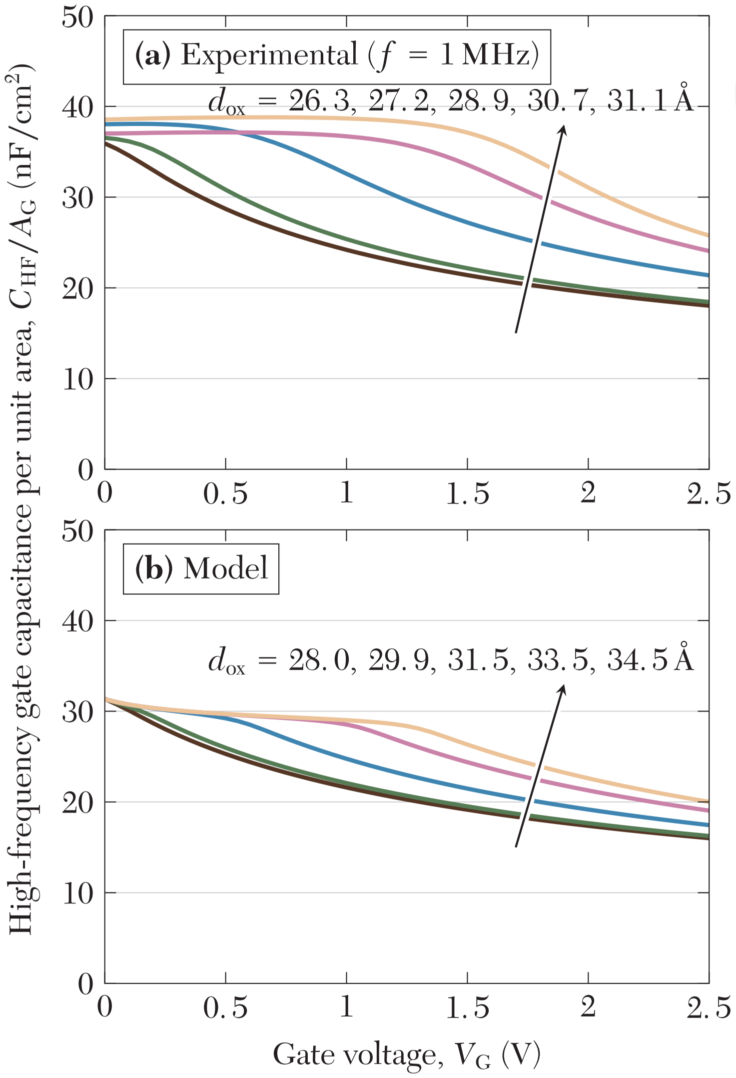


Figure 3–14. **(a)** Experimental high-frequency capacitance per unit area vs. V_{G} curves from the experimental data set. Measured under an AC frequency of 1 MHz. The curves are flat in the CI regime, but descending in the DD regime. **(b)** Modeled curves with like V_{C} values in **(a)**, showing replicated behaviors. However, there is a systematic d_{ox} overestimation ($\sim$ 3 Å) by the model, as every modeled curve is associated with a d_{ox} that is $\sim$ 3 Å thicker than the experimental value to reach the experimental V_{C}.

certain d_{ox}. Still, the deviation appears to be very minor.

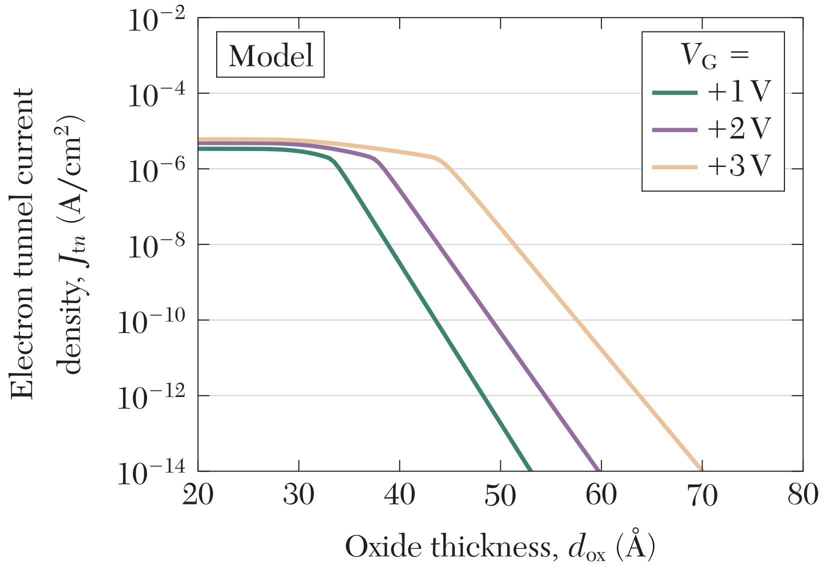


Figure 3–15. Modeled $J_{\mathrm{tn}} - d_{\mathrm{ox}}$ relations at several fixed values of positive V_{G}.

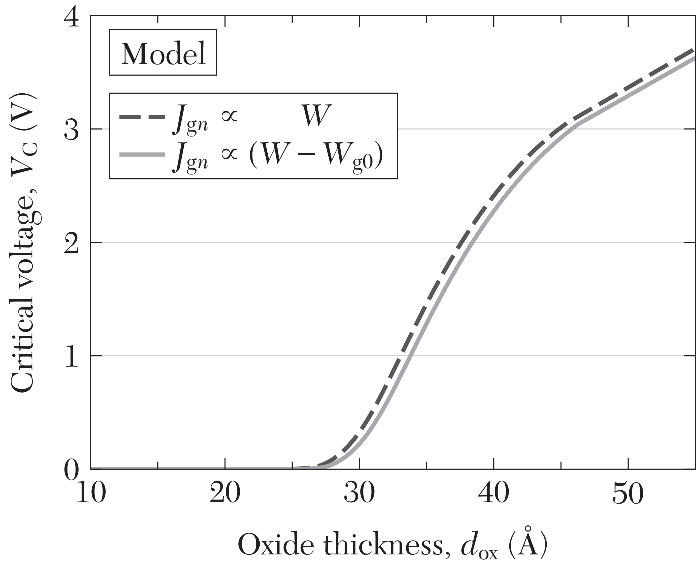


Figure 3–16. $V_{\mathrm{C}} - d_{\mathrm{ox}}$ relations by not considering the W_{g0} correction (so $J_{gn} \propto W$; eq. 3.14), and by including it (so $J_{gn} \propto (W - W_{\mathrm{g0}})$; eq. A.53), in calculating the generation current in the model. The latter has been adopted in all other calculations above.

3.5 A Closed-Form Approximation for the Critical Voltage vs. Oxide Thickness Relation

With the finding that ψ_{S} (as well as other electrostatics-related quantities) is sharply demarcated at V_{C} (3.20), it may be beneficial to find a closed-form approximation for the $V_{\mathrm{C}} - d_{\mathrm{ox}}$ relation to describe the MIS TD electrostatics in a more convenient manner, and with better physical interpretation. To begin with, inspired by **Figure 3–6**, we may define the critical point as where $\Delta\phi_n$ rises from 0 (CI regime) to a small but specific value, say

$$(V_{\mathrm{G}} \equiv V_{\mathrm{C}}) \qquad (\Delta\phi_n)_{\mathrm{C}} = \hat{\alpha}\phi_{\mathrm{t}}, \tag{3.26}$$

where the subscript "C" stands for "at the critical point," and the dimensionless quantity $\hat{\alpha}$ is assumed to be a constant that is irrelevant to d_{ox}, which may later be treated as a fitting parameter. By also noting that $J_{\mathrm{d}n} \ll J_{\mathrm{gn}}$ (see **Appendix A**), the current balance equation (3.2, combined with 3.3 and A.53) at the critical point is now

$$J_{\mathrm{t}n0} P_{\mathrm{tC}} \ln \frac{1 + \exp\left(-\dfrac{\phi_{\mathrm{CF0}} - \psi_{\mathrm{S0}}}{\phi_{\mathrm{t}}} - \hat{\alpha}\right)}{1 + \exp\left(-\dfrac{\phi_{\mathrm{CF0}} - \psi_{\mathrm{S0}} + V_{\mathrm{G}}}{\phi_{\mathrm{t}}}\right)} = \left(1 - e^{-\hat{\alpha}}\right) \frac{q n_{\mathrm{i}}}{2\tau_0}(W_0 - W_{\mathrm{g0}}), \tag{3.27}$$

where

$$W_0 \equiv \sqrt{\frac{2\epsilon_{\mathrm{s}}\psi_{\mathrm{S0}}}{q N_{\mathrm{A}}}} \tag{3.28}$$

is the depletion width at $V_{\mathrm{G}} = 0$, and also at $V_{\mathrm{G}} = V_{\mathrm{C}}$ according to (3.15 and 3.20).

The logarithm term in the left-hand side of (3.27) is identified as the supply function (3.10). We may further assume $V_{\mathrm{G}} = V_{\mathrm{C}}$ to be sufficiently high such that the exponential term in its denominator vanishes:

$$
\begin{aligned}
(\phi_{\mathrm{t}} \ll V_{\mathrm{G}} = V_{\mathrm{C}}) \qquad\qquad S &\approx \ln\left(1 + \exp\left(-\frac{\phi_{\mathrm{CF0}} - \psi_{\mathrm{S0}}}{\phi_{\mathrm{t}}} - \hat{\alpha}\right)\right) \\
&\equiv \hat{S}. \qquad (3.29)
\end{aligned}
$$

Then, (3.27) can be rearranged as

$$P_{\mathrm{tC}} = \frac{(1 - e^{-\hat{\alpha}})qn_{\mathrm{i}}}{2J_{\mathrm{t}n0}\hat{S}\tau_0}(W_0 - W_{\mathrm{g0}}), \qquad (3.30)$$

where we find P_{tC}, the critical tunneling probability, is now a constant independent of d_{ox}. For $\hat{\alpha} = 0.1$, P_{tC} is calculated as 6.16×10^{-15}.

Finally, the $V_{\mathrm{C}} - d_{\mathrm{ox}}$ relation is soluble by algebraically finding the inverse of the P_{t} formulae (3.5 and 3.7) with $P_{\mathrm{t}} = P_{\mathrm{tC}}$, and also by noting that

$$
\begin{aligned}
\mathscr{E}_{\mathrm{ox}} &= \frac{V_{\mathrm{ox}}}{d_{\mathrm{ox}}} \\
&= \frac{V_{\mathrm{C}}}{d_{\mathrm{ox}}} \qquad (3.31)
\end{aligned}
$$

due to (3.19) at the critical point. Moreover, we may define the *characteristic oxide thickness* d_{ox}^{*} as

$$d_{\mathrm{ox}}^{*} \equiv \frac{\Phi_{\mathrm{b}} \ln(1/P_{\mathrm{tC}})}{B}, \qquad (3.32)$$

which is evaluated as $47.3\,\text{Å}$ for $\hat{\alpha} = 0.1$. A further expanded formulation of d_{ox}^{*} obtained by substituting in (3.4) and (3.6) may be useful:

$$d_{\mathrm{ox}}^{*} = \frac{3\hbar}{4\sqrt{2m_{\mathrm{ox}}q\Phi_{\mathrm{b}}}}\ln\left[\frac{\pi^2\hbar^3(1-e^{-\hat{\alpha}})n_{\mathrm{i}}}{m^{*}k_{\mathrm{B}}^2T^2\hat{S}\tau_0}(W_0 - W_{\mathrm{g0}})\right]. \tag{3.33}$$

The definition of d_{ox}^{*} reduces the rectangular tunneling criterion (3.8) to

$$(V_{\mathrm{C}} = 0) \qquad d_{\mathrm{ox}} = \frac{2}{3}d_{\mathrm{ox}}^{*}, \tag{3.34}$$

and the DT-FNT transition point criterion (3.9) as

$$(V_{\mathrm{C}} = \Phi_{\mathrm{b}}) \qquad d_{\mathrm{ox}} = d_{\mathrm{ox}}^{*}. \tag{3.35}$$

In other words, for a device with $d_{\mathrm{ox}} \in \left[\,\frac{2}{3}d_{\mathrm{ox}}^{*},\, d_{\mathrm{ox}}^{*}\,\right]$, the tunneling mechanism at the critical point is DT, with P_{tC} governed by (3.7) that can be rearranged as

$$(\mathrm{DT}) \qquad 1 = \frac{d_{\mathrm{ox}}}{d_{\mathrm{ox}}^{*}}\cdot\frac{1-(1-V_{\mathrm{C}}/\Phi_{\mathrm{b}})^{3/2}}{V_{\mathrm{C}}/\Phi_{\mathrm{b}}}. \tag{3.36}$$

Here, by performing a change of variable (say $V_{\mathrm{C}}/\Phi_{\mathrm{b}} \equiv 1 - x^2$), the numerator $1-(1-V_{\mathrm{C}}/\Phi_{\mathrm{b}})^{3/2} = 1 - x^3 = (1-x)(1+x+x^2)$, and (3.36) eventually reduces to a quadratic equation for x, where V_{C} is finally solved as

$$\begin{aligned}(\mathrm{DT}) \qquad V_{\mathrm{C}} &= \Phi_{\mathrm{b}}\left[1-\left(\frac{(1-\mathscr{D}_{\mathrm{ox}})+\sqrt{(1-\mathscr{D}_{\mathrm{ox}})(1+3\mathscr{D}_{\mathrm{ox}})}}{2\mathscr{D}_{\mathrm{ox}}}\right)^2\right] \\ &= \frac{\Phi_{\mathrm{b}}}{2}\left[3-\frac{1}{\mathscr{D}_{\mathrm{ox}}^2}-\left(\frac{1}{\mathscr{D}_{\mathrm{ox}}}-1\right)^{3/2}\left(\frac{1}{\mathscr{D}_{\mathrm{ox}}}+3\right)^{1/2}\right], \end{aligned} \tag{3.37}$$

where $\mathscr{D}_{\text{ox}}$ is the *characteristically-normalized oxide thickness* defined as

$$\mathscr{D}_{\text{ox}} \equiv \frac{d_{\text{ox}}}{d_{\text{ox}}^{*}}. \tag{3.38}$$

Likewise, for a device with $d_{\text{ox}} \geq d_{\text{ox}}^{*}$, or $\mathscr{D}_{\text{ox}} \geq 1$, the tunneling mechanism at the critical point is FNT, with P_{tC} governed by (3.5) that can be rearranged as

$$\text{(FNT)} \qquad V_{\text{C}} = \Phi_{\text{b}} \mathscr{D}_{\text{ox}}. \tag{3.39}$$

For $d_{\text{ox}} \leq \frac{2}{3} d_{\text{ox}}^{*}$, or $\mathscr{D}_{\text{ox}} \leq \frac{2}{3}$, the tunneling rate is so high that $V_{\text{ox}} = V_{\text{C}} = 0$.

In conclusion, we obtain the following $V_{\text{C}} - d_{\text{ox}}$ relation:

$$V_{\text{C}} \approx \begin{cases} 0 & , \mathscr{D}_{\text{ox}} \leq \dfrac{2}{3}; \\[2ex] \dfrac{\Phi_{\text{b}}}{2} \left[3 - \dfrac{1}{\mathscr{D}_{\text{ox}}^{2}} - \left(\dfrac{1}{\mathscr{D}_{\text{ox}}} - 1 \right)^{3/2} \left(\dfrac{1}{\mathscr{D}_{\text{ox}}} + 3 \right)^{1/2} \right] & , \dfrac{2}{3} \leq \mathscr{D}_{\text{ox}} \leq 1; \\[2ex] \Phi_{\text{b}} \mathscr{D}_{\text{ox}} & , \mathscr{D}_{\text{ox}} \geq 1. \end{cases} \tag{3.40}$$

Figure 3–17 shows the extracted vs. approximated $V_{\text{C}} - d_{\text{ox}}$ curves for the model. While the deviation at low V_{C} is relatively high due to an assumption in simplifying S (3.29), the model is highly accurate for sufficiently high V_{C}. The approximation gives an physical interpretation that d_{ox}^{*} and Φ_{b} are the horizontal and vertical scale factors, respectively, for the $V_{\text{C}} - d_{\text{ox}}$ relation. An increase in substrate generation rate ($\tau_0 \downarrow$) will yield a higher P_{tC} and finally lead to horizontal shrinking of the curve, without changing its base shape.

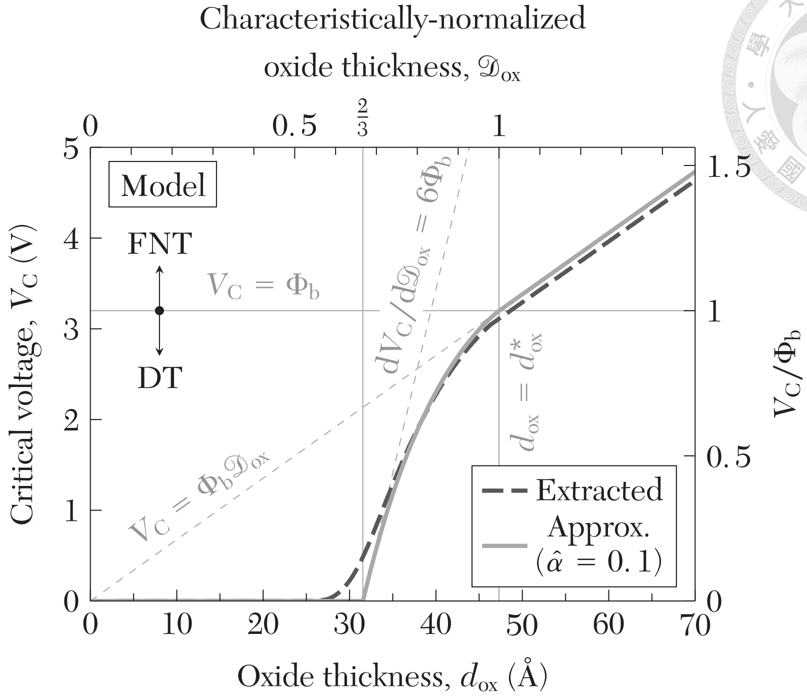


Figure 3–17. Comparison of the $V_C - d_{\mathrm{ox}}$ curves by extraction from the modeled $\psi_S - V_G$ curves, and by the closed-form approximation (3.40) with $\hat{\alpha} = 0.1$, respectively. P_{tC} and d^*_{ox} are calculated as 6.16×10^{-15} and 47.3 Å, respectively. The approximation shows high accuracy for sufficiently high V_C.

3.6 Summary

In this chapter, an analytical electrostatics model for Al/SiO$_2$/Si(p) MIS TDs, applicable under reverse bias, has been proposed by balancing the electron carrier components. The model contains no integrals, differentials and does not rely on mesh-based numerical methods. The modeled $\psi_S - V_G$ curves, in agreement with TCAD simulation results, suggest the existence of a critical voltage, V_C, associated with device d_{ox}, as the demarcation of two electrostatically-distinct regimes, CI and DD. The former is analogous to the CEST, while the latter is the consequence of high tunnel rate and V_{ox} saturation under high positive V_G. Model accuracy was validated by the $V_C - d_{\mathrm{ox}}$

characteristics compared to the TCAD simulation results. Furthermore, modeled electron QFL vs. V_{G} curves were inspected to study the mechanism for the CI and DD operation regimes. V_{ox}, Q_{S}, Q_{d}, Q_{i}, C_{LF}, and C_{HF} characteristics as functions of V_{G} (and/or d_{ox}) were also addressed. Specifically, the model manages to replicate experimental $C_{\mathrm{HF}} - V_{\mathrm{G}}$ curve shapes, while the systematic d_{ox} overestimation ($\sim 3\,\text{Å}$) must be addressed and inspected for more delicate data-fitting purposes for this model in future works. In addition, from the $J_{\mathrm{t}n} - d_{\mathrm{ox}}$ characteristics, device electron current were found to be tunneling-limited in the CI regime and substrate-supply-limited in the DD regime. The effect for W_{g0} correction in $J_{\mathrm{g}n}$ on the $V_{\mathrm{C}} - d_{\mathrm{ox}}$ relation has also been studied. Finally, an accurate and physically-meaningful closed-form approximation for the $V_{\mathrm{C}}\,(d_{\mathrm{ox}})$ relation was derived. This model is believed to be a convenient and comprehensive approach for evaluating MIS TD electrostatics under reverse bias.

4

Electrostatics Theory and Modeling II: Ultra-High-Low MOS(p) Tunnel Structures Under Reverse Bias

4.1 Objective

With the reverse-bias electrostatics model for planar MIS(p) TDs developed in **Chapter 3**, it may be generalized into what applies to UHL structures, in accordance to the qualitative findings in **Chapter 2**. It has been shown in **Figure 2–22** that the $\psi_{\mathrm{S}} - V_{\mathrm{G}}$ characteristics for UHL devices, in analogy to those of planar MIS TDs, demonstrate apparent two-stage behavior around some V_{C}. CI and DD regimes (see **Section 3.4**) are seemingly well-defined in the curves. We may therefore assume similar physical mechanisms that govern both the planar and UHL electrostatics under reverse bias. Moreover, as the UHL low region is structurally equivalent to a planar MIS TD, we may directly apply the planar electrostatics model on the low region, but with its generation current component also including generated electrons originating from the high region. In this chapter, the high-region generation current model and its incorporation into the UHL electrostatics will be discussed in detail.

4.2 High-Region Generation Current Model for Cylindrically-Symmetric UHL Devices

In the scope of the following quantitative discussions, the UHL devices are assumed to be cylindrically-symmetric. Assuming cylindrical symmetry for the UHL devices is convenient for mathematical derivations and comparison purposes with TCAD simulation results in **Chapter 2**. It has also been shown in the chapter that, under reverse bias, ψ_{S} is nearly constant in both the high and low regions, say $\psi_{\mathrm{S}} = \psi_{\mathrm{SL}}$ and $\psi_{\mathrm{S}} = \psi_{\mathrm{SH}}$, respectively:

$$\psi_{\mathrm{S}}(r) \approx \begin{cases} \psi_{\mathrm{SL}} & , r < r_{\mathrm{L}}; \\ \psi_{\mathrm{SH}} & , r_{\mathrm{L}} < r < r_{\mathrm{G}}. \end{cases} \tag{4.1}$$

doi:10.6342/NTU202304241

4.2.1 High-Region Surface Band Bending

With such a strong electrostatics coupling in reverse-biased UHL devices, ψ_{SH} and ψ_{SL} may be related somewhat, and it may be helpful to find ψ_{SH} in terms of ψ_{SL}. From both **Figure 2–18 (b)** and also **Figure 2–23** , we find ψ_{SH} to be clamped by a V_{G}- and d_{H}-dependent upper limit (say $\psi^{\wedge}_{\mathrm{SH}}$) and is otherwise close to ψ_{SL}. That is,

$$\psi_{\mathrm{SH}} \approx \min\{\psi^{\wedge}_{\mathrm{SH}},\ \psi_{\mathrm{SL}}\}. \tag{4.2}$$

In addition, the clamped ψ_{SH}, the scarcity of n_{S}, and DD are all found to be related in **Figure 2–18** . Thus, to find the value of $\psi^{\wedge}_{\mathrm{SH}}$, we may exploit the fact that inversion charge is negligible ($Q_{\mathrm{iH}} \approx 0$) at the high region under DD, in a manner that

$$\text{(DD)} \qquad \begin{aligned} Q_{\mathrm{SH}} &\approx Q_{\mathrm{dH}} \\ &= -\sqrt{2q\epsilon_{\mathrm{s}} N_{\mathrm{A}} \psi^{\wedge}_{\mathrm{SH}}}, \end{aligned} \tag{4.3}$$

where Q_{SH}, Q_{iH} and Q_{dH} are the surface charge, inversion charge, and depletion charge, respectively, per unit area in the high region. Combined with (1.1), this yields

$$\text{(DD)} \qquad V_{\mathrm{G}} = V_{\mathrm{FB}} + \gamma_{\mathrm{H}}\sqrt{\psi^{\wedge}_{\mathrm{SH}}} + \psi^{\wedge}_{\mathrm{SH}}, \tag{4.4}$$

where γ_{H} is the body effect coefficient associated with C_{oxH} (3.18). This is a quadratic equation for $\sqrt{\psi^{\wedge}_{\mathrm{SH}}}$ with the solution

$$\psi^{\wedge}_{\mathrm{SH}} = \left[-\frac{\gamma_{\mathrm{H}}}{2} + \sqrt{\left(\frac{\gamma_{\mathrm{H}}}{2}\right)^2 + (V_{\mathrm{G}} - V_{\mathrm{FB}})}\,\right]^2. \tag{4.5}$$

doi:10.6342/NTU202304241

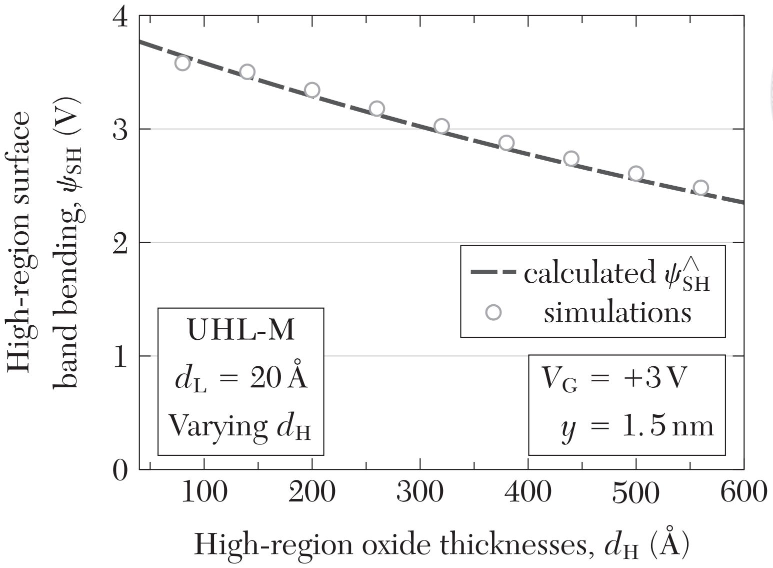


Figure 4–1. $\psi_{\mathrm{SH}} - d_{\mathrm{H}}$ plot for the UHL-M device (see **TABLE 2–I**) under $V_{\mathrm{G}} = +3\,\mathrm{V}$, TCAD-simulated (symbols, from **Figure 2–20**) and calculated $\hat{\psi}_{\mathrm{SH}}$ curve (line, from 4.5). Simulation data extracted at probing point A and $y = 1.5\,\mathrm{nm}$.

It should also be noted here that $\hat{\psi}_{\mathrm{SH}}$ has indeed no lateral dependency in its mathematical form. In other words, we would expect a uniform ψ_{S} in the high region of UHL devices under extreme DD. Also, $\hat{\psi}_{\mathrm{SH}}$ is independent of d_{L} but will decrease as d_{H} rises. **Figure 4–1** demonstrates the TCAD-simulated $\psi_{\mathrm{S}} - d_{\mathrm{H}}$ curve for the UHL-M device under $V_{\mathrm{G}} = +3\,\mathrm{V}$ (symbols, from **Figure 2–20**) compared to the calculated $\hat{\psi}_{\mathrm{SH}} - d_{\mathrm{H}}$ curve (line, from 4.5), showing good consistency in-between.

4.2.2 Electron QFL, Carrier Generation, and Electron Coupling in the High Region

The presence of a radial electric field $\mathscr{E}_r$ in reverse-biased UHL devices has been demonstrated in **Figure 2–10**. The radial field is capable of drifting electrons from the high region to the low region, and may be responsible for the highly-effective coupling of inversion charge, which has been discovered in the experimental and simulation

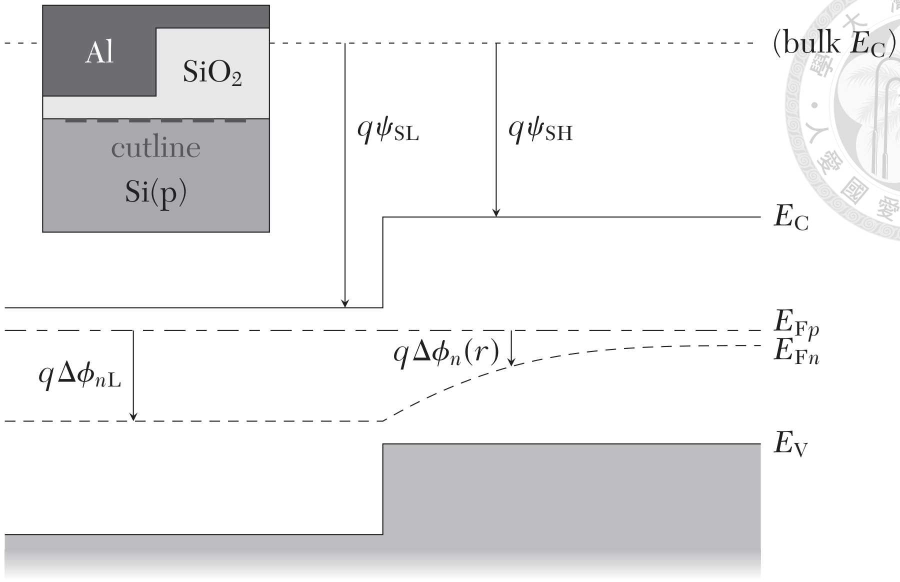


Figure 4–2. Schematic horizontal band diagram along a radius at the surface (see inset) of a UHL device under $V_\mathrm{G} > 0$, assuming an abrupt transition of ψ_S at the high-low boundary. The non-constant E_Fn in the high region gives rise to a lateral electron flux towards the low region.

results. However, to consider the total lateral electron current (drift+diffusion) in the substrate, it may be beneficial to inspect the lateral electron QFL profile instead [19]. **Figure 4–2** depicts the schematic horizontal band diagram along a radius at the surface (see the cutline in the inset) of a reverse-biased UHL device. The band edges are abrupt at the high-low boundary by assuming constant values of ψ_SL and ψ_SH (4.1). In addition, we assume an r-independent electron QFL splitting in the low region ($\Delta\phi_{n\mathrm{L}}$) due to the fact that vertical electron tunneling should be much more significant than lateral electron movement in this region. On the other hand, with vertical tunneling prohibited through the high oxide, $\Delta\phi_n$ in the high region must decay with the radial location, as shown in the figure, so as to give rise to the lateral electron flux towards the low region. Unfortunate, the declined QFL splitting in the high region may jeopardize carrier generation ($G \propto (1 - e^{-\Delta\phi_n/\phi_\mathrm{t}})$, see eq. A.7) and degrade device I_G Modeling of $\Delta\phi_n(r)$ is therefore mandatory for assessing this effect and its influence on the total electron generation current.

From **Appendix A**, $\Delta\phi_n(r)$ in the high region ($r_\mathrm{L} \le r \le r_\mathrm{G}$) of a reverse-biased UHL device is found to satisfy

$$(r_\mathrm{L} \le r \le r_\mathrm{G}) \qquad 1 - e^{-\Delta\phi_n(r)/\phi_\mathrm{t}} \approx \frac{I_1(\rho_\mathrm{G})K_0(\rho) + K_1(\rho_\mathrm{G})I_0(\rho)}{I_1(\rho_\mathrm{G})K_0(\rho_\mathrm{L}) + K_1(\rho_\mathrm{G})I_0(\rho_\mathrm{L})} \times \left(1 - e^{-\Delta\phi_{n\mathrm{L}}/\phi_\mathrm{t}}\right), \qquad \text{(from A.81)}$$

assuming that $\Delta\phi_{n\mathrm{L}}$ is given, where I_ν and K_ν represent the modified Bessel functions of the first and second kinds, respectively, of order ν, and

$$\rho \equiv \frac{r}{\Lambda} \qquad \text{(from A.78)}$$

denotes the *normalized radius* (sic ρ_G and ρ_L denote the normalized gate and low-region radii), normalized by the *Coupling length* Λ that comes with an approximate form for sufficiently large depletion widths:

$$(W_\mathrm{H} > W_\mathrm{g0}) \qquad \Lambda \approx \left(\frac{\sqrt{2\pi}\, q\, \mu_n n_\mathrm{i} \tau_0 \lambda_\mathrm{p}^3 \operatorname{erfi}\sqrt{\psi_\mathrm{SH}/\phi_\mathrm{t}}}{\epsilon_\mathrm{s}(W_\mathrm{H} - W_\mathrm{g0})}\right)^{1/2}, \qquad \text{(from A.76)}$$

where erfi represents the imaginary error function, and $W_\mathrm{H} \equiv (2\epsilon_\mathrm{s}\psi_\mathrm{SH}/qN_\mathrm{A})^{1/2}$ denotes the depletion width in the high region. It can be evaluated from values in **TABLE 3–I** that $\Lambda \approx 0.45\,\mathrm{cm}$ at $\psi_\mathrm{SH} = 2\phi_\mathrm{F}$ and will grow even bigger beyond. That is, Λ is generally far greater than the device dimensions under the inversion regime, rendering $\Delta\phi_n(r) \approx \Delta\phi_{n\mathrm{L}}$ in (A.81). This has been verified in **Figure 2–15** for the UHL-M device.

By also defining the *coupling efficiency* $\eta_{\mathrm{C}} \in [0,\ 1]$ as

$$\eta_{\mathrm{C}} = \frac{2\rho_{\mathrm{L}}}{\rho_{\mathrm{G}}^2 - \rho_{\mathrm{L}}^2} \frac{I_1(\rho_{\mathrm{G}})K_1(\rho_{\mathrm{L}}) - K_1(\rho_{\mathrm{G}})I_1(\rho_{\mathrm{L}})}{I_1(\rho_{\mathrm{G}})K_0(\rho_{\mathrm{L}}) + K_1(\rho_{\mathrm{G}})I_0(\rho_{\mathrm{L}})}, \tag*{(from A.85)}$$

the total generation current from the substrate (high+low) is found to be

$$I_{gn} \approx \frac{qn_{\mathrm{i}}}{2\tau_0}\left[A_{\mathrm{L}}\left(W_{\mathrm{L}} - W_{\mathrm{g0}}\right) + \eta_{\mathrm{C}}A_{\mathrm{H}}\left(W_{\mathrm{H}} - W_{\mathrm{g0}}\right)\right]\left(1 - e^{-\Delta\phi_{n\mathrm{L}}/\phi_{\mathrm{t}}}\right), \tag*{(from A.86)}$$

where $W_{\mathrm{L}} \equiv \left(2\epsilon_{\mathrm{s}}\psi_{\mathrm{SL}}/qN_{\mathrm{A}}\right)^{1/2}$ is the depletion width in the low region. From this result, as we have observed that $W_{\mathrm{L}} \sim W_{\mathrm{H}}$, it follows that $I_{gn} \propto A_{\mathrm{L}}$ if the coupling efficiency is low ($\eta_{\mathrm{C}} \to 0$), and $I_{gn} \propto A_{\mathrm{L}} + A_{\mathrm{H}} = A_{\mathrm{G}}$ if it is high ($\eta_{\mathrm{C}} \to 1$).

Direct evaluation on (A.85) suggests that η_{C} is very close to unity for physical device sizes, which can be attributed to the centimeter-scale Λ that makes ρ_{L}, $\rho_{\mathrm{G}} \ll 1$ in the expression. E.g., for the UHL-M device with $r_{\mathrm{G}} = 20\,\mu\mathrm{m}$ and $r_{\mathrm{L}} = 0.5\,\mu\mathrm{m}$, $\eta_{\mathrm{C}} \approx 0.99997$; for a bigger device that resembles the experimental device dimension (say $r_{\mathrm{G}} = 150\,\mu\mathrm{m}$, $r_{\mathrm{L}} = 10\,\mu\mathrm{m}$), $\eta_{\mathrm{C}} \approx 0.9989$. Indeed, the A_{G}-dependency of I_{G}, observed both experimentally and simulation-wise in **Chapter 2**, already suggest near-unity η_{C}. In conclusion, practically, η_{C} may always be treated as 1. As a result, (A.86) becomes

$$I_{gn} \approx \frac{qn_{\mathrm{i}}}{2\tau_0}\left(A_{\mathrm{L}}W_{\mathrm{L}} + A_{\mathrm{H}}W_{\mathrm{H}} - A_{\mathrm{G}}W_{\mathrm{g0}}\right)\left(1 - e^{-\Delta\phi_{n\mathrm{L}}/\phi_{\mathrm{t}}}\right). \tag{4.6}$$

4.3 Generalization of the Planar Electrostatics Model into Cylindrically-Symmetric UHL Structures

In analogy to the planar MIS TD, the current balance equation (like (3.2), assume negligible $J_{\mathrm{d}n}$) may be written as

$$I_{\mathrm{t}n} = I_{gn}, \tag{4.7}$$

where

$$\begin{aligned} I_{\mathrm{t}n} &= A_{\mathrm{L}} \cdot J_{\mathrm{t}n} \\ &= A_{\mathrm{L}} \cdot J_{\mathrm{t}n0} \cdot P_{\mathrm{t}} \cdot S \end{aligned} \tag{4.8}$$

is the device tunnel current that takes place only in the low region. With identical tunneling mechanism to planar devices at the UHL low region, $J_{\mathrm{t}n}$ and related quantities thereof must abide by the same equations in **Section 3.2**, but with the evaluation performed across the low oxide ($V_{\mathrm{ox}} \to V_{\mathrm{oxL}}$, $d_{\mathrm{ox}} \to d_{\mathrm{L}}$, $Q_{\mathrm{S}} \to Q_{\mathrm{SL}}$, $\cdots$). **Figure 4–3** shows the flowchart for the evaluation procedure (similar to **Figure 3–3**). The procedure is similar to that in **Section 3.2**: A given ψ_{SL}, together with given device dimensions d_{L}, d_{H}, A_{L}, and A_{G}, serves as the input. $\Delta\phi_{n\mathrm{L}}$ is then adaptively chosen to have the current balance equation satisfied. Thereby, V_{G} is uniquely determined by ψ_{SL}, and the desired $\psi_{\mathrm{SL}} - V_{\mathrm{G}}$ relation can be obtained. **Figure 4–4** shows the modeled UHL $\psi_{\mathrm{SL}} - V_{\mathrm{G}}$ characteristics subject to varying $A_{\mathrm{L}} = A_{\mathrm{G}}$, with $d_{\mathrm{H}} = 320\,\text{Å}$, and $d_{\mathrm{L}} = 20\,\text{Å}$ or $30\,\text{Å}$. Two-stage electrostatics behavior is still evident in each curve, and every V_{C} is well-defined. The critical voltage exhibits a positive correlation with d_{L}, similar to that in planar devices. In addition, it also shows a negative correlation with $A_{\mathrm{L}}/A_{\mathrm{G}}$. **Figure 4–5** shows the extracted $V_{\mathrm{C}} - d_{\mathrm{L}}$ plots at several $A_{\mathrm{L}}/A_{\mathrm{G}}$ values. The curve

doi:10.6342/NTU202304241

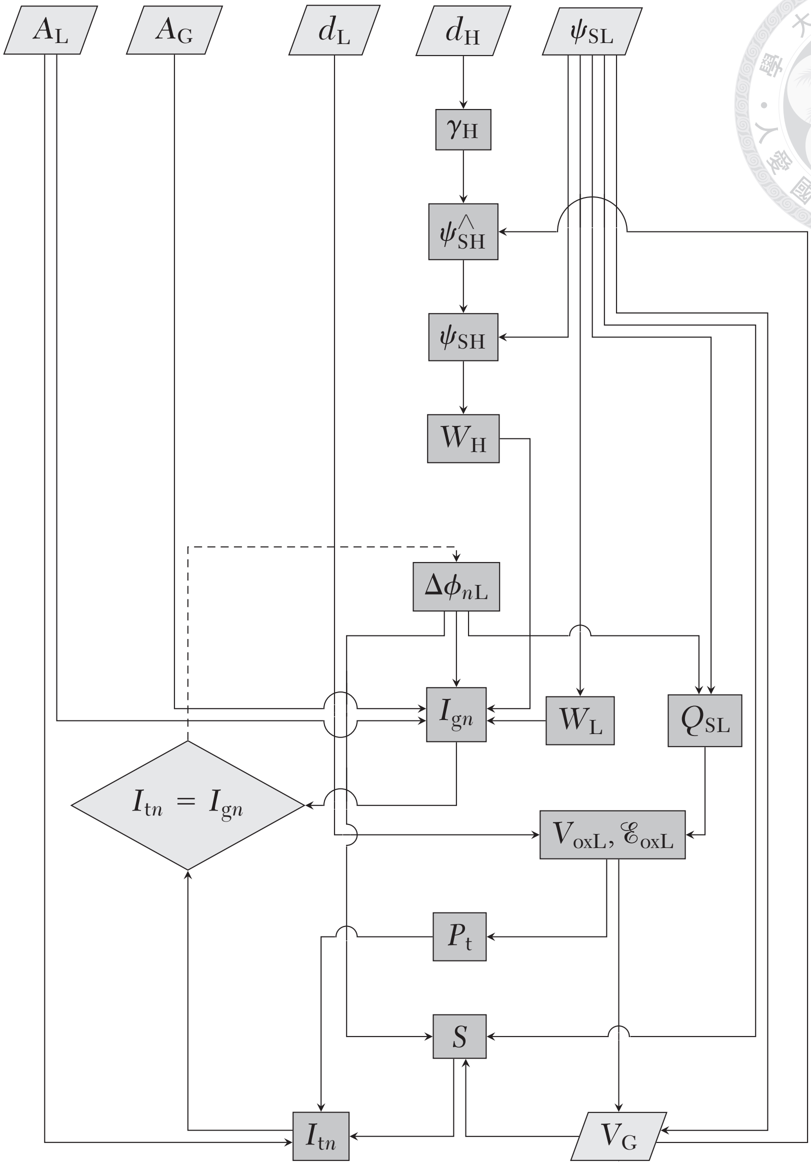


Figure 4–3. Flowchart for the UHL electrostatics evaluation procedure.

shapes resemble **Figure 3–5** and **Figure 3–17** for planar MIS TDs, except for a horizontal shrinkage as A_L reduces, which, according to the discussions in **Section 3.5**, may be attributed to *a reduced characteristic oxide thickness*.

Moreover, a similar procedure to what has been shown in **Section 3.5** may be performed to obtain an approximated closed-form expression for the $V_\mathrm{C} - d_\mathrm{L}$ relation, provided that A_L, A_G, and d_H are fixed. First, by noting the $\psi_{\overset{\wedge}{\mathrm{SH}}}$ values in **Figure 4–1**

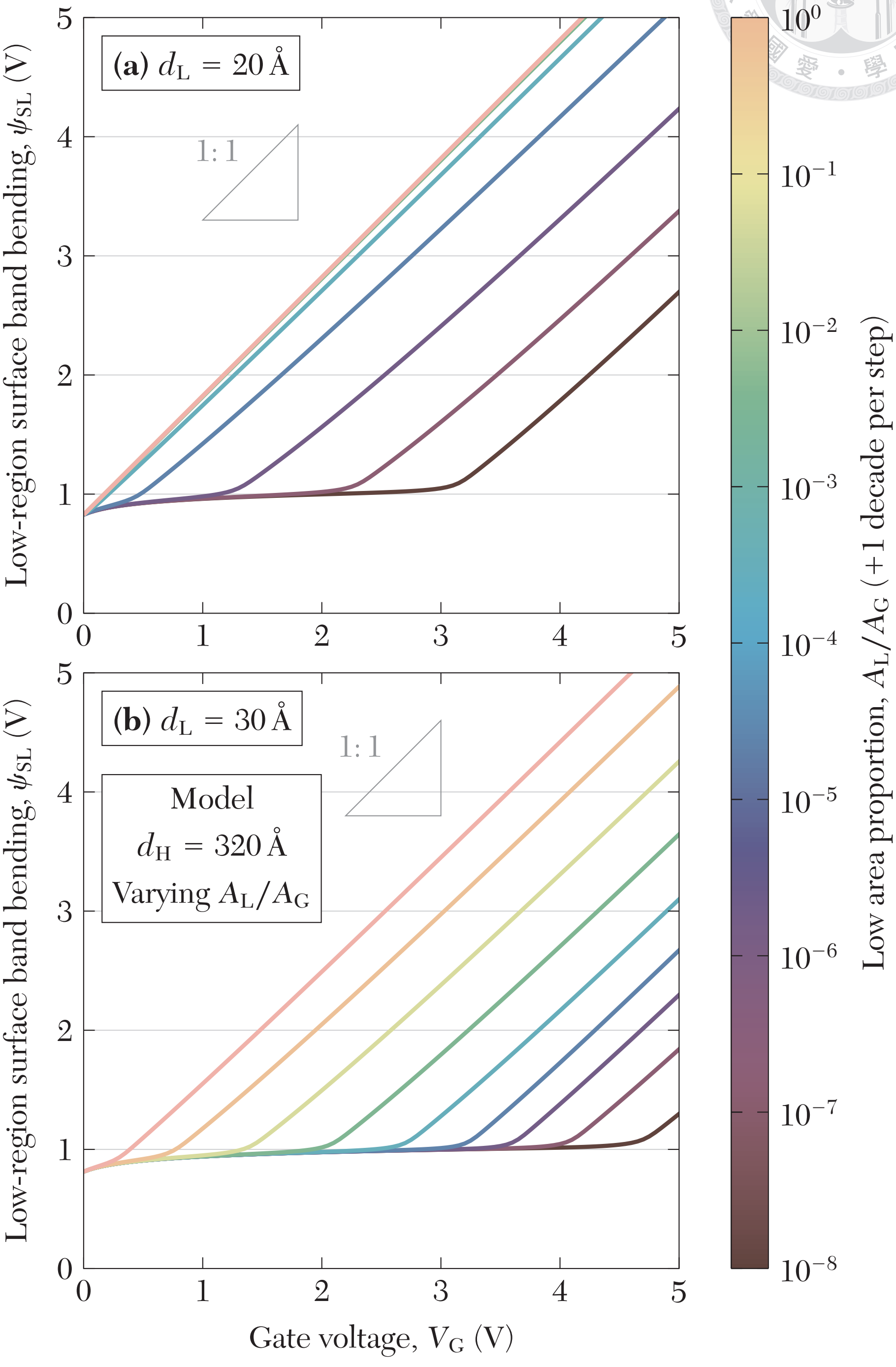


Figure 4–4. Modeled UHL $\psi_{SL} - V_G$ characteristics subject to varying $A_L = A_G$, with d_H = 320 Å, and **(a)** d_L = 20 Å or **(b)** d_L = 30 Å.

doi:10.6342/NTU202304241

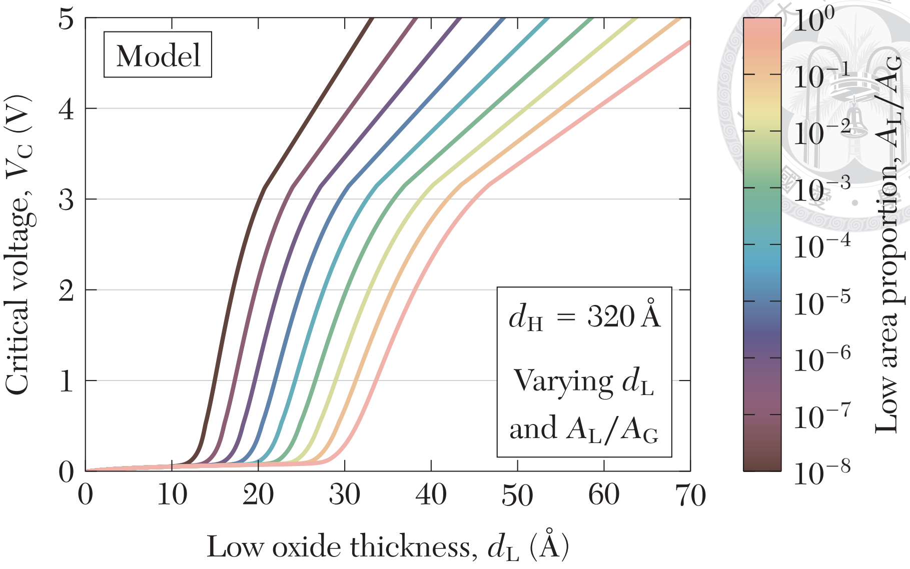


Figure 4–5. Extracted $V_C - d_L$ plots for UHL devices with d_H = 320 Å, at several A_L/A_G values. Obtained by finding the intersections of asymptotes from **Figure 4–4** .

to be comparable to ψ_{SL} anyway, even for d_H being tens of nanometers, we no longer have to distinguish ψ_{SH} from ψ_{SL}, as (4.1) now reduces to

$$\psi_{SH} \approx \psi_{SL}. \tag{4.9}$$

Hence, $W_H \approx W_L$, and (4.6) becomes

$$I_{gn} \approx \frac{qn_i}{2\tau_0} A_G \left(W_L - W_{g0}\right)\left(1 - e^{-\Delta\phi_{nL}/\phi_t}\right), \tag{4.10}$$

showing explicit proportionality to A_G. Next, we may define the critical point by a fixed $\Delta\phi_{nL}$ value, in analogy to (3.26):

$$(V_G \equiv V_C) \qquad (\Delta\phi_{nL})_C = \hat{\alpha}\phi_t. \tag{4.11}$$

doi:10.6342/NTU202304241

The current balance equation at the critical point (and with $W_\mathrm{L} = W_0$,generalized from 3.27) now becomes

$$A_\mathrm{L} J_{\mathrm{t}n0} P_\mathrm{tC}^{(\mathrm{UHL})} \hat{S} = A_\mathrm{G} \left(1 - e^{-\hat{\alpha}}\right) \frac{q n_\mathrm{i}}{2\tau_0} (W_0 - W_{\mathrm{g}0}), \tag{4.12}$$

and the critical tunneling probability is found to be

$$P_\mathrm{tC}^{(\mathrm{UHL})} = \frac{A_\mathrm{G}}{A_\mathrm{L}} \frac{\left(1 - e^{-\hat{\alpha}}\right) q n_\mathrm{i}}{2 J_{\mathrm{t}n0} \hat{S} \tau_0} (W_0 - W_{\mathrm{g}0}). \tag{4.13}$$

Compared to the planar case in (3.30),

$$\begin{aligned} P_\mathrm{tC}^{(\mathrm{UHL})} &= \frac{A_\mathrm{G}}{A_\mathrm{L}} P_\mathrm{tC}^{(\mathrm{planar})} \\ &> P_\mathrm{tC}^{(\mathrm{planar})}. \end{aligned} \tag{4.14}$$

That is, UHL devices, in the DD regime, are equipped with a higher tunneling probability through the thin oxide compared to planar devices. The critical tunneling probability is now dependent on the low area proportion, but still independent of d_L (cf. $P_\mathrm{tC}^{(\mathrm{planar})}$ is independent of d_ox.) Consequently, the definition of the characteristic oxide thickness d_ox^* for planar devices (3.32) is modified to the *characteristic low oxide thickness*:

$$\begin{aligned} d_\mathrm{L}^* &= \frac{\Phi_\mathrm{b} \ln\left(1/P_\mathrm{tC}^{(\mathrm{UHL})}\right)}{B} \\ &= d_\mathrm{ox}^{*(\mathrm{planar})} - \frac{\Phi_\mathrm{b}}{B} \ln \frac{A_\mathrm{G}}{A_\mathrm{L}}, \end{aligned} \tag{4.15}$$

which is smaller than the planar case and would further decreases with decreasing

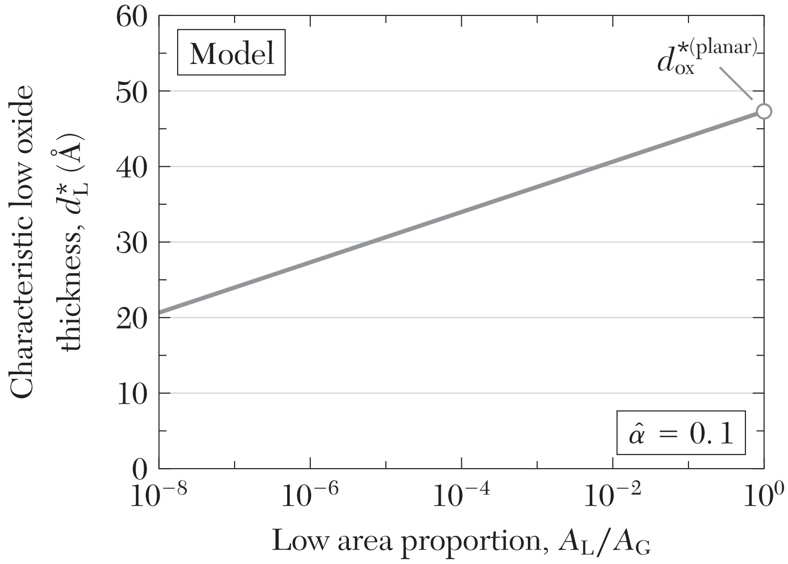


Figure 4–6. Semilogarithmic $d_{\mathrm{L}}^{*} - (A_{\mathrm{L}}/A_{\mathrm{G}})$ plot according to (4.15).

$A_{\mathrm{L}}/A_{\mathrm{G}}$. Thereby, a smaller $A_{\mathrm{L}}/A_{\mathrm{G}}$ corresponds to a larger $\mathscr{D}_{\mathrm{ox}}$ (eq. 3.38, here $\mathscr{D}_{\mathrm{ox}} \equiv d_{\mathrm{L}}/d_{\mathrm{L}}^{*}$) and a greater horizontal shrinkage (curve shrinks to the left) in the $V_{\mathrm{C}} - d_{\mathrm{L}}$ characteristics (**Figure 3–17**). In other words, among UHL devices with the same A_{G}, d_{H}, and d_{L}, those with smaller A_{L} will demonstrate higher V_{C} and mitigated DD, as qualitatively observed in **Figure 2–18** . Furthermore, (4.15) implies linearity between d_{L}^{*} and $\log_{10}(A_{\mathrm{G}}/A_{\mathrm{L}})$:

$$\frac{\Delta d_{\mathrm{L}}^{*}}{\Delta \log_{10}(A_{\mathrm{G}}/A_{\mathrm{L}})} = -\frac{\Phi_{\mathrm{b}} \ln 10}{B}$$

$$= -3.33\,\text{Å/decade}. \tag{4.16}$$

Figure 4–6 depicts the semilogarithmic $d_{\mathrm{L}}^{*} - (A_{\mathrm{L}}/A_{\mathrm{G}})$ plot according to (4.15) that approaches the $d_{\mathrm{ox}}^{*(\mathrm{planar})}$ value in **Section 3.5** as $A_{\mathrm{L}} = A_{\mathrm{G}}$; i.e., as the oxide is planar and low.

For the simulated UHL-M device (**TABLE 2–I**) as an example, d_{L}^{*} is evaluated as 36.6 Å at $\hat{\alpha} = 0.1$, and the $V_{\mathrm{C}} - d_{\mathrm{ox}}$ curve, as mentioned above, will shrink to the

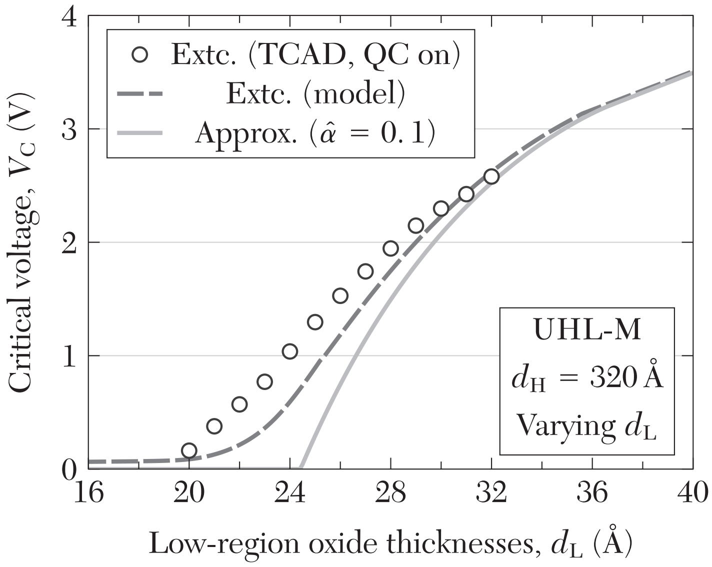


Figure 4–7. Comparison of the $V_C - d_L$ characteristics extracted from the asymptotes of $\psi_{SL} - V_G$ characteristics of TCAD simulation results (**Figure 2–22**), extracted from those of the modeled $\psi_{SL} - V_G$ curves (**Figure 4–3**), and evaluated from the approximate closed-form expression (3.40), respectively, for a UHL-M device. TCAD simulation data were extracted at $r = 0$ and $y = 1.5\,\text{nm}$.

left with respect to the planar MIS TD with $d_{ox}^* = 47.3\,\text{Å}$. **Figure 4–7** compares the $V_C - d_L$ characteristics extracted from the asymptotes of $\psi_{SL} - V_G$ characteristics (see **Figure 3–5**) of TCAD simulation results (**Figure 2–22**), extracted from those of the modeled $\psi_{SL} - V_G$ curves (**Figure 4–3**), and evaluated from the approximate closed-form expression (3.40) using $d_L^* = 36.6\,\text{Å}$ and $\hat{\alpha} = 0.1$, respectively. The TCAD data points' locating on the left of the modeled curve can be attributed to QC and is also visible in the planar MIS TD $V_C - d_{ox}$ plots (**Figure 3–5**). On the other hand, the discrepancy between the modeled and approximated curves at a relatively small V_G has also been observed in the planar model (**Figure 3–17**) and addressed in **Section 3.5**. Nonetheless, the $V_C - d_L$ trend has successfully been replicated by both the modified UHL reverse-bias electrostatics model and the closed-form approximation.

doi:10.6342/NTU202304241

4.4 Summary

Several qualitative findings in **Chapter 2** regarding reverse-biased UHL device electrostatics have been explained and verified in this chapter by means of quantitative modeling. First, there exists an upper limit for ψ_{SH} under extreme DD. Second, the lateral coupling length (Λ) for electron QFL in UHL devices are in the millimeter scale or beyond. As a result, electron coupling is highly effective ($\eta_{\mathrm{C}} \rightarrow 1$) for practical device dimensions, yielding forcibly-aligned electron QFLs across the high and low regions, and also the A_{G}-dependent substrate generation currents.

Next, the reverse-bias electrostatics model for planar devices (**Chapter 3**) has been generalized to apply to the UHL devices. An approximate closed-form $V_{\mathrm{C}} - d_{\mathrm{L}}$ relation has been developed and compared to that of planar MIS TDs. Mathematically, it has been demonstrated that reducing device $A_{\mathrm{L}}/A_{\mathrm{G}}$ ratio will shrink the $V_{\mathrm{C}} - d_{\mathrm{L}}$ horizontally towards the left of the plot, indicating mitigated DD. Finally, the TCAD-extracted, model-extracted, and approximated $V_{\mathrm{C}} - d_{\mathrm{L}}$ curves for the UHL device have been compared in the same plot, showing acceptable accuracy of the developed electrostatics model.

doi:10.6342/NTU202304241

doi:10.6342/NTU202304241

5

Low-Voltage Sensor Applications for Ultra-High-Low MOS(p) Tunnel Structures Utilizing Intensified Schottky Barrier Height Modulation Effect

doi:10.6342/NTU202304241

5.1 Background

For oxide thicknesses being sufficiently thick ($\gtrsim$ 25 Å) to actuate SBHME in MIS TDs, the hole current is considerably intensified under reverse bias (1.3) and dominates the reverse current, leading to the anomalous positive $I_\mathrm{G} - d_\mathrm{ox}$ correlation (see **Section 1.2** and **Figure 1–6**). The hole currents' high responsivity upon perturbed Q_i in MIS TDs is the basis for their sensor applications. Moreover, the current output may be conveniently converted to a voltage output by very few additional circuit components. **Figure 5–1 (a)** depicts the circuit symbol for an MIS TD. **Figure 5–1 (b)** shows a sample circuit for biasing the MIS TD through a load resistor

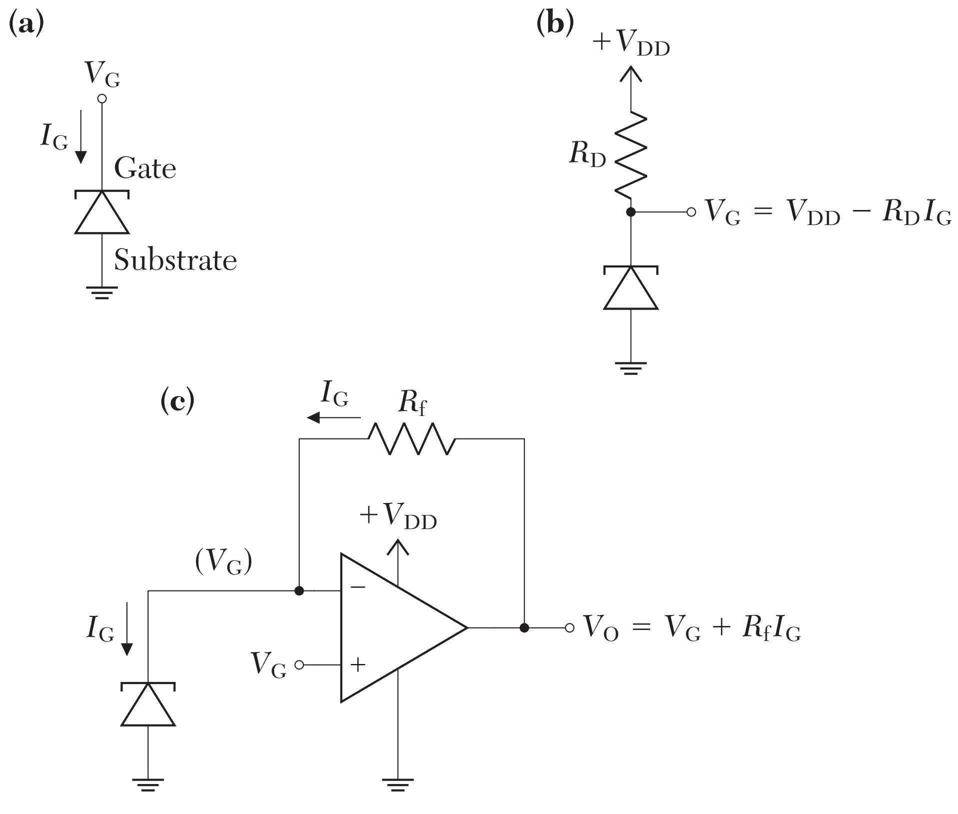


Figure 5–1. **(a)** Circuit symbol for an MIS TD. **(b)** An MIS TD sensor circuit with voltage output by a load resistor. **(c)** An MIS TD sensor circuit with voltage output by a transresistance amplifier from a fixed V_G.

to acquire a voltage output. While V_{G} is not fixed in this case, the circuit may benefit from its simplicity. **Figure 5–1 (c)** shows another sample circuit that utilizes a transresistance amplifier to keep V_{G} constant and also acquire a voltage output from the MIS TD current. MIS TD sensors are especially competitive for their extreme low costs.

Unfortunately, under small V_{G} ($\leq 0.3\,\mathrm{V}$), MIS TDs, regardless of d_{ox}, suffer from low current— either the current saturates to a low magnitude for thinner oxides, or the current is still far below the saturation value (due to high saturation voltages) for thicker oxides, as it can be observed from **Figure 1–6**. Consequently, low-voltage sensor applications may be hampered for MIS TDs.

As will be shown in the following sections, UHL MOS tunnel structures can exhibit comparable saturation currents to a planar MIS TD with the same A_{G}, while the saturation voltages are significantly reduced. Accordingly, under a low V_{G}, both the current and the sensor responsivity are improved. This makes UHL devices a better candidate for low-voltage sensor applications over planar MIS TDs.

5.2 Experimental Details

The entire device fabrication process was identical to that in **Section 2.2**, except that the second ANO process was lengthened to 4.5 minutes to obtain slightly thicker d_{L}. The device shapes and the naming convention follows **Figure 2–1**. There were only the high-only, the low-only, the UHL-100, and the UHL-200 devices fabricated. A cross-sectional TEM image of the UHL-200 device (**Figure 1–11**) reveals that d_{H} = $400\,\text{Å}$ and $d_{\mathrm{L}} = 31\,\text{Å}$.

Device $I - V$ and $C - V$ characteristics were measured with an Agilent® B1500A Semiconductor Device Analyzer. Measurements at elevated temperatures were con-

doi:10.6342/NTU202304241

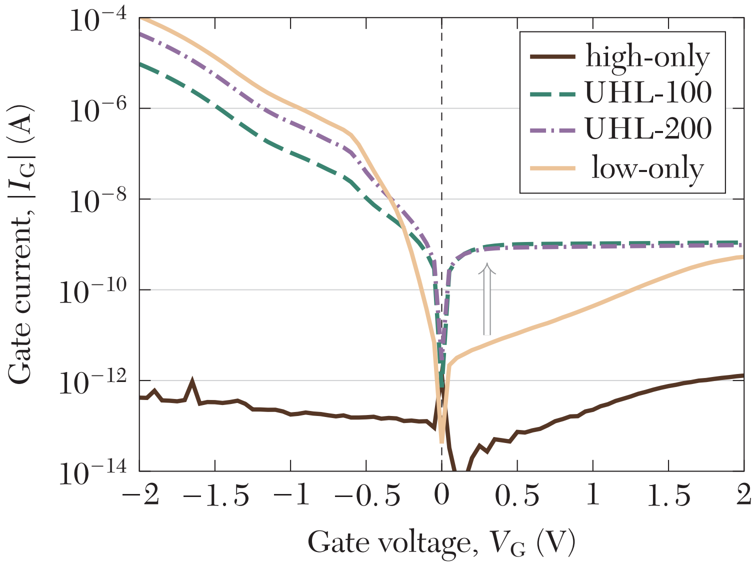


Figure 5–2. Current-voltage characteristics for the high-only, low-only, and UHL devices. The ultrathin oxide (d_L = 31 Å) is sufficiently thick to actuate SBHME for holes, causing late current saturation in the low-only device. However, UHL currents readily saturate to a high magnitude under low V_G.

ducted on a hot chuck with a temperature monitoring and control system. Illumination was achieved by an overhead high-power white light-emitting diode (LED) connected to a constant current source. The light intensity was tuned by varying the current source output. The illuminance was monitored by a lux meter. TCAD simulations were conducted with SILVACO® ATLAS™.

5.3 Results and Discussion

5.3.1 Electrical Characteristics and Schottky Barrier Height Extraction

Figure 5–2 shows the $I - V$ characteristics for the high-only, low-only, UHL-100, and UHL-200 devices. Here the UHL devices are equipped with relatively large A_L/A_G ratios (11 % and 44 %, respectively). Compared to the devices in **Figure 2–2** with thinner d_{L} and thus no SBHME, in this figure, the low-only (planar MIS TD) device demonstrates ascending I_{G} (not yet saturated) under positive V_{G}. This indicates the presence of SBHME and corresponds to a thicker d_{L}, which is consistent with the TEM image showing physical $d_{\mathrm{L}} = 31\,\text{Å}$. The saturation voltage is $\sim +2\,\mathrm{V}$. On the contrary, the high-only device expectedly presents negligible I_{G}.

The UHL devices under negative (forward) bias, together with the low-only device, exhibit a positive correlation between $|I_{\mathrm{G}}|$ and A_{L}, in accordance to **Figure 2–2**. Yet, the UHL devices under reverse bias present similar magnitudes in I_{sat} to the low-only device, but with significantly smaller saturation voltages. Their currents readily saturate like MIS TDs with thinner oxides, but to high magnitudes like those with thicker oxides. That is, compared to the low-only device with SBHME, the UHL devices demonstrate similar $I - V$ characteristics, except that they undergo early current saturation at positive V_{G}.

Such an early current saturation behavior may be attributed to a further reduced hole SBH, in addition to the SBHME for thicker oxides, near the boundary of the low region, which may be traced back to lateral electron coupling at the boundary. **Figure 5–3 (a)** depicts the schematic cross section of a UHL device about the high-low boundary under a very low V_{G}. With the presence of the lateral built-in field $\mathscr{E}_r$ (see **Section 2.3**) that is capable of drifting electrons from the high region, electrons that pile up at the boundary may increase the V_{ox} at the low region boundary with respect to the center value. **Figure 5–3 (b)** and **Figure 5–3 (c)** compare the schematic band diagrams across the low oxide of **Figure 5–3 (a)**, at far away from the boundary (center value), and near the boundary, respectively. The increment of V_{ox} indicates a re-

doi:10.6342/NTU202304241

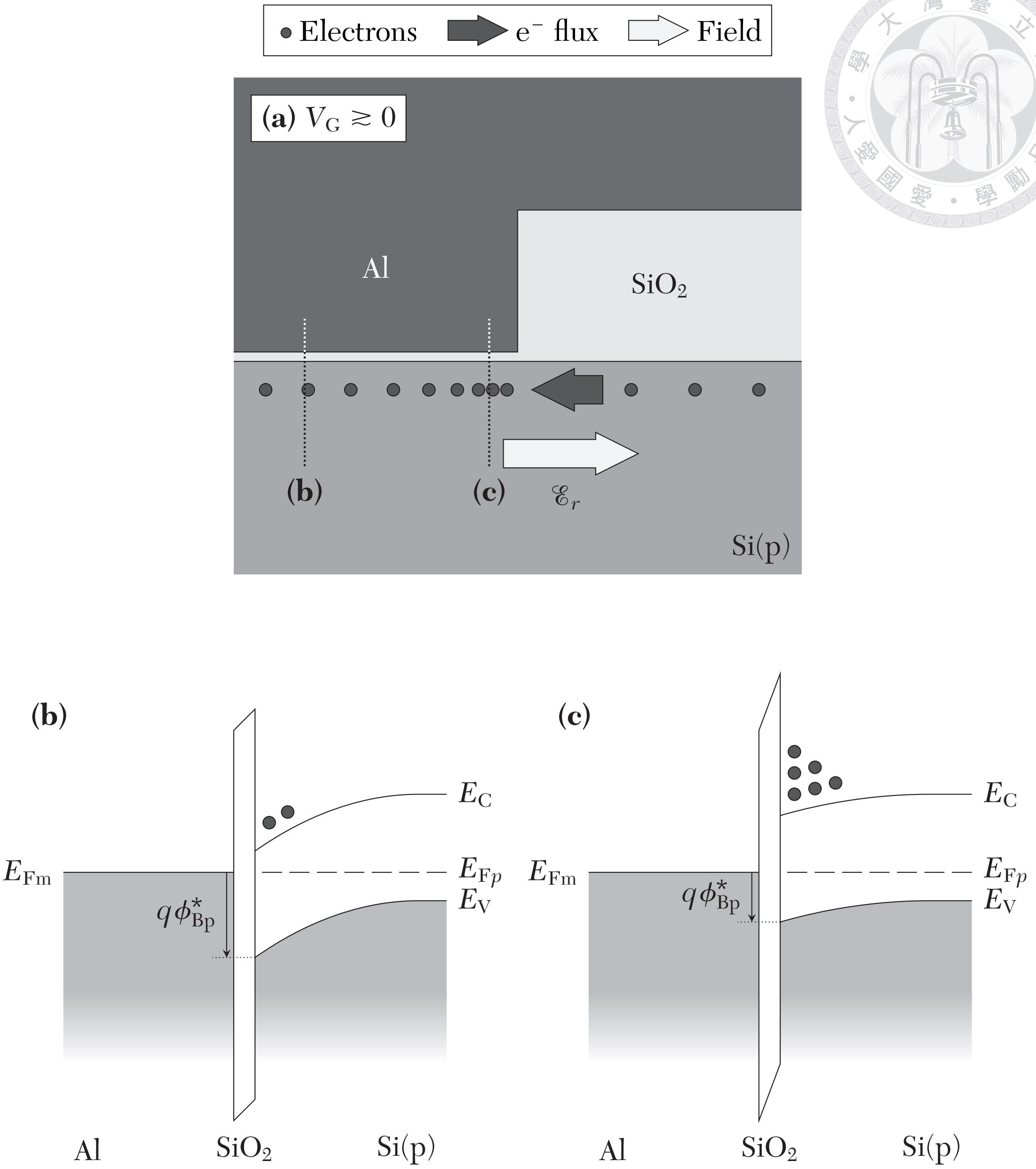


Figure 5–3. **(a)** Schematic cross section of a UHL device about the high-low boundary under a very low V_G, featuring the lateral field $\mathscr{E}_r$ and electron concentrations. **(b)** Schematic band diagram across the low region, far away from the boundary. **(c)** Schematic band diagram across the low region, near the boundary, showing reduced $q\phi^*_{Bp}$.

duction in the hole SBH (1.3), $q\phi^*_{Bp}$, near the boundary. **Figure 5–4** further shows the simulated depletion edge profile in a UHL device under zero applied bias. The smaller depletion width near the high-low boundary (W_B) compared to the center value (W_L)

doi:10.6342/NTU202304241

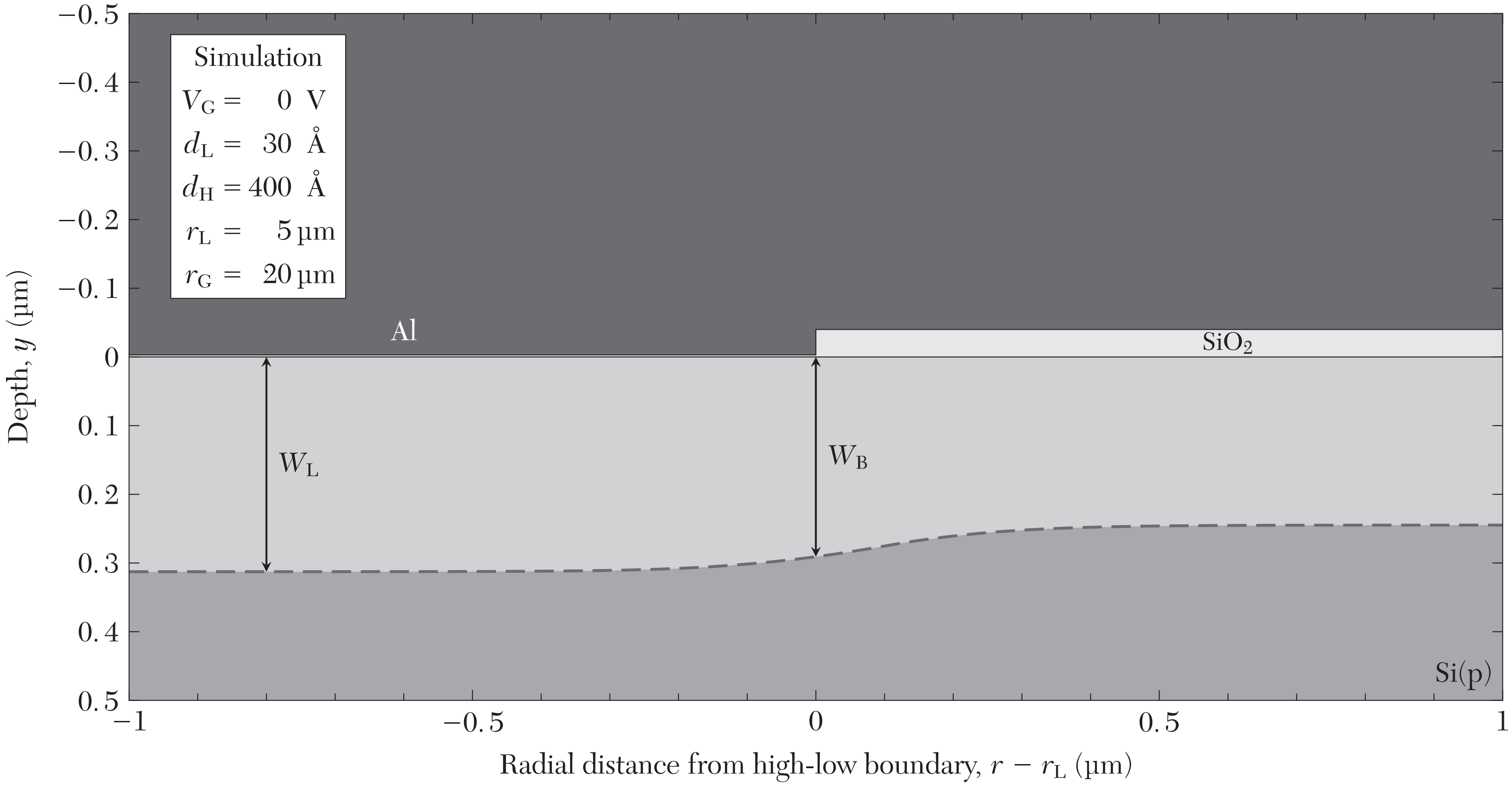


Figure 5–4. Simulated depletion edge profile (dashed line) in a UHL device with $d_L = 30$ Å and $d_H = 400$ Å, under zero applied bias. The depletion width near the high-low boundary, W_B, is less than that in the center of low region, W_L. As a result, ψ_S is also lower at the boundary of the low region. Here the depletion edge is defined as where $p = \frac{1}{2}p_{p0}$.

doi:10.6342/NTU202304241

of the low region consistently implies a smaller ψ_{S}, a larger V_{ox}, and a reduced $q\phi^{*}_{\mathrm{Bp}}$ at the boundary.

Analogously, in planar MIS TDs subject to SBHME, it has been observed that hole SBH is slightly lower at the gate edge with respect to the gate center, known as the fringing field effect (FFE) [60,61]. Such a subtle SBH reduction that increases the hole tunnel current density at the gate edge can lead to the edge hole current's domination over the center hole current. As a consequence, the I_{sat} in an MIS TD with SBHME is primarily the hole tunnel current at the gate edge, and therefore dependency of I_{sat} on the gate perimeter has been observed. Likewise, it is believed that the I_{sat} in a UHL device with SBHME is also dependent on the low-region perimeter to some extent.

Hence, one may conclude that the UHL device current is boosted by the reduction of hole SBH near the boundary, under low V_{G}, due to lateral coupling of electrons. This SBH lowering mechanism is absent in the low-only device, rendering its current to be incomparable. Nonetheless, this mechanism comes to an end as V_{G} continues to rise and the high region undergoes extreme DD, curtailing the electron supply and saturating I_{G}. Finally, under a high positive V_{G}, SBH reduction at the gate edge/high-low boundary is comparable among the low-only/UHL devices. Also, gate and low-region perimeters are also comparable for this specific set of experimental devices. This ultimately leads to like magnitudes of I_{sat} among the UHL-100, UHL-200, and low-only devices under a high V_{G} (**Figure 5–2**).

Moreover, direct extraction of the edge/boundary SBHs (in low-only/UHL devices) is possible. **Figure 5–5** shows the Arrhenius plot [62] for the device conductances ($\mathcal{G} \equiv \mathrm{d}I_{\mathrm{G}}/\mathrm{d}V_{\mathrm{G}}$) at zero applied bias, achieved by measuring the device $I - V$ char-

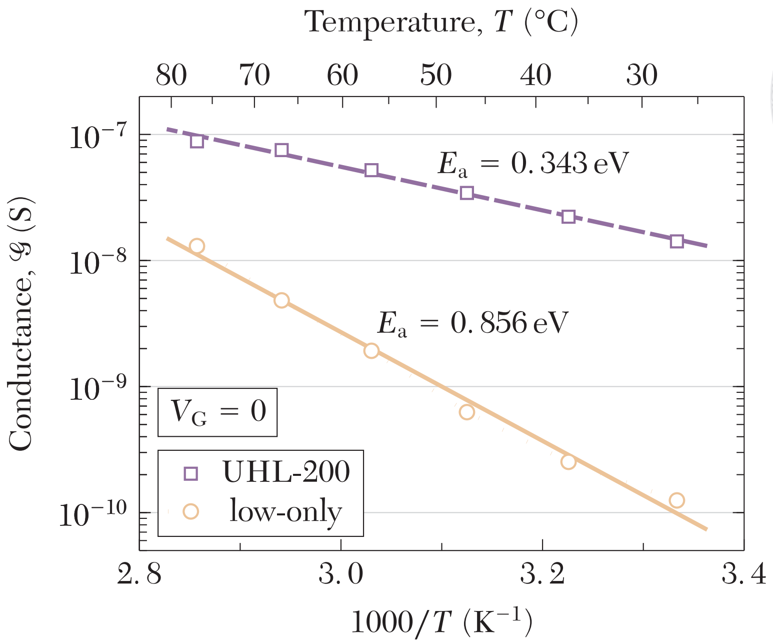


Figure 5–5. Arrhenius plot for the device conductances (UHL-200 and low-only) at zero applied bias. The smaller E_a associated with the UHL device may be attributed to a lower hole SBH.

acteristics under various elevated temperatures. Knowing that the dominant current component is associated with the edge/boundary SBH, it can be estimated from (1.3) that

$$\mathscr{G} \sim \exp\left(-\frac{q\phi^*_{\mathrm{Bp}}}{k_B T}\right), \tag{5.1}$$

and thus $q\phi^*_{\mathrm{Bp}}$ values can be estimated from the activation energies (E_a) in the Arrhenius plot:

$$q\phi^*_{\mathrm{Bp}} \sim E_a. \tag{5.2}$$

Compared to the low-only device, in addition to the significantly higher conductance at all temperatures for the UHL-200 device (thus its higher I_G at low V_G), it also demonstrates a significantly smaller E_a, which is an evidence of its intensified SBHME near the boundary of the low region.

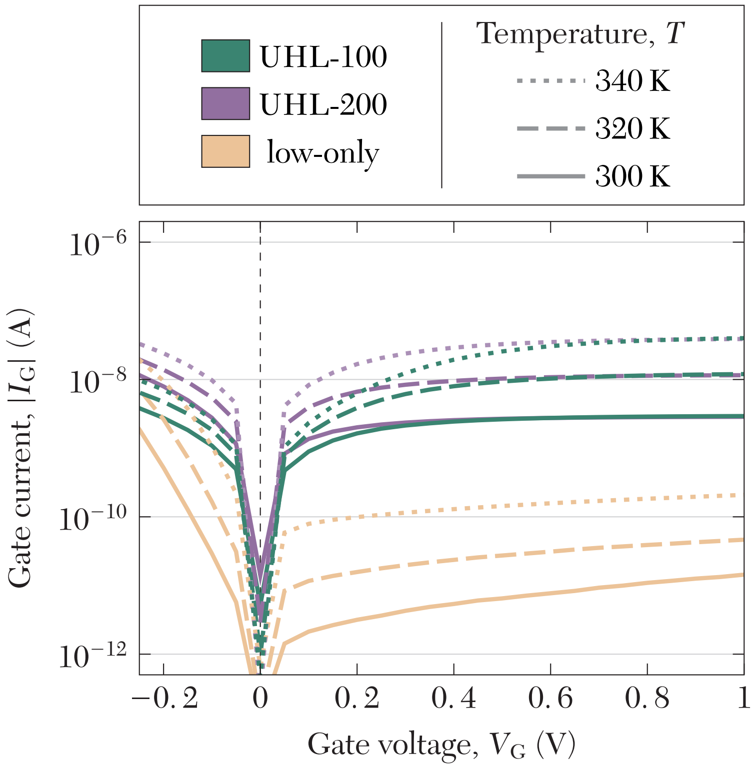


Figure 5–6. $I - V$ characteristics for the low-only and UHL devices upon a temperature rise from 300 K to 340 K.

5.3.2 Temperature Sensor Applications

An MIS TD can be utilized as a temperature sensor [21], and so can a UHL device. The reverse saturation current for an MIS TD is positively correlated and highly sensitive to the temperature. Upon a temperature rise, the increment in Q_i [63] leads to a reduction in hole SBH [21] and thus a rise in I_{sat}. **Figure 5–6** shows the $I - V$ characteristics for the low-only and UHL devices at three different temperatures. All device currents under $V_G > 0$ increase by ~ 1 decade as the temperature rises from 300 K (27 °C) to 340 K (67 °C), showing high sensitivity. However, under a low positive V_G (say $+0.3$ V), the intrinsically larger I_G in a UHL device leads to a correspondingly higher increment in its magnitude (ΔI_G). **Figure 5–7 (a)** shows the I_G vs. temperature plot for the low-only and UHL devices under $V_G = +0.3$ V. The duller I_G sensitivity towards

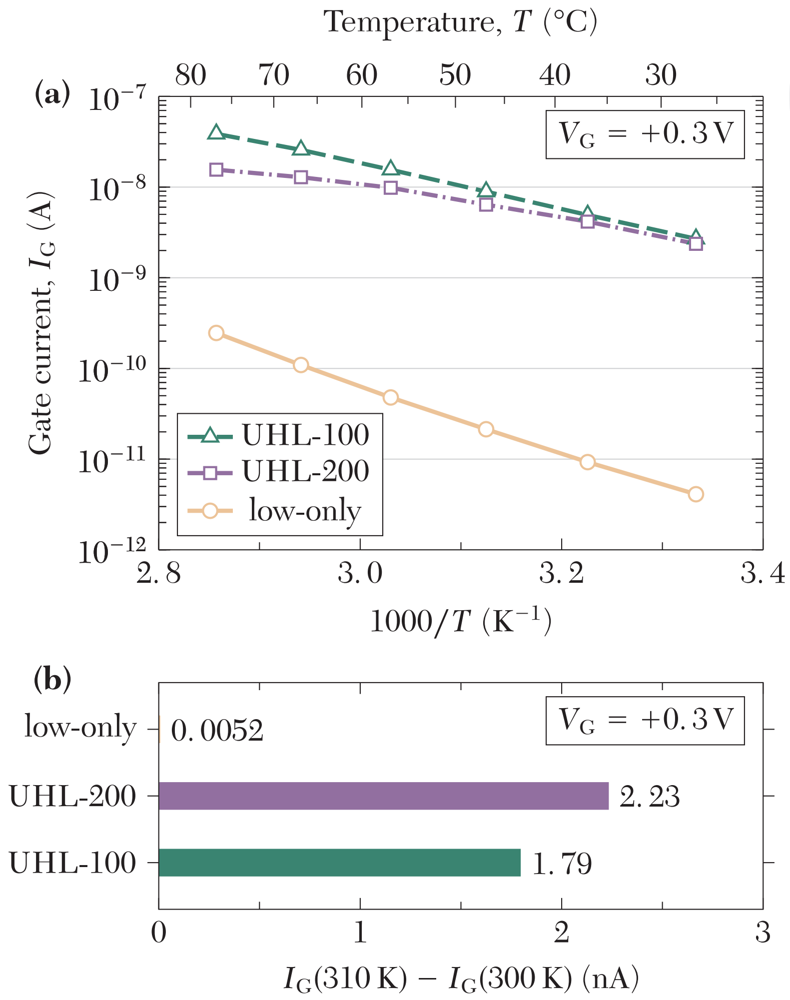


Figure 5–7. **(a)** Device I_G under +0.3 V vs. temperature (Arrhenius plot) for the low-only and UHL devices. **(b)** Comparison of change in device currents in **(a)** upon a temperature rise from 300 K to 310 K, showing $> 300\times$ improvement in ΔI_G for the UHL devices compared to the MIS TD.

temperature change for the UHL devices (smaller slopes in the semilogarithmic plot) may be associated with **Figure 5–5** and attributed to a lower hole SBH with respect to the low-only device. However, their responsivities ($\propto \Delta I_G$) are still considerably higher than that of the low-only device. **Figure 5–7 (b)** compares the change in I_G upon a temperature rise from 300 K to 310 K for the devices in **Figure 5–7 (a)** and the bias condition thereof. The slightly higher ΔI_G for the UHL-200 device compared to UHL-

doi:10.6342/NTU202304241

100 may be ascribed to a larger low region perimeter ($4b = 800\,\mu\mathrm{m}$ vs. $400\,\mu\mathrm{m}$, see **Figure 2–1** .) Overall, a $> 300\times$ improvement in ΔI_{G} (and responsivity) with respect to the low-only structure has been achieved by the UHL structure, while A_{G} remains the same.

5.3.3 Ambient Light Sensor Applications

An MIS TD can be utilized as a light sensor or photodetector [23], and so can a UHL device. In fact, rather than being regarded as a photodiode, the operation of an MIS TD as a light sensor involves minority-controlled majority current amplification and is much like a phototransistor [64]. **Figure 5–8 (a)** depicts the band diagram of an illuminated bipolar homojunction pnp phototransistor, biased under $V_{\mathrm{EC}} > 0$. The photogenerated electrons from the collector-base junction feeds to the base and amplifies the hole emitter current. Correspondingly, **Figure 5–8 (b)** depicts the band diagram of a reverse-biased MIS TD under illumination. The photogenerated electrons from the depletion layer reduces the hole SBH ($q\phi^{*}_{\mathrm{Bp}}$) and analogously boosts the hole current.

Figure 5–9 shows the $I - V$ characteristics for the low-only and UHL devices, under dark condition or illuminated at an illuminance (e_{V}) of 100 lx. Similar to **Figure 5–6** , all device currents increase by ~ 2 decades, and the UHL devices with higher intrinsic I_{G} therefore exhibit the highest "photocurrents", $I_{\mathrm{ph}} \equiv I_{\mathrm{G}}^{(\mathrm{illuminated})} - I_{\mathrm{G}}^{(\mathrm{dark})}$. The similar multiplication factors in I_{G} pre-and-post illumination may be attributed to a similar magnitude in hole SBH reduction among all devices by the illumination. (From eq. (1.3), $I'_{\mathrm{G}}/I_{\mathrm{G}} \sim \exp(-\Delta\phi^{*}_{\mathrm{Bp}}/\phi_{\mathrm{t}})$.) **Figure 5–10 (a, b)** shows in-depth the $I - V$ curves for the low-only and UHL-200 devices, respectively, under various illumination levels. For the low-only device, the dark $I_{\mathrm{G}} - V_{\mathrm{G}}$ curve possesses the steepest slope, and the

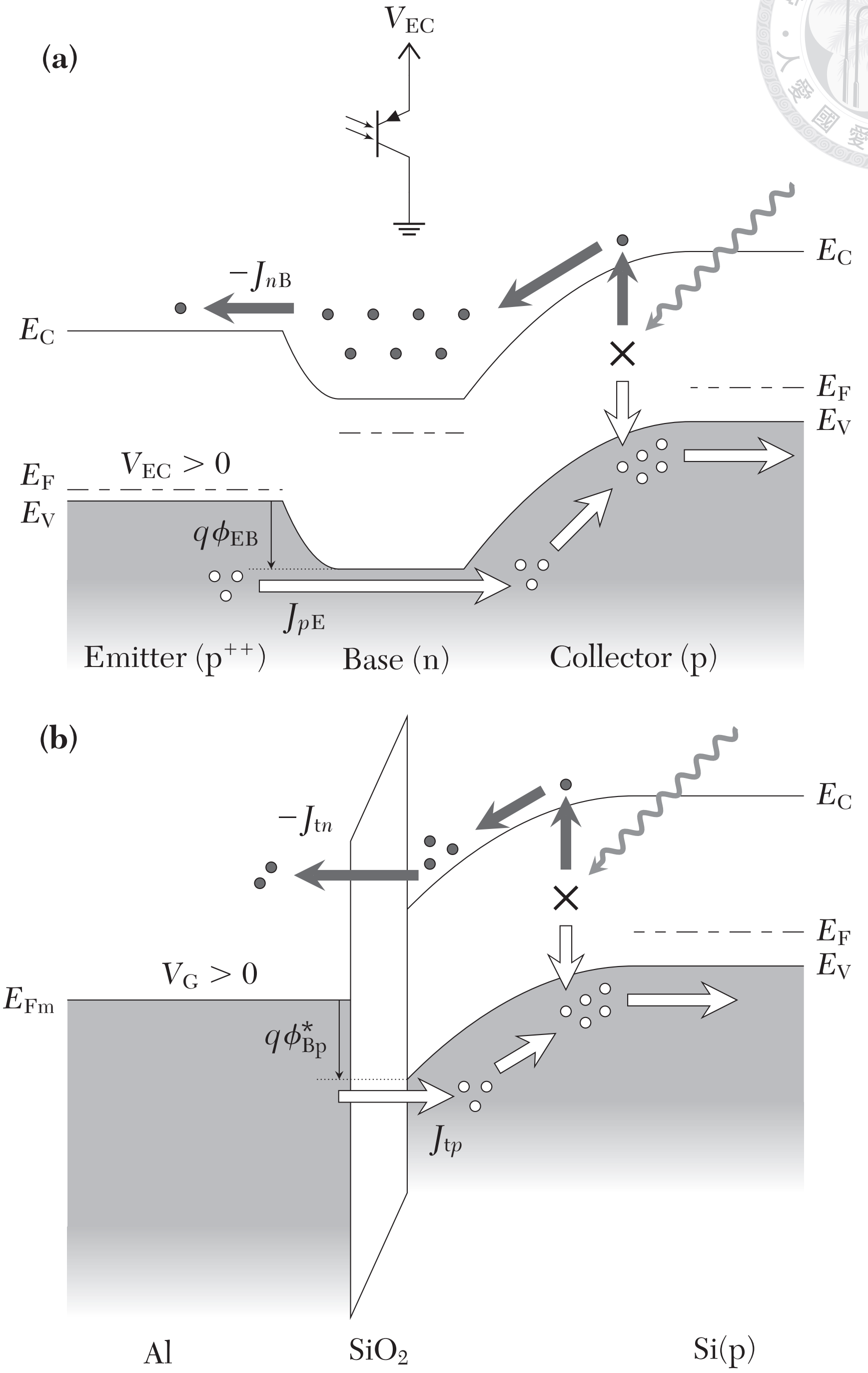


Figure 5–8. **(a)** Schematic band diagram of an illuminated bipolar homojunction pnp phototransistor under $V_{\mathrm{EC}} > 0$. The barrier for holes ($q\phi_{\mathrm{EB}}$) and the hole current are controlled by base electrons. Inset shows the circuit symbol. **(b)** Schematic band diagram of an MIS(p) TD under $V_{\mathrm{G}} > 0$ and illumination. The barrier for holes ($q\phi^{*}_{\mathrm{Bp}}$) and the hole current are controlled by the inversion charge.

doi:10.6342/NTU202304241

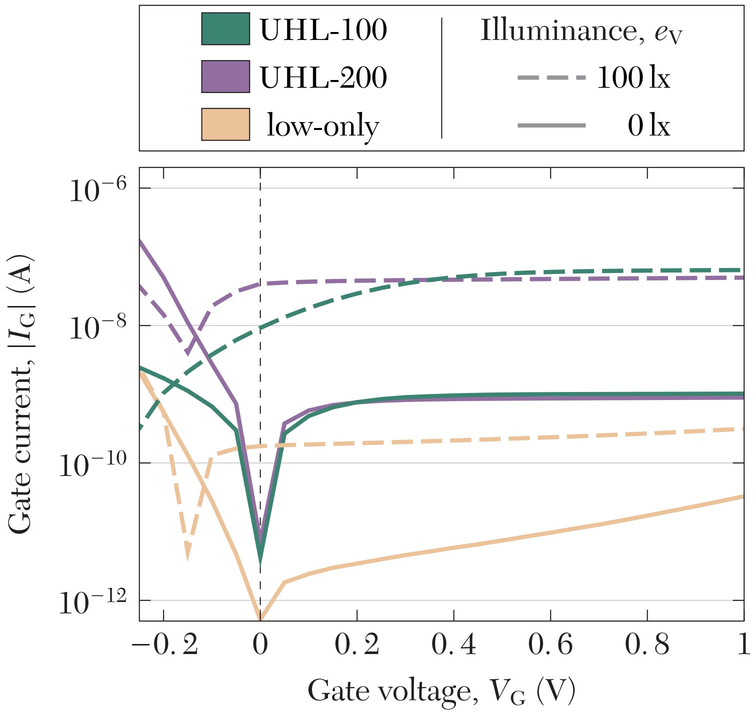


Figure 5–9. $I - V$ characteristics for the low-only and UHL devices, under (solid) dark condition or (dashed) illumination at an illuminance of 100 lx.

reduction of curve slope in the semilogarithmic plot as e_V increases may imply a constant $I_{ph} - V_G$ relation at every e_V value. On the other hand, in the UHL-200 device, both the dark $I_G - V_G$ characteristics and the $I_{ph} - V_G$ characteristics at all e_V values are obviously constant. What's more, photovoltaic effect (zero-current voltage shift under illumination) is detectable in both devices.

Figure 5–11 (a) shows the I_G vs. illuminance plots for the low-only and UHL devices, biased under a low positive voltage, $V_G = +0.2\,\mathrm{V}$. Albeit the sensitivities is slightly lower for the UHL devices, their responsivities is considerably higher, again, due to their intrinsically higher I_G. **Figure 5–11 (b)** compares the photocurrents at an illuminance of 100 lx for the devices in **Figure 5–11 (a)** and also $V_G = +0.2\,\mathrm{V}$. The slightly higher I_{ph} in the UHL-200 device with respect to UHL-100 may also be at-

doi:10.6342/NTU202304241

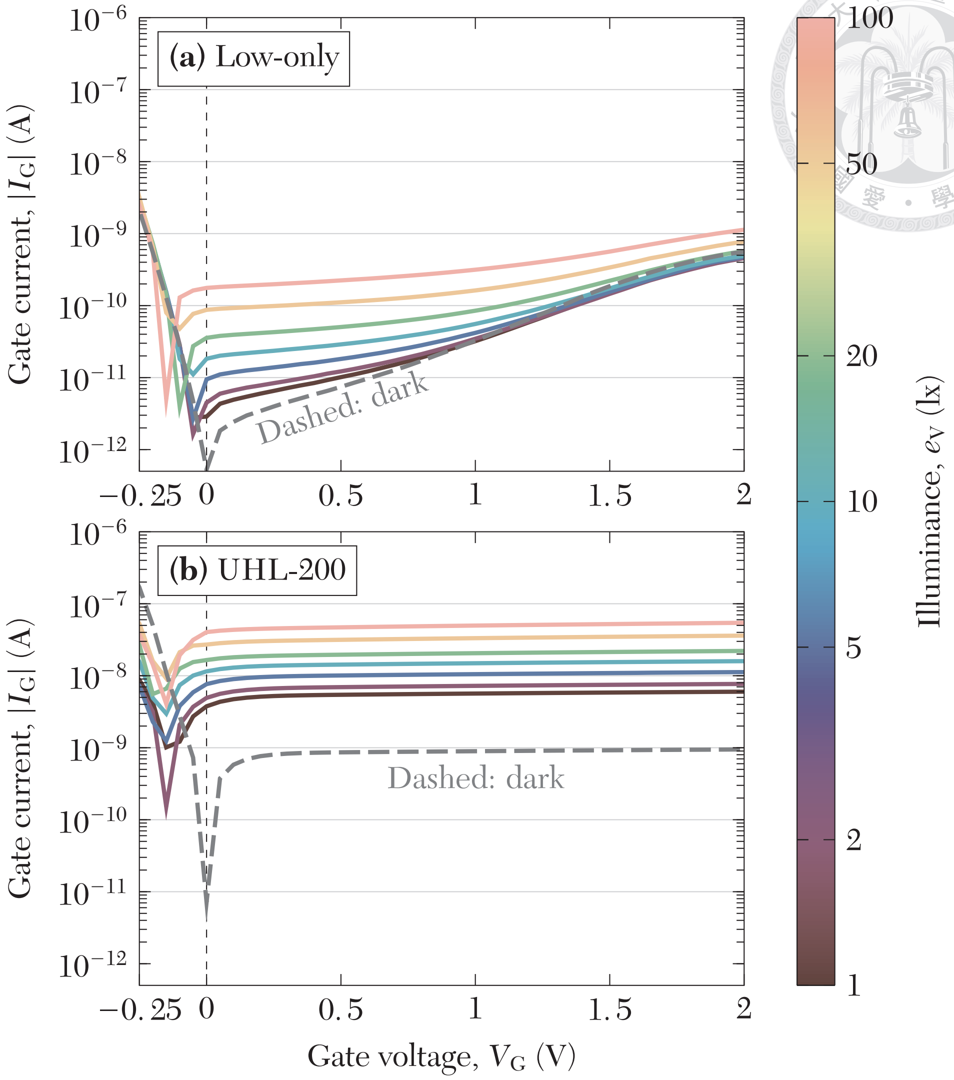


Figure 5–10. $I - V$ characteristics for **(a)** the low-only device, and **(b)** the UHL-200 device, under various illumination levels.

tributed to the longer low-region perimeter, as it has been discussed in **Figure 5–7 (b)**. Overall, compared to the low-only device, I_{ph} is improved by $> 100\times$ in the UHL-100 device, as well as in the UHL-200 device. **Figure 5–11 (b)** compares the photocurrents at an illuminance of 100 lx for the devices in **Figure 5–11 (a)**, except that $V_G = +0.6\,\text{V}$, showing also an $> 100\times$ improvement in I_{ph}. It is remarkable that the UHL-100 device now exhibits a higher I_G with respect to the UHL-200 device. Under such a higher V_G, photogenerated electrons at the gate edge may be more effectively

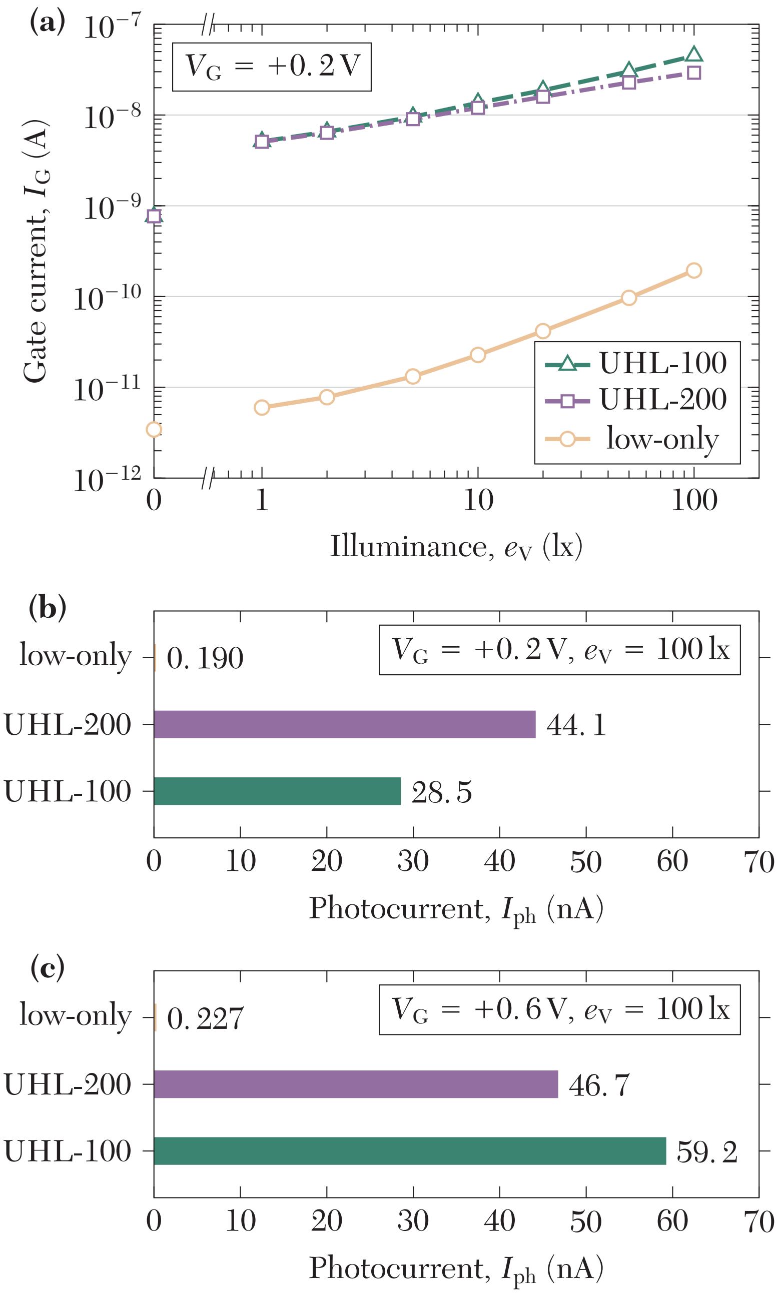


Figure 5–11. **(a)** Device I_G under $+0.2\,V$ vs. illuminance for the low-only and UHL devices. **(b)** Comparison of photocurrent in **(a)** at an illuminance of 100 lx. **(c)** Same as **(b)**, except that $V_G = +0.6\,V$.

coupled to the high-low boundary, and they would be more concentrated at the boundary for smaller A_L. This may lead to stronger SBHME and higher I_G with respect to the big-A_L devices, causing the positive $I_G - A_L$ correlation under a low V_G to revert under a higher V_G.

5.4 Summary

For d_{ox} (and d_L) being sufficiently thick to actuate SBHME in MIS TDs (and UHL MOS tunnel structures), the reverse current is dominated by hole tunnel current, and that in a UHL device exhibits an early saturation behavior compared to that in the planar device. This may be attributed to intensified SBHME near the high-low boundary due to electron coupling from the high region. The reduction of hole SBH in a UHL device, with respect to the low-only MIS TD, has been demonstrated with an Arrhenius plot through measurements at elevated temperatures. All in all, the early saturation behavior significantly boosts UHL device currents under very small gate voltages, making UHL devices a competitive candidate for low-voltage sensor applications.

Temperature and ambient light sensing performances for the experimental devices are then inspected. Upon a temperature rise from 300 K to 310 K, ΔI_G in the UHL-100 and UHL-200 devices under $V_G = +0.3\,\mathrm{V}$ are $> 300\times$ greater than that in the low-only device. Also, upon an illumination at 100 lx, the UHL photocurrents under $+0.2\,\mathrm{V}$ are boosted $> 100\times$. In conclusion, under a small V_G, the I_G responsivities upon temperature and illumination changes are indeed significantly improved in the UHL devices with unchanged A_G.

doi:10.6342/NTU202304241

doi:10.6342/NTU202304241

6

Conclusions and Future Works

6.1 Conclusions

In this dissertation, the electrical characteristics, electrostatics, and carrier transport phenomena in Al/SiO$_2$/Si(p) UHL MOS tunnel structures have been comprehensively studied through physical device characterization and TCAD simulations, specifically under reverse bias. It has been demonstrated how a tiny LOT spot ($A_\mathrm{L}/A_\mathrm{G} \sim 10^{-5}$) can prompt exhaustive DD in the substrate and lead to an I_G comparable to that in planar MIS TDs. An effective lateral inversion charge coupling mechanism, which is believed responsible for these phenomena, has also been proposed and examined.

Next, for quantitative study on the UHL electrostatics, an analytical electrostatics model has been proposed for the planar MIS TDs first and then generalized into the UHL structures. The device electrostatics were found to be demarcated by some critical gate voltage, before which the CEST applies, and after which DD takes over. The modeled $V_\mathrm{C} - d_\mathrm{ox}$ relation for planar devices and $V_\mathrm{C} - d_\mathrm{L}$ relation for UHL devices agrees with respective TCAD simulation results and qualitatively reproduce the experimental device characteristics. Moreover, modeling of the aforementioned charge coupling mechanism suggests a coupling length beyond the millimeter scale and coupling efficiencies greater than 99 % for any practical device dimensions. The models are believed helpful for better understanding of MIS TDs and UHL MOS tunnel structures in a quantitative aspect.

Finally, with appropriate oxide thicknesses that actuate SBHME and make the device reverse currents dominated by the hole tunnel currents, we proposed that the intensified SBHME at the high-low boundary of UHL devices boost their currents under low positive V_G with respect to planar MIS TDs. This drastically improves the low-voltage sensing performance for a UHL device over a planar MIS TD. E.g., upon a temperature rise from 300 K to 310 K, the responsivity can improve over 300×. When illuminated at $e_V = 100\,\mathrm{lx}$, the photocurrent can be intensified over 100×. In conclusion, UHL MOS tunnel structures are competitive candidates for low-voltage sensor applications.

6.2 Suggestions for Future Works

There are several findings and shortcomings in this dissertation that require further studies in future works.

While TCAD simulations have been conducted over many combinations of device dimensions (r_L, r_G, d_L, and d_H), the combinations of experimental device dimensions fall short. Also, in the experimental devices, the metal gates and low regions were in square shapes. These may be replaced with circular shapes for better accordance with the TCAD-simulated devices.

doi:10.6342/NTU202304241

In modeling the electrostatics of planar MIS TDs, some assumptions may be overly ideal; e.g., by not considering QC, D_{it}, Q_{eff}, and effects of holes on the electrostatics. Incorporating the nonidealities into the electrostatics model (and TCAD) may mitigate discrepancies between the model-predicted and experimental characteristics in the cost of model complexity and intuitiveness. In addition, the systematic d_{ox} overestimation by the model ($\sim 3\,\text{Å}$) as indicated in **Figure 3–14** deserves more investigation. Last, the closed-form approximation for the $V_{\mathrm{C}} - d_{\mathrm{ox}}$ relation suffers from relatively high inaccuracy at low V_{C}, which was the consequence of an assumption in simplifying S (3.29). A better closed-form approximation that alleviates the inaccuracy may be developed in future works.

In modeling the electrostatics and electron QFL profiles in the UHL devices, all physical quantities were implicitly assumed to be r-independent within the low region. While this is seemingly true from observing the TCAD simulation results, the assumption's validity requires in-depth inspection in the future. Moreover, the mathematical derivation of η_{C} requires ψ_{SH} to be a constant in the high region. This is true if the high region undergoes extreme DD ($\psi_{\mathrm{SH}} = \psi_{\overset{\wedge}{\mathrm{SH}}}$) but becomes questionable otherwise. A coupled $\psi_{\mathrm{S}}(r)$-and-η_{C} submodel may be developed in future works for better model accuracy.

Albeit it is not out objective in this dissertation to model the forward-bias electrostatics, it may be conducted in future works.

Quantitative analyses on the device hole currents subject to SBHME and other findings in **Chapter 5** are yet to be performed. This may be fulfilled in future works for better understanding of the UHL devices' improved sensing performance under low applied bias.

Appendix A

Minority Carrier Quasi Fermi Level in Space Charge Region Considering Shockley-Read-Hall Generation

---

A.1 Motivation

In the presence of gate leakage in MOS structures, modeling the profile of minority carrier quasi Fermi levels (QFLs) in the space charge region is particularly important. Only from a well-modeled QFL profile can we evaluate the minority carrier current, and ultimately the electrostatics of the device itself. In **Section 3.2**, as in most literature [30], minority carrier QFLs were assumed to be flat at sufficiently long car-

doi:10.6342/NTU202304241

rier lifetimes [56]. However, the definition of the “flatness” were quite vague, and the influence of a presumably non-flat QFL on the minority carrier current is yet to be addressed. What’s worse, in reverse-biased P-N junctions and MOS tunnel structures, the splitting of QFLs facilitate carrier generation, whose feedback influence on the QFLs themselves is still unknown. Sah et al. (1957) proposed a model for variations of QFLs in the space charge region of P-N junctions [18, 57], without considering the feedback influence. That is, infinite carrier lifetime was assumed. Yet, the model predicts non-flat QFLs under moderate reverse bias, making the flat-QFL assumption even more questionable.

As such, in this appendix, we propose a new minority carrier QFL model in the space charge region that takes care of the feedback influence by carrier generation, in hope of answering the puzzles above and modeling the electrostatics (**Chapter 3** and **Chapter 4**) in a more accurate manner.

A.2 Minority Quasi Fermi Level to A Linear Differential Equation Problem

Consider the p-type substrate in MOS(p) tunnel structures under reverse bias as shown in **Figure A–1**. The nonzero current across the device indicates nonequilibrium in the semiconductor and splitting of the electron and hole quasi Fermi levels ($q(\Delta\phi_p - \Delta\phi_n) \neq 0$). However, we assume $\Delta\phi_p = 0$ for the majority carrier and only care about the spatial profile of $\Delta\phi_n$.

The time-dependent diffusion equation for minority carriers (electrons) reads

$$\frac{\partial n}{\partial t} = \frac{1}{q}\nabla \cdot \mathbf{J}_n + G, \tag{A.1}$$

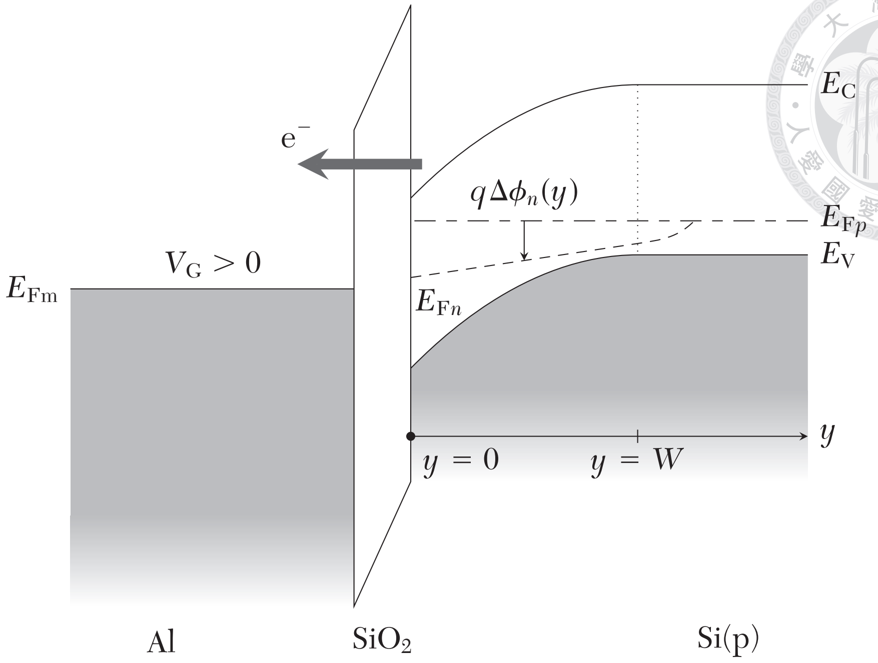


Figure A–1. **(a)** Schematic band diagram of an MOS(p) tunnel structure under reverse bias ($V_G > 0$) along the depth (y) axis. The spatial profile of $\Delta\phi_n(y)$ is to be solved. Assume flat hole quasi Fermi level throughout the substrate.

where n is the electron concentration, $\mathbf{J}_n$ is the electron current, and G is the net generation rate. Expressing $\mathbf{J}_n$ as drift and diffusion current components yields a more familiar form [19]:

$$\frac{\partial n}{\partial t} = D_n \nabla^2 n + \mu_n \left(\mathbf{E} \cdot \boldsymbol{\nabla} n + n \boldsymbol{\nabla} \cdot \mathbf{E} \right) + G, \tag{A.2}$$

where μ_n and D_n are the electron mobility and diffusion coefficient, respectively. By divergence theorem, the integration form of (A.1) may also be useful:

$$\oint_{\partial \mathbb{V}} \mathbf{J}_n \cdot \mathrm{d}\mathbf{A} = q \int_{\mathbb{V}} \left(\frac{\partial n}{\partial t} - G \right) \mathrm{d}V. \tag{A.3}$$

Noting that $\mathbf{J}_n$ can be expressed in terms of $\Delta\phi_n$ as [19]

$$\mathbf{J}_n = -q\mu_n n \boldsymbol{\nabla}(\Delta\phi_n), \tag{A.4}$$

equation (A.1) can be written as

$$\frac{\partial n}{\partial t} = -\mu_n \boldsymbol{\nabla} \cdot (n \boldsymbol{\nabla}(\Delta\phi_n)) + G. \tag{A.5}$$

Now consider the Shockley-Read-Hall process which shall dominate the generation rate in silicon [45]. For simplicity, we consider traps at the intrinsic level, then the net generation rate is given by

$$G = \frac{n_\mathrm{i}^2 - np}{\tau_{p0}(n + n_\mathrm{i}) + \tau_{n0}(p + n_\mathrm{i})}, \tag{from 2.6}$$

where τ_{p0} and τ_{n0} are excess hole and electron lifetimes. Simplifications can be made by plugging in the expressions of n and p:

$$n = n_\mathrm{p0} e^{(\psi - \Delta\phi_n)/\phi_\mathrm{t}}, \tag{A.6a}$$

$$p = p_\mathrm{p0} e^{-\psi/\phi_\mathrm{t}}, \tag{A.6b}$$

where $n_\mathrm{p0} = n_\mathrm{i}^2/N_\mathrm{A}$ and $p_\mathrm{p0} = N_\mathrm{A}$ are the electron and hole concentrations in the neutral region, respectively, ψ is the semiconductor band bending at the location, and $\phi_\mathrm{t} \equiv k_\mathrm{B}T/q$ is the thermal voltage. By also assuming low electron concentration ($n \ll$

doi:10.6342/NTU202304241

n_i in the denominator) throughout the space charge region, (2.6) becomes

$$\begin{aligned} G &\approx \frac{n_\mathrm{i}^2\left(1 - e^{-\Delta\phi_n/\phi_\mathrm{t}}\right)}{\tau_{p0}n_\mathrm{i} + \tau_{n0}(n_\mathrm{i} + N_\mathrm{A}e^{-\psi/\phi_\mathrm{t}})} \\ &= \frac{n_\mathrm{i}}{2\tau_0}\frac{1 - e^{-\Delta\phi_n/\phi_\mathrm{t}}}{1 + \kappa e^{-\psi/\phi_\mathrm{t}}}, \end{aligned} \tag{A.7}$$

in which

$$\tau_0 \equiv \frac{1}{2}\left(\tau_{p0} + \tau_{n0}\right), \tag{A.8}$$

and

$$\kappa \equiv \frac{N_\mathrm{A}\tau_{n0}}{2n_\mathrm{i}\tau_0} \tag{A.9}$$

is a relatively large dimensionless material constant. e.g., for $N_\mathrm{A} = 10^{16}\,\mathrm{cm}^{-3}$ that corresponds to the experimental devices in this study and $\tau_{n0} \approx \tau_{p0}$, $\kappa \approx 3.3 \times 10^5$.

Overall, (A.5) would now become

$$\frac{\partial n}{\partial t} = -\mu_n \nabla \cdot \left(n_\mathrm{p0}e^{(\psi-\Delta\phi_n)/\phi_\mathrm{t}}\nabla(\Delta\phi_n)\right) + \frac{n_\mathrm{i}}{2\tau_0}\frac{1 - e^{-\Delta\phi_n/\phi_\mathrm{t}}}{1 + \kappa e^{-\psi/\phi_\mathrm{t}}}. \tag{A.10}$$

This is a nonlinear, inhomogeneous differential equation for $\Delta\phi_n$ from which its spatial profile can be solved. However, linearization of this equation can be achieved by performing a change of variable

$$\boxed{u \equiv 1 - e^{-\Delta\phi_n/\phi_\mathrm{t}}}, \tag{A.11}$$

which converts (A.6a) to

$$n = n_\mathrm{p0}e^{\psi/\phi_\mathrm{t}}(1 - u), \tag{A.12}$$

doi:10.6342/NTU202304241

and therefore

$$\frac{\partial n}{\partial t} = -n_{\mathrm{p0}} e^{\psi/\phi_{\mathrm{t}}} \frac{\partial u}{\partial t}. \tag{A.13}$$

More importantly, notice that

$$\boldsymbol{\nabla} u = \frac{1}{\phi_{\mathrm{t}}} e^{-\Delta\phi_n/\phi_{\mathrm{t}}} \boldsymbol{\nabla} \left(\Delta\phi_n\right). \tag{A.14}$$

which also converts (A.4) into

$$\mathbf{J}_n = -q\mu_n n_{\mathrm{p0}} \phi_{\mathrm{t}} e^{\psi/\phi_{\mathrm{t}}} \boldsymbol{\nabla} u. \tag{A.15}$$

Therefore, (A.10) becomes

$$n_{\mathrm{p0}} e^{\psi/\phi_{\mathrm{t}}} \frac{\partial u}{\partial t} = \mu_n n_{\mathrm{p0}} \phi_{\mathrm{t}} \boldsymbol{\nabla} \cdot \left(e^{\psi/\phi_{\mathrm{t}}} \boldsymbol{\nabla} u\right) - \frac{n_{\mathrm{i}}}{2\tau_0} \frac{1}{1+\kappa e^{-\psi/\phi_{\mathrm{t}}}} u. \tag{A.16}$$

Dividing both sides by $\mu_n n_{\mathrm{p0}} \phi_{\mathrm{t}} e^{\psi/\phi_{\mathrm{t}}}$ renders

$$\frac{1}{D_n} \frac{\partial u}{\partial t} = e^{-\psi/\phi_{\mathrm{t}}} \boldsymbol{\nabla} \cdot \left(e^{\psi/\phi_{\mathrm{t}}} \boldsymbol{\nabla} u\right) - \frac{1}{2\lambda_{\mathrm{p}}^2} \frac{\alpha}{\kappa + e^{\psi/\phi_{\mathrm{t}}}} u. \tag{A.17}$$

Now, it turns into a linear differential equation for u. This is very beneficial for upcoming discussions with regards to mathematical complexity. The term

$$\lambda_{\mathrm{p}} = \sqrt{\frac{\epsilon_{\mathrm{s}} \phi_{\mathrm{t}}}{q N_{\mathrm{A}}}} \tag{A.18}$$

is the extrinsic Debye length in the p-type semiconductor [24], and

$$\alpha \equiv \frac{\epsilon_\mathrm{s}}{q\mu_n n_\mathrm{i}\tau_0} \tag{A.19}$$

is defined as a dimensionless, material-dependent constant. With typical values $\mu_n = 1000\,\mathrm{cm}^2/\mathrm{V}\cdot\mathrm{s}$ and $\tau_0 = 10^{-7}\,\mathrm{s}$, $\alpha \approx 4.3$.

We are specifically interested in the steady-state ($\partial n/\partial t = 0$) electron QFL profile, where (A.17) becomes

$$\boxed{\nabla\cdot\left(e^{\psi/\phi_\mathrm{t}}\nabla u\right) = \frac{1}{2\lambda_\mathrm{p}^2}\frac{\alpha}{1+\kappa e^{-\psi/\phi_\mathrm{t}}}u}. \tag{A.20}$$

A.3 Electron Quasi Fermi Level in One-Dimensional MIS(p) Tunnel Diodes Under Reverse Bias

Consider a one-dimensional MIS(p) tunnel diode with known values of $\Delta\phi_n$ ($\equiv \Delta\phi_{n\mathrm{S}}$) and u ($\equiv u_\mathrm{S} = 1 - e^{-\Delta\phi_{n\mathrm{S}}/\phi_\mathrm{t}}$) at the surface. Equation (A.20) in 1D is

$$\frac{\mathrm{d}}{\mathrm{d}y}\left(e^{\psi/\phi_\mathrm{t}}\frac{\mathrm{d}u}{\mathrm{d}y}\right) = \frac{1}{2\lambda_\mathrm{p}^2}\frac{\alpha}{1+\kappa e^{-\psi/\phi_\mathrm{t}}}u, \tag{A.21}$$

where y is the depth beneath the semiconductor starting from the surface. We may utilize the depletion approximation

$$\psi(y) = \psi_\mathrm{S}\left(1-\frac{y}{W}\right)^2 \tag{A.22}$$

for $0 \le y \le W$ where ψ_S is the surface band bending, and

$$W = \sqrt{\frac{2\epsilon_\mathrm{s}\psi_\mathrm{S}}{qN_\mathrm{A}}} \tag{A.23}$$

is the depletion width. Next, we perform another change of variable from y to a dimensionless quantity z.

$$z \equiv \beta\left(1 - \frac{y}{W}\right), \tag{A.24}$$

in which

$$\beta \equiv \left(\frac{\psi_\mathrm{S}}{\phi_\mathrm{t}}\right)^{1/2} \tag{A.25}$$

is yet another dimensionless constant determined by the band bending. It is noteworthy that

$$\frac{W}{\beta} = \sqrt{2}\lambda_\mathrm{p}. \tag{A.26}$$

The definition of (A.24) turns (A.21) into the following second-order linear ordinary differential equation:

$$\frac{\mathrm{d}}{\mathrm{d}z}\left(e^{z^2}\frac{\mathrm{d}u}{\mathrm{d}z}\right) = \frac{\alpha}{1 + \kappa e^{-z^2}}u. \tag{A.27}$$

Solutions to The Differential Equation One may find any pair of linearly-independent solutions to the differential equation and express every other solution as their superposition due to the equation's linearity. Here we define $\Phi_{\alpha,\kappa}(z)$ and $\Gamma_{\alpha,\kappa}(z)$ to be *the* two solutions to (A.27) that satisfy the initial conditions

$$\Phi_{\alpha,\kappa}(0) \equiv 1, \quad \Phi'_{\alpha,\kappa}(0) \equiv 0; \tag{A.28a}$$

$$\Gamma_{\alpha,\kappa}(0) \equiv 0, \quad \Gamma'_{\alpha,\kappa}(0) \equiv 1. \tag{A.28b}$$

Then u can be expressed as

$$u\,(z) = c_1\Phi_{\alpha,\kappa}(z) + c_2\Gamma_{\alpha,\kappa}(z). \tag{A.29}$$

Plugging (A.24) back in,

$$u\,(y) = c_1\Phi_{\alpha,\kappa}\left(\beta\left(1 - \frac{y}{W}\right)\right) + c_2\Gamma_{\alpha,\kappa}\left(\beta\left(1 - \frac{y}{W}\right)\right). \tag{A.30}$$

Properties of The Functions There are no known expressions for $\Phi_{\alpha,\kappa}(z)$ and $\Gamma_{\alpha,\kappa}(z)$ in terms of standard functions. However, numerical solutions can be obtained for better comprehension of their properties. Defining $v \equiv \mathrm{d}u/\mathrm{d}z$, (A.27) is reduced to a first-order differential equation:

$$\frac{\mathrm{d}}{\mathrm{d}z}\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \dfrac{\alpha}{\kappa + e^{z^2}} & -2z \end{bmatrix}\begin{bmatrix} u \\ v \end{bmatrix}. \tag{A.31}$$

Utilizing Runge-Kutta 4th-order method [65,66] with the pre-defined initial conditions (A.28), $\Phi_{\alpha,\kappa}(z)$ and $\Gamma_{\alpha,\kappa}(z)$ can be plotted, as shown in **Figure A–2** . For lower values of α (i.e., longer lifetimes) and higher values of κ, the functions seem to approach a

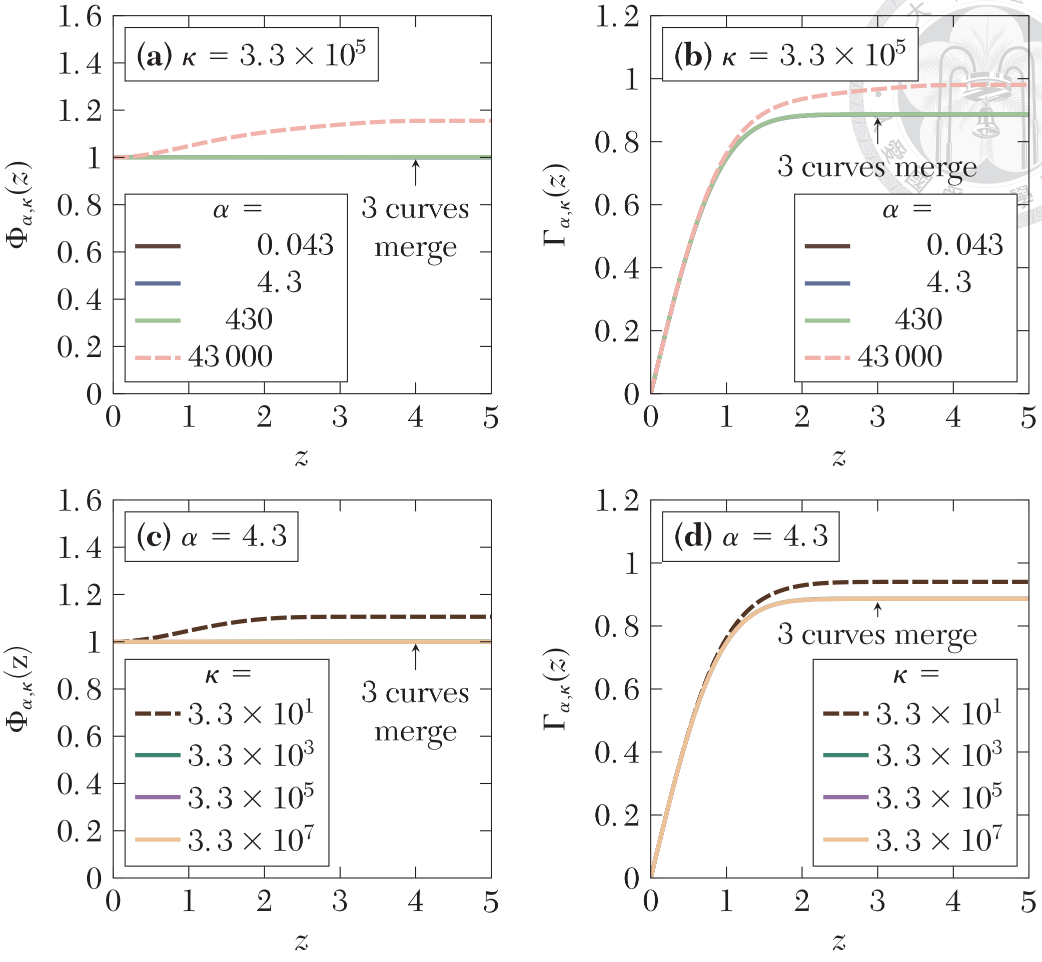


Figure A–2. **(a)** $\Phi_{\alpha,\kappa}(z)$ and **(b)** $\Gamma_{\alpha,\kappa}(z)$ plots under several values of α and a fixed value of $\kappa = 3.3 \times 10^5$. Higher values of α correspond to shorter lifetimes according to (A.19). **(c)** $\Phi_{\alpha,\kappa}(z)$ and **(d)** $\Gamma_{\alpha,\kappa}(z)$ plots under a fixed value of $\alpha = 4.3$ and several values of κ. κ is proportional to N_A (A.9).

certain profile. This is attributed to the right-hand side of (A.27) being negligible:

$$\frac{\mathrm{d}}{\mathrm{d}z}\left(e^{z^2}\frac{\mathrm{d}u}{\mathrm{d}z}\right) \approx 0. \tag{A.32}$$

From here, approximate expressions for $\Phi_{\alpha,\kappa}(z)$ and $\Gamma_{\alpha,\kappa}(z)$ can be solved with an aid of (A.28):

$$\Phi_{\alpha,\kappa}(z) \approx 1, \tag{A.33a}$$

$$\Gamma_{\alpha,\kappa}(z) \approx \frac{\sqrt{\pi}}{2}\operatorname{erf} z. \tag{A.33b}$$

Boundary Condition Matching To find the coefficients c_1 and c_2 in (A.30), we exploit the fact that

$$\begin{aligned} u\,(y = 0) &= c_1\Phi_{\alpha,\kappa}(\beta) + c_2\Gamma_{\alpha,\kappa}(\beta) \\ &\equiv u_{\mathrm{S}}. \end{aligned} \tag{A.34}$$

Also, recall the decay profile of minority carrier concentration in the neutral region $(y \geq W)$ [19]:

$$n_{\mathrm{p}0} - n\,(y) = \left[n_{\mathrm{p}0} - n\,(W)\right]\exp\left(-\frac{y - W}{L_n}\right), \tag{A.35}$$

where L_n is the electron diffusion length, which is typically tens to hundreds of micrometers. Taking the derivative at $y = W$ and plugging in (A.6a) renders

$$u'(y = W) = -\frac{u\,(y = W)}{L_n}. \tag{A.36}$$

Substituting (A.30) into this equation,

$$-\frac{\beta}{W}\left[c_1\Phi'_{\alpha,\kappa}(0) + c_2\Gamma'_{\alpha,\kappa}(0)\right] = -\frac{1}{L_n}\left[c_1\Phi_{\alpha,\kappa}(0) + c_2\Gamma_{\alpha,\kappa}(0)\right]. \tag{A.37}$$

Combining this with (A.26) and (A.28) renders

$$\frac{c_2}{\sqrt{2}\lambda_{\mathrm{p}}} = \frac{c_1}{L_n}. \tag{A.38}$$

Solving c_1 and c_2 from (A.34) and (A.38), (A.30) becomes

$$u\,(y) = \frac{L_n \Phi_{\alpha,\kappa}\left(\beta\left(1 - \frac{y}{W}\right)\right) + \sqrt{2}\lambda_{\mathrm{p}}\Gamma_{\alpha,\kappa}\left(\beta\left(1 - \frac{y}{W}\right)\right)}{L_n \Phi_{\alpha,\kappa}(\beta) + \sqrt{2}\lambda_{\mathrm{p}}\Gamma_{\alpha,\kappa}(\beta)} \cdot u_{\mathrm{S}}. \tag{A.39}$$

However, noting that $\lambda_{\mathrm{p}} \ll L_n$, this reduces to

$$\boxed{u\,(y) \approx \frac{\Phi_{\alpha,\kappa}\left(\beta\left(1 - \frac{y}{W}\right)\right)}{\Phi_{\alpha,\kappa}(\beta)} \cdot u_{\mathrm{S}}}. \tag{A.40}$$

The explicit expression for $\Delta\phi_n(y)$ can also be acquired from the definition of u (A.11):

$$\Delta\phi_n(y) = \phi_{\mathrm{t}} \ln\left[1 - \frac{\Phi_{\alpha,\kappa}\left(\beta\left(1 - \frac{y}{W}\right)\right)}{\Phi_{\alpha,\kappa}(\beta)}\left(1 - e^{-\Delta\phi_{n\mathrm{S}}/\phi_{\mathrm{t}}}\right)\right]^{-1}. \tag{A.41}$$

Figure A–3 (a) and **Figure A–3 (b)** depict the calculated $\Delta\phi_n(y)$ and $u\,(y)$ profiles, respectively, under $\psi_{\mathrm{S}} = 1.0\,\mathrm{V}$ ($\beta = 6.2$) under some given values of $\Delta\phi_{n\mathrm{S}}$. While $\Delta\phi_{n\mathrm{S}}$ curves do not appear to be flat, especially for higher $\Delta\phi_{n\mathrm{S}}$, $u\,(y)$ is always flat. For (realistic) sufficiently low α and high κ, (A.33a) and (A.40) implies

$$u\,(y) \approx u_{\mathrm{S}}. \tag{A.42}$$

The "Diffusion" and "Generation" Currents **Figure A–4** shows the J_n vs. y plot evaluated from (A.15) and (A.40) under $\psi_{\mathrm{S}} = 1.0\,\mathrm{V}$ and sufficiently high $\Delta\phi_{n\mathrm{S}}$ (i.e.,

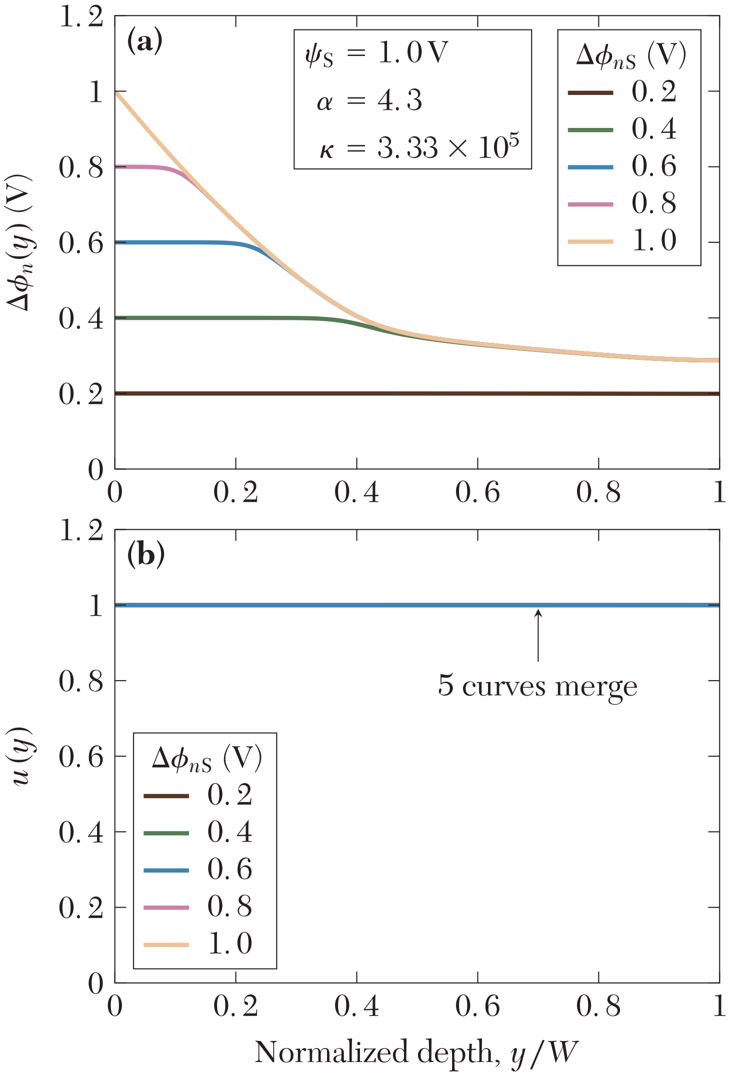


Figure A–3. Calculated one-dimensional **(a)** $\Delta\phi_{nS}(y)$ and **(b)** $u(y)$ profiles under $\psi_S = 1.0\,V$ using (A.41), starting from a collection of initial $\Delta\phi_n$ values at the surface, $\Delta\phi_{nS}$. The values of α and κ were calculated using $\tau_{n0} = \tau_{p0} = 10^{-7}$ s, $\mu_n = 1000\,\text{cm}^2/\text{V}\cdot\text{s}$, and $N_A = 10^{16}\,\text{cm}^{-3}$.

$u_S \approx 1$). The electron current grows as y approaches the surface, which can be clearly attributed to carrier generation in the depletion layer. We may therefore define the "diffusion current" of this structure as what originated from the neutral region:

$$J_{dn} \equiv J_n(y = W), \tag{A.43a}$$

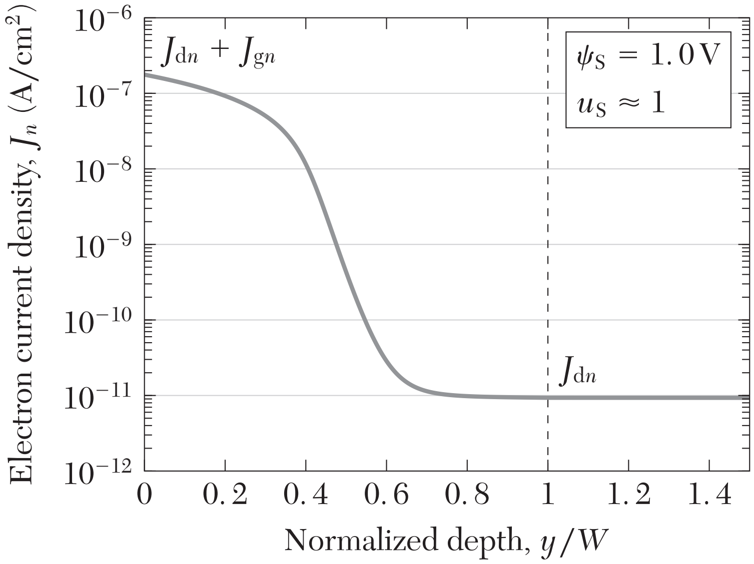


Figure A–4. Calculated J_n vs. y plot for a planar MOS(p) tunnel device under $\psi_S = 1.0\,\mathrm{V}$ and sufficiently high $\Delta\phi_{nS}$ (i.e., $u_S \approx 1$).

and the "generation current" as what originated in the depletion layer:

$$J_{gn} \equiv J_n(y = 0) - J_n(y = W). \tag{A.43b}$$

From the figure one concludes that $J_{dn} \lll J_{gn}$.

According to (A.3), J_{gn} at steady state can be expressed as

$$\begin{aligned} J_{gn} &= -\big(J_n(y = W) - J_n(y = 0)\big) \\ &= q \int_0^W G \,\mathrm{d}y. \end{aligned} \tag{A.44}$$

Substituting in (A.7) and (A.42),

$$\begin{aligned} J_{gn} &= \frac{qn_\mathrm{i}}{2\tau_0} \int_0^W \frac{u_\mathrm{S}}{1 + \kappa e^{-\psi/\phi_\mathrm{t}}} \,\mathrm{d}y \\ &= \frac{qn_\mathrm{i}}{2\tau_0} \left(\sqrt{2}\lambda_\mathrm{p} \mathscr{I}_\kappa(\beta)\right) u_\mathrm{S} \end{aligned} \tag{A.45}$$

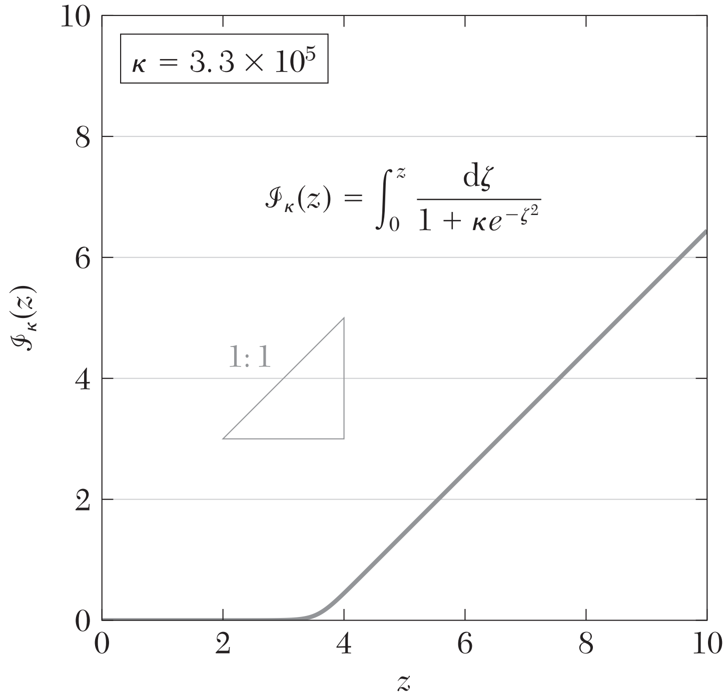


Figure A–5. $\mathscr{I}_\kappa(z)$ vs. z plot (A.46) at $\kappa = 3.3 \times 10^5$.

where the integral $\mathscr{I}_\kappa(z)$ is defined as

$$\mathscr{I}_\kappa(z) \equiv \int_0^z \frac{\mathrm{d}\zeta}{1+\kappa e^{-\zeta^2}}. \tag{A.46}$$

Figure A–5 shows the $\mathscr{I}_\kappa(z)$ vs. z plot at $\kappa = 3.3 \times 10^5$ ($N_\mathrm{A} = 10^{16}\,\mathrm{cm}^{-3}$). Intriguingly, the plot exhibits unity slope above some z, say, at $z > \sigma_\kappa$.

$$(z > \sigma_\kappa) \qquad \mathscr{I}_\kappa(z) \approx z - \sigma_\kappa. \tag{A.47}$$

doi:10.6342/NTU202304241

Mathematically, we find

$$\begin{aligned}\sigma_\kappa &= \lim_{z\to\infty}\left(z - \mathscr{I}_\kappa(z)\right)\\ &= \lim_{z\to\infty}\int_0^z\left(1-\frac{1}{1+\kappa e^{-\zeta^2}}\right)\mathrm{d}\zeta\\ &= \int_0^\infty \frac{\mathrm{d}\zeta}{(1/\kappa)+e^{\zeta^2}}. \end{aligned} \tag{A.48}$$

Some other equivalent representations for σ_κ might also be useful:

$$\begin{aligned}\sigma_\kappa &= \frac{\sqrt{\pi}}{2}\mathscr{F}_{-1/2}(\ln\kappa)\\ &= \frac{\sqrt{\pi}}{2}\left[-\operatorname{Li}_{1/2}(-\kappa)\right], \end{aligned} \tag{A.49}$$

where $\mathscr{F}_j(z)$ is the complete Fermi-Dirac integral, and

$$\begin{aligned}\operatorname{Li}_s(z) &\equiv \sum_{k=1}^{\infty}\frac{z^k}{k^s}\\ &= \frac{1}{\Gamma(s)}\int_0^\infty \frac{t^{s-1}}{e^t/z-1}\,\mathrm{d}z \end{aligned} \tag{A.50}$$

is the polylogarithm. Additionally, by exploiting the asymptotic expansion of $\mathscr{F}_{-1/2}(z)$ at big z [67],

$$(z \gg 1) \qquad \mathscr{F}_{-1/2}(z) \approx \frac{2z^{1/2}}{\sqrt{\pi}}, \tag{A.51}$$

σ_κ may also be approximated, for very high κ, as

$$\sigma_\kappa \approx \sqrt{\ln\kappa}. \tag{A.52}$$

TABLE A–I
Values of σ_κ (unapproximated) and W_{g0} under several doping concentrations. $\tau_{n0} = \tau_{p0}$ was assumed while evaluating κ.

N_{A} (cm^{-3})	κ	σ_κ	λ_{p} (nm)	W_{g0} (nm)
10^{14}	3.33×10^{3}	2.828	413	1650
10^{15}	3.33×10^{4}	3.214	130.	593
10^{16}	3.33×10^{5}	3.557	41.3	208
10^{17}	3.33×10^{6}	3.868	13.0	71.4
10^{18}	3.33×10^{7}	4.157	4.13	24.3

TABLE A–I lists the values of σ_κ (unapproximated) at several κ values. What's more, combining (A.45) and (A.47) yields

$$(W > W_{\mathrm{g0}}) \qquad \boxed{J_{\mathrm{g}n} \approx \frac{qn_{\mathrm{i}}(W - W_{\mathrm{g0}})}{2\tau_0}u_{\mathrm{S}}}, \tag{A.53}$$

where we define the *generation-depletion width offset* as

$$\begin{aligned} W_{\mathrm{g0}} &= \sqrt{2}\sigma_\kappa\lambda_{\mathrm{p}} \\ &\approx (2\ln\kappa)^{1/2}\,\lambda_{\mathrm{p}}. \end{aligned} \tag{A.54}$$

The values of W_{g0} (using unapproximated σ_κ values) are also listed in **TABLE A–I** .

At high $\Delta\phi_{n\mathrm{S}}$, $J_{\mathrm{gn}} \rightarrow qn_{\mathrm{i}}(W - W_{\mathrm{g0}})/2\tau_0$. The extra offset term W_{g0} compared to the Sah-Noyce-Shockley model ($J_{\mathrm{g}n} \approx qn_{\mathrm{i}}W/2\tau_0$) [57] arises from the fact that the old model assumes negligible hole concentration in the space charge region, which overestimates G at $n_{\mathrm{i}}/2\tau_0$ compared to that of the new model (A.7).

A.4 Electron Quasi Fermi Level in Cylindrically-Symmetric Ultra-High-Low MOS(p) Devices Under Reverse Bias

For ultra-high-low MOS(p) devices, tunneling of the generated electrons through the high oxide is forbidden. In other words, lateral current must be present under the high oxide, and the electron QFL must vary across the lateral direction according to (A.4), blundering the one-dimensional model. While the 1D model may apply to the low region, appropriate modeling of the high-region electron QFL profile is still mandatory for developing current and electrostatic theories for the device. For simplicity, cylindrically-symmetric devices are considered, with the low oxide located at the center ($r \le r_{\mathrm{L}}$), encircled by the high oxide ($r_{\mathrm{L}} \le r \le r_{\mathrm{G}}$). The schematic cross section is shown in **Figure A–6 (a)**. Under sufficiently high gate voltage such that the high region also undergoes deep depletion, and the inversion charge barely influence the electrostatics, the band bending ψ shall have no radial dependence.

$$(r_{\mathrm{L}} \le r \le r_{\mathrm{G}}) \qquad \frac{\partial \psi(r,\, y)}{\partial r} = 0. \tag{A.55}$$

Therefore, the depletion approximation shall also apply under the high region:

$$(r_{\mathrm{L}} \le r \le r_{\mathrm{G}}) \qquad \psi(r,\, y) = \psi_{\mathrm{SH}}\left(1 - \frac{y}{W_{\mathrm{H}}}\right)^2. \tag{A.56}$$

The divergence term in (A.17) with ψ having no radial dependence now becomes:

$$\boldsymbol{\nabla} \cdot \left(e^{\psi(y)/\phi_{\mathrm{t}}} \boldsymbol{\nabla} u\right) = \frac{\partial}{\partial y}\left(e^{\psi/\phi_{\mathrm{t}}} \frac{\partial u}{\partial y}\right) + e^{\psi/\phi_{\mathrm{t}}} \frac{1}{r}\frac{\partial}{\partial r}\left(r \frac{\partial u}{\partial r}\right) \tag{A.57}$$

doi:10.6342/NTU202304241

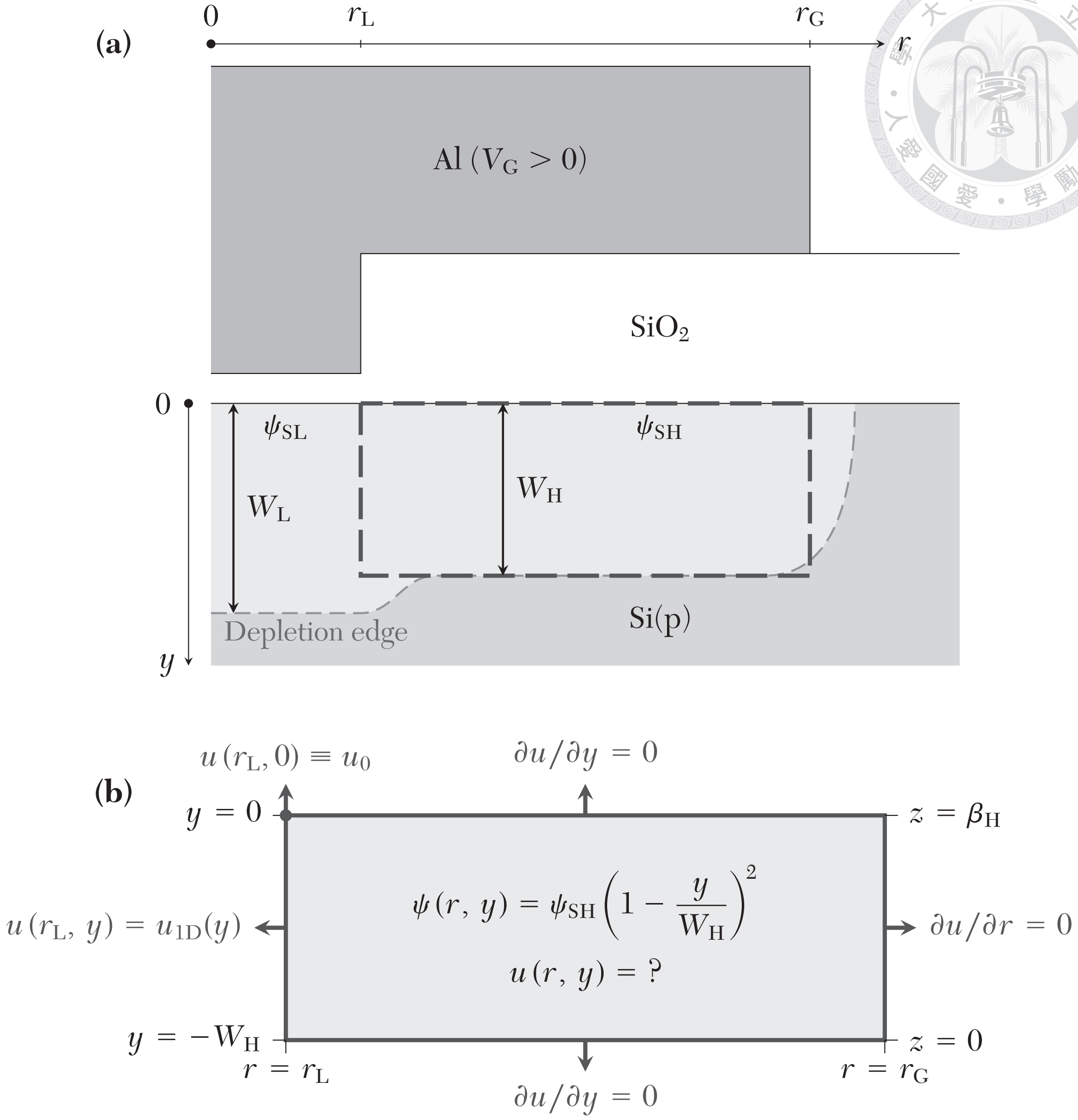


Figure A–6. **(a)** Schematic cross section of a cylindrically-symmetric ultra-high-low MOS(p) tunnel device and its dimensions, featuring the depletion edge. **(b)** a simplified geometry for solving $u(r, y)$ profile in the box of **(a)**, with boundary conditions specified.

Also, change of variables in analogy with the one-dimensional model may be helpful:

$$z \equiv \beta_H\left(1 - \frac{y}{W_H}\right), \tag{A.58a}$$

$$\beta_H \equiv \left(\frac{\psi_{SH}}{\phi_t}\right)^{1/2}. \tag{A.58b}$$

Substituting these into (A.17) renders

$$\frac{2\lambda_{\mathrm{p}}^2}{D_n}\frac{\partial u}{\partial t} = e^{-z^2}\frac{\partial}{\partial z}\left(e^{z^2}\frac{\partial u}{\partial z}\right) + \left(2\lambda_{\mathrm{p}}^2\right)\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial u}{\partial r}\right) - \frac{\alpha}{\kappa + e^{z^2}}u. \tag{A.59}$$

Under steady state ($\partial u/\partial t = 0$), this simplifies to

$$\boxed{e^{-z^2}\frac{\partial}{\partial z}\left(e^{z^2}\frac{\partial u}{\partial z}\right) + \left(2\lambda_{\mathrm{p}}^2\right)\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial u}{\partial r}\right) = \frac{\alpha}{\kappa + e^{z^2}}u}. \tag{A.60}$$

The solution to $u\,(r,\ z)$ can be solved numerically within a simplified rectangular region in the space charge region under the high oxide ($r_{\mathrm{L}} \le r_{\mathrm{G}}$, $0 \le y \le W_{\mathrm{H}}$) as depicted in **Figure A–6 (b)**, subject to the following mixed boundary conditions:

1. At $r = r_{\mathrm{L}}$, assume the low region's influence is high and dominates the electron QFL at the high-low boundary. By also assuming the one-dimensional model (A.40) to work in the low region, we have

   $$\begin{aligned} u\,(r_{\mathrm{L}},\ y) &\approx u_{\mathrm{1D}}(y) \\ &= \frac{\Phi_{\alpha,\kappa}\left(\beta_{\mathrm{H}}\left(1 - \frac{y}{W_{\mathrm{H}}}\right)\right)}{\Phi_{\alpha,\kappa}(\beta_{\mathrm{H}})} \cdot u_0, \end{aligned} \tag{A.61}$$

   where $u_0 \equiv u\,(r_{\mathrm{L}},\ 0)$.

2. At $y = 0$, the absence of vertical current requires that $\partial u/\partial y = 0$.

3. At $y = W_{\mathrm{H}}$, the "diffusion current" is negligible due to high L_n. Therefore, $\partial u/\partial y = 0$.

4. For the same reason, at $r = r_{\mathrm{G}}$, $\partial u/\partial r = 0$.

The Depth-Average Approximation The necessity of solving (A.60) numerically makes the $u\,(r,\,y)$ profile not readily comprehensible. Still, further approximations can be achieved for better physical interpretation. More specifically speaking, to what extent the low region laterally affect the high-region electron QFL profile is more of our interest. Formulation of the $u\,(r)$ profile with eliminated y-dependency is preferable.

We may eliminate the y- (or z-) dependency in (A.60) by first multiplying both sides by e^{z^2},

$$\frac{\partial}{\partial z}\left(e^{z^2}\frac{\partial u}{\partial z}\right)+\left(2\lambda_{\mathrm{p}}^2\right)e^{z^2}\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial u}{\partial r}\right)=\frac{\alpha}{1+\kappa e^{-z^2}}u, \tag{A.62}$$

and by taking its *depth average*, defined by

$$\begin{aligned}\langle * \rangle &\equiv \frac{1}{\mathrm{W_H}}\int_0^{\mathrm{W_H}}(*)\ \mathrm{d}y\\ &=\frac{1}{\beta_{\mathrm{H}}}\int_0^{\beta_{\mathrm{H}}}(*)\ \mathrm{d}z. \end{aligned}\tag{A.63}$$

Noting that the term

$$\begin{aligned}e^{z^2}\frac{\partial u}{\partial z} &= -\left(\sqrt{2}\lambda_{\mathrm{p}}\right)e^{\psi/\phi_{\mathrm{t}}}\frac{\partial u}{\partial y}\\ &=\frac{\sqrt{2}\lambda_{\mathrm{p}}}{q\,\mu_n n_{\mathrm{p}0}\phi_{\mathrm{t}}}J_{ny}\end{aligned}\tag{A.64}$$

according to (A.15), the average of the leftmost term in (A.62) is

$$\begin{aligned}\left\langle\frac{\partial}{\partial z}\left(e^{z^2}\frac{\partial u}{\partial z}\right)\right\rangle &= \left\langle-\sqrt{2}\lambda_{\mathrm{p}}\frac{\partial}{\partial y}\left(\frac{\sqrt{2}\lambda_{\mathrm{p}}}{q\,\mu_n n_{\mathrm{p}0}\phi_{\mathrm{t}}}J_{ny}\right)\right\rangle\\ &=-\frac{2\lambda_{\mathrm{p}}^2}{q\,\mu_n n_{\mathrm{p}0}\phi_{\mathrm{t}}}\frac{1}{\mathrm{W_H}}\int_0^{W_H}\frac{\partial J_{ny}}{\partial y}\,\mathrm{d}y\\ &=\frac{2\lambda_{\mathrm{p}}^2}{q\,\mu_n n_{\mathrm{p}0}\phi_{\mathrm{t}}}\frac{J_{ny}(y=0)-J_{ny}(y=\mathrm{W_H})}{\mathrm{W_H}}.\end{aligned}\tag{A.65}$$

doi:10.6342/NTU202304241

Now that there is no vertical current at the surface (blocked by the high oxide) and the depletion edge (with negligible diffusion current),

$$\left\langle \frac{\partial}{\partial z}\left(e^{z^2}\frac{\partial u}{\partial z}\right)\right\rangle = 0. \tag{A.66}$$

Therefore, averaging both sides in (A.62) yields

$$\begin{aligned}\left(2\lambda_{\mathrm{p}}^2\right)\left\langle e^{z^2}\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial u}{\partial r}\right)\right\rangle &= \left(2\lambda_{\mathrm{p}}^2\right)\frac{1}{r}\frac{\mathrm{d}}{\mathrm{d}r}\left(r\left\langle e^{z^2}\frac{\partial u}{\partial r}\right\rangle\right)\\ &= \alpha\left\langle\frac{u}{1+\kappa e^{-z^2}}\right\rangle.\end{aligned} \tag{A.67}$$

Inspired by (A.42), we further assume negligible variation of $u(r,\, z)$ along z for any given r. That is, we may write

$$u(r,\, z) \approx \bar{u}(r). \tag{A.68}$$

Then, for a function $f(z)$ with no r-dependency, it follows that

$$\langle f(z)u(r,\, z)\rangle \approx \langle f(z)\rangle\, \bar{u}(r). \tag{A.69}$$

This transforms (A.67) into

$$\left(2\lambda_{\mathrm{p}}^2\right)\left\langle e^{z^2}\right\rangle\frac{1}{r}\frac{\mathrm{d}}{\mathrm{d}r}\left(r\frac{\mathrm{d}\bar{u}}{\mathrm{d}r}\right) = \alpha\left\langle\frac{1}{1+\kappa e^{-z^2}}\right\rangle\bar{u}. \tag{A.70}$$

doi:10.6342/NTU202304241

The average terms in the equations are evaluated as

$$\begin{aligned}\left\langle e^{z^2} \right\rangle &= \frac{1}{\beta_\mathrm{H}} \int_0^{\beta_\mathrm{H}} e^{z^2}\,\mathrm{d}z \\ &= \frac{\sqrt{\pi}}{2\beta_\mathrm{H}} \operatorname{erfi} \beta_\mathrm{H},\end{aligned} \tag{A.71}$$

where erfi is the imaginary error function, and from (A.46),

$$\left\langle \frac{1}{1+\kappa e^{-z^2}} \right\rangle = \frac{\mathscr{I}_\kappa(\beta_\mathrm{H})}{\beta_\mathrm{H}}. \tag{A.72}$$

Therefore, (A.70) can be written as

$$\frac{1}{r}\frac{\mathrm{d}}{\mathrm{d}r}\left(r\frac{\mathrm{d}\bar{u}}{\mathrm{d}r}\right) = \frac{\bar{u}}{\Lambda^2}, \tag{A.73}$$

with the *coupling length* Λ defined as

$$\boxed{\Lambda \equiv \left(\frac{\sqrt{\pi}\operatorname{erfi}\beta_\mathrm{H}}{\alpha \mathscr{I}_\kappa(\beta_\mathrm{H})}\right)^{1/2} \lambda_\mathrm{p}}. \tag{A.74}$$

To make it more clear by expanding α, κ, β_H, and $\mathscr{I}_\kappa$,

$$\Lambda(\psi_\mathrm{SH}) = \left(\frac{\sqrt{\pi} q \mu_n \operatorname{erfi}\sqrt{\psi_\mathrm{SH}/\phi_\mathrm{t}}}{2\epsilon_\mathrm{s}\displaystyle\int_0^{\sqrt{\psi_\mathrm{SH}/\phi_\mathrm{t}}} \frac{\mathrm{d}z}{2n_\mathrm{i}\tau_0 + N_\mathrm{A}\tau_{n0}e^{-z^2}}}\right)^{1/2} \lambda_\mathrm{p}. \tag{A.75}$$

Alternatively, from (A.47) and (A.54), as $W_\mathrm{H} > W_\mathrm{g0}$, Λ is approximated as

$$(W_\mathrm{H} > W_\mathrm{g0}) \qquad \boxed{\Lambda \approx \left(\frac{\sqrt{2\pi} q \mu_n n_\mathrm{i} \tau_0 \lambda_\mathrm{p}^3 \operatorname{erfi}\sqrt{\psi_\mathrm{SH}/\phi_\mathrm{t}}}{\epsilon_\mathrm{s}(W_\mathrm{H} - W_\mathrm{g0})}\right)^{1/2}}. \tag{A.76}$$

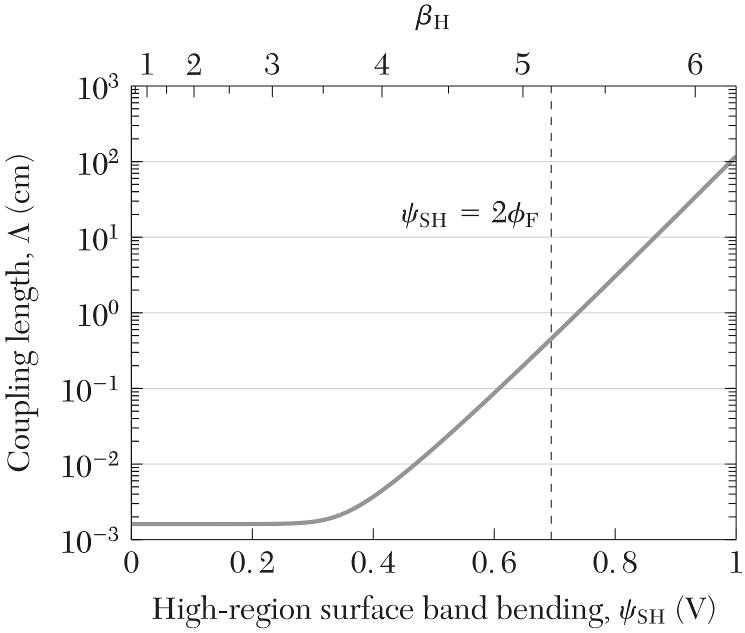


Figure A–7. Plot of the coupling length Λ vs. ψ_{SH} at $\mu_n = 1000\,\mathrm{cm^2/V\cdot s}$, $\tau_{n0} = \tau_{p0} = 10^{-7}\,\mathrm{s}$, and $N_{\mathrm{A}} = 10^{16}\,\mathrm{cm^{-3}}$.

Figure A–7 shows the Λ vs. ψ_{SH} plot at $\mu_n = 1000\,\mathrm{cm^2/V\cdot s}$, $\tau_{n0} = \tau_{p0} = 10^{-7}\,\mathrm{s}$, and $N_{\mathrm{A}} = 10^{16}\,\mathrm{cm^{-3}}$. Λ is relatively small ($\sim 16\,\mathrm{\mu m}$) as $\psi_{\mathrm{SH}} < 0.4\,\mathrm{V}$. However, under inversion or deep depletion ($\psi_{\mathrm{SH}} > 2\phi_{\mathrm{F}} \approx 0.69\,\mathrm{V}$), Λ grows beyond the centimeter scale.

The solution to (A.73) is

$$\bar{u}\,(r) = c_I I_0(r/\Lambda) + c_K K_0(r/\Lambda). \tag{A.77}$$

where c_I and c_K are constants, I_ν is the modified Bessel function of the first kind, and K_ν, that of the second kind. It may also be convenient to define the normalized radii

$$\boxed{\rho \equiv r/\Lambda} \tag{A.78}$$

doi:10.6342/NTU202304241

to tidy things up. The constants c_I and c_K can be found by boundary condition matching. First, at $r = r_\mathrm{L}$, $\bar{u}\,(r) \approx u_0$; thus,

$$c_I I_0(\rho_\mathrm{L}) + c_K K_0(\rho_\mathrm{L}) = u_0. \tag{A.79}$$

Next, at $r = r_\mathrm{G}$, the negligibility of lateral diffusion current suggests

$$\begin{aligned} 0 &= \left(\frac{\mathrm{d}\bar{u}}{\mathrm{d}r}\right)_{r_\mathrm{G}} \\ &= \Lambda^{-1}\left(c_I I_1(\rho_\mathrm{G}) - c_K K_1(\rho_\mathrm{G})\right). \end{aligned} \tag{A.80}$$

The result is

$$\boxed{\bar{u}\,(\rho) = \frac{I_1(\rho_\mathrm{G})K_0(\rho) + K_1(\rho_\mathrm{G})I_0(\rho)}{I_1(\rho_\mathrm{G})K_0(\rho_\mathrm{L}) + K_1(\rho_\mathrm{G})I_0(\rho_\mathrm{L})} \cdot u_0}. \tag{A.81}$$

Total Electron Current and The Coupling Coefficient Assuming negligible diffusion current, the total electron current across the device at steady state reads (A.3)

$$\begin{aligned} I_n &= q\int_\mathrm{SCR} G\,\mathrm{d}V \\ &= \frac{qn_\mathrm{i}}{2\tau_0}\left(\int_0^{W_\mathrm{L}}\int_0^{r_\mathrm{L}} + \int_0^{W_\mathrm{H}}\int_{r_\mathrm{L}}^{r_\mathrm{G}}\right)\frac{u\,(r,\,y)}{1+\kappa e^{-\psi/\phi_\mathrm{t}}}\,2\pi r\,\mathrm{d}r\,\mathrm{d}y, \end{aligned} \tag{A.82}$$

doi:10.6342/NTU202304241

where SCR stands for the space charge region. Exploiting the facts that $u\,(r,\,y) \approx u_0$ as $r \leq r_\mathrm{L}$ and $u\,(r,\,y) \approx \bar{u}\,(r)$ as $r_\mathrm{L} \leq r \leq r_\mathrm{G}$, the integrals are separable:

$$\begin{aligned} I_n &= \frac{qn_\mathrm{i}}{2\tau_0} 2\pi \left[\left(\int_0^{W_\mathrm{L}} \frac{\mathrm{d}y}{1+\kappa e^{-\psi/\phi_\mathrm{t}}} \right) \left(\int_0^{r_\mathrm{L}} u_0\, r\, \mathrm{d}r \right) + \left(\int_0^{W_\mathrm{H}} \frac{\mathrm{d}y}{1+\kappa e^{-\psi/\phi_\mathrm{t}}} \right) \left(\int_{r_\mathrm{L}}^{r_\mathrm{G}} \bar{u}\,(r)\, r\, \mathrm{d}r \right) \right] \\ &= \frac{qn_\mathrm{i}}{2\tau_0} \left[\left(W_\mathrm{L} - W_\mathrm{g0}\right) \pi r_\mathrm{L}^2 u_0 + \left(W_\mathrm{H} - W_\mathrm{g0}\right) 2\pi \left(\int_{r_\mathrm{L}}^{r_\mathrm{G}} \bar{u}\,(r)\, r\, \mathrm{d}r \right) \right]. \end{aligned} \tag{A.83}$$

From (A.81), the remaining integral is evaluated as

$$\int_{r_\mathrm{L}}^{r_\mathrm{G}} \bar{u}\,(r)\, r\, \mathrm{d}r = \Lambda^2 \rho_\mathrm{L} \frac{I_1(\rho_\mathrm{G})K_1(\rho_\mathrm{L}) - K_1(\rho_\mathrm{G})I_1(\rho_\mathrm{L})}{I_1(\rho_\mathrm{G})K_0(\rho_\mathrm{L}) + K_1(\rho_\mathrm{G})I_0(\rho_\mathrm{L})} \cdot u_0. \tag{A.84}$$

If we define the *coupling efficiency* (η_C) as

$$\boxed{\eta_\mathrm{C} = \frac{2\rho_\mathrm{L}}{\rho_\mathrm{G}^2 - \rho_\mathrm{L}^2} \frac{I_1(\rho_\mathrm{G})K_1(\rho_\mathrm{L}) - K_1(\rho_\mathrm{G})I_1(\rho_\mathrm{L})}{I_1(\rho_\mathrm{G})K_0(\rho_\mathrm{L}) + K_1(\rho_\mathrm{G})I_0(\rho_\mathrm{L})}}, \tag{A.85}$$

then (A.83) becomes

$$\boxed{I_n = \frac{qn_\mathrm{i}}{2\tau_0} \left[A_\mathrm{L} \left(W_\mathrm{L} - W_\mathrm{g0}\right) + \eta_\mathrm{C} A_\mathrm{H} \left(W_\mathrm{H} - W_\mathrm{g0}\right) \right] u_0}, \tag{A.86}$$

where $A_\mathrm{L} \equiv \pi r_\mathrm{L}^2$ and $A_\mathrm{H} \equiv \pi \left(r_\mathrm{G}^2 - r_\mathrm{L}^2\right)$ are the low- and high-region areae, respectively.

From here the physical meaning of η_C is clear. The minority current consists of a component originating from carrier generation beneath the low oxide, and another from that beneath the high oxide. The fact that $u < u_0$ (the high region having smaller electron QFL splitting than the low region) degrades the generation rate and impairs the minority carrier current. η_C serves as a quantitative indication of how much the

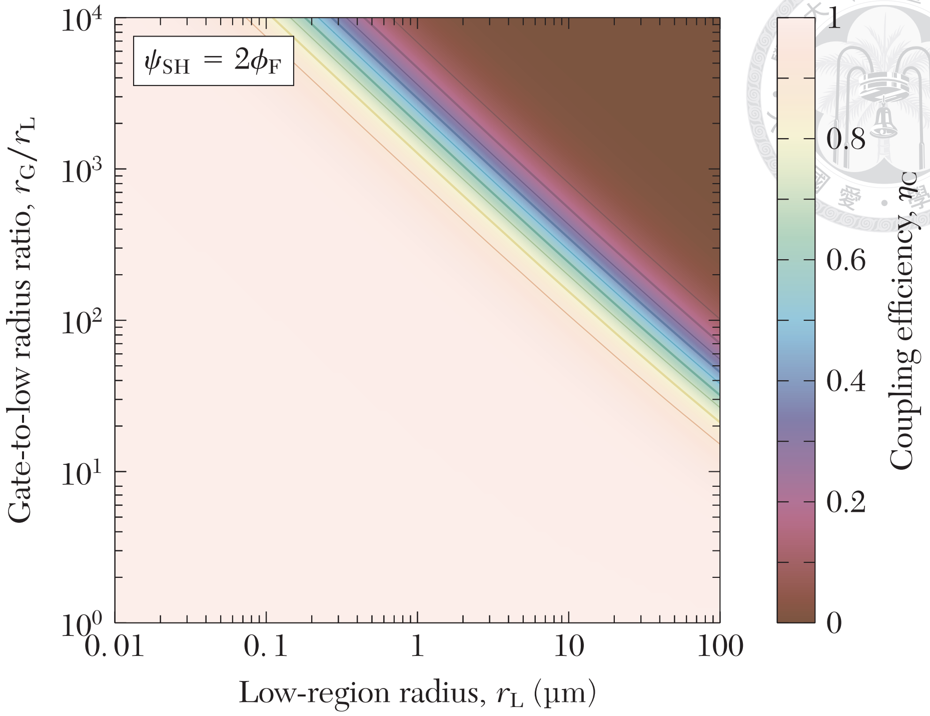


Figure A–8. Contour plot of η_{C} vs. r_{L} and r_{G} under bias condition $\psi_{\mathrm{SH}} = 2\phi_{\mathrm{F}}$.

current of "high" origin is degraded. If $\eta_{\mathrm{C}} \approx 1$, the electron current would show proportionality to the total gate area, $A_{\mathrm{G}} = A_{\mathrm{L}} + A_{\mathrm{H}}$. Contrarily, if $\eta_{\mathrm{C}} \approx 0$, proportionality to A_{L} is expected. The $\eta_{\mathrm{C}} - (r_{\mathrm{L}}, r_{\mathrm{G}})$ plots (**Figure A–8**) at $\psi_{\mathrm{SH}} = 2\phi_{\mathrm{F}} = 0.69\,\mathrm{V}$ calculated from (A.85) reveals $\eta_{\mathrm{C}} \approx 1$ under any reasonable device dimension. Consonantly, experimental ultra-high-low devices exhibit $I_n \propto A_{\mathrm{G}}$, but not $\propto A_{\mathrm{L}}$, indicating near-unity experimental η_{C}. **Figure A–9** shows the calculated $\eta_{\mathrm{C}} - \psi_{\mathrm{SH}}$ plots for a collection of devices with identical $r_{\mathrm{G}} = 150\,\mu\mathrm{m}$ but different values of r_{L}. While the efficient is degraded under low band bending, for the bias region of our interest ($\psi_{\mathrm{SH}} \geq 2\phi_{\mathrm{F}}$), η_{C} is very close to unity for every demonstrated r_{L}.

Total Inversion Charge Under The High Region Assuming low electron concentration under the high region under $V_{\mathrm{G}} > 0$ (i.e., subthreshold condition), the inversion

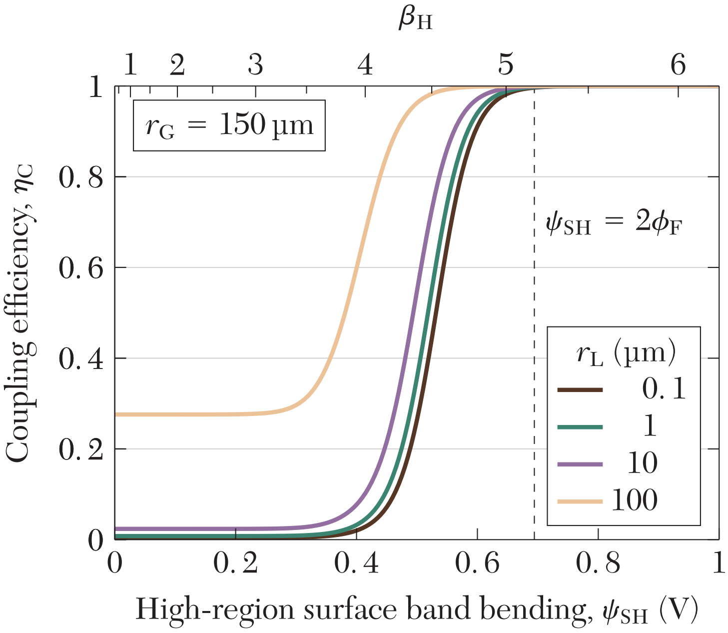


Figure A–9. $\eta_C - \psi_{SH}$ plots for a collection of devices with identical r_G (= 150 µm) but different values of r_L.

charge per unit area, Q_{iH}, at a given location, is given by [18]

$$\begin{aligned} Q_{iH}(r) &\approx -\sqrt{\frac{q\epsilon_s N_A}{2\psi_{SH}}}\left(\frac{n_i}{N_A}\right)^2 \phi_t e^{(\psi_{SH}-\Delta\phi_n)/\phi_t} \\ &= -\sqrt{\frac{q\epsilon_s N_A}{2\psi_{SH}}}\left(\frac{n_i}{N_A}\right)^2 \phi_t e^{\psi_{SH}/\phi_t}\left(1-\bar{u}\left(r\right)\right). \end{aligned} \qquad \text{(A.87)}$$

The total inversion charge under the high region, Q_{IH}, is therefore

$$\begin{aligned} Q_{IH} &= \int_{r_L}^{r_G} Q_{iH}(r)\, 2\pi r\, \mathrm{d}r \\ &\approx -(1-\eta_C u_0) A_H \left[\sqrt{\frac{q\epsilon_s N_A}{2\psi_{SH}}}\left(\frac{n_i}{N_A}\right)^2 \phi_t e^{\psi_{SH}/\phi_t}\right]. \end{aligned} \qquad \text{(A.88)}$$

For sufficiently high $\Delta\phi_n$ beneath the thin region ($u_0 \approx 1$),

$$\boxed{Q_{\mathrm{IH}} \approx -(1-\eta_{\mathrm{C}})A_{\mathrm{H}}\left[\sqrt{\frac{q\epsilon_{\mathrm{s}}N_{\mathrm{A}}}{2\psi_{\mathrm{SH}}}}\left(\frac{n_{\mathrm{i}}}{N_{\mathrm{A}}}\right)^2\phi_{\mathrm{t}}e^{\psi_{\mathrm{SH}}/\phi_{\mathrm{t}}}\right]}. \tag{A.89}$$

As $\eta_{\mathrm{C}} \approx 1$ for reasonable device dimensions, this model predicts little inversion charge.

A.5 Summary

To solve the spatial profile of $\Delta\phi_n$ in the space charge region of MOS(p) tunnel structures under reverse bias, linear differential equations for u where $u \equiv 1 - e^{-\Delta\phi_n/\phi_{\mathrm{t}}}$ have been derived and solved. In planar devices, $\Delta\phi_n$ is not necessarily flat along the space charge region, but u is very flat. The generation current has little to do with the non-flat $\Delta\phi_n$ profile — It still adopts a simple mathematical form thanks to the flatness of u. The generation current differs from the Sah-Noyce-Shockley model by a negative offset term $-W_{\mathrm{g0}}$. For ultra-high-low devices, u still appears to be near constant under the entire gate area. The coupling efficiency η_{C} is quantitatively defined and related to the total electron current and total inversion charge. The mathematical fact that $\eta_{\mathrm{C}} \approx 1$ under every reasonable gate and low-region radii, r_{G} and r_{L}, above threshold, manages to predict the A_{G}-dependent electron current and the absence of inversion charge in experimental devices.

doi:10.6342/NTU202304241

doi:10.6342/NTU202304241

References

[1] S. Solomon, "Breakdown in Silicon Oxide — A Review," *J. Vac. Sci. Technol.* **14** (5): 1123–1130, September/October 1977, doi:10.5116/1.569344.

[2] G. Swartz, "Gate Oxide Integrity of MOS/SOS Devices," *IEEE Trans. Electron Dev.* **33** (1): 119–125, January 1986, doi:10.1109/T-ED.1986.22447.

[3] S. Mei, N. Raghavan, M. Bosman, D. Linten, G. Groeseneken, N. Horiguchi, & K. L. Pey. "New Understanding of Dielectric Breakdown in Advanced FinFET Devices — Physical, Electrical, Statistical, and Multiphysics Study," in *2016 IEEE International Electron Devices Meeting (IEDM)*, December 2016, doi:10.1109/IEDM.2016.7838424.

[4] M. C. Chen, S. H. Ku, C. T. Chan, & T. Wang, "Comparison of Oxide Breakdown Progression in Ultra-Thin Oxide Silicon-on-Insulator and Bulk Metal-Oxide-Semiconductor Field-Effect Transistors," *J. Appl. Phys.* **96**: 3473–3477, 2004, doi:10.1063/1.1776640.

[5] C. Zhao, C. Z. Zhao, S. Taylor, & P. R. Chalker, "Review on Non-Volatile Memory with High-*k* Dielectrics Flash for Generation Beyond 32 nm," *Materials* **2014** (7): 5117–5145, July 2014, doi:10.3390/ma7075117.

doi:10.6342/NTU202304241

[6] K. K. Kim, “Analysis of Gate Oxide Breakdown in CMOS Combinational Logics,” *J. Sens. Sci. Technol.* **28** (1): 17–22, 2019, doi:10.5369/JSST.2019.28.1.17.

[7] M. A. Alam, B. Weir, J. Bude, P. Silverman, & D. Monroe. “Explanation of Soft and Hard Breakdown and Its Consequences for Area Scaling,” in *International Electron Devices Meeting 1999. Technical Digest (Cat. No. 99CH36318)*, pages 449–452, 1999, doi:10.1109/IEDM.1999.824190.

[8] S. Lombardo, A. La Magna, I. Crupi, C. Gerardi, & F. Crupi, “Reduction of Thermal Damage in Ultrathin Gate Oxides After Intrinsic Dielectric Breakdown,” *Appl. Phys. Lett.* **79** (10): 1522–1524, September 2001, doi:10.1063/1.1400083.

[9] D. K. Schroder. *Semiconductor Material and Device Characterization*, 3rd ed. John Wiley & Sons, New Jersey, 2006.

[10] S. Takagi & M. Takayanagi, “Carrier Transport Properties of Thin Gate Oxides after Soft and Hard Breakdown,” *Microelectronic Eng.* **59** (1-4): 5–15, November 2001, doi:10.1016/S0167-9317(01)00624-4.

[11] H. Jin, S. Dong, M. Miao, J. J. Liou, & C. Y. Yang, “Breakdown Voltage of Ultrathin Dielectric Film Subject to Electrostatic Discharge Stress,” *J. Appl. Phys.* **110** (5): 054516, September 2011, doi:10.1063/1.3633527.

[12] B. Greenwood, A. Suhwanov, D. Daniel, S. Menon, P. Price, S. Hose, J. Guo, G. Piatt, M. Lu, Y. Watanabe, Y. Kamura, R. Takada, L. Sheng, & J. P. Gambino. “Gate Oxide Yield Improvement for 0.18 µm Power Semiconductor Devices with Deep Trenches,” in *2017 40th International Convention on Information*

and Communication Technology, Electronics and Microelectronics (MIPRO), May 2017, doi:10.23919/MIPRO.2017.7966605.

[13] Y.-P. Deng, H.-R. Ren, Q. Fan, H.-M. Ho, Q. Xu, Z.-Y. Zhang, K. Zheng, & Y.-J. Wu, "GOI Improvement in 65 nm Sacrificial Oxide Free Process Integration," *ECS Trans.* **27** (1): 383–387, January 2010, doi:10.1149/1.3360648.

[14] K. L. Pey, C. H. Tung, L. J. Tang, W. H. Lin, & M. K. Radhakrishnan, "Size Difference in Dielectric-Breakdown-Induced Epitaxy in Narrow n- and p-Metal Oxide Semiconductor Field Effect Transistors," *Appl. Phys. Lett.* **83** (14): 2940–2942, October 2003, doi:10.1063/1.1616195.

[15] G. Ribes, J. Mitard, M. Denais, S. Bruyere, F. Monsieur, C. Parthasarathy, E. Vincent, & G. Ghibaudo, "Review on High-k Dielectrics Reliability Issues," *IEEE Trans. Device Mater. Rel.* **5** (1): 5–19, April 2005, doi:10.1109/TDMR.2005.845236.

[16] C.-S. Liao & J.-G. Hwu, "Subthreshold Swing Reduction by Double Exponential Control Mechanism in an MOS gated-MIS Tunnel Transistor," *IEEE Trans. Electron Dev.* **62** (6): 2061–2065, June 2015, doi:10.1109/TED.2015.2424245.

[17] C.-S. Liao & J.-G. Hwu, "Remote Gate-Controlled Negative Transconductance in Gated MIS Tunnel Diode," *IEEE Trans. Electron Dev.* **63** (7): 2864–2870, July 2016, doi:10.1109/TED.2016.2565688.

[18] Y. Taur & T. H. Ning. *Fundamentals of Modern VLSI Devices*, 3rd ed. Cambridge University Press, Cambridge, 2022.

[19] D. A. Neamen. *Semiconductor Physics and Devices: Basic Principles*, 4th ed. McGraw Hill, Singapore, 2012.

[20] P. A. Clifton. *Characterisation of Silicon MIS Negative Resistance Devices*. PhD thesis, Durham University, June 1989.

[21] Y.-H. Shih & J.-G. Hwu, "An On-Chip Temperature Sensor by Utilizing an MOS Tunneling Diode," *IEEE Electron Dev. Lett.* **22** (6): 299–301, June 2001, doi:10.1109/55.924848.

[22] Li Zhu & S. McNamara, "Low Power Tunneling Current Strain Sensor Using MOS Capacitors," *J. Microelectromech. Syst.* **24** (3): 755–762, June 2015, doi:10.1109/JMEMS.2014.2351778.

[23] C.-W. Liu, W.-T. Liu, M.-H. Lee, W.-S. Kuo, & B.-C. Hsu, "A Novel Photodetector Using MOS Tunneling Structures," *IEEE Electron Dev. Lett.* **21** (6): 307–300, June 2000, doi:10.1109/55.843159.

[24] E. H. Nicollian & J. R. Brews. *MOS (Metal Oxide Semiconductor) Physics and Technology*. John Wiley & Sons, 1982.

[25] C.-C. Lin, P.-L. Hsu, L. Lin, & J.-G. Hwu, "Investigation on Edge Fringing Field Effect and Oxide Thickness Dependence of Inversion Current in MOS Tunneling Diodes with Comb-Shaped Electrodes," *J. Appl. Phys.* **115** (12): 124109, March 2014, doi:10.1063/1.4870186.

[26] H.-Y. Chen & J.-G. Hwu, "Photo Sensitivity Enhanced by the Modulation of Oxide Thickness in MIS(p) Structure," *ECS Trans.* **85** (13): 1441–1448, May 2018, doi:10.1149/08513.1441ecst.

[27] Y.-K. Lin & J.-G. Hwu, "Minority Carriers Induced Schottky Barrier Height Modulation in Current Behavior of Metal-Oxide-Semiconductor Tunneling Diode," *ECS J. Solid State Sci. Technol.* **3** (6): Q132–Q135, May 2014, doi:10.1149/2.019406jss.

[28] B. E. Cross, W.-Y. Loh, R. M. Wallace, J. Kim, P. Majhi, & R. Jammy, "Near Band Edge Schottky Barrier Height Modulation Using High-k Dielectric Dipole Tuning Mechanism," *Appl. Phys. Lett.* **95** (22): 222015, December 2009, doi:10.1063/1.3263719.

[29] D. Connelly, C. Faulkner, D. E. Grupp, & J. S. Harris, "A New Route to Zero-Barrier Metal Source-Drain MOSFETs," *IEEE Trans. Nanotech.* **3** (1): 98–104, March 2014, doi:10.1109/TNANO.2003.820774.

[30] M. Y. Doghish & F. D. Ho, "A Comprehensive Analytical Model for Metal-Insulator-Semiconductor (MIS) Devices," *IEEE Trans. Electron Dev.* **39** (12): 2771–2780, December 1992, doi:10.1109/16.168723.

[31] C. R. Crowell, "The Richardson Constant for Thermionic Emission in Schottky Barrier Diodes," *Solid-State Electron.* **8** (4): 395–399, April 1965, doi:10.1016/0038-1101(65)90116-4.

[32] K.-C. Chen, K.-W. Lin, & J.-G. Hwu, "Role of Schottky Barrier Height Modulation on the Reverse Bias Current Behavior of MIS(p) Tunnel Diodes," *IEEE Access* **9**: 163929–163937, 2021, doi:10.1109/ACCESS.2021.3133575.

[33] S. Morita, A. Shinozaki, Y. Morita, K. Nishimura, T. Okazaki, S. Urabe, & M. Morita, "Tunneling Current Through Ultrathin Silicon Dioxide Films Under Light Exposure," *Jpn. J. Appl. Phys.* **43** (11B): 7857–7860, 2004, doi:10.1143/JJAP.43.7857.

[34] J.-Y. Cheng, C.-T. Huang, & J.-G. Hwu, "Comprehensive Study on the Deep-Depletion Capacitance-Voltage Behavior for Metal-Oxide-Semiconductor Capacitor with Uitra-thin Oxides," *J. Appl. Phys.* **106** (7): 074507, October 2009, doi:10.1063/1.3226853.

[35] H.-H. Lin & J.-G. Hwu, "Local Thinning Induced Less Oxide Breakdown in MOS Structures Due to Lateral Non-uniformity Effect," *ECS Trans.* **75** (5): 63–67, 2016, doi:10.1149/07505.0063ecst.

[36] Y.-C. Yang, Kuan-Wun Lin, & J.-G. Hwu, "Transient Two-State Characteristics in MIS(p) Tunnel Diode with Edge-Thickened Oxide (ETO) Structure," *ECS J. Solid State Sci. Technol.* **9** (10): 103006, October 2020, doi:10.1149/2162-8777/abc576.

[37] S.-W. Huang & J.-G. Hwu, "Transient Current Enhancement in MIS Tunnel Diodes with Lateral Electric Field Induced by Designed High-Low Oxide Layers," *IEEE Trans. Electron Dev.* **68** (12): 6580–6585, December 2021, doi:10.1109/TED.2021.3122814.

doi:10.6342/NTU202304241

[38] T.-H. Chiang & J.-G. Hwu, "Ultra-Low Subthreshold Swing in Gated MIS(p) Tunnel Diodes with Engineered Oxide Local Thinning Layers," *IEEE Trans. Electron Dev.* **67** (4): 1887–1893, April 2020, doi:10.1109/TED.2020.2976119.

[39] C.-F. Yang, B.-J. Chen, W.-C. Chen, K.-W. Lin, & J.-G. Hwu, "Gate Oxide Local Thinning Mechanism Induced Sub-60 mV/Decade Subthreshold Swing on Charge-Coupled Tunnel Transistors," *IEEE Trans. Electron Dev.* **61** (1): 279–285, January 2019, doi:10.1109/TED.2018.2879654.

[40] G. C. Jain, A. Prasad, & B. C. Chakravarty, "On the Mechanism of the Anodic Oxidation of Si at Constant Voltage," *J. Electrochem. Soc.* **126** (1): 89–92, January 1979, doi:10.1149/1.2128996.

[41] M. Grecea, G. Rotaru, N. Nastase, & G. Graciun, "Physical Properties of SiO_2 Thin Films Obtained by Anodic Oxidation," *J. Mol. Struct.* **480–481**: 607–610, 1999, doi:10.1016/S0022-2860(99)00017-4.

[42] J. Appels, E. Kooi, M. M. Paffen, J. J. H. Schatorje, & W. H. C. G. Verkuylen, "Local Oxidation of Silicon and Its Application in Semiconductor-Device Technology," *Philips Res. Repts.* **25** (2): 118–132, 1970.

[43] W. Kern & D. A. Puotinen, "Cleaning Solutions Based on Hydrogen Peroxide for Use in Silicon Semiconductor Technology," *RCA Rev.* **31** (2): 187–206, 1970.

[44] H.-W. Lu & J.-G. Hwu, "Roles of Interface and Oxide Trap Density on the Kinked Current Behavior of Al/SiO_2/Si(p) Structures with Ultra-thin Oxides," *Appl. Phys. A* **115** (3): 837–842, June 2014, doi:10.1007/s00339-013-7873-2.

[45] W. Shockley, "The Theory of p-n Junction in Semiconductors and p-n Junction Transistors," *The Bell System Technical Journal* **28** (3): 435–489, 1949, doi:10.1002/j.1538-7305.1949.tb03645.x.

[46] J. C. Ranuárez, M. J. Deen, & C.-H. Chen, "A Review of Gate Tunneling Current in MOS Devices," *Microelectron. Reliab.* **46**: 1939–1956, 2006, doi:10.1016/j.microrel.2005.12.006.

[47] R. Tsu & L. Esaki, "Tunneling in Finite Superlattice," *Appl. Phys. Lett.* **22** (11): 562–564, June 1973, doi:10.1063/1.1654509.

[48] C. B. Duke. *Tunneling in Solids*. Academic Press, New York, 1969.

[49] R. H. Fowler & L. Nordheim, "Electron Emission in Intense Electric Fields," *Proc. R. Soc. A* **119** (781): 173–181, 1928, doi:10.1098/rspa.1928.0091.

[50] M. Lenzinger & E. H. Snow, "Fowler-Nordheim Tunneling into Thermally-Grown SiO_2," *J. Appl. Phys.* **40**: 278–283, 1969, doi:10.1063/1.1657043.

[51] K. F. Schuegraf & C. Hu, "Hole Injection SiO_2 Breakdown Model for Very Low Voltage Lifetime Extrapolation," *IEEE Trans. Electron Dev.* **41** (5): 761–767, May 1994, doi:10.1109/16.285029.

[52] X. Gu, T.-L. Chen, G. Gildenblat, G. O. Workman, S. Veeraraghavan, S. Shapira, & K. Stiles, "A Surface Potential-Based Compact Model of n-MOSFET Gate-Tunneling Current," *IEEE Trans. Electron Dev.* **51** (1): 127–135, January 2004, doi:10.1109/TED.2003.820652.

[53] Y. Tsividis & C. McAndrew. *Operation and Modeling of the MOS Transistor*. Oxford University Press, Oxford, 2011.

[54] G. Iannaccone, G. Curatola, & G. Fiori. “Effective Bohm Quantum Potential for Device Simulators Based on Drift-Diffusion and Energy Transport,” in *International Conference on Simulation of Semiconductor Processes and Devices: 2004, Munich*, January 2004, doi:10.1007/978-3-7091-0624-2_64.

[55] R. Brendel. *Thin-Film Crystalline Silicon Solar Cells: Physics and Technology*. Wiley-VCH, Weinheim, 2003.

[56] J. Singh. *Semiconductor Devices: An Introduction*. McGraw-Hill, New York, 1994.

[57] C.-T. Sah, R. N. Noyce, & W. Shockley, “Carrier Generation and Recombination in P-N Junctions and P-N Junction Characteristics,” *Proc. IRE* **45** (9): 1228–1243, 1957, doi:10.1109/JRPROC.1957.278528.

[58] M. Hirose, “Electron Tunneling Through Ultrathin SiO_2,” *Mater. Sci. Eng. B* **41** (1): 35–38, 1996, doi:10.1016/S0921-5107(96)01619-4.

[59] A. Hartstein & N. F. Albert, “Determination of the Inversion-Layer Thickness from Capacitance Measurements of Metal-Oxide-Semiconductor Field-Effect Transistors with Ultrathin Oxide Layers,” *Phys. Rev. B* **38** (2): 1235–1240, July 1988, doi:10.1103/PhysRevB.38.1235.

[60] H.-W. Lu & J.-G. Hwu, "Lateral Nonuniformity of the Tunneling Current of Al/SiO_2/p-Si Capacitor in Inversion Region due to Edge Fringing-Field Effect," *ECS Trans.* **58** (7): 339–344, October 2013, doi:10.1149/05807.0339ecst.

[61] C.-F. Yang & J.-G. Hwu, "Role of Fringing Field on the Electrical Characteristics of Metal-Oxide-Semiconductor Capacitors with Co-Planar and Edge-Removed Oxides," *AIP Advances* **6**: 125017, 2016, doi:10.1063/1.4791845.

[62] P. Atkins & L. Jones. *Chemical Principles: The Quest for Insight*, 5th ed. W. H. Freeman and Company, New York, 2010.

[63] S. Hlali, N. Hizem, & A. Kalboussi, "Investigation of Capacitance Characteristics in Metal/High-k Semiconductor Devices at Different Parameters and With and Without Interface State Density (Traps)," *Bull. Mater. Sci.* **40** (5): 1035–1041, September 2017, doi:10.1007/s12034-017-1443-8.

[64] M. Bansal & R. R. Maiya, "Phototransistor: The Story So Far," *J. Electron. Inform.* **2** (4): 202–210, December 2020, doi:10.36548/jei.2020.4.002.

[65] W. Kutta, "Beitrag zur Näherungsweisen Integration Totaler Differentialgleichungen," *Z. Math. Phys.* **46**: 434–453, 1901.

[66] E. Hairer, S. P. Nørsett, & G. Wanner. *Solving Ordinary Differential Equations I: Nonstiff Problems*. Springer Verlag, Berlin, 1993.

[67] J. McDougall & E. C. Stoner, "The Computation of Fermi-Dirac Functions," *Philos. Trans. R. Soc. A* **237** (773): 67–104, February 1938, doi:10.1098/rsta.1938.0004.

Excursus

General Modeling of p-n Diode Electrostatics and Currents Under Forward Bias for All Injection Levels

This chapter summarizes my work as a visiting scholar in University of California San Diego, USA, with Prof. Yuan Taur, from May 2023 to December 2023. This chapter is a separate topic from the main dissertation.

---

doi:10.6342/NTU202304241

doi:10.6342/NTU202304241

Chapter Abstract

A general model for the electrostatics and current components of one-dimensional p-n diodes under forward bias with arbitrary doping concentrations, device dimensions, and injection level has been established in this work and shown to be consistent to TCAD simulation results. The model transcends the commonly-perceived depletion approximation and diffusion-only current model that are only valid under the low-level injection limit. As the applied voltage exceeds the diode built-in potential, it is shown that the band diagram flattens out near the junction, and the net charge density profile shrinks to two antisymmetric spikes with the magnitude given by the average of the doping concentrations. The distinction of the applied voltage, the junction quasi Fermi level splitting Δ, and the lowering of junction barrier is thus crucial, as the quantities divert at higher injection levels. It is also found that high carrier concentrations under high applied voltage give rise to non-negligible electric fields at the depletion edges, and therefore drift currents comparable or even prevailing diffusion currents, both modeled analytically in this work. As a result, the $\propto \exp(q\Delta/2k_\mathrm{B}T)$ dependency of diode currents under high injection levels of common practice can revert to $\propto \exp(q\Delta/k_\mathrm{B}T)$ under even higher injection levels. Finally, we noted that the fields and currents are overestimated by a factor in the high-level injection regime when low-level-injection values of carrier concentration decay lengths are utilized in the model. This is attributed to longer decay lengths in the high-level injection regime, and is then remedied by modeling and recalculating their new values.

EXCURSUS

EX.1 Introduction

Modeling of p-n junction electrostatics and currents has been an imperative topic ever since the structure has been proposed [Ex1, Ex2]. However, only theories under the low-level injection limit have been robustly established [Ex1, Ex3]. Minority carrier drift currents are frequently excluded from discussions in low-level injection theories [Ex2, Ex4]. Prior works that aimed for establishing high-level injection models [Ex4, Ex5, Ex6, Ex7] have addressed the non-negligibility of drift currents under the high-level injection limit, but the scopes of discussion were limited to short n^+/p diodes due to their prominent application in bipolar junction transistors (BJTs). In fact, as it will be shown later, what the works referred to as the “high-level injection regime” is associated only with high-level injection on the lightly-doped side of the diode, but low-level injection on the lightly-doped side, coined the “mixed-level injection regime” in our work. Besides, discussions subject to low-level injection usually carry the implicit assumption for the following 3 quantities to be identical: the applied voltage (V_a), the quasi Fermi level (QFL) splitting at the junction, and the lowering of the energy barrier across the junction (the “junction barrier”). However, the distinction of these quantities at higher injection levels are crucial (as they divert) but overlooked in the prior works, raising controversies in model formulation and leading to suboptimal modeling results [Ex4, Ex5]. Such distinction is to be addressed in our work.

In this work, general modeling of p-n diodes under forward bias with arbitrary doping concentrations, device dimension and injection level were achieved by solving the Poisson’s equation and exploiting the current continuity conditions. The junction QFL splitting, out of the 3 aforementioned quantities, is chosen to be the basis of our framework of modeling. Constant mobilities, flat-QFL assumption in the space charge region [Ex4], charge neutrality outside the space charge region (i.e., quasi-neutral regions), Boltzmann statistics, steady-state condition, absence of high field effects, Shockley-Read-Hall (SRH) process for carrier generation and recombination, neutral contacts, negligible depletion width compared to quasi-neutral region widths or minority carrier diffusion lengths, and complete ionization of the dopants are assumed. Important findings in the high-level injection regime include that the junction barrier never drops below zero, that the net charge density profile develops antisymmetry, that drift currents become the dominant current components, and that decay lengths become longer than in the low-level injection regime. Our general model is consistent with TCAD simulation results.

TABLE EX–I lists the adopted values of constants for model evaluation and TCAD simulation.

EX.2 Modeling of Band Diagram, Electrostatics, and Carrier Concentrations in the Depletion Region

EX.2.1 Analytical Modeling of Carrier Concentrations and the Junction Barrier

Figure EX–1 depicts the schematic band diagram of a p-n diode under forward bias with uniform acceptor and donor doping concentrations of N_a and N_d in the p- and n-regions, respectively, where $q\Delta$ and $q\phi_\mathrm{j}$ represent the junction QFL splitting and junction barrier, respectively (q: elementary charge); $-W_\mathrm{p}$ and W_n are the coordinates of the contacts, and $-x_\mathrm{p}$ and x_n are the coordinates of the (hypothetical) depletion

TABLE EX–I
List of the adopted values of constants for model evaluation and TCAD simulation.

Quantity	Unit	Value	Note
n_{i}	cm^{-3}	1.5×10^{10}	(Si @ 300 K)
$\epsilon_{\mathrm{S}}/\epsilon_0$	–	11.9	(Si @ 300 K)
T	K	300	
μ_n	$\mathrm{cm}^2/\mathrm{V\,s}$	1000	
μ_p	$\mathrm{cm}^2/\mathrm{V\,s}$	500	
τ_p	s	1×10^{-6}	
τ_n	s	1×10^{-6}	
D_n	cm^2/s	26	$= (k_{\mathrm{B}}T/q)\mu_n$
D_p	cm^2/s	13	$= (k_{\mathrm{B}}T/q)\mu_p$
L_n	µm	51	$= (D_n\tau_n)^{1/2}$
L_p	µm	36	$= (D_p\tau_p)^{1/2}$
L^*	µm	42	(EX.78)

edges (also indicated by subscripts “pE” and “nE”) where both charge neutrality and flat QFLs are assumed, as in the quasi-neutral and depletion regions, respectively. In the large-scale view of **Figure EX–1 (a)**, the bands are tilted in the quasi-neutral regions ($-W_{\mathrm{p}} < x < -x_{\mathrm{p}}$ and $x_{\mathrm{n}} < x < W_{\mathrm{n}}$) due to $I \times R$ drop under sufficiently high injection level, and thus $V_{\mathrm{a}} > \Delta$. Only under low-level injection, where the bands and majority carrier QFLs are flat in the quasi-neutral regions, are V_{a} and Δ indistinguishable. The value of ϕ_{j} at zero applied voltage is clearly the built-in potential, $\psi_{\mathrm{bi}} = (k_{\mathrm{B}}T/q)\ln(N_{\mathrm{a}}N_{\mathrm{d}}/n_{\mathrm{i}}^2)$, where k_{B} is the Boltzmann constant, T is the temperature, and n_{i} is the intrinsic carrier concentration. As an example, **Figure EX–2** shows the TCAD-simulated band diagram of a p-n diode under forward bias, in which $I \times R$ drop is apparent in the quasi-neutral regions whereas ϕ_{j} is strictly positive. In the small-scale view of **Figure EX–1 (b)**, the smooth transitions of bands are shown, and flat

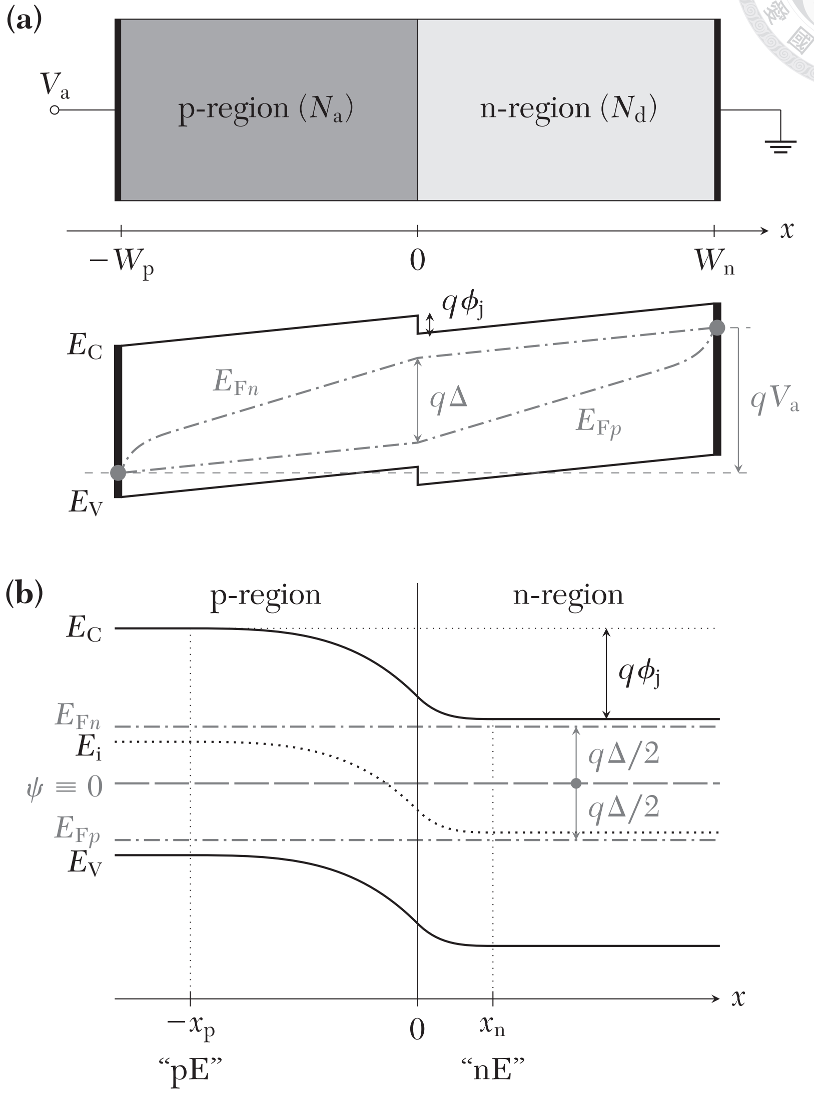


Figure EX–1. **(a)** Schematic large-scale structure and band diagram of a p-n diode under forward bias. $x = 0$ is the metallurgical junction. (N_a and N_d: doping concentrations; W_p and W_n: p- and n-region widths; V_a: applied voltage; $q\Delta$: junction QFL splitting; $q\phi_j$: junction barrier height.) **(b)** Schematic small-scale band diagram near the junction, assuming flat QFLs. ($-x_p$ and x_n: locations of the depletion edges; E_{Fn} and E_{Fp}: electron and hole QFLs.) The zero reference of the electrostatic potential ψ is set at the right middle between E_{Fn} and E_{Fp}.

doi:10.6342/NTU202304241

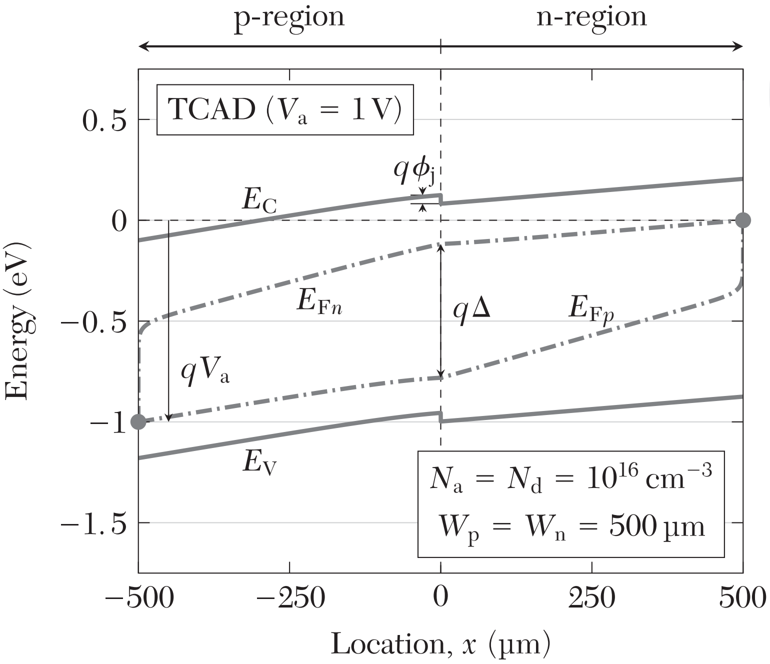


Figure EX–2. TCAD-simulated band diagram of a p-n diode with $N_a = N_d = 10^{16}\,\mathrm{cm}^{-3}$, $W_p = W_n = 500\,\mu\mathrm{m}$, and under $V_a = 1\,\mathrm{V}$.

QFLs are depicted according to the assumption. The carrier concentrations can be expressed by [Ex8]

$$\begin{aligned} n(x) &= n_i e^{(E_{Fn}-E_i)/k_BT} \\ &= n_i e^{q\psi(x)/k_BT} e^{q\Delta/2k_BT}; \qquad \text{(EX.1a)} \\ p(x) &= n_i e^{(E_i-E_{Fp})/k_BT} \\ &= n_i e^{-q\psi(x)/k_BT} e^{q\Delta/2k_BT}; \qquad \text{(EX.1b)} \end{aligned}$$

where $\psi(x) \equiv -E_i(x)/q$ is the electrostatic potential (E_i: intrinsic Fermi level), with its zero reference set at the right middle between the QFLs. Meanwhile, the Poisson's

doi:10.6342/NTU202304241

equation reads

$$\begin{aligned}\frac{\mathrm{d}^2\psi}{\mathrm{d}x^2} &= -\frac{\rho\,(x)}{\epsilon_\mathrm{S}} \\ &= -\frac{q}{\epsilon_\mathrm{S}}\left[p(x) - n(x) + N_\mathrm{d} - N_\mathrm{a}\right], \end{aligned} \tag{EX.2}$$

where $\rho\,(x)$ is the net charge density, and ϵ_S is the permittivity of semiconductor. Combined with (EX.1), the electrostatic potential in the p- and n-regions, $\psi_\mathrm{p}(x)$ and $\psi_\mathrm{n}(x)$, satisfy

$$(x < 0) \qquad \frac{\mathrm{d}^2\psi_\mathrm{p}}{\mathrm{d}x^2} = -\frac{q}{\epsilon_\mathrm{S}}\left[2n_\mathrm{i}\sinh\frac{q\psi_\mathrm{p}}{k_\mathrm{B}T}e^{q\Delta/2k_\mathrm{B}T} - N_\mathrm{a}\right]; \tag{EX.3a}$$

and

$$(x > 0) \qquad \frac{\mathrm{d}^2\psi_\mathrm{n}}{\mathrm{d}x^2} = -\frac{q}{\epsilon_\mathrm{S}}\left[2n_\mathrm{i}\sinh\frac{q\psi_\mathrm{n}}{k_\mathrm{B}T}e^{q\Delta/2k_\mathrm{B}T} + N_\mathrm{d}\right]; \tag{EX.3b}$$

respectively. The charge neutrality assumption at depletion edges ($x = -x_\mathrm{p}$ where $\psi_\mathrm{p}(-x_\mathrm{p}) \equiv \psi_\mathrm{pE}$, and $x = x_\mathrm{n}$ where $\psi_\mathrm{n}(x_\mathrm{n}) \equiv \psi_\mathrm{nE}$) requires the right-hand side of the Poisson's equation to vanish at these locations. Thus we obtain the following expression for the potential at depletion edges:

$$\psi_\mathrm{pE} = -\frac{k_\mathrm{B}T}{q}\,\mathrm{arsinh}\left(\frac{N_\mathrm{a}}{2n_\mathrm{i}}e^{-q\Delta/2k_\mathrm{B}T}\right); \tag{EX.4a}$$

$$\psi_\mathrm{nE} = \frac{k_\mathrm{B}T}{q}\,\mathrm{arsinh}\left(\frac{N_\mathrm{d}}{2n_\mathrm{i}}e^{-q\Delta/2k_\mathrm{B}T}\right); \tag{EX.4b}$$

doi:10.6342/NTU202304241

where arsinh $u = \ln(u + \sqrt{1+u^2})$ is the inverse hyperbolic sine function. Combined again with (EX.1), we acquire the expressions for carrier concentrations at $x = -x_\mathrm{p}$:

$$n_\mathrm{pE} = -\frac{N_\mathrm{a}}{2} + n_\mathrm{i}\sqrt{\left(\frac{N_\mathrm{a}}{2n_\mathrm{i}}\right)^2 + e^{q\Delta/k_\mathrm{B}T}}; \tag{EX.5a}$$

$$p_\mathrm{pE} = \frac{N_\mathrm{a}}{2} + n_\mathrm{i}\sqrt{\left(\frac{N_\mathrm{a}}{2n_\mathrm{i}}\right)^2 + e^{q\Delta/k_\mathrm{B}T}}; \tag{EX.5b}$$

and at $x = x_\mathrm{n}$:

$$n_\mathrm{nE} = \frac{N_\mathrm{d}}{2} + n_\mathrm{i}\sqrt{\left(\frac{N_\mathrm{d}}{2n_\mathrm{i}}\right)^2 + e^{q\Delta/k_\mathrm{B}T}}; \tag{EX.6a}$$

$$p_\mathrm{nE} = -\frac{N_\mathrm{d}}{2} + n_\mathrm{i}\sqrt{\left(\frac{N_\mathrm{d}}{2n_\mathrm{i}}\right)^2 + e^{q\Delta/k_\mathrm{B}T}}. \tag{EX.6b}$$

Figure EX–3 shows an example plot of the carrier concentrations as functions of Δ at the depletion edges of a p-n diode with mismatched doping concentrations ($N_\mathrm{d} > N_\mathrm{a}$). In the case that $N_\mathrm{d} \geq N_\mathrm{a}$, we may define two knee values of Δ from the plot as

$$(N_\mathrm{d} \geq N_\mathrm{a}) \qquad \Delta_\mathrm{L} \equiv \frac{2k_\mathrm{B}T}{q}\ln\left(\frac{N_\mathrm{a}}{2n_\mathrm{i}}\right); \tag{EX.7a}$$

$$\Delta_\mathrm{H} \equiv \frac{2k_\mathrm{B}T}{q}\ln\left(\frac{N_\mathrm{d}}{2n_\mathrm{i}}\right); \tag{EX.7b}$$

that demarcate the low- and high-level injection behaviors of the carrier concentrations in the p- and n-regions, respectively. ($\Delta_\mathrm{L} \leq \Delta_\mathrm{H}$.) Such definition arises from the relative magnitudes of the terms in the square roots of (EX.5) and (EX.6). Specifically, Δ_L at where the two terms in the square root have equal magnitudes in (EX.5), and likewise for Δ_H with (EX.6). From here we can further define 3 operation regimes for the p-n diode:

doi:10.6342/NTU202304241

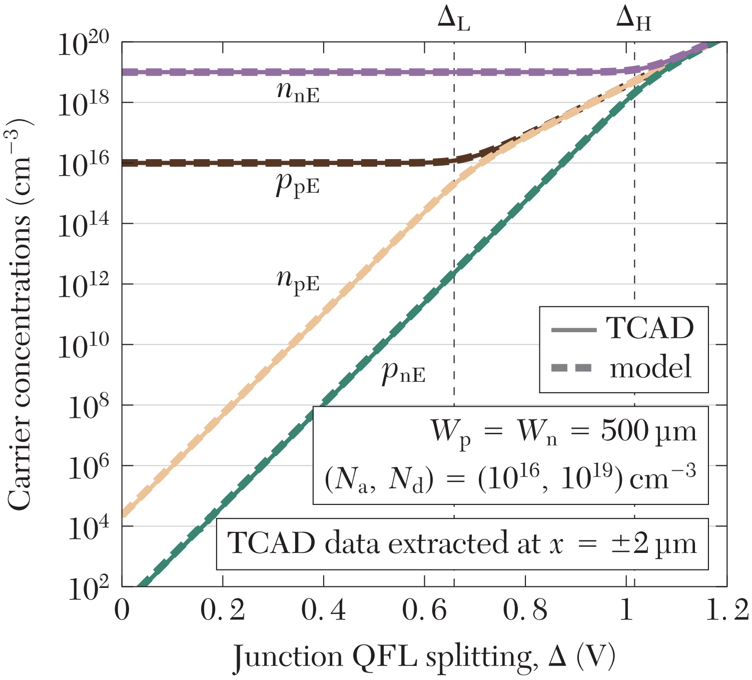


Figure EX–3. TCAD-simulated and modeled plots of the carrier concentrations at the depletion edges, as functions of Δ, in a p-n diode with mismatched doping concentrations ($N_d > N_a$) and far contacts ($W_p = W_n = 500\,\mu m$.)

- *Low-level injection regime* ($\Delta < \Delta_L$): the operation regime of common practice, where both the p- and n-regions are under low-level injection. The majority carrier concentrations are $p_{pE} \approx N_a$, $n_{pE} \approx N_d$, and the minority carrier concentrations are $n_{pE} \approx (n_i^2/N_a)e^{q\Delta/k_BT}$, $p_{nE} \approx (n_i^2/N_d)e^{q\Delta/k_BT}$.

- *High-level injection regime* ($\Delta > \Delta_H$): where both the p- and n-regions are under high-level injection. All carrier concentrations assimilate to p_{pE}, n_{pE}, p_{nE}, $n_{nE} \approx n_i e^{q\Delta/2k_BT}$.

- *Mixed-level injection regime* ($\Delta_L < \Delta < \Delta_H$): where the lightly-doped region (p) is under high-level injection, but the heavily-doped side (n) is under low-level injection. The carrier concentrations are p_{pE}, $n_{pE} \approx n_i e^{q\Delta/2k_BT}$, $n_{nE} \approx N_d$,

doi:10.6342/NTU202304241

and $p_{\mathrm{nE}} \approx (n_{\mathrm{i}}^2/N_{\mathrm{d}})e^{q\Delta/k_{\mathrm{B}}T}$. Note the absence of mixed-level injection regime in diodes with matched doping concentrations, as $\Delta_{\mathrm{L}} = \Delta_{\mathrm{H}}$.

As for the junction barrier ϕ_{j}, it is given by (EX.4):

$$\begin{aligned} \phi_{\mathrm{j}} &\equiv \psi_{\mathrm{nE}} - \psi_{\mathrm{pE}} \\ &= \frac{k_{\mathrm{B}}T}{q}\left[\operatorname{arsinh}\left(\frac{N_{\mathrm{a}}}{2n_{\mathrm{i}}}e^{-q\Delta/2k_{\mathrm{B}}T}\right) + \operatorname{arsinh}\left(\frac{N_{\mathrm{d}}}{2n_{\mathrm{i}}}e^{-q\Delta/2k_{\mathrm{B}}T}\right)\right]. \end{aligned} \tag{EX.8}$$

It can also be shown from (EX.5b) and (EX.6a) that

$$\boxed{\phi_{\mathrm{j}} = \psi_{\mathrm{bi}} - \Delta + \frac{k_{\mathrm{B}}T}{q}\ln\left(\frac{p_{\mathrm{pE}}n_{\mathrm{nE}}}{N_{\mathrm{a}}N_{\mathrm{d}}}\right).} \tag{EX.9}$$

In the low-level injection regime, $\phi_{\mathrm{j}} \approx \psi_{\mathrm{bi}} - \Delta$. That is, the lowering of ϕ_{j} from its value at zero applied voltage (ψ_{bi}) is equal to Δ, making V_{a}, Δ, and $(\psi_{\mathrm{bi}} - \phi_{\mathrm{j}})$ indistinguishable. As we enter the mixed-level injection regime, however, $\phi_{\mathrm{j}} \approx (k_{\mathrm{B}}T/q)\ln(N_{\mathrm{d}}/n_{\mathrm{i}}) - \Delta/2$, impeding its lowering. As we finally enter the high-level injection regime, $\phi_{\mathrm{j}} \to 0$ whereas Δ is increasing unboundedly, and the deviation between V_{a}, Δ, and $(\psi_{\mathrm{bi}} - \phi_{\mathrm{j}})$ is even more pronounced. **Figure EX–4** shows the $\phi_{\mathrm{j}} - \Delta$ plot, indicating that $\phi_{\mathrm{j}} \approx \psi_{\mathrm{bi}} - \Delta$ in the low-level injection and never drops below 0 even in the high-level injection regime. In other words, the band diagram near the junction flattens as Δ increases beyond the low-level injection limit, but never really flips the direction of bending. **Figure EX–5** shows the TCAD-simulated $\phi_{\mathrm{j}} - V_{\mathrm{a}}$ and $\Delta - V_{\mathrm{a}}$ plots, again demonstrating the identicality of V_{a}, Δ, and $\psi_{\mathrm{bi}} - \phi_{\mathrm{j}}$ in the low-level injection regime. Beyond this regime, Δ does not catch up V_{a} due to $I \times R$ drop, but can still increase unboundedly beyond ψ_{bi}. Contrarily, the lowering of ϕ_{j} is never beyond ψ_{bi}, and thus

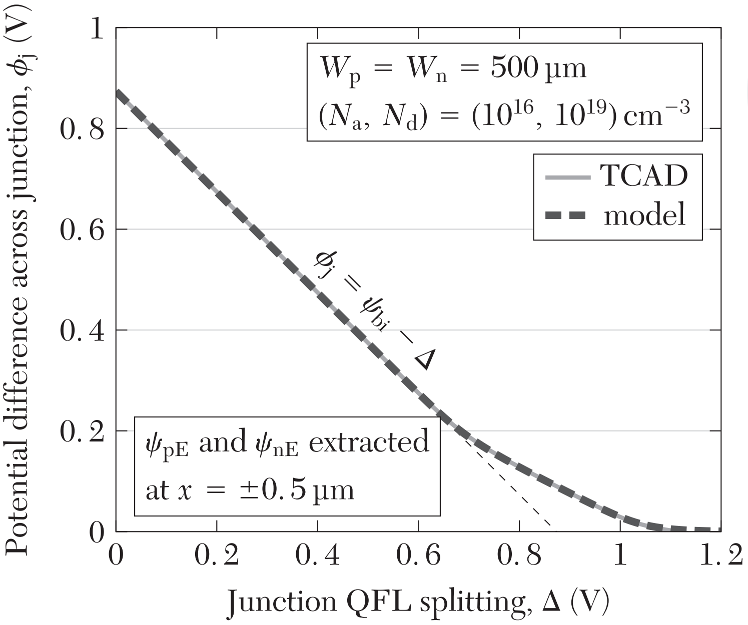


Figure EX–4. TCAD-simulated and modeled plot of ϕ_j as a function of Δ in a p-n diode with mismatched doping concentrations and far contacts.

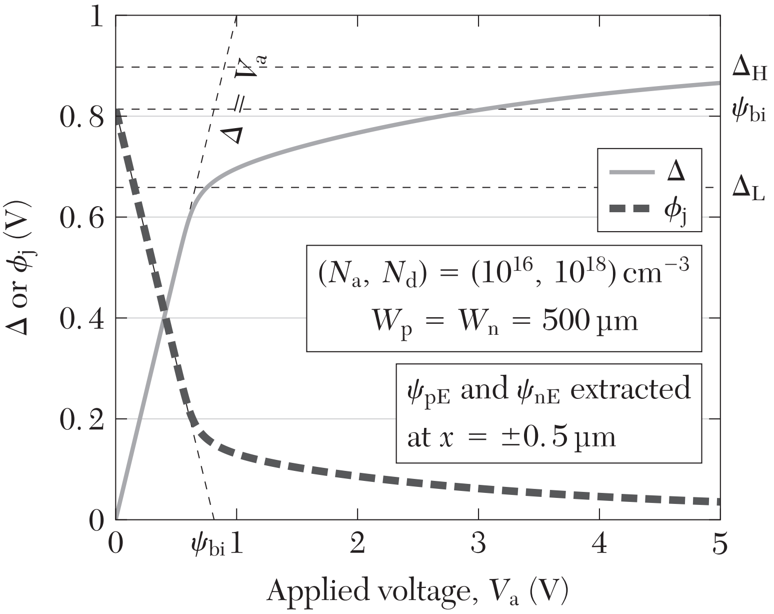


Figure EX–5. TCAD-simulated plot of ϕ_j and Δ as functions of V_a in a p-n diode with mismatched doping concentrations and far contacts.

$V_a > \Delta > (\psi_{bi} - \phi_j)$.

In some instances, it may also be useful to evaluate the potential drops in the quasi-neutral p- and n-regions, defined respectively as

$$\begin{aligned} \phi_\mathrm{p} &\equiv \psi(-W_\mathrm{p}) - \psi_\mathrm{pE} \\ &= \int_{-W_\mathrm{p}}^{-x_\mathrm{p}} \mathscr{E}_\mathrm{p}(x)\,\mathrm{d}x; \end{aligned} \tag{EX.10a}$$

and

$$\begin{aligned} \phi_\mathrm{n} &\equiv \psi_\mathrm{nE} - \psi(W_\mathrm{n}) \\ &= \int_{x_\mathrm{n}}^{W_\mathrm{n}} \mathscr{E}_\mathrm{n}(x)\,\mathrm{d}x. \end{aligned} \tag{EX.10b}$$

From the geometry of the band diagram (**Figure EX–1 (a)**),

$$V_\mathrm{a} = \phi_\mathrm{p} + (\psi_\mathrm{bi} - \phi_\mathrm{j}) + \phi_\mathrm{n}. \tag{EX.11}$$

EX.2.2 Modeling of the Potential and Net Charge Density Profiles

The Poisson's equations ((EX.3a), (EX.3b)) are second-order autonomous differential equation for $\psi_\mathrm{p}(x)$ and $\psi_\mathrm{n}(x)$, which can be integrated once [Ex9] to obtain the analytical expressions for $\mathrm{d}\psi_\mathrm{p}(x)/\mathrm{d}x$ and $\mathrm{d}\psi_\mathrm{n}(x)/\mathrm{d}x$:

$$(x \le 0) \qquad \begin{aligned} \left(\frac{\mathrm{d}\psi_\mathrm{p}}{\mathrm{d}x}\right)^2 = \frac{2k_\mathrm{B}T}{\epsilon_\mathrm{S}}\Bigg[& 2n_\mathrm{i}\cosh\frac{q\psi_\mathrm{p}}{k_\mathrm{B}T}e^{q\Delta/2k_\mathrm{B}T} + N_\mathrm{a}\frac{q\psi_\mathrm{p}}{k_\mathrm{B}T} \\ & - 2n_\mathrm{i}\sqrt{\left(\frac{N_\mathrm{a}}{2n_\mathrm{i}}\right)^2 + e^{q\Delta/k_\mathrm{B}T}} \\ & + N_\mathrm{a}\,\mathrm{arsinh}\left(\frac{N_\mathrm{a}}{2n_\mathrm{i}}e^{-q\Delta/2k_\mathrm{B}T}\right)\Bigg]; \end{aligned} \tag{EX.12a}$$

doi:10.6342/NTU202304241

and

$$(x \geq 0) \qquad \left(\frac{\mathrm{d}\psi_{\mathrm{n}}}{\mathrm{d}x}\right)^{2} = \frac{2k_{\mathrm{B}}T}{\epsilon_{\mathrm{S}}}\Bigg[2n_{\mathrm{i}}\cosh\frac{q\psi_{\mathrm{n}}}{k_{\mathrm{B}}T}e^{q\Delta/2k_{\mathrm{B}}T} - N_{\mathrm{d}}\frac{q\psi_{\mathrm{n}}}{k_{\mathrm{B}}T} - 2n_{\mathrm{i}}\sqrt{\left(\frac{N_{\mathrm{d}}}{2n_{\mathrm{i}}}\right)^{2} + e^{q\Delta/k_{\mathrm{B}}T}} + N_{\mathrm{d}}\,\mathrm{arsinh}\left(\frac{N_{\mathrm{d}}}{2n_{\mathrm{i}}}e^{-q\Delta/2k_{\mathrm{B}}T}\right)\Bigg]. \tag{EX.12b}$$

Here we assumed the boundary condition of zero electric field at the depletion edges, $(\mathrm{d}\psi_{\mathrm{p}}/\mathrm{d}x)_{-x_{\mathrm{p}}} = (\mathrm{d}\psi_{\mathrm{n}}/\mathrm{d}x)_{x_{\mathrm{n}}} = 0$. In practice, the electric fields can be nonzero under sufficiently high injection levels (see **Section EX.3**), but this effect can be negligible as the fields are generally not comparable to the built-in field.

Moreover, through matching the first-order derivatives at the metallurgical junction, $\psi'_{\mathrm{p}}(0^{-}) \equiv \psi'_{\mathrm{n}}(0^{+})$, we can acquire the analytical expressions for the potential and electric field ($\mathscr{E} = -\mathrm{d}\psi/\mathrm{d}x$) at $x = 0$:

$$\psi(0) = \frac{k_{\mathrm{B}}T}{q}\frac{1}{N_{\mathrm{a}} + N_{\mathrm{d}}}\Bigg[2n_{\mathrm{i}}\left(\sqrt{\left(\frac{N_{\mathrm{a}}}{2n_{\mathrm{i}}}\right)^{2} + e^{q\Delta/k_{\mathrm{B}}T}} - \sqrt{\left(\frac{N_{\mathrm{d}}}{2n_{\mathrm{i}}}\right)^{2} + e^{q\Delta/k_{\mathrm{B}}T}}\right) + N_{\mathrm{d}}\,\mathrm{arsinh}\left(\frac{N_{\mathrm{d}}}{2n_{\mathrm{i}}}e^{-q\Delta/2k_{\mathrm{B}}T}\right) - N_{\mathrm{a}}\,\mathrm{arsinh}\left(\frac{N_{\mathrm{a}}}{2n_{\mathrm{i}}}e^{-q\Delta/2k_{\mathrm{B}}T}\right)\Bigg]. \tag{EX.13}$$

$$-\mathscr{E}(0) = \sqrt{\frac{2k_{\mathrm{B}}T}{\epsilon_{\mathrm{S}}}}\Bigg\{2n_{\mathrm{i}}\cosh\frac{q\psi(0)}{k_{\mathrm{B}}T}e^{q\Delta/2k_{\mathrm{B}}T} + \frac{N_{\mathrm{a}}N_{\mathrm{d}}}{N_{\mathrm{a}} + N_{\mathrm{d}}}\times \Bigg[\mathrm{arsinh}\left(\frac{N_{\mathrm{a}}}{2n_{\mathrm{i}}}e^{-q\Delta/2k_{\mathrm{B}}T}\right) + \mathrm{arsinh}\left(\frac{N_{\mathrm{d}}}{2n_{\mathrm{i}}}e^{-q\Delta/2k_{\mathrm{B}}T}\right) - \sqrt{1 + \left(\frac{2n_{\mathrm{i}}}{N_{\mathrm{a}}}e^{q\Delta/2k_{\mathrm{B}}T}\right)^{2}} - \sqrt{1 + \left(\frac{2n_{\mathrm{i}}}{N_{\mathrm{d}}}e^{q\Delta/2k_{\mathrm{B}}T}\right)^{2}}\Bigg]\Bigg\}^{1/2}. \tag{EX.14}$$

We can immediately compare the low-level injection limit of (EX.14),

$$\text{(Low-level inj.)} \qquad -\mathscr{E}(0) = \sqrt{\frac{2q}{\epsilon_{\mathrm{S}}}}\left[\frac{2k_{\mathrm{B}}T}{q}\cosh\frac{q\psi(0)}{k_{\mathrm{B}}T}e^{q\Delta/2k_{\mathrm{B}}T} + \frac{N_{\mathrm{a}}N_{\mathrm{d}}}{N_{\mathrm{a}}+N_{\mathrm{d}}}\left(V_{\mathrm{bi}} - \Delta - \frac{2k_{\mathrm{B}}T}{q}\right)\right]^{1/2}, \tag{EX.15}$$

with the depletion approximation [Ex2, Ex8],

$$\text{(Depl. approx.)} \qquad -\mathscr{E}(0) = \sqrt{\frac{2q}{\epsilon_{\mathrm{S}}}}\left[\frac{N_{\mathrm{a}}N_{\mathrm{d}}}{N_{\mathrm{a}}+N_{\mathrm{d}}}(V_{\mathrm{bi}} - \Delta)\right]^{1/2}. \tag{EX.16}$$

to see the absence of the skew term (first bracketed term in (EX.14)) in the depletion approximation. This can render severe inaccuracy from the depletion approximation even in the low-level injection regime, especially when the doping concentrations are highly mismatched [Ex3]. As a demonstration, **Figure EX–6** compares the TCAD-simulated and modeled results of the $\psi(0) - \Delta$ and $\mathscr{E}(0) - \Delta$ relations in p-n diodes with several configurations of doping concentrations. In both **Figure EX–6 (a)** showing the $\psi(0) - \Delta$ and **Figure EX–6 (b)** showing the $\mathscr{E}(0) - \Delta$ curves, our model is congruent with TCAD simulation results. Also shown in **Figure EX–6 (b)** are the $\mathscr{E}(0) - \Delta$ curves given by the depletion approximation (EX.16). Even at zero applied voltage, the magnitude of $\mathscr{E}(0)$ is terribly underestimated by the depletion approximation when $N_{\mathrm{d}} \gg N_{\mathrm{a}}$, not to mention its complete collapse when Δ exceeds ψ_{bi}.

The profile of electrostatic potential, $\psi(x)$, can be numerically integrated from (EX.12a) and (EX.12b). Subsequently, the profile of net charge density $\rho(x)$ can be calculated from (EX.1). **Figure EX–7** overlays the plots of the modeled and TCAD-simulated $\rho(x)$ profiles in a long diode with slightly mismatched doping concentrations

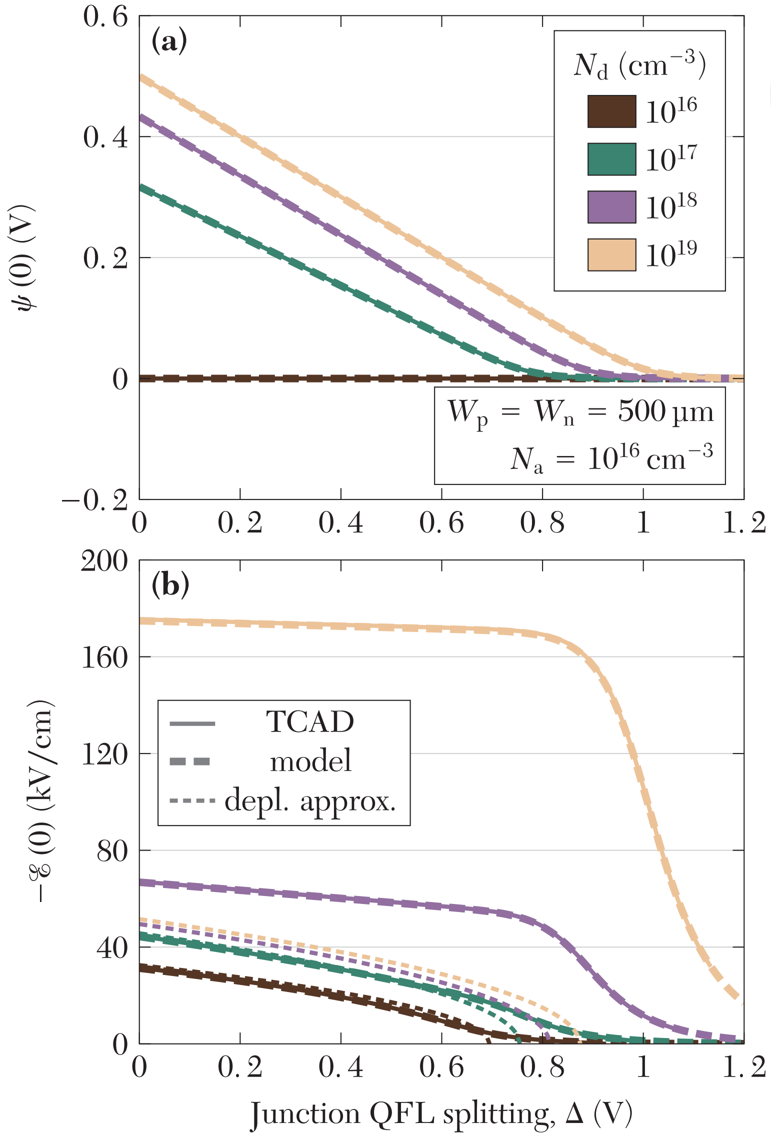


Figure EX–6. **(a)** TCAD-simulated and modeled plots for the potential at the metallurgical junction, $\psi\,(0)$, as a function of Δ, with fixed $N_\mathrm{a} = 10^{16}\,\mathrm{cm}^{-3}$ and various N_d. $W_\mathrm{p} = W_\mathrm{n} = 500\,\mu\mathrm{m}$. **(b)** Same plots for the electric field at the metallurgical junction, $\mathscr{E}\,(0)$. Curves given by depletion approximation are also included. Depletion approximation is terribly inaccurate, even under low-level injection, when the doping concentrations are highly asymmetric. (Compare $\mathscr{E}\,(0) = -175\,\mathrm{kV/cm}$ for TCAD & model vs. $-52\,\mathrm{kV/cm}$ for depletion approximation with $N_\mathrm{d} = 10^{19}\,\mathrm{cm}^{-3}$ at $\Delta = 0$.)

($N_\mathrm{d} \gtrsim N_\mathrm{a}$), showing model and TCAD agreement. Note the curve tails in the vicinity of depletion edges, which are beyond the prediction of depletion approximation.

doi:10.6342/NTU202304241

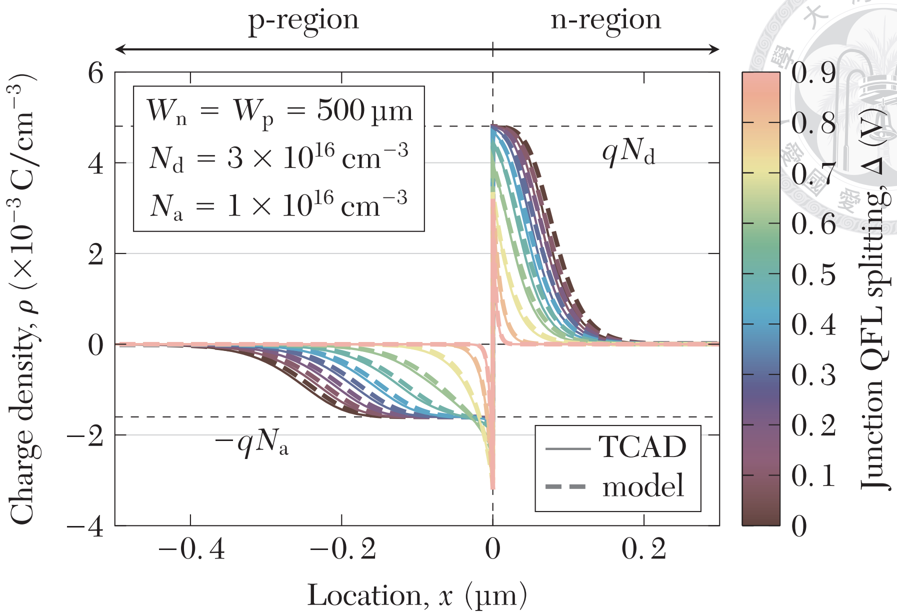


Figure EX–7. TCAD-simulated and modeled plots of the charge density profiles, under different values of Δ, in a diode with slightly mismatched doping concentrations, $N_d = 3 \times 10^{16}\,\text{cm}^{-3}$, $N_a = 1 \times 10^{16}\,\text{cm}^{-3}$, and $W_p = W_n = 500\,\mu\text{m}$.

Note also the shrinking of $\rho(x)$ into two antisymmetric spikes at higher injection levels, with the peak magnitudes also seemingly identical, despite the fact that $N_a \neq N_d$. (In contrast, $\rho(x)$ is not totally antisymmetric under low-level injection, in terms of both peak magnitudes and widths.) **Figure EX–8** sequentially shows the evolution of band diagram and $\rho(x)$ in the same diode, as Δ increases from 0 to high injection levels. Following the increment of Δ, the band eventually flattens out, and $\rho(x)$ shrinks to two antisymmetric spikes. The total antisymmetry of $\rho(x)$ under high-level injection can be understood from the space charge terms N_a and N_d being less prominent in the Poisson's equations ((EX.3a), (EX.3b)), so their mismatch takes less effect in $\rho(x)$. The identicality of peak magnitudes across the metallurgical junction can be understood

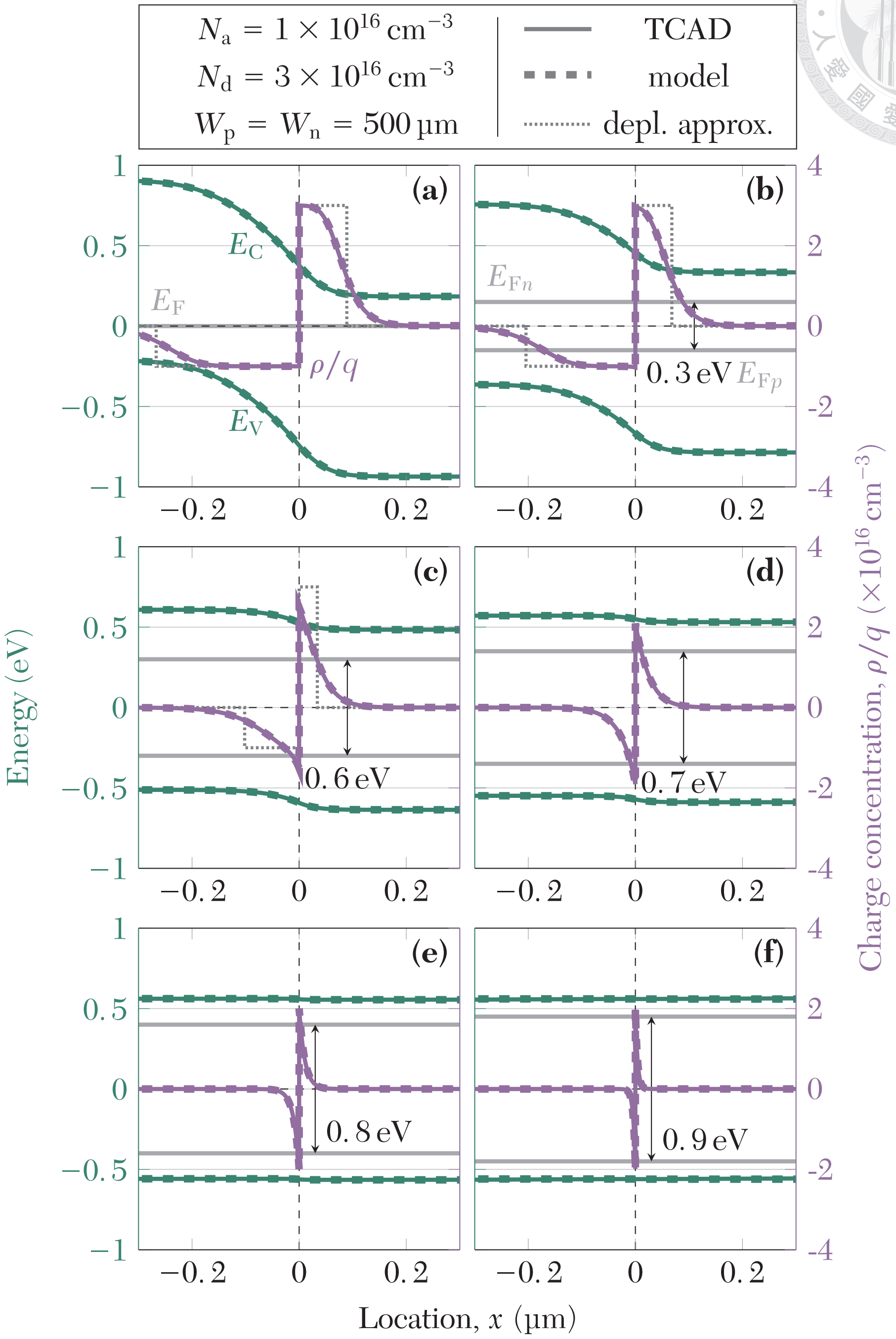


Figure EX–8. Evolution of the band diagram and $\rho(x)$ as Δ increases, in a diode with slightly mismatched doping concentrations, $N_d = 3\times10^{16}$ cm^{-3} ($\Delta_H = 0.72$ V), $N_a = 1\times10^{16}$ cm^{-3} ($\Delta_L = 0.66$ V), and $W_p = W_n = 500$ µm. Model vs. TCAD. From **(a)** to **(f)**: $q\Delta = 0$ eV, 0.3 eV, 0.6 eV, 0.7 eV, 0.8 eV and 0.9 eV.

from the limiting behavior of $\psi\,(0)$ (EX.13) under high-level injection:

$$\text{(High-level inj.)} \qquad \psi\,(0) \approx \frac{k_\mathrm{B}T}{q}\frac{N_\mathrm{d}-N_\mathrm{a}}{4n_\mathrm{i}}e^{-q\Delta/2k_\mathrm{B}T}. \tag{EX.17}$$

The carrier concentrations at $x = 0$ (EX.1) are therefore given by

$$\text{(High-level inj.)} \qquad p\,(0) \approx n_\mathrm{i}e^{q\Delta/2k_\mathrm{B}T} + \frac{N_\mathrm{a}-N_\mathrm{d}}{4}; \tag{EX.18a}$$

$$n\,(0) \approx n_\mathrm{i}e^{q\Delta/2k_\mathrm{B}T} + \frac{N_\mathrm{d}-N_\mathrm{a}}{4}; \tag{EX.18b}$$

yielding equal magnitudes of the peaks of $\rho\,(x)$ under high-level injection, with its value given by the arithmetic mean of N_a and N_d:

$$\text{(High-level inj.)} \qquad \rho\,(0^+) = -\rho(0^-) = \frac{q(N_\mathrm{a}+N_\mathrm{d})}{2}. \tag{EX.19}$$

EX.2.3 Universal Modeling of Potential and Net Charge Density Profiles Under Low-Level Injection

In **Figure EX–7**, at each depletion edge among different values of Δ under low-level injection, the $\rho\,(x)$ profiles seemingly undergo only horizontal shift, without shape changes. Moreover, the curve shapes of the $x < 0$ and $x > 0$ portions also look alike, except with 180° rotation and with horizontal and vertical scaling. In fact, we will demonstrate in this subsection that all $\rho\,(x)$ profiles under low-level injection, regardless of the doping concentrations and Δ, follow a universal shape, and can therefore be described by the translation and scaling of a universal function. Likewise, this statement is also true for $\psi_\mathrm{p}(x)$ and $\psi_\mathrm{n}(x)$.

doi:10.6342/NTU202304241

Taking the n-region as an example (p-region can be treated similarly.) Starting from

$$\begin{aligned}\left(\frac{\mathrm{d}\psi_\mathrm{n}}{\mathrm{d}x}\right)^2 = \frac{2k_\mathrm{B}T}{\epsilon_\mathrm{S}}\Bigg[&2n_\mathrm{i}\cosh\frac{q\psi_\mathrm{n}}{k_\mathrm{B}T}e^{q\Delta/2k_\mathrm{B}T} - N_\mathrm{d}\frac{q\psi_\mathrm{n}}{k_\mathrm{B}T}\\ &- 2n_\mathrm{i}\sqrt{\left(\frac{N_\mathrm{d}}{2n_\mathrm{i}}\right)^2 + e^{q\Delta/k_\mathrm{B}T}}\\ &+ N_\mathrm{d}\operatorname{arsinh}\left(\frac{N_\mathrm{d}}{2n_\mathrm{i}}e^{-q\Delta/2k_\mathrm{B}T}\right)\Bigg];\end{aligned} \qquad \text{(from EX.12b)}$$

Under low-level injection, all bracketed terms except for the first term can be simplified and combined, factoring out N_d. Also expanding the first term yields,

$$\text{(Low-level inj.)}\quad \begin{aligned}\left(\frac{\mathrm{d}\psi_\mathrm{n}}{\mathrm{d}x}\right)^2 \approx \frac{2k_\mathrm{B}T}{\epsilon_\mathrm{S}}\Bigg[&n_\mathrm{i}e^{q\psi_\mathrm{n}/k_\mathrm{B}T}e^{q\Delta/2k_\mathrm{B}T} + n_\mathrm{i}e^{-q\psi_\mathrm{n}/k_\mathrm{B}T}e^{q\Delta/2k_\mathrm{B}T}\\ &+ N_\mathrm{d}\left(\ln\frac{N_\mathrm{d}}{n_\mathrm{i}} - 1 - \frac{k_\mathrm{B}T}{q}\left(\psi_\mathrm{n} + \frac{\Delta}{2}\right)\right)\Bigg].\end{aligned} \qquad \text{(EX.20)}$$

In this expression, the first, second, and third bracketed terms are clearly associated with the electron concentration, hole concentration, and space charge density, respectively. Arguably, the second bracketed term may be omitted, since the hole concentration in the n-region is negligible compared either to N_d at lower ψ_n, or to the electron concentration at higher ψ_n. The equation therefore simplifies to

$$\left(\frac{\mathrm{d}\psi_\mathrm{n}}{\mathrm{d}x}\right)^2 \approx \frac{2k_\mathrm{B}T}{\epsilon_\mathrm{S}}\left[n_\mathrm{i}e^{q(\psi_\mathrm{n}+\Delta/2)/k_\mathrm{B}T} + N_\mathrm{d}\left(\ln\frac{N_\mathrm{d}}{n_\mathrm{i}} - 1 - \frac{k_\mathrm{B}T}{q}\left(\psi_\mathrm{n} + \frac{\Delta}{2}\right)\right)\right]. \qquad \text{(EX.21)}$$

Next, we perform the following change of variable:

$$\boxed{\omega_\mathrm{n}(\psi_\mathrm{n}) \equiv \ln\left(\frac{N_\mathrm{d}}{n_\mathrm{i}}\right) - \frac{q}{k_\mathrm{B}T}\left(\psi_\mathrm{n} + \frac{\Delta}{2}\right),} \qquad \text{(EX.22)}$$

then (EX.21) becomes

$$\left(\frac{\mathrm{d}\omega_{\mathrm{n}}}{\mathrm{d}x}\right)^2 = \frac{2}{\lambda_{\mathrm{D}}^2}(e^{-\omega_{\mathrm{n}}} + \omega_{\mathrm{n}} - 1); \tag{EX.23}$$

where $\lambda_{\mathrm{D}} \equiv (k_{\mathrm{B}} T \epsilon_{\mathrm{S}} / q^2 N_{\mathrm{d}})^{1/2}$ is the extrinsic Debye length. Taking care of the sign, the $x - \omega_{\mathrm{n}}$ relation is obtained by integration:

$$x = \frac{\lambda_{\mathrm{D}}}{\sqrt{2}} \int_{\omega_{\mathrm{n}}}^{\omega_{\mathrm{n0}}} \frac{\mathrm{d}\omega'}{\sqrt{e^{-\omega'} + \omega' - 1}}, \tag{EX.24}$$

in which

$$\begin{aligned} \omega_{\mathrm{n0}} &\equiv \omega_n(x = 0) \\ &= \ln\left(\frac{N_{\mathrm{d}}}{n_{\mathrm{i}}}\right) - \frac{q}{k_{\mathrm{B}} T}\left(\psi_{\mathrm{n}}(0) + \frac{\Delta}{2}\right). \end{aligned} \tag{EX.25}$$

If we define a universal function $I(\omega)$ from the following integral,

$$\boxed{I(\omega) \equiv \frac{1}{\sqrt{2}} \int_{\ln 2}^{\omega} \frac{\mathrm{d}\omega'}{\sqrt{e^{-\omega'} + \omega' - 1}}}, \tag{EX.26}$$

where we conveniently set $I(\ln 2) = 0$, then (EX.24) becomes

$$\boxed{x(\omega_{\mathrm{n}}) = \lambda_{\mathrm{D}}\left[I(\omega_{\mathrm{n0}}) - I(\omega_{\mathrm{n}})\right]}. \tag{EX.27}$$

To inspect the properties of the function $I(\omega)$, we may develop an approximate

doi:10.6342/NTU202304241

expression for it by inserting cancelling integral terms into (EX.26):

$$
\begin{aligned}
I(\omega) &= \frac{1}{\sqrt{2}}\Bigg[\int_{\ln 2}^{1} \frac{d\omega'}{\sqrt{e^{-\omega'}+\omega'-1}} - \int_{1}^{\omega}\left(\frac{1}{\sqrt{\omega'-1}} - \frac{1}{\sqrt{e^{-\omega'}+\omega'-1}}\right) d\omega' \\
&\qquad + \int_{1}^{\omega} \frac{d\omega'}{\sqrt{\omega'-1}}\Bigg] \\
&= \frac{1}{\sqrt{2}}\Bigg[\int_{\ln 2}^{1} \frac{d\omega'}{\sqrt{e^{-\omega'}+\omega'-1}} - \int_{1}^{\omega}\left(\frac{1}{\sqrt{\omega'-1}} - \frac{1}{\sqrt{e^{-\omega'}+\omega'-1}}\right) d\omega'\Bigg] \\
&\qquad + \sqrt{2(\omega-1)}.
\end{aligned}
\tag{EX.28}
$$

At locations within the depletion region and sufficiently distant from the depletion edge ($\psi_n(x) \ll \psi_{nE} - k_B T/q$, or equivalently $\omega \gg 1$), the first term in this expression converges rapidly, rendering the following approximation:

$$
(\omega \gg 1) \qquad \boxed{I(\omega) \approx \sqrt{2(\omega-1)} - C_I}, \tag{EX.29}
$$

with the constant C_I given by

$$
\begin{aligned}
C_I &\equiv \frac{1}{\sqrt{2}}\Bigg[-\int_{\ln 2}^{1} \frac{d\omega'}{\sqrt{e^{-\omega'}+\omega'-1}} + \int_{1}^{\infty}\left(\frac{1}{\sqrt{\omega'-1}} - \frac{1}{\sqrt{e^{-\omega'}+\omega'-1}}\right) d\omega'\Bigg] \\
&\approx 0.160\,840.
\end{aligned}
\tag{EX.30}
$$

Figure EX–9 plots the function $I(\omega)$ (EX.26) along with its approximation (EX.29), showing its validity for $\omega \gg 1$. Subject to such approximation, (EX.27) then becomes

$$
(\omega_n,\ \omega_{n0} \gg 1) \qquad x(\omega_n) \approx \lambda_D\left[\sqrt{2(\omega_{n0}-1)} - \sqrt{2(\omega_n-1)}\right], \tag{EX.31}
$$

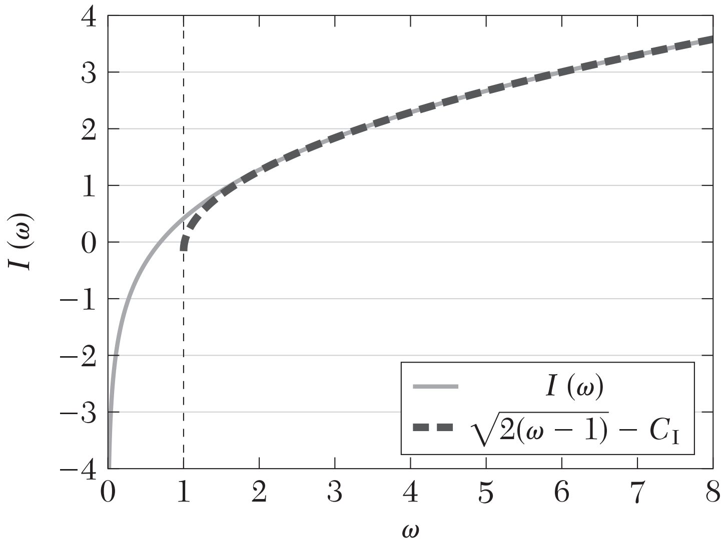


Figure EX–9. Plot of the function $I\,(\omega)$ as defined by (EX.26), along with its approximation (EX.29).

or equivalently,

$$(\omega_\mathrm{n},\,\omega_\mathrm{n0}\gg 1)\qquad \psi_\mathrm{n}(x)\approx\psi_\mathrm{n}(0)+\frac{k_\mathrm{B}T}{q}\frac{x}{\lambda_\mathrm{D}}\left(\sqrt{2(\omega_\mathrm{n0}-1)}-\frac{x}{2\lambda_\mathrm{D}}\right).\qquad\text{(EX.32)}$$

The quadratic approximate formulation for $\psi_\mathrm{n}(x)$ resembles the depletion approximation.

For the $\rho_\mathrm{n}\,(x)$ profile, we define the half-depletion coordinate, $x_{1/2}$, as where $\rho_\mathrm{n}\,(x)$ is half its maximum value under the low-level injection limit, $\rho_\mathrm{n}(0^+)=qN_\mathrm{d}$.

$$\boxed{\rho_\mathrm{n}(x_{1/2})\equiv\frac{qN_\mathrm{d}}{2}}.\qquad\text{(EX.33)}$$

doi:10.6342/NTU202304241

Utilizing (EX.1a), we can express the charge density in terms of ω_{n}:

$$\begin{aligned} \rho_{\mathrm{n}}(x) &\approx q[-n_{\mathrm{i}}e^{q\psi_{\mathrm{n}}(x)/k_{\mathrm{B}}T}e^{q\Delta/2k_{\mathrm{B}}T} + N_{\mathrm{d}}] \\ &= qN_{\mathrm{d}}\Big[1 - e^{-\omega_{\mathrm{n}}(x)}\Big]. \end{aligned} \tag{EX.34}$$

This requires $\omega\,(x_{1/2}) = \ln 2$, making $I\,(\omega) = 0$ at this location by definition. (EX.27) then conveniently relates $x_{1/2}$ and $\omega_{\mathrm{n}0}$ as

$$\boxed{x_{1/2} = \lambda_{\mathrm{D}} I\,(\omega_{n0})}, \tag{EX.35}$$

and the equation itself may be rewritten as

$$x\,(\omega_{\mathrm{n}}) = x_{1/2} - \lambda_{\mathrm{D}} I\,(\omega_{\mathrm{n}}). \tag{EX.36}$$

In other words,

$$\omega_{\mathrm{n}}(x) = I^{-1}\left(-\frac{x - x_{1/2}}{\lambda_{\mathrm{D}}}\right), \tag{EX.37}$$

where I^{-1} is the inverse function of I. Equivalently is the universal expression for $\psi_{\mathrm{n}}(x)$ under low-level injection, in terms of the function I^{-1}:

$$\boxed{\frac{q}{k_{\mathrm{B}}T}\,[\psi_{\mathrm{nE}} - \psi_{\mathrm{n}}(x)] = I^{-1}\left(-\frac{x - x_{1/2}}{\lambda_{\mathrm{D}}}\right)}. \tag{EX.38}$$

We can tell that λ_{D} and $x_{1/2}$ serve as the horizontal shift and scale factors, respectively. Since λ_{D} is independent of Δ, changes in Δ will only shift the curves horizontally (by affecting $x_{1/2}$) without horizontal scaling. **Figure EX–10 (a)** shows the model pre-

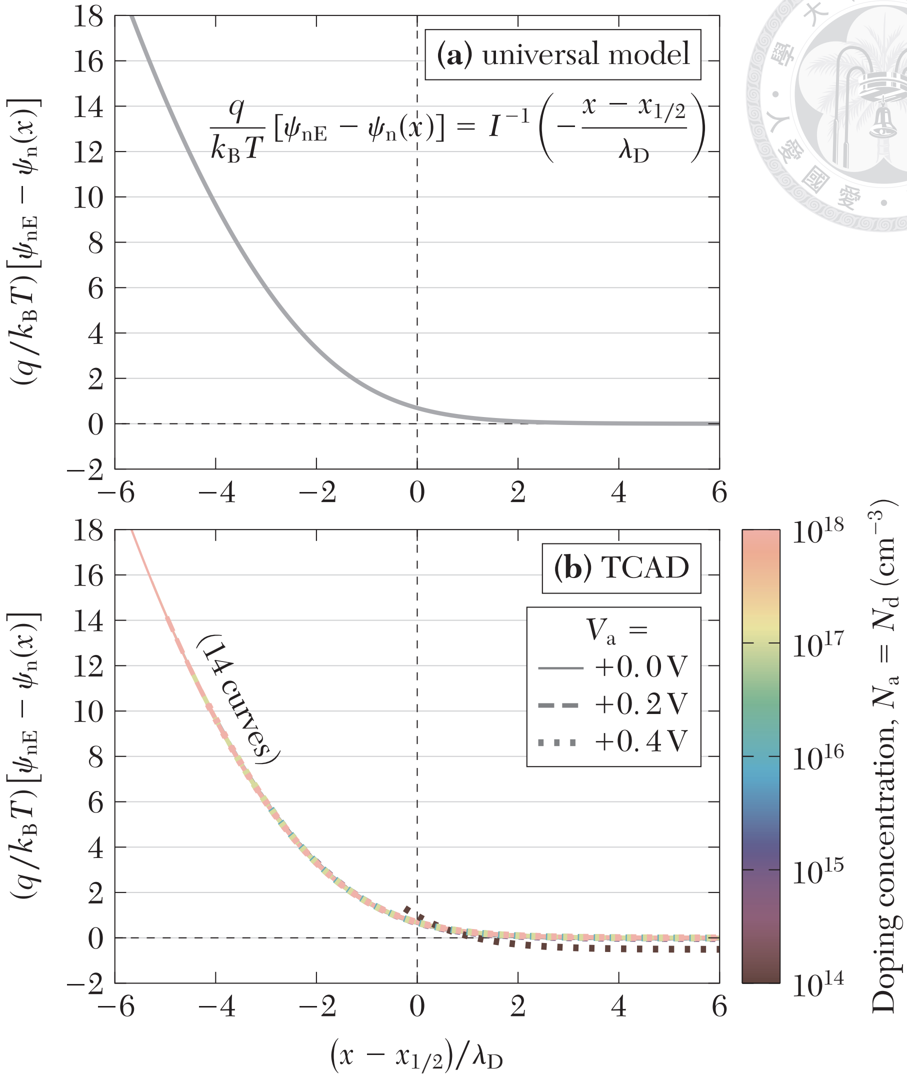


Figure EX–10. Universal plots of the $\psi_n(x) - x$ profiles given by **(a)** the model (EX.38), and **(b)** TCAD simulation results with 5 different doping concentrations and 3 different values of V_a (Δ under low-level injection), each curve associated with a different value of $x_{1/2}$. All TCAD curves merge in accordance with the universal model, except the one with the lowest doping concentration and the highest V_a, where the model validity range of low-level injection is exceeded.

diction of the universal $\psi_n(x) - x$ profile given by (EX.38), which shall be good for arbitrary N_d and Δ as long as the diode is biased in the low-level injection regime. **Figure EX–10 (b)** overlays multiple TCAD-simulated $\psi_n(x) - x$ curves with a wide range of N_d and Δ in a p-n diode with matched doping concentrations. As predicted, all curves merge after the x- and y- axes are manipulated according to the theory (EX.38), except for the curves with the lowest value of N_d and highest values of Δ, as the bias condition has gone beyond the model validity range.

Likewise, the net charge density profile can also be expressed in terms of a universal function:

$$\frac{\rho_\mathrm{n}(x)}{qN_\mathrm{d}} = F\left(-\frac{x - x_{1/2}}{\lambda_\mathrm{D}}\right), \tag{EX.39}$$

where

$$F\,(u) \equiv 1 - \exp\left[-I^{-1}(u)\right]. \tag{EX.40}$$

This is derived from (EX.34). In this case, qN_d is the vertical scale factor, with $x_{1/2}$ and λ_D still being the horizontal shift and scale factors, respectively. **Figure EX–11 (a)** shows the model prediction of the universal $\rho_\mathrm{n}(x) - x$ profile given by (EX.39), which shall be good for arbitrary N_d and Δ as long as the diode is biased in the low-level injection regime. **Figure EX–11 (b)** overlays multiple TCAD-simulated $\rho_\mathrm{n}(x) - x$ curves with a wide range of N_d and Δ in a p-n diode with matched doping concentrations, akin to **Figure EX–10 (b)**. All curves merge except for the one with the lowest N_d and the highest Δ for the same reason.

The $\rho_n(x)$ profiles in **Figure EX–11** undergo gradual changes around $x = x_{1/2}$, in contrast to the depletion approximation that assumes abrupt profiles. From (EX.26), (EX.39), and (EX.40), the slope of $\rho_\mathrm{n}(x)$ at the half-depletion edge is

$$\begin{aligned} \left.\frac{\mathrm{d}\rho_\mathrm{n}}{\mathrm{d}x}\right|_{x=x_{1/2}} &= -\frac{\sqrt{-1+2\ln 2}}{2}\frac{qN_\mathrm{d}}{\lambda_\mathrm{D}} \\ &\approx -0.310\,763\,\frac{qN_\mathrm{d}}{\lambda_\mathrm{D}}. \end{aligned} \tag{EX.41}$$

doi:10.6342/NTU202304241

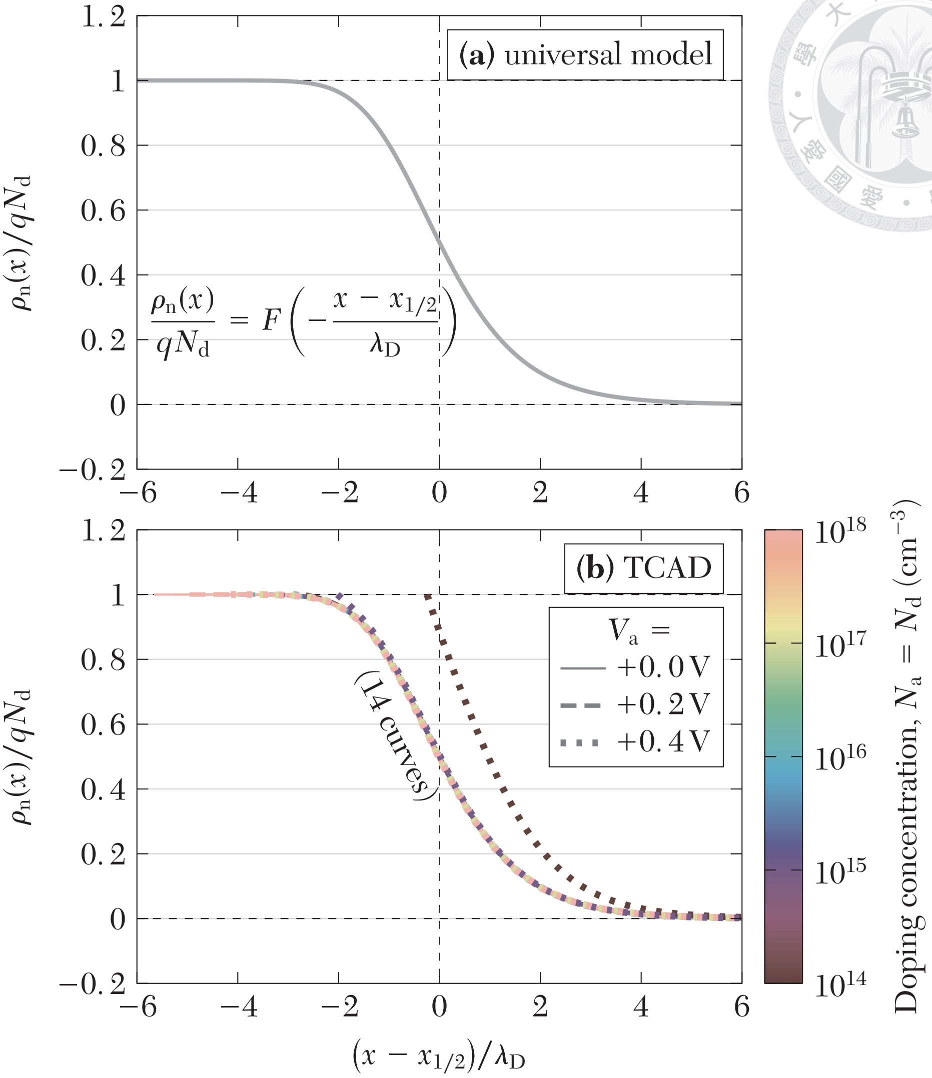


Figure EX–11. Universal plots of the $\rho_n(x) - x$ profiles given by **(a)** the model (EX.39), and **(b)** TCAD simulation results with 5 different doping concentrations and 3 different values of V_a (= Δ under low-level injection), each curve associated with a different value of $x_{1/2}$. All TCAD curves merge in accordance with the universal model, except the one with the lowest doping concentration and the highest V_a, where the model validity range of low-level injection is exceeded.

doi:10.6342/NTU202304241

EX.3 Analytical Modeling of Diffusion and Drift Current Components Across the Junction

EX.3.1 Expressions for the Electric Fields at the Depletion Edges

Under high applied voltage, there must be voltage drops and fields in the quasi-neutral regions, since the lowering of ϕ_j will eventually saturate. The nonzero electric fields give rise to drift currents, which is usually considered negligible under low-level injection. It is of commonly practice to evaluate the minority carrier diffusion currents at the depletion edges upon deriving the diode equation [Ex1, Ex2]. In this section, we aim to evaluate the fields and thus the minority carrier drift currents at the depletion edges, which we can compare with the diffusion current counterparts, or adding them up to obtain the total current. We may exploit the current continuity equation for electron and holes across the junction:

$$(J_n)_\mathrm{pE} = (J_n)_\mathrm{nE} - J_\mathrm{SC}; \tag{EX.42a}$$

$$(J_p)_\mathrm{pE} = (J_p)_\mathrm{nE} + J_\mathrm{SC}; \tag{EX.42b}$$

where $J_n \equiv J_{n,\,\mathrm{drift}} + J_{n,\,\mathrm{diff}}$ is the total electron current, $J_p \equiv J_{p,\,\mathrm{drift}} + J_{p,\,\mathrm{diff}}$ is the total hole current, and J_SC is the space charge current density [Ex8] that arises from carrier recombination in the depletion region in the case of forward bias and contributes to an ideality factor of 2 in the diode current under small V_a [Ex2]. As J_SC is negligible under sufficiently high V_a, usually still far below the onset of high-level injection, it may be neglected in our scope of interest. Expanding all the drift and diffusion current

doi:10.6342/NTU202304241

components [Ex8],

$$qn_{\mathrm{pE}}\mu_n\mathscr{E}_{\mathrm{pE}} + qD_n\left(\frac{\mathrm{d}n}{\mathrm{d}x}\right)_{\mathrm{pE}} = qn_{\mathrm{nE}}\mu_n\mathscr{E}_{\mathrm{nE}} + qD_{\mathrm{n}}\left(\frac{\mathrm{d}n}{\mathrm{d}x}\right)_{\mathrm{nE}}; \tag{EX.43a}$$

$$qp_{\mathrm{pE}}\mu_p\mathscr{E}_{\mathrm{pE}} - qD_p\left(\frac{\mathrm{d}p}{\mathrm{d}x}\right)_{\mathrm{pE}} = qp_{\mathrm{nE}}\mu_p\mathscr{E}_{\mathrm{nE}} - qD_{\mathrm{p}}\left(\frac{\mathrm{d}p}{\mathrm{d}x}\right)_{\mathrm{nE}}; \tag{EX.43b}$$

where μ_n and μ_p are the electron and hole mobilities, respectively; $D_n = (k_{\mathrm{B}}T/q)\mu_n$ and $D_p = (k_{\mathrm{B}}T/q)\mu_p$ are the electron and hole diffusion coefficients, respectively; $\mathscr{E}_{\mathrm{pE}}$ and $\mathscr{E}_{\mathrm{nE}}$ are the electric fields at the p- and n-depletion edges, respectively. We may now recall the quasi-neutrality assumption in the quasi-neutral regions and at the depletion edges that facilitates $p_{\mathrm{p}}(x) = n_{\mathrm{p}}(x) + N_{\mathrm{a}}$ and $n_{\mathrm{n}}(x) = p_{\mathrm{n}}(x) + N_{\mathrm{d}}$. As a result, the electron and hole concentrations shall possess the same derivative: $\mathrm{d}p_{\mathrm{p}}/\mathrm{d}x = \mathrm{d}n_{\mathrm{p}}/\mathrm{d}x$ and $\mathrm{d}n_{\mathrm{n}}/\mathrm{d}x = \mathrm{d}p_{\mathrm{n}}/\mathrm{d}x$ at each depletion edges. If we further define the *minority carrier decay lengths*, λ_n at the p-depletion edge and λ_p at the n-depletion edge, as

$$\left(\frac{\mathrm{d}n}{\mathrm{d}x}\right)_{\mathrm{pE}} \equiv \frac{n_{\mathrm{pE}}}{\lambda_n}; \tag{EX.44a}$$

$$\left(\frac{\mathrm{d}p}{\mathrm{d}x}\right)_{\mathrm{nE}} \equiv -\frac{p_{\mathrm{nE}}}{\lambda_p}; \tag{EX.44b}$$

with the low-level-injection values (associated with hyperbolic minority carrier concentration profiles) given by [Ex8]

(Low-level inj.)
$$\lambda_n = L_n \tanh(W_{\mathrm{p}}/L_n); \tag{EX.45a}$$

$$\lambda_p = L_p \tanh(W_{\mathrm{n}}/L_p); \tag{EX.45b}$$

doi:10.6342/NTU202304241

as an example, where $L_n = \sqrt{D_n \tau_n}$ and $L_p = \sqrt{D_p \tau_p}$ are the minority electron and hole diffusion lengths (τ_n and τ_p: minority excess electron and hole lifetimes,) the analytical expressions for $\mathscr{E}_{\mathrm{pE}}$ and $\mathscr{E}_{\mathrm{nE}}$, in terms of carrier concentrations and decay lengths, can be solved from (EX.43) as

$$\mathscr{E}_{\mathrm{pE}} = \frac{k_{\mathrm{B}}T}{q} \frac{p_{\mathrm{nE}} + n_{\mathrm{nE}}}{p_{\mathrm{pE}}n_{\mathrm{nE}} - p_{\mathrm{nE}}n_{\mathrm{pE}}} \left(\frac{n_{\mathrm{pE}}}{\lambda_n} + \frac{p_{\mathrm{nE}}}{\lambda_n} \right); \tag{EX.46a}$$

$$\mathscr{E}_{\mathrm{nE}} = \frac{k_{\mathrm{B}}T}{q} \frac{p_{\mathrm{pE}} + n_{\mathrm{pE}}}{p_{\mathrm{pE}}n_{\mathrm{nE}} - p_{\mathrm{nE}}n_{\mathrm{pE}}} \left(\frac{n_{\mathrm{pE}}}{\lambda_n} + \frac{p_{\mathrm{nE}}}{\lambda_p} \right). \tag{EX.46b}$$

Figure EX–12 shows the modeled and TCAD-simulation $\mathscr{E}_{\mathrm{pE}}$ and $\mathscr{E}_{\mathrm{nE}}$ plots as functions of Δ in a p-n diode with far contacts, with two configurations of doping concentrations: **(a)** $N_{\mathrm{a}} = N_{\mathrm{d}}$ and **(b)** $N_{\mathrm{a}} \lll N_{\mathrm{d}}$. The modeled curves utilize low-level-injection values of λ_n and λ_p (EX.45). $\mathscr{E}_{\mathrm{pE}}$ and $\mathscr{E}_{\mathrm{nE}}$ in both cases exhibit proportionality to $e^{q\Delta/k_{\mathrm{B}}T}$ in the low-level injection regime, and proportionality to $e^{q\Delta/2k_{\mathrm{B}}T}$ in the high-level injection regime. In **Figure EX–12 (a)**, the matched doping concentrations result in symmetric $\mathscr{E}_{\mathrm{pE}}$ and $\mathscr{E}_{\mathrm{nE}}$. In **Figure EX–12 (b)**, the electric field at the depletion edge of the lightly-doped side ($\mathscr{E}_{\mathrm{pE}}$) is greater than that of the heavily-doped side in the low- and mixed-level injection regimes. However, $\mathscr{E}_{\mathrm{pE}}$ and $\mathscr{E}_{\mathrm{nE}}$ exhibit identical magnitudes in the high-level injection regime. In the mixed-level injection regime, the greater electric field of the two ($\mathscr{E}_{\mathrm{pE}}$) appears to be a constant. These findings can be understood quantitatively from (EX.46): The denominator $p_{\mathrm{pE}}n_{\mathrm{nE}} - p_{\mathrm{nE}}n_{\mathrm{pE}}$ in the expressions can be approximated as $N_{\mathrm{a}}N_{\mathrm{d}}$ in the low-level injection regime, and if $N_{\mathrm{d}} \ggg N_{\mathrm{a}}$, as $n_{\mathrm{i}}N_{\mathrm{d}}e^{q\Delta/2k_{\mathrm{B}}T}$ in the mixed-level injection regime. As for the high-level

doi:10.6342/NTU202304241

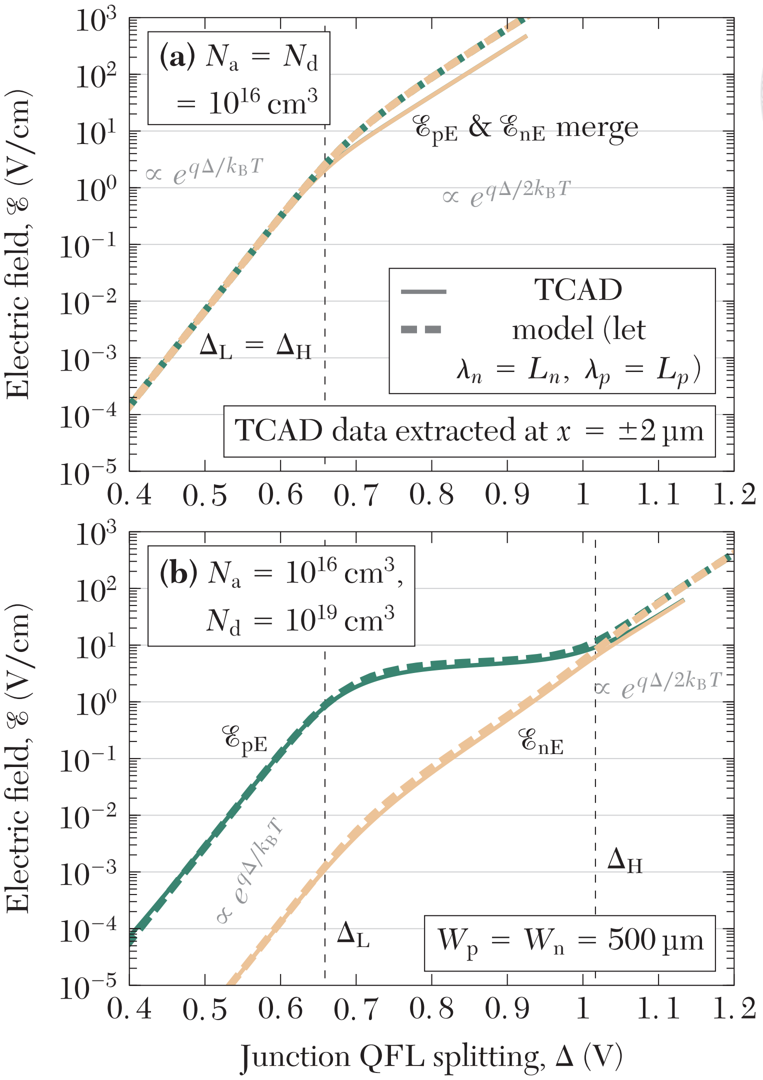


Figure EX–12. Modeled and TCAD-simulated $\mathcal{E}_{pE} - \Delta$ and $\mathcal{E}_{nE} - \Delta$ plots in a p-n diode with far contacts and with **(a)** $N_a = N_d$ and **(b)** $N_a \ll N_d$. The modeled curves utilize low-level-injection values of λ_n and λ_p (EX.45).

injection regime,

(High-level inj.)

$$\begin{aligned} p_{pE}n_{nE} - p_{nE}n_{pE} &= (n_{pE} + N_a)(p_{nE} + N_d) - p_{nE}n_{pE} \\ &= N_d n_{pE} + N_a p_{nE} + N_a N_d \\ &\approx (N_a + N_d) n_i e^{q\Delta/2k_BT}. \end{aligned} \tag{EX.47}$$

doi:10.6342/NTU202304241

This leads to the following regional approximations for $\mathscr{E}_{\mathrm{pE}}$ for $N_{\mathrm{a}} \sim N_{\mathrm{d}}$,

$$(N_{\mathrm{a}} \sim N_{\mathrm{d}}) \quad \mathscr{E}_{\mathrm{pE}} \approx \begin{cases} \dfrac{k_{\mathrm{B}}T}{q}\dfrac{n_{\mathrm{i}}^2}{N_{\mathrm{a}}}\left(\dfrac{1}{N_{\mathrm{a}}\lambda_n}+\dfrac{1}{N_{\mathrm{d}}\lambda_p}\right)e^{q\Delta/k_{\mathrm{B}}T} & \text{, low-level inj.;} \\ \dfrac{k_{\mathrm{B}}T}{q}\dfrac{2n_{\mathrm{i}}}{N_{\mathrm{a}}+N_{\mathrm{d}}}\left(\dfrac{1}{\lambda_n}+\dfrac{1}{\lambda_p}\right)e^{q\Delta/2k_{\mathrm{B}}T} & \text{, high-level inj.;} \end{cases} \tag{EX.48}$$

the following regional approximations for $\mathscr{E}_{\mathrm{nE}}$ for $N_{\mathrm{a}} \sim N_{\mathrm{d}}$,

$$(N_{\mathrm{a}} \sim N_{\mathrm{d}}) \quad \mathscr{E}_{\mathrm{nE}} \approx \begin{cases} \dfrac{k_{\mathrm{B}}T}{q}\dfrac{n_{\mathrm{i}}^2}{N_{\mathrm{d}}}\left(\dfrac{1}{N_{\mathrm{a}}\lambda_n}+\dfrac{1}{N_{\mathrm{d}}\lambda_p}\right)e^{q\Delta/k_{\mathrm{B}}T} & \text{, low-level inj.;} \\ \dfrac{k_{\mathrm{B}}T}{q}\dfrac{2n_{\mathrm{i}}}{N_{\mathrm{a}}+N_{\mathrm{d}}}\left(\dfrac{1}{\lambda_n}+\dfrac{1}{\lambda_p}\right)e^{q\Delta/2k_{\mathrm{B}}T} & \text{, high-level inj.;} \end{cases} \tag{EX.49}$$

the following regional approximations for $\mathscr{E}_{\mathrm{pE}}$ for $N_{\mathrm{d}} \gg N_{\mathrm{a}}$,

$$(N_{\mathrm{d}} \gg N_{\mathrm{a}}) \qquad \mathscr{E}_{\mathrm{pE}} \approx \begin{cases} \dfrac{k_{\mathrm{B}}T}{q}\left(\dfrac{n_{\mathrm{i}}}{N_{\mathrm{a}}}\right)^2\dfrac{1}{\lambda_n}e^{q\Delta/k_{\mathrm{B}}T} & \text{, low-level inj.;} \\ \dfrac{k_{\mathrm{B}}T}{q}\dfrac{1}{\lambda_n} & \text{, mixed-level inj.;} \\ \dfrac{k_{\mathrm{B}}T}{q}\dfrac{2n_{\mathrm{i}}}{N_{\mathrm{d}}}\left(\dfrac{1}{\lambda_n}+\dfrac{1}{\lambda_p}\right)e^{q\Delta/2k_{\mathrm{B}}T} & \text{, high-level inj.;} \end{cases} \tag{EX.50}$$

and the following regional approximations for $\mathscr{E}_{\mathrm{pE}}$ for $N_{\mathrm{d}} \gg N_{\mathrm{a}}$,

$$(N_{\mathrm{d}} \gg N_{\mathrm{a}}) \qquad \mathscr{E}_{\mathrm{nE}} \approx \begin{cases} \dfrac{k_{\mathrm{B}}T}{q}\dfrac{n_{\mathrm{i}}^2}{N_{\mathrm{a}}N_{\mathrm{d}}}\dfrac{1}{\lambda_n}e^{q\Delta/k_{\mathrm{B}}T} & \text{, low-level inj.;} \\ \dfrac{k_{\mathrm{B}}T}{q}\dfrac{2n_{\mathrm{i}}}{N_{\mathrm{d}}}\dfrac{1}{\lambda_n}e^{q\Delta/2k_{\mathrm{B}}T} & \text{, mixed-level inj.;} \\ \dfrac{k_{\mathrm{B}}T}{q}\dfrac{2n_{\mathrm{i}}}{N_{\mathrm{d}}}\left(\dfrac{1}{\lambda_n}+\dfrac{1}{\lambda_p}\right)e^{q\Delta/2k_{\mathrm{B}}T} & \text{, high-level inj.} \end{cases} \tag{EX.51}$$

Note in **Figure EX–12** the discrepancies of a certain factor between the modeled and TCAD-simulated magnitudes of electric fields in the high-level injection regimes. The model that employs low-level-injection values of λ_n and λ_p tends to overestimate the fields. This suggests greater values of λ_n and λ_p beyond what are given by (EX.45) in the high-level injection regimes. The evaluation of correct high-level-injection decay lengths will be tackled in **Section EX.4**.

EX.3.2 Revisiting the Negligibility of Space Charge Current Effect

We have claimed the negligibility of the space charge current, J_{SC}, in (EX.42) upon deriving $\mathscr{E}_{\mathrm{pE}}$ and $\mathscr{E}_{\mathrm{nE}}$. In fact, we could still proceed a derivation inclusive of J_{SC} to obtain correction terms for (EX.46):

$$\mathscr{E}_{\mathrm{pE}}\,(\text{with } J_{\mathrm{SC}}) = \mathscr{E}_{\mathrm{pE}}\,(\text{no } J_{\mathrm{SC}}) + \frac{p_{\mathrm{nE}}/q\mu_n + n_{\mathrm{nE}}/q\mu_{\mathrm{p}}}{p_{\mathrm{pE}}n_{\mathrm{nE}} - p_{\mathrm{nE}}n_{\mathrm{pE}}} \cdot J_{\mathrm{SC}}; \tag{EX.52a}$$

$$\mathscr{E}_{\mathrm{nE}}\,(\text{with } J_{\mathrm{SC}}) = \mathscr{E}_{\mathrm{nE}}\,(\text{no } J_{\mathrm{SC}}) + \frac{p_{\mathrm{pE}}/q\mu_n + n_{\mathrm{pE}}/q\mu_{\mathrm{p}}}{p_{\mathrm{pE}}n_{\mathrm{nE}} - p_{\mathrm{nE}}n_{\mathrm{pE}}} \cdot J_{\mathrm{SC}}. \tag{EX.52b}$$

J_{SC} arises from excess carrier recombination, with its magnitude (if subject to SRH process) in the order of [Ex8]

$$(q\Delta/k_{\mathrm{B}}T \gg 1) \qquad \begin{aligned} J_{\mathrm{SC}} &\sim q\sqrt{\sigma_n\sigma_p}\,v_{\mathrm{th}}n_{\mathrm{i}}e^{q\Delta/2k_{\mathrm{B}}T}N_{\mathrm{t}}W_{\mathrm{d}} \\ &= \frac{qn_{\mathrm{i}}W_{\mathrm{d}}}{\sqrt{\tau_n\tau_p}}e^{q\Delta/2k_{\mathrm{B}}T}, \end{aligned} \tag{EX.53}$$

where σ_n and σ_p are the electron and hole capture cross-sections, respectively; v_{th} is the thermal velocity, N_{t} is the concentration of trap centers, and $W_{\mathrm{d}} = x_{\mathrm{p}} + x_{\mathrm{n}}$ is the depletion layer width. The proportion of this correction term in, for example $\mathscr{E}_{\mathrm{pE}}$, can be estimated as

$$\text{(low-level inj.)}\qquad \begin{aligned}\frac{\mathscr{E}_{\mathrm{pE}}\,(J_{\mathrm{SC}}\text{ term})}{\mathscr{E}_{\mathrm{pE}}\,(\text{no }J_{\mathrm{SC}})} &= \frac{(p_{\mathrm{nE}}/q\mu_n + n_{\mathrm{nE}}/q\mu_p)J_{\mathrm{SC}}}{(k_{\mathrm{B}}T/q)(p_{\mathrm{nE}} + n_{\mathrm{nE}})(n_{\mathrm{pE}}/\lambda_n + p_{\mathrm{nE}}/\lambda_p)} \\ &\approx \frac{(n_{\mathrm{nE}}/q\mu_p)J_{\mathrm{SC}}}{(k_{\mathrm{B}}T/q)n_{\mathrm{nE}}(n_{\mathrm{i}}^2/\lambda_n N_{\mathrm{a}} + n_{\mathrm{i}}^2/\lambda_p N_{\mathrm{d}})e^{q\Delta/k_{\mathrm{B}}T}} \\ &\sim \frac{J_{\mathrm{SC}}/q\mu}{(k_{\mathrm{B}}T/q)(n_{\mathrm{i}}^2/\lambda N)e^{q\Delta/k_{\mathrm{B}}T}} \\ &\sim \left(\frac{n_{\mathrm{i}}}{N}\right)\left(\frac{W_{\mathrm{d}}}{L}\right)\left(\frac{\lambda}{L}\right)e^{-q\Delta/2k_{\mathrm{B}}T}. \end{aligned} \tag{EX.54}$$

Where N, L, and λ represent the order of the doping concentration, the minority carrier diffusion length, and the minority carrier decay length, respectively. Even under low-level injection, the first and the second terms are far less than 1, the third term is strictly less than 1 (EX.45), and the fourth term does not exceed unity, indicating that the effect of J_{SC} is truly negligible in our scope of interest.

EX.3.3 Analytical Expressions for the Minority Carrier Diffusion and Drift Currents

We can acquire analytical expressions for the minority carrier drift currents from $\mathscr{E}_{\mathrm{pE}}$ and $\mathscr{E}_{\mathrm{nE}}$ (EX.46) as

$$\begin{aligned}(J_{n,\,\mathrm{drift}})_{\mathrm{pE}} &= q\mu_n n_{\mathrm{pE}}\mathscr{E}_{\mathrm{pE}} \\ &= qD_n\frac{(p_{\mathrm{nE}} + n_{\mathrm{nE}})n_{\mathrm{pE}}}{p_{\mathrm{pE}}n_{\mathrm{nE}} - p_{\mathrm{nE}}n_{\mathrm{pE}}}\left(\frac{n_{\mathrm{pE}}}{\lambda_n} + \frac{p_{\mathrm{nE}}}{\lambda_p}\right);\end{aligned} \tag{EX.55a}$$

$$\begin{aligned}(J_{p,\,\mathrm{drift}})_{\mathrm{nE}} &= q\mu_p p_{\mathrm{nE}}\mathscr{E}_{\mathrm{nE}} \\ &= qD_p\frac{(p_{\mathrm{pE}} + n_{\mathrm{pE}})p_{\mathrm{nE}}}{p_{\mathrm{pE}}n_{\mathrm{nE}} - p_{\mathrm{nE}}n_{\mathrm{pE}}}\left(\frac{n_{\mathrm{pE}}}{\lambda_n} + \frac{p_{\mathrm{nE}}}{\lambda_p}\right);\end{aligned} \tag{EX.55b}$$

whereas the minority carrier diffusion currents at the depletion edges are well-known as

$$\begin{aligned}(J_{n,\,\mathrm{diff}})_{\mathrm{pE}} &= qD_n\left(\frac{\mathrm{d}n}{\mathrm{d}x}\right)_{\mathrm{pE}} \\ &= qD_n\frac{n_{\mathrm{pE}}}{\lambda_n}; \qquad \text{(EX.56a)} \\ (J_{p,\,\mathrm{diff}})_{\mathrm{nE}} &= -qD_p\left(\frac{\mathrm{d}p}{\mathrm{d}x}\right)_{\mathrm{nE}} \\ &= qD_p\frac{p_{\mathrm{nE}}}{\lambda_p}. \qquad \text{(EX.56b)}\end{aligned}$$

Figure EX–13 plots the modeled and TCAD-simulated minority carrier current components, at respective depletion edges, in a long p-n diode with $N_{\mathrm{d}} \gg N_{\mathrm{a}}$. The diffusion currents, proportional to n_{pE} and p_{nE} (EX.56), exhibit proportionality to $e^{q\Delta/k_{\mathrm{B}}T}$ and $e^{q\Delta/2k_{\mathrm{B}}T}$ in the low- and high-level injection regimes, respectively. On the other hand, the drift currents, with considerably smaller magnitudes compared to the diffusion currents, exhibit proportionality to $e^{2q\Delta/kT}$ in the low-level injection regime and eventually catch up the diffusion counterparts. The factor of $e^{q\Delta/2k_{\mathrm{B}}T}$ arises from the minority carrier concentrations ((EX.5), (EX.6)) that contribute a factor of $e^{q\Delta/k_{\mathrm{B}}T}$, and also from $\mathscr{E}_{\mathrm{pE}}$ and $\mathscr{E}_{\mathrm{nE}}$ (**Figure EX–12**) which contribute another factor of $e^{q\Delta/k_{\mathrm{B}}T}$. In the high-level injection regime, the drift currents dominates over and grow faster than the diffusion currents, with a factor of $e^{q\Delta/k_{\mathrm{B}}T}$, in which a factor of $e^{q\Delta/2k_{\mathrm{B}}T}$ comes from the carrier concentrations, and another identical factor comes also from the fields. In the mixed-level injection regime, the minority electron drift current exhibits a proportionality to $e^{q\Delta/2k_{\mathrm{B}}T}$, which is identical to the diffusion counterpart. This can be attributed to constant $\mathscr{E}_{\mathrm{pE}}$ ($\mathscr{E}_{\mathrm{pE}} = k_{\mathrm{B}}T/q\lambda_n$, (EX.50)), with the proportionality solely

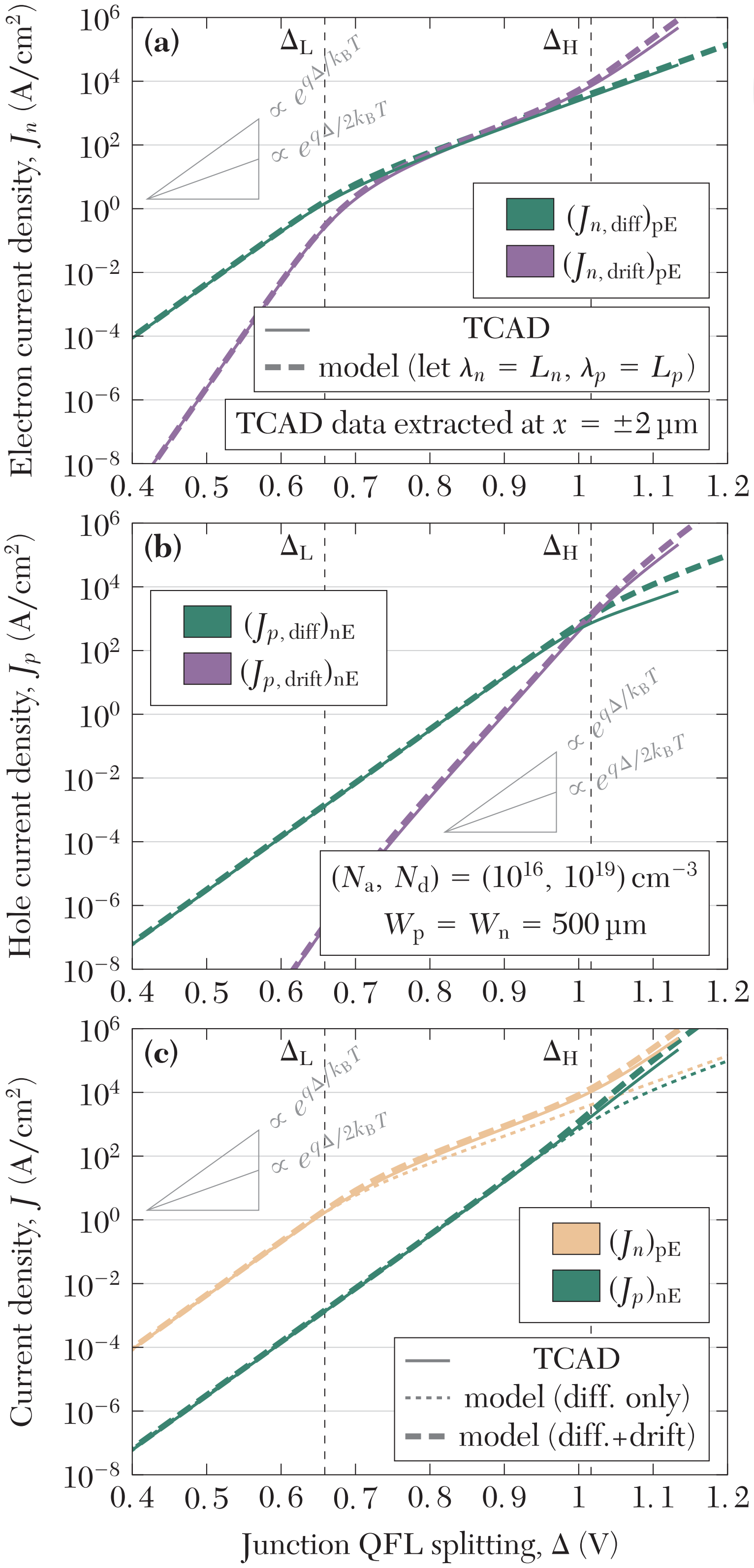


Figure EX–13. Modeled and TCAD-simulated minority **(a)** electron and **(b)** hole drift and diffusion current components, as functions of Δ, in a long diode with $N_\mathrm{d} \gg N_\mathrm{a}$. **(c)** Comparison of total electron and hole currents. Also shown are the diffusion-only modeled curves. Low-level-injection decay lengths (EX.45) are employed in the modeled curves.

governed by n_{pE} while the p-region is subject to high-level injection in this operation regime. In fact, $(J_{n,\,\mathrm{drift}})_{\mathrm{pE}}$ and $(J_{n,\,\mathrm{diff}})_{\mathrm{pE}}$ have equal magnitudes in the mixed-level injection regime, according to (EX.50) and (EX.55):

$$\begin{aligned}(J_{n,\,\mathrm{drift}})_{\mathrm{pE}} &\approx q\mu_n n_{\mathrm{pE}}\left(\frac{k_{\mathrm{B}}T}{q}\frac{1}{\lambda_n}\right)\\ &= qD_n\frac{n_{\mathrm{pE}}}{\lambda_n}\\ &= (J_{n,\,\mathrm{diff}})_{\mathrm{pE}}.\end{aligned} \tag{EX.57}$$

This is known as the Webster effect [Ex7] in the emitter-base junction of bipolar junction transistors.

Adding up the drift and diffusion current components in **Figure EX–13 (c)**, the total hole current exhibits proportionality to $e^{q\Delta/k_{\mathrm{B}}T}$ through all operation regimes, whereas the total electron current is $\propto e^{q\Delta/k_{\mathrm{B}}T}$ in the low-level injection regime, $\propto e^{q\Delta/2k_{\mathrm{B}}T}$ in the mixed-level injection regime, and reverts back to $\propto e^{q\Delta/k_{\mathrm{B}}T}$ in the high-level injection regime. This lends support to our previous statement in **Section EX.1** that the commonly-perceived "high-level injection regime" with $J_{\mathrm{tot}} \propto e^{q\Delta/2k_{\mathrm{B}}T}$ actually corresponds to the mixed-level injection regime in our work, beyond which there is a true high-level injection regime with $J_{\mathrm{tot}} \propto e^{q\Delta/k_{\mathrm{B}}T}$. **TABLE EX–II** summarizes the proportionality of each current component in a p-n diode with $N_{\mathrm{d}} \gg N_{\mathrm{a}}$.

Also by noting the total hole current to be negligible in the low- and mixed-level injection regimes but comparable in the high-level injection regime for the $N_{\mathrm{d}} \gg N_{\mathrm{a}}$ case, compared to the total electron current, we may compose the approximate regional expressions for the total diode current, $J_{\mathrm{tot}} = (J_n)_{\mathrm{pE}} + (J_p)_{\mathrm{nE}}$, as

$$(N_{\mathrm{d}} \gg N_{\mathrm{a}}) \qquad J_{\mathrm{tot}} \approx \begin{cases} qD_n \dfrac{n_{\mathrm{i}}^2}{\lambda_n N_{\mathrm{a}}} e^{q\Delta/k_{\mathrm{B}}T} & \text{, low-level inj.;} \\ 2qD_n \dfrac{n_{\mathrm{i}}}{\lambda_n} e^{q\Delta/2k_{\mathrm{B}}T} & \text{, mixed-level inj.;} \\ q(D_n + D_p) \dfrac{2n_{\mathrm{i}}^2}{N_{\mathrm{d}}} \left(\dfrac{1}{\lambda_n} + \dfrac{1}{\lambda_p} \right) e^{q\Delta/k_{\mathrm{B}}T} & \text{, high-level inj.} \end{cases} \tag{EX.58}$$

This is the combined result of (EX.5), (EX.6), (EX.50), (EX.51), (EX.55), (EX.56), and (EX.57).

Figure EX–14 plots the modeled and TCAD-simulated minority carrier current components, at respective depletion edges, in a long p-n diode with $N_{\mathrm{a}} = N_{\mathrm{d}}$. There is no mixed-level injection regime in this case. In **Figure EX–14 (a, b)** showing the drift and diffusion current components, for the same reasons as the $N_{\mathrm{d}} \gg N_{\mathrm{a}}$ case, minority carrier diffusion and drift currents exhibit proportionality to $e^{q\Delta/k_{\mathrm{B}}T}$ and $e^{2q\Delta/k_{\mathrm{B}}T}$ in the low-level injection, respectively, and to $e^{q\Delta/2k_{\mathrm{B}}T}$ and $e^{q\Delta/k_{\mathrm{B}}T}$ in the high-level injec-

TABLE EX–II
Proportionality of each current component in a p-n diode with highly mismatched doping concentrations ($N_{\mathrm{d}} \gg N_{\mathrm{a}}$).

Regime		low-level	mixed-level	high-level
Proportionality	$(J_{n,\,\mathrm{diff}})_{\mathrm{pE}}$	$e^{q\Delta/k_{\mathrm{B}}T}$	$e^{q\Delta/2k_{\mathrm{B}}T}$	$e^{q\Delta/2k_{\mathrm{B}}T}$
	$(J_{n,\,\mathrm{drift}})_{\mathrm{pE}}$	$e^{2q\Delta/k_{\mathrm{B}}T}$	$e^{q\Delta/2k_{\mathrm{B}}T}$	$e^{q\Delta/k_{\mathrm{B}}T}$
	$(J_n)_{\mathrm{pE}}$	$e^{q\Delta/k_{\mathrm{B}}T}$	$e^{q\Delta/2k_{\mathrm{B}}T}$	$e^{q\Delta/k_{\mathrm{B}}T}$
	$(J_{p,\,\mathrm{diff}})_{\mathrm{nE}}$	$e^{q\Delta/k_{\mathrm{B}}T}$	$e^{q\Delta/k_{\mathrm{B}}T}$	$e^{q\Delta/2k_{\mathrm{B}}T}$
	$(J_{p,\,\mathrm{drift}})_{\mathrm{nE}}$	$e^{2q\Delta/k_{\mathrm{B}}T}$	$e^{2q\Delta/k_{\mathrm{B}}T}$	$e^{q\Delta/k_{\mathrm{B}}T}$
	$(J_p)_{\mathrm{nE}}$	$e^{q\Delta/k_{\mathrm{B}}T}$	$e^{q\Delta/k_{\mathrm{B}}T}$	$e^{q\Delta/k_{\mathrm{B}}T}$

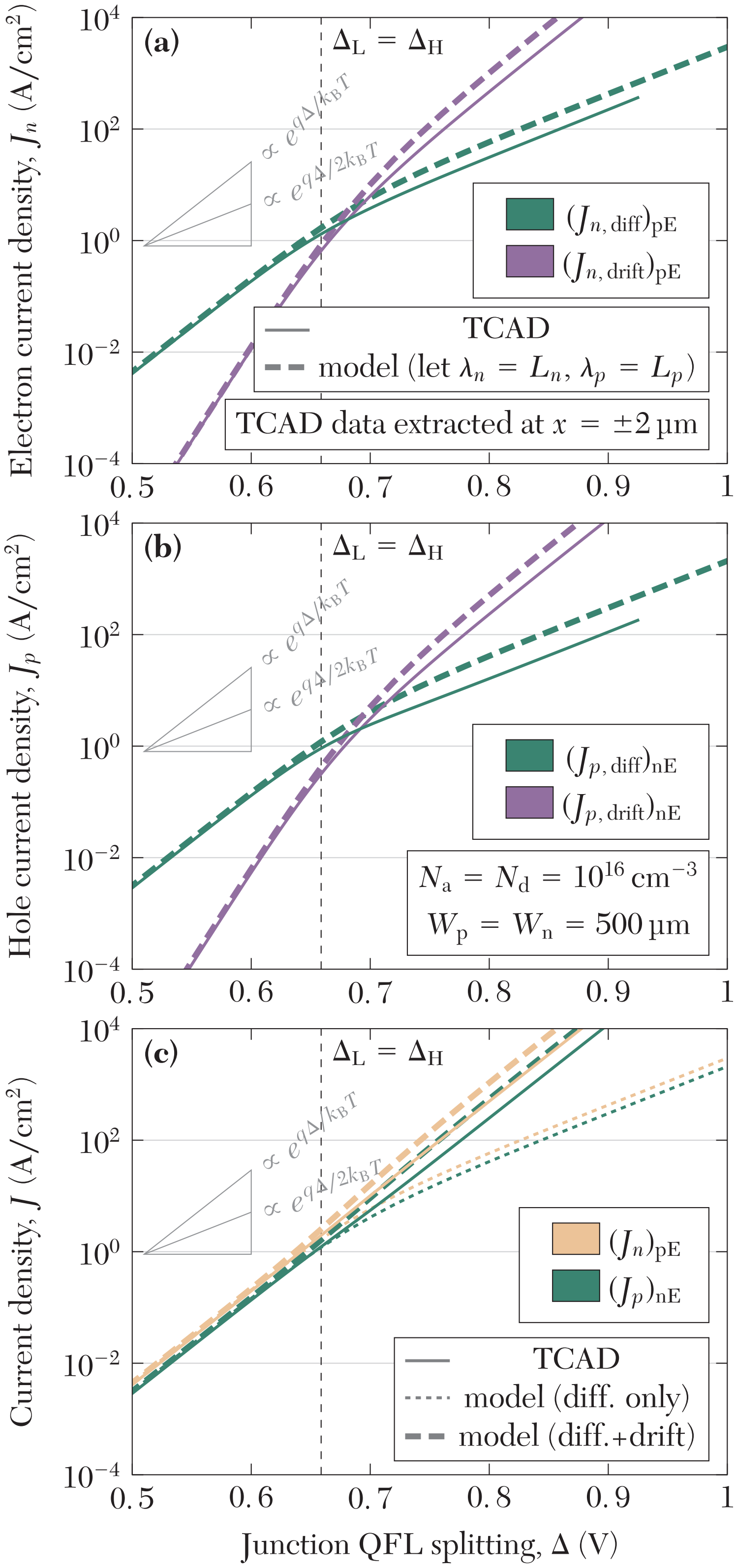


Figure EX–14. Modeled and TCAD-simulated minority **(a)** electron and **(b)** hole drift and diffusion current components, as functions of Δ, in a long diode with $N_a = N_d$. **(c)** Comparison of total electron and hole currents. Also shown are the diffusion-only modeled curves. Low-level-injection decay lengths (EX.45) are employed in the modeled curves.

doi:10.6342/NTU202304241

tion regime, respectively. Since the diffusion currents are prominent in the low-level injection regime, whereas the drift currents are prominent in the high-level injection regime, the total current has a $\propto e^{q\Delta/k_\mathrm{B}T}$ dependency throughout all operation regimes, with the complete absence of the $\propto e^{q\Delta/2k_\mathrm{B}T}$ dependency as commonly perceived in diodes under higher injection levels. **Figure EX–14 (c)** plots the total electron and hole currents as functions of Δ; none of which is negligible in all operation regimes due to the comparable magnitudes. Combining (EX.5), (EX.6), (EX.48), (EX.49), (EX.55), and (EX.56), we obtain the approximate regional expressions for J_tot in this case (or cases with only slightly mismatched doping concentrations) as

$$(N_\mathrm{a} \sim N_\mathrm{d}) \qquad J_\mathrm{tot} \approx \begin{cases} qD_n n_\mathrm{i}^2 \left(\dfrac{1}{\lambda_n N_\mathrm{a}} + \dfrac{1}{\lambda_p N_\mathrm{d}} \right) e^{q\Delta/k_\mathrm{B}T} & \text{, low-level inj.;} \\ q(D_n + D_p) \dfrac{2n_\mathrm{i}^2}{N_\mathrm{a} + N_\mathrm{d}} \left(\dfrac{1}{\lambda_n} + \dfrac{1}{\lambda_p} \right) e^{q\Delta/k_\mathrm{B}T} & \text{, high-level inj.} \end{cases} \tag{EX.59}$$

The modeled curves from **Figure EX–13** and **Figure EX–14** suffer from having overestimated the TCAD simulation results in the high-level injection regimes. This is attributed to incorrect values of λ_n and λ_p under high-level injection, as mentioned above in **Figure EX–12**. By employing in the modeled expressions proper values of the decay lengths (**Section EX.4**) which are shown to be longer than the low-level injection values (EX.45), such discrepancies between modeled and TCAD curves are eliminated.

EX.4 Carrier Recombination and Diode Current Evaluation in the High-Level Injection Regime

EX.4.1 Net Recombination Rate Under High-Level Injection

The concepts of decay lengths (EX.45) known to Shockley's diode equation [Ex1, Ex8] arises from the decaying profiles of excess carriers from the depletion edges towards the contacts. The decaying profiles of excess carriers can be attributed to net recombination of carriers in the quasi-neutral regions, which we assumed is governed by the SRH process and is subject to the net recombination rate [Ex8]:

$$U = \frac{np - n_{\mathrm{i}}^2}{\tau_p(n + n_{\mathrm{t}}) + \tau_n(p + n_{\mathrm{t}})}, \tag{EX.60}$$

where n_{t} is the carrier concentration associated with the trap level. Here we always assume midgap traps and therefore $n_{\mathrm{t}} = n_{\mathrm{i}}$. The well-known limits of U under low-level injection are $U \approx \delta n/\tau_n$ in the quasi-neutral p-region (δn: excess carrier concentration) where $p \gg n_{\mathrm{i}} \gg n$, and $U \approx \delta n/\tau_p$ in the quasi-neutral n-region where $n \gg n_{\mathrm{i}} \gg p$.

We may inspect the formulation of (EX.60) in depth to obtain approximations for U under higher injection levels. By temporarily letting $\alpha \equiv (\tau_p n + \tau_n p)/(\tau_p + \tau_n)$, the expression can be factored as

$$\begin{aligned} U &= \frac{\left[\alpha + \frac{\tau_n}{\tau_p+\tau_n}(n - p)\right]\left[\alpha + \frac{\tau_p}{\tau_p+\tau_n}(p - n)\right] - n_{\mathrm{i}}^2}{(\tau_p + \tau_n)(\alpha + n_{\mathrm{i}})} \\ &= \frac{\tau_p(p - n_{\mathrm{i}}) + \tau_n(n - n_{\mathrm{i}})}{(\tau_p + \tau_n)^2} + \frac{(\tau_n - \tau_p)(p - n) + \frac{\tau_p\tau_n}{\tau_p+\tau_n}(p - n)^2}{(\tau_p + \tau_n)[\tau_p(n + n_{\mathrm{i}}) + \tau_n(p + n_{\mathrm{i}})]}. \end{aligned} \tag{EX.61}$$

Under high-level injection, $n,\ p \gg |p - n| = N_{\mathrm{a}}$ or N_{d}, so the second term in (EX.61)

doi:10.6342/NTU202304241

becomes negligible, and the first term remains as

(High-level inj.) $$U \approx \frac{\delta n}{\tau_p + \tau_n}. \tag{EX.62}$$

Its proportionality to δn is about a half of those of the low-level injection limits. This result was also addressed in [Ex5].

EX.4.2 Electric field as a function of *x* in the quasi-neutral regions

To achieve our goal of obtaining the high-level injection decay lengths, we may first solve for the minority carrier profiles in the quasi-neutral regions and compute the decay lengths from their spatial derivatives at respective depletion edges. As will be demonstrated later, we will solve the minority carrier profiles from the carrier continuity equations that involve the profiles of electric fields, $\mathscr{E}_\mathrm{p}(x)$ in the quasi-neutral p-region and $\mathscr{E}_\mathrm{n}(x)$ in the quasi-neutral n-region. We may therefore start from deriving the expressions for $\mathscr{E}_\mathrm{p}(x)$ and $\mathscr{E}_\mathrm{n}(x)$.

For the conciseness of symbols, we define the following constants:

$$N_\mathrm{d}^* \equiv \frac{\mu_n N_\mathrm{d}}{\mu_n + \mu_p}; \tag{EX.63a}$$

$$N_\mathrm{a}^* \equiv \frac{\mu_p N_\mathrm{a}}{\mu_n + \mu_p}; \tag{EX.63b}$$

and also a skew term from the constant mobilities as

$$\begin{aligned}\sigma &\equiv \frac{\mu_n - \mu_p}{\mu_n + \mu_p} \\ &= \frac{D_n - D_p}{D_n + D_p}. \end{aligned} \tag{EX.64}$$

doi:10.6342/NTU202304241

The current continuity equation, valid at any location of the diode, and expanded into drift and diffusion current components, reads

$$J_n(x) + J_p(x) = q\mu_n n(x)\mathscr{E}(x) + q\mu_p p(x)\mathscr{E}(x) + qD_n\frac{\mathrm{d}n}{\mathrm{d}x} - qD_p\frac{\mathrm{d}p}{\mathrm{d}x}$$
$$= \text{constant}. \tag{EX.65}$$

In the quasi-neutral p-region, by the quasi-neutrality assumption $p_\mathrm{p}(x) = n_\mathrm{p}(x) + N_\mathrm{a}$ and $\mathrm{d}p_\mathrm{p}/\mathrm{d}x = \mathrm{d}n_\mathrm{p}/\mathrm{d}x$, this becomes

$$\begin{aligned} & q(\mu_p n_\mathrm{p}(x) + \mu_p N_\mathrm{a} + \mu_n n_\mathrm{p}(x))\mathscr{E}_\mathrm{p}(x) + q(D_n - D_p)\frac{\mathrm{d}n_\mathrm{p}}{\mathrm{d}x} \\ & = q(\mu_n + \mu_p)\left[(n_\mathrm{p}(x) + N_\mathrm{a}^\star)\mathscr{E}_\mathrm{p}(x) + \frac{k_\mathrm{B}T}{q}\sigma\frac{\mathrm{d}n_\mathrm{p}}{\mathrm{d}x}\right] \\ & = \text{constant}. \end{aligned} \tag{EX.66}$$

The constant can be the expression itself evaluated at any reference location. If we choose $x = -x_\mathrm{p}$ as the reference location and drop the leading constant $q(\mu_n + \mu_p)$, it is evaluated as

$$[n_\mathrm{p}(x) + N_\mathrm{a}^\star]\mathscr{E}_\mathrm{p}(x) + \frac{k_\mathrm{B}T}{q}\sigma\frac{\mathrm{d}n_\mathrm{p}}{\mathrm{d}x} = [n_\mathrm{pE} + N_\mathrm{a}^\star]\mathscr{E}_\mathrm{pE} + \frac{k_\mathrm{B}T}{q}\sigma\frac{n_\mathrm{pE}}{\lambda_n}$$

$$\Rightarrow \boxed{\mathscr{E}_\mathrm{p}(x < 0) = \frac{[n_\mathrm{pE} + N_\mathrm{a}^\star]\mathscr{E}_\mathrm{pE} + \frac{k_\mathrm{B}T}{q}\sigma\left[\frac{n_\mathrm{pE}}{\lambda_n} - \frac{\mathrm{d}n_\mathrm{p}}{\mathrm{d}x}\right]}{n_\mathrm{p}(x) + N_\mathrm{a}^\star}}; \tag{EX.67}$$

yielding the expression for $\mathscr{E}_\mathrm{p}(x)$ in terms of n_pE, λ_n, the known value of $\mathscr{E}_\mathrm{pE}$ (EX.46), and the unsolved $n_{\mathrm{p}}(x)$ profile. Likewise, in the quasi-neutral n-region, we obtain

$$\boxed{\mathscr{E}_{\mathrm{n}}(x>0)=\frac{[p_{\mathrm{nE}}+N_{\mathrm{d}}^{*}]\mathscr{E}_{\mathrm{nE}}+\frac{k_{\mathrm{B}}T}{q}\sigma\left[-\frac{p_{\mathrm{nE}}}{\lambda_p}-\frac{\mathrm{d}p_{\mathrm{n}}}{\mathrm{d}x}\right]}{p_{\mathrm{n}}(x)+N_{\mathrm{d}}^{*}}}. \tag{EX.68}$$

EX.4.3 Derivation of the Differential Equations for Minority Carrier Concentration Profiles

In the quasi-neutral p-region, the minority carrier continuity equation at steady state ($\partial n/\partial t=0$) reads [Ex8]

$$\begin{aligned}0&=\frac{1}{q}\frac{\mathrm{d}J_n(x)}{\mathrm{d}x}-U(x)\\&=\mu_n\frac{\mathrm{d}}{\mathrm{d}x}[n_{\mathrm{p}}(x)\mathscr{E}_{\mathrm{p}}(x)]+D_n\frac{\mathrm{d}^2n_{\mathrm{p}}}{\mathrm{d}x^2}-U(x),\end{aligned} \tag{EX.69}$$

where, from (EX.67),

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}x}[n_{\mathrm{p}}(x)\mathscr{E}_{\mathrm{p}}(x)]\\&=\frac{\left\{[n_{\mathrm{pE}}+N_{\mathrm{a}}^{*}]\mathscr{E}_{\mathrm{pE}}+\frac{k_{\mathrm{B}}T}{q}\sigma\left[\frac{n_{\mathrm{pE}}}{\lambda_n}-\frac{\mathrm{d}n_{\mathrm{p}}}{\mathrm{d}x}\right]\right\}N_{\mathrm{a}}^{*}\frac{\mathrm{d}n_{\mathrm{p}}}{\mathrm{d}x}-\frac{k_{\mathrm{B}}T}{q}\sigma[n_{\mathrm{p}}(x)+N_{\mathrm{a}}^{*}]n_{\mathrm{p}}(x)\frac{\mathrm{d}^2n_{\mathrm{p}}}{\mathrm{d}x^2}}{[n_{\mathrm{p}}(x)+N_{\mathrm{a}}^{*}]^2}.\end{aligned} \tag{EX.70}$$

Therefore, (EX.69) becomes

$$\begin{aligned}0=&\frac{\left\{\mu_n[n_{\mathrm{pE}}+N_{\mathrm{a}}^{*}]\mathscr{E}_{\mathrm{pE}}+D_n\sigma\left[\frac{n_{\mathrm{pE}}}{\lambda_n}-\frac{\mathrm{d}n_{\mathrm{p}}}{\mathrm{d}x}\right]\right\}N_{\mathrm{a}}^{*}\frac{\mathrm{d}n_{\mathrm{p}}}{\mathrm{d}x}-D_n\sigma[n_{\mathrm{p}}(x)+N_{\mathrm{a}}^{*}]n_{\mathrm{p}}(x)\frac{\mathrm{d}^2n_{\mathrm{p}}}{\mathrm{d}x^2}}{[n_{\mathrm{p}}(x)+N_{\mathrm{a}}^{*}]^2}\\&+D_n\frac{\mathrm{d}^2n_{\mathrm{p}}}{\mathrm{d}x^2}-U(x).\end{aligned} \tag{EX.71}$$

Similarly, in the quasi-neutral n-region,

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}x}[p_{\mathrm{n}}(x)\mathscr{E}_{\mathrm{n}}(x)] \\
&= \frac{\left\{[p_{\mathrm{nE}} + N_{\mathrm{d}}^{*}]\mathscr{E}_{\mathrm{nE}} + \frac{k_{\mathrm{B}}T}{q}\sigma\left[-\frac{p_{\mathrm{nE}}}{\lambda_{\mathrm{p}}} - \frac{\mathrm{d}p_{\mathrm{n}}}{\mathrm{d}x}\right]\right\} N_{\mathrm{d}}^{*}\frac{\mathrm{d}p_{\mathrm{n}}}{\mathrm{d}x} - \frac{k_{\mathrm{B}}T}{q}\sigma[p_{\mathrm{n}}(x) + N_{\mathrm{d}}^{*}]p_{\mathrm{n}}(x)\frac{\mathrm{d}^2 p_{\mathrm{n}}}{\mathrm{d}x^2}}{[p_{\mathrm{n}}(x) + N_{\mathrm{d}}^{*}]^2}.
\end{aligned} \tag{EX.72}$$

Hence, the minority carrier continuity equation at steady state:

$$\begin{aligned}
0 =& -\frac{1}{q}\frac{\mathrm{d}J_p(x)}{\mathrm{d}x} - U(x) \\
=& -\mu_p\frac{\mathrm{d}}{\mathrm{d}x}[p_{\mathrm{n}}(x)\mathscr{E}_{\mathrm{n}}(x)] + D_p\frac{\mathrm{d}^2 p_{\mathrm{n}}}{\mathrm{d}x^2} - U(x) \\
=& -\frac{\left\{\mu_p[p_{\mathrm{nE}} + N_{\mathrm{d}}^{*}]\mathscr{E}_{\mathrm{nE}} + D_p\sigma\left[-\frac{p_{\mathrm{nE}}}{\lambda_p} - \frac{\mathrm{d}p_{\mathrm{n}}}{\mathrm{d}x}\right]\right\} N_{\mathrm{d}}^{*}\frac{\mathrm{d}p_{\mathrm{n}}}{\mathrm{d}x} - D_p\sigma[p_{\mathrm{n}}(x) + N_{\mathrm{d}}^{*}]p_{\mathrm{n}}(x)\frac{\mathrm{d}^2 p_{\mathrm{n}}}{\mathrm{d}x^2}}{[p_{\mathrm{n}}(x) + N_{\mathrm{d}}^{*}]^2} \\
& + D_p\frac{\mathrm{d}^2 p_{\mathrm{n}}}{\mathrm{d}x^2} - U(x).
\end{aligned} \tag{EX.73}$$

The equations (EX.71) and (EX.73) are second-order differential equations for $n_{\mathrm{p}}(x)$ and $p_{\mathrm{n}}(x)$, respectively, valid for any injection level, from which we can solve the $n_{\mathrm{p}}(x)$ and $p_{\mathrm{n}}(x)$ profiles by also incorporating the expression for $U(x)$ (EX.60).

EX.4.4 Differential Equations for Decay Lengths in High-Level Injection Regime

The messy differential equations (EX.71) and (EX.73) are barely useful due to their complexity and non-homogeneous nature. However, we may consider extremely high-level injection cases and assume that $n(x)$, $p(x) \gg N_{\mathrm{d}}$, N_{a} everywhere in the quasi-neutral regions. This assumption may not be true near the contacts as δn shall finally approach 0 from the neutral contact assumptions. Regardless of this fact, however, this problematic range in the vicinity of the contacts may only account for a tiny portion of the quasi-neutral p- and n-region widths (see the high QFL splitting throughout the quasi-neutral regions in **Figure EX–2**).

In the quasi-neutral p-region, as now $n_\mathrm{p}(x) \gg N_\mathrm{a} > N_\mathrm{a}^*$, we may replace every $[n_\mathrm{p}(x) + N_\mathrm{a}^*]$ in (EX.71) with $n_\mathrm{p}(x)$ and get

$$0 = \mu_n n_\mathrm{pE} \mathscr{E}_\mathrm{pE} N_\mathrm{a}^* \frac{\frac{\mathrm{d}n_\mathrm{p}}{\mathrm{d}x}}{[n_\mathrm{p}(x)]^2} + D_n \sigma N_\mathrm{a}^* \frac{\frac{n_\mathrm{pE}}{\lambda_n} - \frac{\mathrm{d}n_\mathrm{p}}{\mathrm{d}x}}{[n_\mathrm{p}(x)]^2} \frac{\mathrm{d}n_\mathrm{p}}{\mathrm{d}x} + D_n(1-\sigma)\frac{\mathrm{d}^2 n_\mathrm{p}}{\mathrm{d}x^2} - U(x). \quad \text{(EX.74)}$$

Furthermore, from (EX.47) and (EX.62), it becomes

$$\begin{aligned} 0 =& D_n n_\mathrm{pE} \frac{2N_\mathrm{a}^*}{N_\mathrm{a} + N_\mathrm{d}} \left[\frac{n_\mathrm{pE}}{\lambda_n} + \frac{p_\mathrm{nE}}{\lambda_p}\right] \frac{\frac{\mathrm{d}n_\mathrm{p}}{\mathrm{d}x}}{[n_\mathrm{p}(x)]^2} + D_n \sigma N_\mathrm{a}^* \left[\frac{n_\mathrm{pE}}{\lambda_n} - \frac{\mathrm{d}n_\mathrm{p}}{\mathrm{d}x}\right] \frac{\frac{\mathrm{d}n_\mathrm{p}}{\mathrm{d}x}}{[n_\mathrm{p}(x)]^2} \\ &+ D_n(1-\sigma)\frac{\mathrm{d}^2 n_\mathrm{p}}{\mathrm{d}x^2} - \frac{n_\mathrm{p}(x)}{\tau_p + \tau_n}. \end{aligned} \quad \text{(EX.75)}$$

Noting that the second term ($\sim N \cdot D(n'/n)^2$) is negligible compared to the first term ($\sim n \cdot D(n'/n)^2$) under high-level injection (regardless of the possibly large $n'(x)$ near the contact), we may drop the second term. The differential equation (divided by D_n on both sides) then becomes

$$(1-\sigma)\frac{\mathrm{d}^2 n_\mathrm{p}}{\mathrm{d}x^2} + \frac{2N_\mathrm{a}^*}{N_\mathrm{a} + N_\mathrm{d}} n_\mathrm{pE} \left[\frac{n_\mathrm{pE}}{\lambda_n} + \frac{p_\mathrm{nE}}{\lambda_p}\right] \frac{1}{[n_\mathrm{p}(x)]^2} \frac{\mathrm{d}n_\mathrm{p}}{\mathrm{d}x} - \frac{n_\mathrm{p}(x)}{D_n(\tau_p + \tau_n)} = 0. \quad \text{(EX.76)}$$

Lastly, multiplying the equation by $(D_n + D_p)/2D_p$ on both sides will yield

$$\boxed{\frac{\mathrm{d}^2 n_\mathrm{p}}{\mathrm{d}x^2} + \frac{N_\mathrm{a}}{N_\mathrm{a} + N_\mathrm{d}} n_\mathrm{pE} \left[\frac{n_\mathrm{pE}}{\lambda_n} + \frac{p_\mathrm{nE}}{\lambda_p}\right] \frac{1}{[n_\mathrm{p}(x)]^2} \frac{\mathrm{d}n_\mathrm{p}}{\mathrm{d}x} - \frac{n_\mathrm{p}(x)}{2L^{*2}} = 0}. \quad \text{(EX.77)}$$

with the *composite diffusion length* defined as

$$\boxed{L^* \equiv \sqrt{\frac{D_p D_n}{D_p + D_n}(\tau_p + \tau_n)}.} \tag{EX.78}$$

Following the same derivation for the n-region, we obtain

$$\boxed{\frac{\mathrm{d}^2 p_\mathrm{n}}{\mathrm{d}x^2} - \frac{N_\mathrm{d}}{N_\mathrm{a} + N_\mathrm{d}} p_\mathrm{nE} \left[\frac{n_\mathrm{pE}}{\lambda_n} + \frac{p_\mathrm{nE}}{\lambda_p}\right] \frac{1}{[p_\mathrm{n}(x)]^2} \frac{\mathrm{d}p_\mathrm{n}}{\mathrm{d}x} - \frac{p_\mathrm{n}(x)}{2L^{*2}} = 0.} \tag{EX.79}$$

Derivation of these equation assumes possibly mismatched $(N_\mathrm{a},\ N_\mathrm{d})$, $(\mu_p,\ \mu_n)$, and $(\tau_p,\ \tau_n)$. The formulation is independent of $(W_\mathrm{p},\ W_\mathrm{n})$, but they may play a role in the boundary conditions. The formulation is also independent of Δ.

These equations (EX.77) and (EX.79) are unfortunately coupled. However, they become homogeneous differential equations when we exploit the fact that $n_\mathrm{pE} \approx n_\mathrm{i} e^{q\Delta/2k_\mathrm{B}T} \approx p_\mathrm{nE}$:

(High-level inj. reg.)

$$\frac{\mathrm{d}^2 n_\mathrm{p}}{\mathrm{d}x^2} + \left(\frac{1}{\lambda_n} + \frac{1}{\lambda_p}\right) \frac{N_\mathrm{a}}{N_\mathrm{a} + N_\mathrm{d}} \left[\frac{n_\mathrm{pE}}{n_\mathrm{p}(x)}\right]^2 \frac{\mathrm{d}n_\mathrm{p}}{\mathrm{d}x} - \frac{n_\mathrm{p}(x)}{2L^{*2}} = 0; \tag{EX.80a}$$

$$\frac{\mathrm{d}^2 p_\mathrm{n}}{\mathrm{d}x^2} - \left(\frac{1}{\lambda_n} + \frac{1}{\lambda_p}\right) \frac{N_\mathrm{d}}{N_\mathrm{a} + N_\mathrm{d}} \left[\frac{p_\mathrm{nE}}{p_\mathrm{n}(x)}\right]^2 \frac{\mathrm{d}p_\mathrm{n}}{\mathrm{d}x} - \frac{p_\mathrm{n}(x)}{2L^{*2}} = 0. \tag{EX.80b}$$

The boundary conditions are $(\mathrm{d}n_\mathrm{p}/\mathrm{d}x)_\mathrm{pE} = n_\mathrm{pE}/\lambda_n$, $(\mathrm{d}p_\mathrm{n}/\mathrm{d}x)_\mathrm{nE} = -p_\mathrm{nE}/\lambda_p$ at the depletion edges, and $n_\mathrm{p}(-W_\mathrm{p}) \to 0$, $p_\mathrm{n}(W_\mathrm{n}) \to 0$ at the contacts. Here it makes no difference to distinguish $x = -x_\mathrm{p}$ from $x = 0^-$, as well as $x = x_\mathrm{n}$ from $x = 0^+$, since x_p and x_n are generally considerably smaller than L^*. The differential equations are

doi:10.6342/NTU202304241

only loosely coupled by the coefficient $1/\lambda_p + 1/\lambda_n$ that is dependent on the initial shapes of the profiles. This coefficient would require extra work to be evaluated if we solve the differential equations as boundary value problems. Instead, we treat them as initial value problems and adjust the values of λ_n and λ_p such that the aforementioned boundary conditions at the contacts are matched. Note that L^* can be absorbed into x thanks to the formulation of the differential equations. Therefore, upon numerically solving them, we may set $L^* = 1$ (with all dimensions normalized by L^*), and also $n_{\mathrm{pE}} = p_{\mathrm{nE}} = 1$ due to their homogeneous nature. The solutions to λ_n and λ_p are really only dependent on N_{a} and N_{d} from the coefficients of the differential equations, as well as W_{p} and W_{n} from the boundary conditions. They are independent of n_{pE} and p_{nE} (i.e., Δ) as long as the injection level is sufficiently high, and are unaffected by the possibly mismatched μ_n and μ_p.

Treated as initial value problems ($n_{\mathrm{p}}(x)$ spanning from 0^- towards the $-x$ direction; $p_{\mathrm{n}}(x)$ spanning from 0^+ towards the $+x$ direction), we can evaluate W_{n} and W_{p} as functions of λ_n and λ_p by generating $n_{\mathrm{p}}(x)$ and $p_{\mathrm{n}}(x)$ from the initial conditions with numerical methods such as the Runge-Kutta method of order 4 [Ex10], and by finding the W_{p} where $n_{\mathrm{p}}(x)$ drops to 0, as well as the W_{n} where $p_{\mathrm{n}}(x)$ drops to 0. **Figure EX–15 (a, b)** show the numerical results of the contour plots of W_{p} and W_{n}, respectively, as functions of λ_n and λ_p, with a doping concentration ratio of $N_{\mathrm{d}}/N_{\mathrm{a}} = 3$ as an example. Both W_{p} and W_{n} increase with λ_n and λ_p, and both W_{p} and W_{n} explode at sufficiently high λ_n and λ_p. In other words, λ_n and λ_p increase with W_{p} and W_{n} when they are small, whereas saturate when they are large. This is similar to the low-level-injection behavior of λ_n and λ_p (EX.45) governed by the hyperbolic tangent functions.

doi:10.6342/NTU202304241

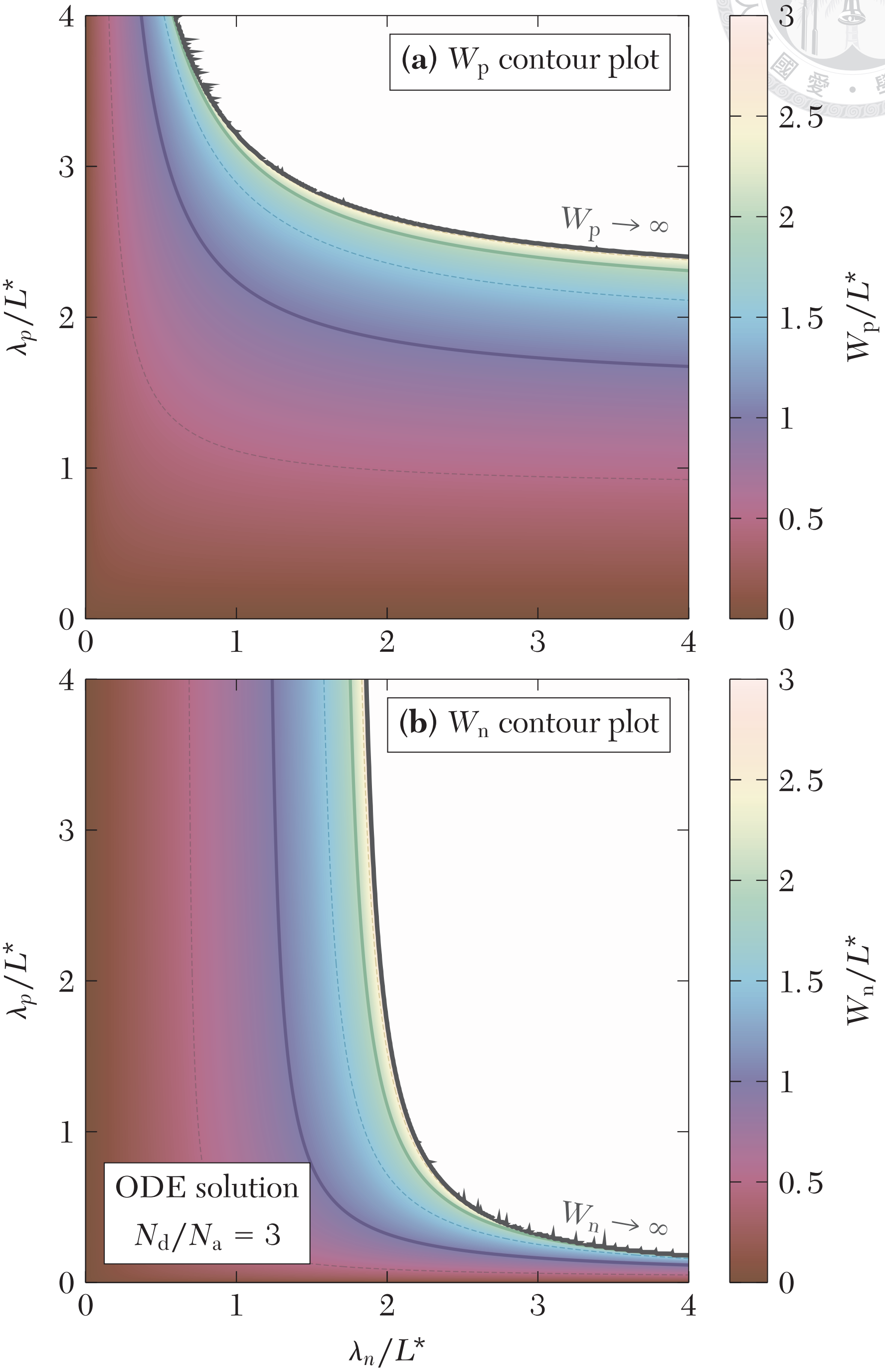


Figure EX–15. Contour plots of **(a)** W_p and **(b)** W_n as functions of λ_n and λ_p with $N_d/N_a = 3$. Numerical results from solving (EX.80a) and (EX.80b). All dimensions normalized by L^*.

doi:10.6342/NTU202304241

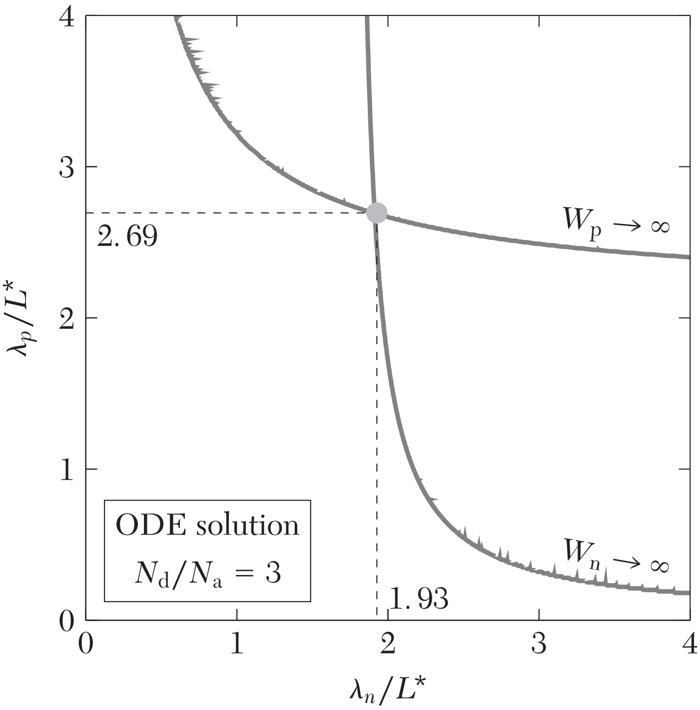


Figure EX–16. Demonstration of solving λ_n and λ_p from the intersection of the contours of $W_\mathrm{p} \to \infty$ and $W_\mathrm{n} \to \infty$ in **Figure EX–15**.

To solve for λ_n and λ_p given the desired values of W_p and W_n, an algorithm was implemented to find the contours of the given W_p and W_n in **Figure EX–15 (a)** and **Figure EX–15 (b)**, respectively, and then locate the intersection of the two contours. **Figure EX–16** demonstrates how λ_n and λ_p are solved from the contours of, for example, $W_\mathrm{p} \to \infty$ and $W_\mathrm{n} \to \infty$, the long diode case. By repeating the process for all doping concentration ratios, **Figure EX–17 (a)** plots λ_n and λ_p as functions of $N_\mathrm{d}/(N_\mathrm{a} + N_\mathrm{d})$ in p-n diodes with infinitely far contacts under high-level injection. λ_n and λ_p exhibit antisymmetry with respect to the doping concentration ratio $N_\mathrm{d}/(N_\mathrm{a} + N_\mathrm{d}) = 1/2$. λ_n decreases monotonically as the doping concentration ratio $N_\mathrm{d}/(N_\mathrm{a} + N_\mathrm{d})$ increases, whereas λ_p increases monotonically. For the matched doping concentration ($N_\mathrm{a} = N_\mathrm{d}$) case, $\lambda_n = \lambda_p = 2.30L^*$; while for the $N_\mathrm{d} \gg N_\mathrm{a}$ case, $\lambda_n = \sqrt{2}L^*$ and

doi:10.6342/NTU202304241

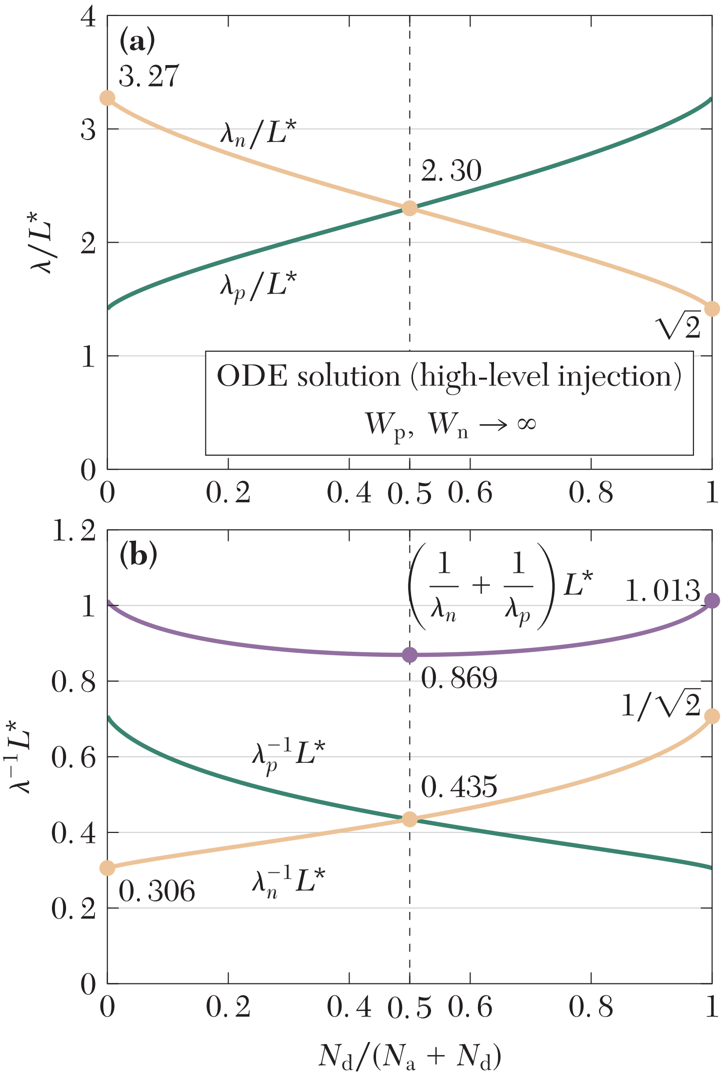


Figure EX–17. **(a)** Plots of λ_n and λ_p as functions of $N_d/(N_a+N_d)$ in p-n diodes with infinitely far contacts under high-level injection. Numerical solutions to (EX.80a) and (EX.80b). **(b)** Plots of $1/\lambda_n$, $1/\lambda_p$, and $1/\lambda_n + 1/\lambda_p$.

$\lambda_p = 3.27L^*$. **Figure EX–17 (b)** plots $1/\lambda_n$ and $1/\lambda_p$, along with their sum that is found in the expressions of total current under high-level injection. According to this plot, $1/\lambda_n + 1/\lambda_p \approx 1/L^*$ might be a decent approximation regardless of the doping concentration ratio.

Figure EX–18 (a, b) plots the extracted values of λ_n and λ_p at the depletion edges, as functions of Δ, from TCAD simulation results in long p-n diodes with $N_a = N_d$ and

doi:10.6342/NTU202304241

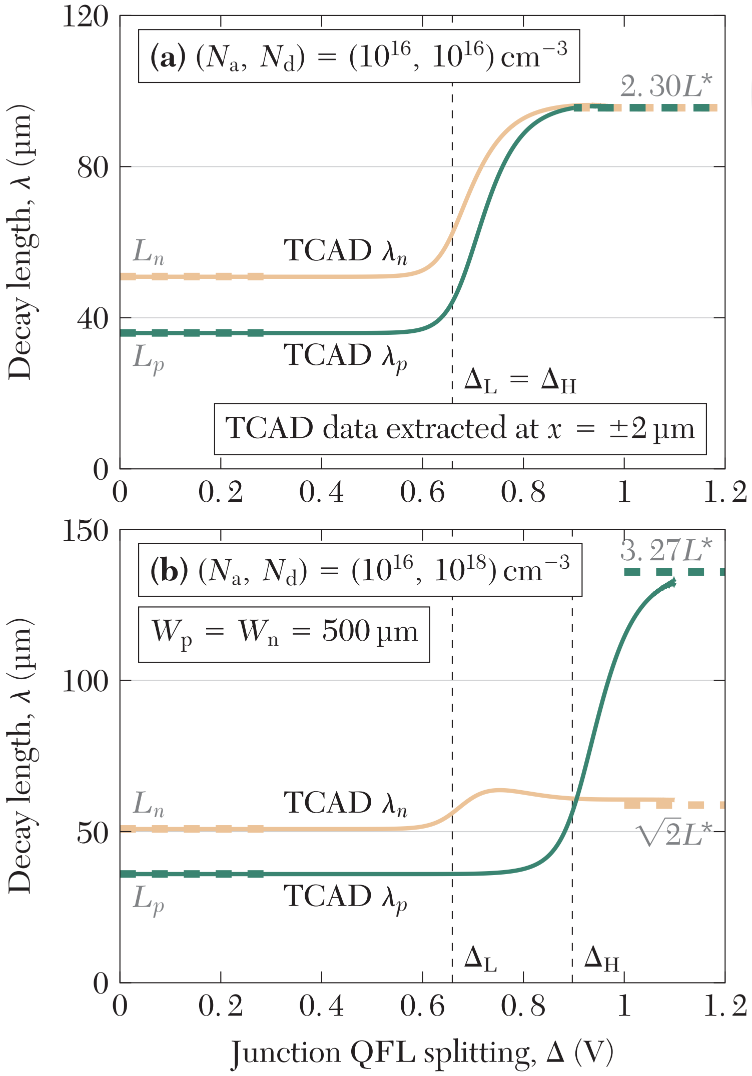


Figure EX–18. TCAD-extracted $\lambda_n - \Delta$ and $\lambda_p - \Delta$ plots in p-n diodes with far contacts and with **(a)** $N_a = N_d$ and **(b)** $N_d \gg N_a$. Also marked are the theoretical low-level injection values (i.e., diffusion lengths) and the modeled high-level injection values as shown in **Figure EX–17**.

$N_d \gg N_a$, respectively. Also marked in the figures are the theoretical low-level injection values of the decay lengths (i.e., L_n and L_p) and the modeled high-level injection values as given by **Figure EX–17**. While the transitional behavior across regimes is beyond the scope of this work that only deals with regional approximations, our modeled values show strong agreement with the TCAD simulated results. Traversing from the low-level to the high-level injection regime, the decay lengths indeed grow longer.

EX.4.5 Mixed-Level-Injection Decay Lengths In p-n Diodes with Highly Mismatched Doping Concentration

In p-n diodes with highly mismatched doping concentrations (say $N_\mathrm{d} \gg N_\mathrm{a}$), the p-region is under high-level injection in the mixed-level injection regime, whereas the n-region is under low-level injection. Thus, the low-level-injection limit of the minority carrier decay length (EX.45) is applicable in the quasi-neutral n-region:

$$(N_\mathrm{d} \gg N_\mathrm{a};\ \text{mixed-level inj. reg.}) \qquad \lambda_p = L_p \tanh(W_\mathrm{n}/L_p). \tag{EX.81}$$

Meanwhile, to obtain λ_n, we can derive the differential equation for $n_\mathrm{p}(x)$ similar to (EX.77), except that we utilizing the mix-level-injection expression of $\mathscr{E}_\mathrm{pE}$ in (EX.50):

$$(N_\mathrm{d} \gg N_\mathrm{a};\ \text{mixed-level inj. reg.}) \qquad \begin{aligned} &\frac{\mathrm{d}^2 n_\mathrm{p}}{\mathrm{d}x^2} + \frac{1}{\lambda_n}\frac{N_\mathrm{a}}{2n_\mathrm{i}e^{q\Delta/2k_\mathrm{B}T}}\left[\frac{n_\mathrm{pE}}{n_\mathrm{p}(x)}\right]^2 - \frac{n_\mathrm{p}(x)}{2L^*} \\ =&\frac{\mathrm{d}^2 n_\mathrm{p}}{\mathrm{d}x^2} + \frac{e^{-q(\Delta-\Delta_\mathrm{L})/k_\mathrm{B}T}}{\lambda_n}\left[\frac{n_\mathrm{pE}}{n_\mathrm{p}(x)}\right]^2 - \frac{n_\mathrm{p}(x)}{2L^*} \\ =&0. \end{aligned} \tag{EX.82}$$

However, the first-order term vanishes as $e^{-q(\Delta-\Delta_\mathrm{L})/k_\mathrm{B}T} \ll 1$ in the mixed-level injection regime where $\Delta > \Delta_\mathrm{L}$. This results in

$$(N_\mathrm{d} \gg N_\mathrm{a};\ \text{mixed-level inj. reg.}) \qquad \frac{\mathrm{d}^2 n_\mathrm{p}}{\mathrm{d}x^2} - \frac{n_\mathrm{p}(x)}{2L^*} = 0. \tag{EX.83}$$

doi:10.6342/NTU202304241

The solution to $n_\mathrm{p}(x)$ subject to boundary conditions $n_\mathrm{p}(0^-) = n_\mathrm{pE}$ and $n_\mathrm{p}(-W_\mathrm{p}) \to 0$ is

$$(N_\mathrm{d} \gg N_\mathrm{a};\ \text{mixed-level inj. reg.}) \quad n_\mathrm{p}(x) = n_\mathrm{pE}\left[\cosh\left(\frac{x}{\sqrt{2}L^*}\right) + \frac{\sinh(x/\sqrt{2}L^*)}{\tanh(W_\mathrm{p}/\sqrt{2}L^*)}\right]. \tag{EX.84}$$

In any case, λ_n (EX.44a) is now

$$(N_\mathrm{d} \gg N_\mathrm{a};\ \text{mixed-level inj. reg.}) \qquad \lambda_n = \sqrt{2}L^* \tanh(W_\mathrm{p}/\sqrt{2}L^*). \tag{EX.85}$$

The total diode current in the mixed-level injection regime (EX.58) can therefore be approximated as

$$(N_\mathrm{d} \gg N_\mathrm{a};\ \text{mixed-level inj. reg.}) \qquad J_\mathrm{tot} \approx \sqrt{2}qD_n \frac{n_\mathrm{i}}{L^* \tanh(W_\mathrm{p}/\sqrt{2}L^*)} e^{q\Delta/2k_\mathrm{B}T}. \tag{EX.86}$$

EX.4.6 Correction on Diode Minority Currents by Regional Decay Lengths.

In the mixed- and high-level injection regimes, the aforementioned overestimation of currents by the model can be attributed to the wrongful employment of low-level-injection decay lengths (L_n and L_p for long diodes) throughout all operation regimes. **Figure EX–19** is a revisit of **Figure EX–13 (c)** and **Figure EX–14 (c)**, the total minority electron and hole currents at the depletion edge of p-n diodes, except that we employ the newly-evaluated regional decay lengths in the modeled plots. With the correction on λ_n and λ_p, the discrepancy between the modeled and TCAD curves are now

doi:10.6342/NTU202304241

eliminated.

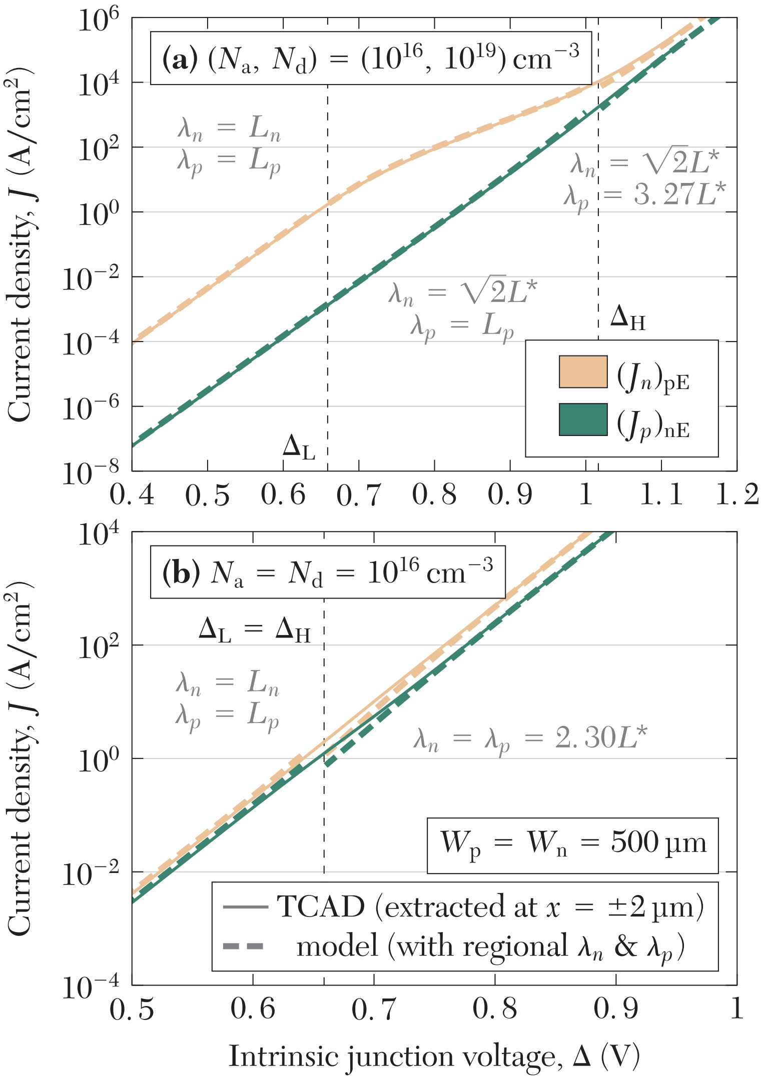


Figure EX–19. Modeled and TCAD $(J_n)_{pE} - \Delta$ and $(J_p)_{nE} - \Delta$ plots in p-n diodes with far contacts and with **(a)** $N_a = N_d$ and **(b)** $N_d \gg N_a$. The regional decay lengths evaluated in **Section EX.4** are utilized in the modeled curves. (Compare with **Figure EX–13 (c)** and **Figure EX–14 (c)** where low-level-injection decay lengths are assumed throughout all injection levels.)

doi:10.6342/NTU202304241

EX.5 Potential Drop Across the Quasi-Neutral Regions

EX.5.1 Potential Drop Across Quasi-Neutral Regions with Hyperbolic Carrier Concentration Profiles

So far, the framework of our discussion is based on the junction QFL splitting, Δ. To relate it with the most external and direct quantity, the applied voltage V_{a}, one may exploit (EX.11) to decompose V_{a} in terms of the junction barrier (EX.8) which is Δ-dependent, and the potential drops across the quasi-neutral regions ϕ_{p} and ϕ_{n} (EX.10a), (EX.10b) which can be integrated from $\mathscr{E}_{\mathrm{p}}(x)$ and $\mathscr{E}_{\mathrm{n}}(x)$, respectively. The expressions for the electric fields in the quasi-neutral regions have been derived in **Section EX.4** ((EX.67), (EX.68)), showing dependencies on $n_{\mathrm{p}}(x)$ and $p_{\mathrm{n}}(x)$. Therefore, we can generally calculate ϕ_{p} and ϕ_{n} once the $n_{\mathrm{p}}(x)$ and $p_{\mathrm{n}}(x)$ profiles have been numerically solved in **Section EX.4**.

A special case for the minority carrier concentration profile (which we refer to as the hyperbolic concentration case), say $p_{\mathrm{n}}(x)$ in the n-region, is where the first-order derivative term in the carrier continuity equation vanishes, resulting in the form of

$$\text{(Hyp. conc.)} \qquad \frac{\mathrm{d}^2 p_{\mathrm{n}}}{\mathrm{d}x^2} - \frac{p_{\mathrm{n}}(x)}{L_p^{\dagger\,2}} = 0, \qquad \text{(EX.87)}$$

where $L_p^{\dagger}$ is a length constant. As discussed earlier, this takes place in the low-level injection regime ($L_n^{\dagger} = L_n$, $L_p^{\dagger} = L_p$), the mixed-level injection regime (for $N_{\mathrm{d}} > N_{\mathrm{a}}$, $L_n^{\dagger} = \sqrt{2}L^{*}$ and $L_p^{\dagger} = L_p$), or on the lightly-doped side in the high-level injection regime (for $N_{\mathrm{d}} > N_{\mathrm{a}}$, $L_n^{\dagger} = \sqrt{2}L^{*}$). The solution to this differential equation, subject

doi:10.6342/NTU202304241

to the boundary conditions $p_\mathrm{n}(0^+) = p_\mathrm{nE}$ and $p_\mathrm{n}(W_\mathrm{n}) \to 0$, is

$$\text{(Hyp. conc.)} \qquad p_\mathrm{n}(x) = p_\mathrm{nE}\left[\cosh\left(\frac{x}{L_p^\dagger}\right) - \frac{\sinh(x/L_p^\dagger)}{\tanh(W_\mathrm{n}/L_p^\dagger)}\right]. \tag{EX.88}$$

An analytical expression for ϕ_n can thus be obtained by plugging this into (EX.68) and by integrating $\mathscr{E}_\mathrm{n}(x)$. The integration utilizes the following identities [Ex11]:

$$\int \frac{\sinh u}{1 + b\cosh u}\mathrm{d}u = \frac{1}{b}\ln[1 + b\cosh u] + C, \tag{EX.89}$$

and

$$\int \frac{\mathrm{d}u}{1 + \cos a \cosh u} = \frac{2}{\sin a}\operatorname{artanh}\left(\tan\frac{a}{2}\cdot\tanh\frac{u}{2}\right) + C. \tag{EX.90}$$

The result is

$$\begin{aligned}\text{(Hyp. conc.)} \qquad \phi_\mathrm{n} = \frac{kT}{q}\Bigg\{&\left[\left(1 + \frac{N_\mathrm{d}^*}{p_\mathrm{nE}}\right)\frac{q}{kT}\mathscr{E}_\mathrm{nE}L_p^\dagger\sinh(W_\mathrm{n}/L_p^\dagger) - \sigma\cosh(W_\mathrm{n}/L_p^\dagger)\right]\\ &\times\frac{2\operatorname{artanh}\left(\frac{\sqrt{1+(N_\mathrm{d}^*/p_\mathrm{nE})^2\sinh^2(W_\mathrm{n}/L_p^\dagger)}}{1+(N_\mathrm{d}^*/p_\mathrm{nE})(1+\cosh(W_\mathrm{n}/L_p^\dagger))}\right)}{\sqrt{1 + (N_\mathrm{d}^*/p_\mathrm{nE})^2\sinh^2(W_\mathrm{n}/L_p^\dagger)}}\\ &+ \sigma\ln\left(1 + \frac{p_\mathrm{nE}}{N_\mathrm{d}^*}\right)\Bigg\}.\end{aligned} \tag{EX.91}$$

We can also derive an analytical expression for ϕ_p akin to this one if the p-region possesses a hyperbolic $n_\mathrm{p}(x)$ profile as well.

EX.5.2 Estimation of the Ideality Factor in the Low-Level Injection Regime

It is commonly perceived that the presence of potential drop across the quasi-neutral regions, referred to as the series resistance effect, is pronounced under higher injection

doi:10.6342/NTU202304241

levels. To justify its effect even in the low-level injection regime, we may inspect the ideality factor (η) by assuming a locally exponential dependence of the diode current, J_{tot}, on the applied voltage [Ex2]:

$$J_{\mathrm{tot}} \propto \exp\left(\frac{qV_{\mathrm{a}}}{\eta k_{\mathrm{B}}T}\right). \tag{EX.92}$$

Then the ideality factor can be evaluated as

$$\eta = \frac{q}{k_{\mathrm{B}}T}\left(\frac{\mathrm{d}J_{\mathrm{tot}}}{\mathrm{d}V_{\mathrm{a}}}\right)^{-1} J_{\mathrm{tot}}. \tag{EX.93}$$

Under forward bias, $\mathscr{E}_{\mathrm{p}}(x)$ and $\mathscr{E}_{\mathrm{n}}(x)$ in the quasi-neutral regions are always minimized at the depletion edges, with the minima of $\mathscr{E}_{\mathrm{pE}}$ and $\mathscr{E}_{\mathrm{nE}}$, respectively, since the carrier concentrations throughout the quasi-neutral regions are the highest at the depletion edges, and because of (EX.67) and (EX.68). As such, ϕ_{p} and ϕ_n (EX.10a), (EX.10b) cannot be less than $\mathscr{E}_{\mathrm{pE}}W_{\mathrm{p}}$ and $\mathscr{E}_{\mathrm{nE}}W_{\mathrm{n}}$, respectively. Also recalling that $\psi_{\mathrm{bi}} - \phi_{\mathrm{j}} \approx \Delta$ under low-level injection, (EX.11) becomes

$$\text{(Low-level inj. reg.)} \qquad V_{\mathrm{a}} \geq \mathscr{E}_{\mathrm{pE}}W_{\mathrm{p}} + \Delta + \mathscr{E}_{\mathrm{nE}}W_{\mathrm{n}}. \tag{EX.94}$$

Utilizing the low-level-injection limits of $\mathscr{E}_{\mathrm{pE}}$ and $\mathscr{E}_{\mathrm{nE}}$ in (EX.48) and (EX.49), this becomes

$$\text{(Low-level inj. reg.)} \qquad V_{\mathrm{a}} \geq \Delta + \frac{k_{\mathrm{B}}T}{q}\left(\frac{1}{N_{\mathrm{a}}\lambda_n} + \frac{1}{N_{\mathrm{d}}\lambda_p}\right)\left(\frac{W_{\mathrm{p}}}{N_{\mathrm{a}}} + \frac{W_{\mathrm{n}}}{N_{\mathrm{d}}}\right) n_{\mathrm{i}}^2 e^{q\Delta/k_{\mathrm{B}}T}; \tag{EX.95}$$

doi:10.6342/NTU202304241

with its derivative with respect to Δ being

$$\text{(Low-level inj. reg.)} \qquad \frac{\mathrm{d}V_\mathrm{a}}{\mathrm{d}\Delta} \geq 1 + \left(\frac{1}{N_\mathrm{a}\lambda_n} + \frac{1}{N_\mathrm{d}\lambda_p}\right)\left(\frac{W_\mathrm{p}}{N_\mathrm{a}} + \frac{W_\mathrm{n}}{N_\mathrm{d}}\right) n_\mathrm{i}^2 e^{q\Delta/k_\mathrm{B}T}. \qquad \text{(EX.96)}$$

For $N_\mathrm{d} \geq N_\mathrm{a}$, at the limit of low-level injection regime ($\Delta = \Delta_\mathrm{L}$) where $\exp(q\Delta_\mathrm{L}/2k_\mathrm{B}T) = N_\mathrm{a}/2n_\mathrm{i}$, this is evaluated as

$$\begin{aligned} (\Delta = \Delta_\mathrm{L}) \qquad \left(\frac{\mathrm{d}V_\mathrm{a}}{\mathrm{d}\Delta}\right)_{\Delta_\mathrm{L}} &\geq 1 + \frac{1}{4}\left(\frac{1}{N_\mathrm{a}\lambda_n} + \frac{1}{N_\mathrm{d}\lambda_p}\right)\left(\frac{W_\mathrm{p}}{N_\mathrm{a}} + \frac{W_\mathrm{n}}{N_\mathrm{d}}\right) N_\mathrm{a}^2 \\ &\geq 1 + \frac{W_\mathrm{p}}{4\lambda_n}. \qquad \text{(EX.97)} \end{aligned}$$

Thus, the ideality factor (EX.93) at $\Delta = \Delta_\mathrm{L}$ is

$$\begin{aligned} (\Delta = \Delta_\mathrm{L}) \qquad \eta &= \frac{q}{k_\mathrm{B}T}\left(\frac{\mathrm{d}J_\mathrm{tot}}{\mathrm{d}\Delta}\right)^{-1}\left(\frac{dV_\mathrm{a}}{d\Delta}\right)_{\Delta_\mathrm{L}} J_\mathrm{tot} \\ &= \left(\frac{\mathrm{d}V_\mathrm{a}}{\mathrm{d}\Delta}\right)_{\Delta_\mathrm{L}} \\ &\geq 1 + \frac{W_\mathrm{p}}{4\lambda_n}. \qquad \text{(EX.98)} \end{aligned}$$

This utilizes the exponential dependence of $J_\mathrm{tot} \propto \exp(q\Delta/k_\mathrm{B}T)$ to Δ in the low-level injection regime such that $\mathrm{d}J_\mathrm{tot}/\mathrm{d}\Delta = qJ_\mathrm{tot}/k_\mathrm{B}T$. In any case, $\lambda_n < W_\mathrm{p}$ under low-level injection (EX.45), making $\eta \geq 5/4$ at $\Delta = \Delta_\mathrm{L}$. This lower bound can even be considerably underestimated, regarding that the N_d terms are dropped (which will yield $\eta \gtrsim 2$ if reconsidered in a perfectly symmetric p-n diode), and that ϕ_p and ϕ_n have been underestimated. In conclusion, the series resistance effect, as indicated by $\eta > 1$, has already taken place even before Δ reaches the limit of the low-level injection regime ($\Delta = \Delta_\mathrm{L}$).

EX.6 Conclusion

A general model for the electrostatics and currents of an abrupt p-n junction, consistent with TCAD simulation results, has been established in this work. We can summarize several important findings as follows:

1. The quantities V_{a}, Δ, and $(\psi_{\mathrm{bi}} - \phi_{\mathrm{j}})$ are indistinguishable only in the low-level injection regime, above which distinction between them is crucial for meaningful upcoming discussions.

2. In the high-level injection regime, the bands of the junction region flattens while ϕ_{j} retains strictly positive, and the profile of net charge density shrinks to two antisymmetric spikes with magnitudes of $\pm q(N_{\mathrm{a}} + N_{\mathrm{d}})/2$.

3. In the low-level injection regime, the potential and net charge density profiles on each side of the p-n junction follow universal shapes, with the horizontal translation of the shapes determined by Δ (in the form of $x_{1/2}$), the horizontal scaling determined by the doping concentration (in the form of extrinsic Debye length), and the vertical scaling of the $\rho(x)$ profile determined also by the doping concentration.

4. The dominant current components in the low- and high-level injection regimes are the diffusion and drift currents, respectively. In the mixed-level injection regime, the drift and diffusion currents of the dominating carrier type are comparable.

5. The total junction current exhibits proportionality to $\exp(q\Delta/k_{\mathrm{B}}T)$ in the low-

and high-level injection regimes, and only to $\exp(q\Delta/2k_\mathrm{B}T)$ in, if present, the mixed-level injection regime.

6. It must be taken into account upon evaluating the currents that λ_n and λ_p grow longer in the high-level injection regime than the commonly-perceived low-level injection values. The latter are given by (EX.45), whereas the former can be solved from a pair of ordinary differential equations ((EX.80a), (EX.80b)) with the solutions only dependent on the doping concentration ratio, but independent of Δ and possible mobility mismatch.

7. The series resistance effect on the ideality factor of the diode already takes place before $\Delta = \Delta_\mathrm{L}$, the upper limit of the low-level injection regime.

Chapter References

[Ex1] W. Shockley, “The Theory of *p-n* Junction in Semiconductors and *p-n* Junction Transistors,” *The Bell System Technical Journal* **28** (3): 435–489, 1949, doi:10.1002/j.1538-7305.1949.tb03645.x.

[Ex2] S. M. Sze & M. K. Lee. *Semiconductor Devices Physics and Technology*, 3 ed. John Wiley & Sons, 2012.

[Ex3] D. P. Kennedy, “The Potential and Electric Field at the Metallurgical Boundary of An Abrupt p-n Semiconductor Junction,” *IEEE Trans. Electron Dev.* **22** (11): 988–994, 1975, doi:10.1109/T-ED.1975.18258.

[Ex4] Y. Yue, J. J. Liou, & A. Ortiz-Conde, "High-Level Injection in Quasi-Neutral Region of n/p Junction," *J. Appl. Phys.* **77** (4): 1611–1615, February 1995, doi:10.1063/1.358915.

[Ex5] J. C. Manifacier, R. Ardebili, & C. Popescu, "High-Level Injection Phenomena in P-N Junctions," *J. Appl. Phys.* **80** (5): 2838–2846, September 1996, doi:10.1063/1.363134.

[Ex6] L. D. Edmonds, "High Level Injection in n^+–p Junction Silicon Devices," *J. Appl. Phys.* **97** (12): 124506–1–124506–7, June 2005, doi:10.1063/1.1940134.

[Ex7] W. M. Webster, "On the Variation of Junction-Transistor Current-Amplification Factor with Emitter Current," *Proc. IRE* **42** (6): 914–920, 1954, doi:10.1109/JRPROC.1954/274751.

[Ex8] Y. Taur & T. H. Ning. *Fundamentals of Modern VLSI Devices*. Cambridge University Press, Cambridge, 1998.

[Ex9] A. D. Polyanin & V. F. Zaitsev. *Handbook of Ordinary Differential Equations: Exact Solutions, Methods, and Problems*, 3 ed. CRC Press, 2017.

[Ex10] W. Kutta, "Beitrag zur Näherungsweisen Integration Totaler Differentialgleichungen," *Z. Math. Phys.* **46**: 434–453, 1901.

[Ex11] G. F. Becker & C. E. van Orstrand. *Smithsonian Mathematical Tables: Hyperbolic Functions*. University of Michigan Library, 1909.