Investigation of electrically active defects in GaN, AlGaN, and AlGaN/GaN high electron mobility transistors

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree
Doctor of Philosophy in the Graduate School of The Ohio State University

By

Aaron R. Arehart, B.S., M.S.

Graduate Program of Electrical and Computer Engineering

The Ohio State University

2009

Dissertation Committee:
Professor Steven A. Ringel, Adviser
Professor Leonard Brillson
Professor Siddharth Rajan
Abstract

Although nitride electronics have matured rapidly, the performance and reliability of nitride high electron mobility transistors (HEMT) and other electronic devices have been hampered by electrically active defects that manifest as deep levels in the bandgap and/or as trap states. To alleviate these problems, not only is a fundamental understanding of the defects in GaN and AlGaN necessary for the continued development of nitride electronics, but also correlations of these defects to the performance and reliability limiting problems is required. Using a multipronged effort, defects were quantitatively studied at the component layer level (i.e. GaN and AlGaN) and also in operational HEMTs using techniques uniquely designed to quantitatively characterize defects as deep levels and traps in these devices.

Deep level optical spectroscopy and related methods of trap spectroscopy are applied to several sets of systematically varied GaN and AlGaN materials. Traps in GaN were typically located at \( E_C - 0.25, E_C - 0.60, E_C - 0.90, E_C - (1.28 - 1.35), E_C - 2.6, \) and \( E_C - 3.22/3.28 \) and for AlGaN at \( E_C - 0.87, E_C - 1.5, E_C - 3.10, \) and \( E_C - 3.93 \). It was determined that several traps showed specific dependencies on variations in growth parameters, substrate orientation, dislocation density, and growth method. Physical sources were attributed to most of these states for the first time, and this taxonomy is essential for analysis of trap effects in working AlGaN/GaN transistors, which constitutes the second focus of this research.

To relate defect incorporation with HEMT performance and reliability, constant drain-current deep level optical/transient spectroscopies using gate or drain voltage as the feedback
mechanism are developed. This enables simultaneous and quantitative measurement of
defect energies and concentrations of individual defects throughout the bandgap in HEMTs,
measurement of device relevant parameters (threshold voltage shift and the change in gate-
drain access resistance), separation of gate and access regions, and comparison of other
HEMT results because of the absolute nature of these measurements. This novel set of
techniques is applied to a HEMT for the first time revealing a 105 mV shift in the threshold
voltage corresponding to a ~3.5 × 10^{11} \text{ cm}^{-2} trap concentration under the gate and 6 \Omega
variation in access resistance in which initial calculations reveal a total access region defect
concentration of ~8 × 10^{12} \text{ cm}^{-2}. States detected under the gate at \( E_C \sim 0.59 \) and \( E_C \sim 3.3 \text{ eV} \)
correlate with known dislocation-related and carbon acceptor defects in GaN, respectively,
while the ~\( E_C \sim 2.3 \) and AlGaN-related \( E_C \sim 3.7 \text{ eV} \) levels observed under the gate and in the
access region are not localized only to the surface. The access region also exhibited a distinct
level at \( E_C \sim 0.47 \) not previously seen in thick GaN or AlGaN. The conclusion of this work is
that the time dependencies of these deep levels are likely sources of dispersion in AlGaN/GaN
HEMTs.

To further refine the capabilities to quantitatively measure defect energies and concen-
trations in the access regions of HEMTs, an atomic force microscope is adapted to perform
nanometer-scale defect characterization. Using scanning Kelvin probe microscopy, evidence
of the spatial and time-dependent measurement capabilities is demonstrated. Initial HEMT
results are presented and suggest the total trap concentration of ~10^{12} \text{ cm}^{-2} consistent with
previous results.
This is dedicated to my family
I am indebted to my adviser, Prof. Steve Ringel, for supporting and leading my work throughout my graduate career. In addition, much of this dissertation would not be possible without the financial freedom provided by Prof. Ringel that allowed new equipment to be designed and custom systems to be assembled. I would also like to thank my committee members and specifically Prof. Rajan for his discussions during the development of the transistor-based measurement techniques. I also greatly appreciate the work of Prof. Speck and Prof. Mishra’s groups at the University of California, Santa Barbara not only for growing nearly all of the samples in this work but also their insight into the workings of nitride transistors. Over the years there have been a number of people I have interacted with including: Man Hoi Wong, Andrea Corrion, Christy Poblenz, Yi Pei, Rongming Chu, and all the other students and former students at UCSB whom I have had the fortune to work with.

The facilities for nitride processing did not exist at OSU when I began processing, but Air Force Research Laboratory at Wright Patterson Air Force Base graciously allowed me to use their facilities and served as a vast resource of knowledge that allowed the successful processing of devices even for novel, one-of-a-kind structures. To Bob Fitch, Gregg Jessen, Jim Gillespie, Antonio Crespo, and Dave Via, I am greatly beholden.

Our collaboration with Prof. Pelz has recently enabled the development of new nano-scale defect spectroscopies. Drew Cardwell has been a great person to work with, and I look forward
to continued collaboration as we work with the AFM to push the envelop of measurement capabilities.

I have worked with a great bunch of people (past and present) in the Electronic Materials and Devices Laboratory and especially my fellow utopians María González and Andy Armstrong. The group dynamic has changed greatly over the past few years, but I will fondly remember the social aspects including parties, volleyball, and lunches as well as professional aspects. To Yong Lin, I have enjoyed our interaction and hope to continue to build our friendship. A special thanks to John Boeckl for lending his microscopy skills, base access, and the great friendship he has offered. Drew Malonis has performed many of the transistor-based spectroscopy measurements and often served as the Devil’s advocate to keep me on my toes. To Mark Brenner, Andrew Carlin, Krishna Swaminathan, Chris Ratcliff, and Tyler Grassman, I look forward to your tutelage in all MBE matters. John Carlin, the old MBEer, it is always enjoyable to hang out with you. Aimee Bross, thanks for the short notice e-beam lithography. Finally to Qilin Gu and Tony Homan, I have enjoyed sharing the DLOS systems with you guys.

Most importantly, I would like to thank my parents who have always valued knowledge and education and my family, Kelly and Nathanael, for supporting me through the long hours.
Vita

July 4, 1976 ........................................... Born - Toledo, OH

1999 .................................................... B.S. Electrical and Computer Engineering, The Ohio State University

1999-2000 ........................................ Distinguished University Fellow, The Ohio State University

2000 .................................................... M.S. Electrical and Computer Engineering, The Ohio State University

2000-2006 .......................................... Graduate Research Associate, The Ohio State University

2006-2007 .......................................... Distinguished University Fellow, The Ohio State University

2007-2009 .......................................... Graduate Research Associate, The Ohio State University

Publications

Research Publications


**Fields of Study**

Major Field: Electrical and Computer Engineering
# Table of Contents

Abstract ........................................................................................................... ii  
Dedication ......................................................................................................... iv  
Acknowledgments ............................................................................................... v  
Vita .................................................................................................................. vii  
List of Tables .................................................................................................... xiv  
List of Figures ................................................................................................... xv  

Chapters:

1. Introduction .................................................................................................... 1  
   1.1 Motivation ............................................................................................... 1  
   1.2 Objective ................................................................................................ 4  
   1.3 Overview of the thesis ......................................................................... 4  
   1.4 References ........................................................................................... 11  

2. Background .................................................................................................... 13  
   2.1 Crystal structure ............................................................................... 15  
   2.2 Polarization .......................................................................................... 17  
   2.3 2DEG formation .............................................................................. 18  
   2.4 Defects and HEMT specific problems .............................................. 20  
   2.5 References ........................................................................................... 35
5.3.3 Conclusion ................................................................. 135
5.4 Correlation of deep levels in ammonia-MBE n-GaN grown with increasing V/III flux ratio ................................................................. 136
5.5 Comparison of defect incorporation in N-face and Ga-face polarity growth of PAMBE n-GaN ................................................................. 143
5.6 Impact of dislocation density on trap concentrations in MOCVD-grown n- GaN Schottky diodes ......................................................... 152
5.7 Defect identification in thick, MOCVD-grown n-AlGaN on AlN ................. 162
  5.7.1 Experimental details .................................................... 163
  5.7.2 Results and discussion .................................................. 164
  5.7.3 Conclusion .............................................................. 173
5.8 Summary ................................................................. 173
5.9 References .............................................................. 177

6. Deep level characterization of HEMT transistors ........................................... 185
  6.1 Theory of constant drain-current and constant conductance DLOS/DLTS ... 187
    6.1.1 Theory for gate-controlled measurements ............................ 187
    6.1.2 Drain-controlled CI-DLOS theory .................................. 197
    6.1.3 Measurement procedure ............................................. 207
  6.2 Experimental Setup ...................................................... 213
  6.3 Experimental verification ................................................ 221
  6.4 Initial results .......................................................... 226
  6.5 Conclusion ............................................................. 238
  6.6 References ............................................................ 242

7. Next generation of quantitative nano-scale defect characterization in AlGaN/GaN HEMTs ................................................................. 245
  7.1 Near-field scanning optical microscopy .................................... 246
  7.2 Lighted capacitance microscopy .......................................... 247
  7.3 Lighted Kelvin probe microscopy ......................................... 252
  7.4 Summary ............................................................... 270
  7.5 References ............................................................. 273

8. Conclusions and future work .......................................................... 275
  8.1 Continuing studies of component materials in nitride HEMTs ................. 276
  8.2 New frontiers in defect characterization of HEMTs ......................... 279
  8.3 Providing quantitative defect characterization of semiconductor devices with nano-scale resolution ................................................. 281
  8.4 Future work ............................................................. 283

xi
8.5 References ....................................................... 290

Appendices:

A. Nitride Schottky diode processing .................................. 293
   A.1 Introduction ..................................................... 293
   A.2 Wafer Clean ...................................................... 296
   A.3 Ni Schottky Contact ............................................... 297
   A.4 Mesa Etch .......................................................... 299
   A.5 Ohmic Contact ..................................................... 302
   A.6 Thick Front Contact ............................................... 304

B. The appendix of Labview programs written to support the research ...... 306
   B.1 Stage operation ..................................................... 307
      B.1.1 Vacuum procedures ........................................... 307
      B.1.2 Temperature procedures ....................................... 308
   B.2 Current-voltage measurements ...................................... 309
      B.2.1 General I-V theory ............................................ 309
      B.2.2 Schottky I-V-T theory ......................................... 310
      B.2.3 I-V circuit diagram ............................................ 318
      B.2.4 I-V program ................................................... 319
      B.2.5 I-V analysis ................................................... 321
      B.2.6 Other I-V programs ............................................ 322
   B.3 Capacitance-voltage ................................................ 325
      B.3.1 C-V circuit diagram ............................................ 325
      B.3.2 C-V program ................................................... 326
      B.3.3 Doping calculator program ..................................... 328
   B.4 Deep level transient spectroscopy .................................. 330
      B.4.1 DLTS circuit diagram .......................................... 330
      B.4.2 DLTS program .................................................. 331
      B.4.3 DLTS analysis .................................................. 335
   B.5 Capacitance-temperature ........................................... 337
      B.5.1 C-T circuit ..................................................... 338
      B.5.2 C-T program ................................................... 338
   B.6 Deep level optical spectroscopy ..................................... 340
      B.6.1 DLOS setup ..................................................... 340
      B.6.2 DLOS program .................................................. 341
      B.6.3 DLOS analysis .................................................. 346
   B.7 Internal photoemission ............................................... 349
      B.7.1 IPE theory ..................................................... 349
B.7.2 IPE setup  ......................................................... 351
B.7.3 IPE program  .................................................... 353
B.8 Photon flux measurement ............................................ 355
  B.8.1 Photon flux circuit ............................................. 355
  B.8.2 Photon flux program .......................................... 356
B.9 References ............................................................. 358

C. Impact of gate-source access resistance variation on CI\textsubscript{D}-DLOS measurements ........................................... 360
  C.1 References ........................................................... 363

Bibliography ................................................................. 364
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Figure of merits from Ref. 8 for various competing material systems. Johnson's figure of merit (JFOM) predicts device performance based on the breakdown field and maximum saturation velocity while Baliga's high frequency figure of merit (BHFFOM) bases device performance solely on switching losses due to the charging and discharging of the input capacitance of power transistors.[8] Shenai et al. developed several figures of merit incorporating thermal effects one of which is QF1.[8]</td>
<td>14</td>
</tr>
<tr>
<td>5.1 Table of trap energy levels and concentration in MBE-grown $n$-GaN for varying nitrogen rf-plasma power.</td>
<td>121</td>
</tr>
<tr>
<td>5.2 Summary of deep levels and concentrations in $n$-GaN films grown using various growth conditions. The MOCVD results were previously reported in Ref. 7.</td>
<td>134</td>
</tr>
<tr>
<td>5.3 Summary of deep levels and concentrations in ammonia-MBE $n$-GaN films. Below detection limit is signified by BDL.</td>
<td>141</td>
</tr>
<tr>
<td>5.4 Summary of deep level concentrations as a function of threading dislocation density (TDD) in MOCVD $n$-GaN. BDL signifies below detection limits.</td>
<td>155</td>
</tr>
</tbody>
</table>
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Unit cell of a wurtzite crystal structure.</td>
</tr>
<tr>
<td>2.2</td>
<td>Representation of the a-, c-, m-, and r-planes of wurtzite crystals.</td>
</tr>
<tr>
<td>2.3</td>
<td>Ball-and-stick model of wurtzite crystal structure for the Ga-face and N-face polarities of GaN.</td>
</tr>
<tr>
<td>2.4</td>
<td>Depiction of spontaneous and piezoelectric polarization in Ga- and N-face heterostructures.</td>
</tr>
<tr>
<td>2.5</td>
<td>Schematic layer structure, device geometry, charge distribution, and energy band diagram of an AlGaN/GaN HEMT.</td>
</tr>
<tr>
<td>2.6</td>
<td>Leakage current of heterojunction bipolar transistors on lateral epitaxial overgrowth (LEO) window and wing regions.</td>
</tr>
<tr>
<td>2.7</td>
<td>GaAs MESFET DC and pulsed I-V curves compared an a curve of constant power dissipation.</td>
</tr>
<tr>
<td>2.8</td>
<td>Experimental drain-current versus time waveform in response to a constant drain voltage and stepped gate voltage.</td>
</tr>
<tr>
<td>2.9</td>
<td>Evidence of drain-lag phenomenon on a 0.6 mm AlGaN/GaN HEMT.</td>
</tr>
<tr>
<td>2.10</td>
<td>Comparison of the dynamic load-lines between continuous wave RF, pulsed-IV pulsed-RF, and pulsed I-V measurements.</td>
</tr>
<tr>
<td>2.11</td>
<td>Arrhenius plot of the mean time to failure (MTTF) for AlGaN/GaN HEMTs on Si.</td>
</tr>
<tr>
<td>2.12</td>
<td>Electroluminescence images of HEMT before and after high-field stressing.</td>
</tr>
</tbody>
</table>
4.10 Theoretical and experimental results of interacting defect behavior. 96
4.11 EBIC image of \( n \)-GaN. 101
4.12 Schematic diagram of Park Systems’ atomic force microscope. 102
4.13 Tip-sample force interaction versus tip-sample separation for AFM. 103

5.1 Growth diagrams for the MOCVD and MBE \( n \)-GaN films whose defect spectra are compared. 111
5.2 DLTS spectra of the MOCVD and MBE samples. 113
5.3 Steady-state photocapacitance results for the MBE and MOCVD samples. 114
5.4 Steady-state photocapacitance spectra of rf-plasma assisted MBE-grown \( n \)-GaN Schottky diodes grown with 150, 300, and 400 W plasma powers. 120
5.5 EBIC performed on the samples grown with different nitrogen rf-plasma powers. 123
5.6 Trap concentrations plotted against rf-plasma power in an Arrhenius relationship exhibits a linear relationship. 126
5.7 DLTS spectra of the rf-plasma assisted MBE-grown \( n \)-GaN Schottky diodes grown with 150, 300, and 400 W plasma powers. 128
5.8 Normalized steady-state photocapacitance spectra of ammonia-based and N-plasma MBE \( n \)-GaN Schottky diodes. 132
5.9 DLTS spectra of \( n \)-GaN Schottky diodes grown using ammonia and rf-plasma nitrogen sources. 134
5.10 Steady-state photocapacitance spectra of ammonia-based MBE \( n \)-GaN Schottky diodes for higher and lower \( \text{NH}_3 / \text{Ga} \) flux ratios. 139
5.11 DLTS spectra of ammonia-MBE \( n \)-GaN Schottky diodes for the higher and lower \( \text{NH}_3 / \text{Ga} \) flux ratios. 141
5.12 Depiction of spontaneous and piezoelectric polarization in Ga- and N-face heterostructures. 144
5.13 Ball-and-stick model of wurtzite crystal structure for the Ga-face and N-face polarities of GaN. ......................................................... 144
5.14 Growth diagrams for the (a) N-face and (b) Ga-face $n$-GaN samples grown by plasma-assisted MBE. ......................................................... 146
5.15 Steady-state photocapacitance and DLOS spectra of rf-plasma assisted MBE-grown N-face and Ga-face $n$-GaN Schottky diodes. ......................... 148
5.16 DLTS spectra of rf-plasma assisted MBE-grown N-face and Ga-face $n$-GaN Schottky diodes. ................................................................. 150
5.17 Device structures for the MOCVD threading dislocation density study. ....... 154
5.18 I-V curves of three MOCVD $n$-GaN films grown on different GaN template substrates. ................................................................. 155
5.19 DLTS spectra for the Type “A” and LEO MOCVD $n$-GaN samples. .............. 156
5.20 DLTS experiments performed with different pulse times on the $E_{C}-0.90$ eV level.157
5.21 DLTS experiments performed with different pulse times on the $E_{C}-0.60$ eV level.158
5.22 Steady-state photocapacitance spectra of the Type “A” and LEO MOCVD-grown $n$-GaN Schottky diodes. ................................................. 159
5.23 Measured trap concentration of the $E_{C}-1.35$ eV level versus the residual carbon concentration measured by secondary ion mass spectroscopy. ................. 160
5.24 DLTS spectrum for the $n$-AlGaN Schottky diode. ................................... 165
5.25 Arrhenius plot for the $n$-AlGaN Schottky diode. .................................... 165
5.26 Arrhenius plot of the $E_{C}-0.87$ eV $n$-AlGaN level compared with similar $n$-GaN levels grown by MOCVD and MBE. .............................. 166
5.27 Steady-state photocapacitance spectrum of the thick MOCVD-grown $n$-AlGaN Schottky diode. ............................................................... 167
5.28 LCV data for the $n$-AlGaN Schottky diode. ...................................... 168
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.29</td>
<td>DLOS spectrum of the thick MOCVD-grown $n$-AlGaN Schottky diode.</td>
</tr>
<tr>
<td>5.30</td>
<td>Band diagram assuming a vacuum referred binding energy model for the cation vacancy in AlGaN.</td>
</tr>
<tr>
<td>6.1</td>
<td>Band diagram of a HEMT.</td>
</tr>
<tr>
<td>6.2</td>
<td>Model of variable gate-drain access region resistance.</td>
</tr>
<tr>
<td>6.3</td>
<td>Charge distribution and resulting band diagram for a Ga-face 30 nm 30% AlGaN/GaN heterostructure.</td>
</tr>
<tr>
<td>6.4</td>
<td>Schematic depiction of a cross section of a HEMT with the gate and drain shown.</td>
</tr>
<tr>
<td>6.5</td>
<td>Plot of $V_{DS}$ vs. $V_{GS} - V_T$ to maintain a fixed drain current.</td>
</tr>
<tr>
<td>6.6</td>
<td>Circuit diagram of a HEMT with the intrinsic device separated from the parasitic gate-drain access resistance.</td>
</tr>
<tr>
<td>6.7</td>
<td>Depiction of timing diagram and biasing schemes for the gate- and drain-controlled DLTS/DLOS measurements.</td>
</tr>
<tr>
<td>6.8</td>
<td>Circuit diagram for gate-controlled Cl$_D$-DLTS/DLOS measurements.</td>
</tr>
<tr>
<td>6.9</td>
<td>Circuit diagram for drain-controlled Cl$_D$-DLTS/DLOS measurements.</td>
</tr>
<tr>
<td>6.10</td>
<td>Variation of gate bias during a gate-controlled DLOS measurement looking at the initial bias before the opening of the shutter.</td>
</tr>
<tr>
<td>6.11</td>
<td>Block diagram of FPGA-based constant current, conductance, or capacitance measurement system.</td>
</tr>
<tr>
<td>6.12</td>
<td>The front panel of the FPGA PID Control.vi program.</td>
</tr>
<tr>
<td>6.13</td>
<td>The front panel of the FPGA Initialization.vi program used to setup constant drain current experiments.</td>
</tr>
<tr>
<td>6.14</td>
<td>Two gate controlled Cl$_D$-DLOS measurements performed in the saturation regime on an unpassivated PAMBE grown HEMT with different fill pulse conditions.</td>
</tr>
</tbody>
</table>
6.15 Gate-controlled DLOS experiment performed on an AlGaN/GaN HEMT with and without passivation. ................................................................. 223

6.16 Two gate controlled CI_D-DLOS measurement performed on an unpassivated PAMBE grown HEMT in the saturation and triode regimes. .................. 224

6.17 Initial drain-controlled CI_D-DLOS result of an unpassivated PAMBE-grown HEMT. 226

6.18 Sample structure of an unpassivated Ga-face, Ga-rich HEMT grown by PAMBE on SiC at UCSB. ................................................................. 227

6.19 Band diagram of the unpassivated 30nm, 30% AlGaN/GaN HEMT device under the gate. ................................................................. 228

6.20 Bias conditions and gate-controlled DLTS results of an unpassivated AlGaN/ GaN HEMT. ................................................................. 230

6.21 Gate-controlled DLOS spectra and pulsing/measuring conditions on the transistor I-V curve for an unpassivated AlGaN/GaN HEMT. .................. 231

6.22 Gate-controlled DLOS experiment performed on an AlGaN/GaN HEMT without passivation. ................................................................. 233

6.23 Measurement conditions and Drain-controlled DLTS results on an unpassivated AlGaN/GaN HEMT. ................................................................. 235

6.24 Drain-controlled DLOS spectra on an unpassivated AlGaN/GaN HEMT. .... 236

7.1 Typical high-frequency C-V and dC/dV curves for a metal-oxide-semiconductor structure with a p-type semiconductor. ......................................... 248

7.2 Mask set designed to explore the minimum capacitance measurable for future absolute capacitance measurements in the AFM. .................. 250

7.3 Schematic charge distribution and band diagram of an AlGaN/GaN heterostructure. ................................................................. 254

7.4 AFM image of the gate-drain access region of an AlGaN/GaN HEMT. .... 259

7.5 Biasing scheme for measuring surface Schottky barrier lowering in the access region of nitride HEMTs. ................................................................. 260

xx
7.6 Measured change in barrier height using SKPM at a fixed location using different wavelengths of light. .......................................................... 261

7.7 Measured change in barrier height using SKPM at a fixed location in the dark. . 262

7.8 Schematic band diagram of a 30 nm, 30% AlGaN/GaN heterojunction with the influence of carrier emission from a surface trap. ........................... 263

7.9 Typical surface barrier and current transients measured for an AlGaN (35%, 202 Å)/GaN heterostructure sample before, during, and after illumination. . . . 264

7.10 AFM topography and SKPM images of the gate-drain access region under illumination from different wavelengths of light. ................................. 267

7.11 Line scans of the dark and 3.80 eV SKPM images. ................................. 269

A.1 Process flow to fabricate N-face n-GaN Schottky contacts. .......................... 295

B.1 Example of Cheung method of extracting series resistance from a diode. .... 310

B.2 Thin surface barrier model band diagram. ............................................ 314

B.3 Current-voltage circuit connection diagram. ........................................... 318

B.4 Screenshot of the I-V program used for the DLOS systems. ....................... 320

B.5 I-V analysis program screenshot. ......................................................... 322

B.6 Screenshot of I-V program for use with the Keithley 2400 SourceMeter. .... 323

B.7 Screenshot of I-V-T program for use with the HP 4145B parametric analyzer. 324

B.8 C-V circuit diagram. ................................................................. 326

B.9 Screenshot of the C-V program....................................................... 327

B.10 Screenshot of the doping calculator program. ........................................ 329

B.11 DLTS circuit diagram. ............................................................... 331
B.12 DLTS program screenshot. .................................................. 332
B.13 DLTS analysis program. ..................................................... 336
B.14 Capacitance-temperature circuit diagram. .......................... 338
B.15 Capacitance-temperature program screenshot. ..................... 339
B.16 DLOS circuit diagram. ...................................................... 340
B.17 Schematic of DLOS optical setup. ..................................... 341
B.18 Initialization.vi screenshot. ............................................... 342
B.19 DLOS program screenshot. ............................................... 344
B.20 Screenshot of the DLOS analysis program. ......................... 347
B.21 Band diagram showing the IPE process. ............................. 349
B.22 IPE results of an $n$-type AlGaN sample with semitransparent 80 Å Ni Schottky contact. ......................................................... 351
B.23 Schematic representation of optical illumination and circuit used for IPE measurements. ................................................. 352
B.24 Electrical circuit for IPE measurements. ............................. 352
B.25 Screenshot of the IPE program. ......................................... 354
B.26 Circuit diagram for photon flux measurement. ..................... 356
B.27 Photon flux program screenshot. ....................................... 357
Chapter 1

Introduction

1.1 Motivation

GaN, AlN, InN, and their alloys form a promising semiconductor material system where an intense research effort has been focused on understanding, improving, and finding application of the various Group III-N devices. Although the first research into single crystal GaN occurred in the late 1960’s,[1] it did not gain global popularity until the 1990’s when Nakamura et al. developed and commercialized the blue LED.[2] At odds with conventional III-V semiconductors, the performance of first InGaN LEDs was far in excess of what would be predicted given the number and degree of structural imperfections compared with other III-V optoelectronic devices.[3] This piqued the interest of many researchers to both understand the reason for this seeming contradiction and to exploit the unique properties of the nitride system. Since the development of the blue LED, nitrides have been explored for many new applications in both the optoelectronic and electric device arenas. The nitride bandgap can be varied from $\sim 0.7$ eV for InN to 6.2 eV for AlN theoretically allowing for bandgap tuning from infrared to visible to ultraviolet to suit the desired application. This ability makes AlGaN an ideal and unique material as both a light emitter and detector. The success of nitrides as light emitters has led to great commercial success in the lighting and display application and
may one day allow for nitride-based room lighting. A new generation of long life, low-lasing threshold, GaN-based ultraviolet laser diodes are finding applications in a new generation of high-density DVD players (Blu-Ray) where the shorter wavelength allow for increased storage density of up to 50.4 GB/in\(^2\).\[^4\] Furthermore, AlGaN-based solar blind detectors capable of detecting jet aircraft exhaust, forest fire, and artillery/missile launch are also of interest because the short-wavelength solar radiation (\(\lambda \sim 290\) nm) from the sun is absorbed in the atmosphere and not detected by the detector allowing high signal-to-noise ratio.\[^5–7\]

Beyond optoelectronic devices, many of the other nitride devices rely in some way on the large piezoelectricity present in hexagonal \(\text{[Al,In,Ga]}\)N. Because of piezoelectricity and high in-plane acoustic velocity, surface acoustic wave (SAW) devices that can be used as filters and resonators, for example, can achieve high frequencies in AlGaN-based films. \[^8, 9\] In the realm of micro-electromechanical systems (MEMS), nitrides are being explored for use as resonators, pressure sensors, ultrasonic transducers, and piezoelectric speakers.\[^9\] Nitrides generally feature high melting temperature and chemical inertness making them ideal candidates for multifunctional biological and gas sensors as well.\[^10, 11\] Finally, lattice-mismatch, piezoelectricity, and spontaneous polarization can lead to formation of two-dimensional electron gases (2DEG) with high charge densities and high mobilities. In addition, high electron saturation velocity, large breakdown fields and wide bandgap enable high frequency operation. The moderate thermal conductivity, compatibility with highly thermally conductive SiC, and high melting temperature allow for high power operation. For these reasons, nitride high electron mobility transistors (HEMT), sometimes referred to as heterojunction field effect transistors (HFET), offer the unique opportunity to provide both high power and high frequency operation beyond the range of more traditional transistor material systems such as GaAs, Si, and InP.
Despite the many advances in the nitride material system, electrically-active defects are limiting device lifetime and performance. Phenomena that generally influence the dynamic, or rf, performance of the device compared with static characteristics are known as dispersion. Dispersion includes gate lag, drain lag, knee walkout, current collapse, and transconductance dispersion can only be caused by defects and self-heating. Other effects such as device degradation and unreliable device lifetime are all suspected to be caused by defect formation and/or migration. Improved lifetime, reliability, and device performance require identification of deep levels and correlation to the various device degrading phenomena. Nitride HEMTs are targeted for various frequencies in the 4+ GHz range, but the defects suspected to influence performance are relatively slow trap capture or emission processes. Identifying the defects responsible has led to conflicting reports and mostly indirect evidence. To bridge the gap between basic science (fundamental research) and device performance, new thinking is required to develop new methods that can more directly measure defect behavior in devices to correlate with the fundamental research.

Much of the research is focused on material properties of GaN and AlGaN with HEMTs in mind, but much of the results are generic can be applied to optoelectronics and other devices as well. Optoelectronic devices, especially laser diodes, require low defect densities to achieve long life, high radiative recombination efficiency, and low lasing current density among other things. Nevertheless, defects are still pervasive in laser diodes and other devices. Their elimination or mitigation in these devices, HEMTs, and all the other applications of nitrides will be necessary for continued technological improvement. Thus, the study of deep levels is likely to remain an important aspect of nitride research for a long time to come.
1.2 Objective

It has been established that defects are responsible for device degradation in many respects including acting as a noise source, limiting device operating range because of leakage current, causing dispersion, and causing eventual device failure, so their reduction or elimination is necessary for the continued advancement of Group III-nitrides. The primary goal of the research is to characterize the deep levels in various nitride films – whether they be thick GaN or AlGaN films or operational HEMTs. Although the techniques to measure defect incorporation in thick films using diode structure is relatively well established, the same measurements on HEMTs are much less mature or non-existent in some cases. The second goal is to formulate methods to quantitatively measure both defect energy levels, concentrations, and spatial location in operating HEMTs. The motivation for this is both to provide a more direct correlation between device measurements and defect measurements by allowing both to be done on the same transistor devices but also identify the trap levels responsible for HEMT specific device degrading phenomena such as hot carrier effects and virtual gating.

1.3 Overview of the thesis

Chapter 2 details the properties of nitrides including breakdown voltage, saturation velocity, piezoelectric and spontaneous polarization, crystal orientation, HEMT specific properties such as 2DEG formation, the issue of foreign substrates, and performance degrading effects caused by defects. This provides knowledge of fundamental properties of nitrides and motivates the continued need for defect identification with an emphasis on HEMT-related issues.

In Chapter 3, the growth and processing of the samples used in this research is described. A large percentage of the growth space is covered in this research including metal-organic
chemical vapor deposition (MOCVD), plasma-assisted molecular beam epitaxy (PAMBE), and ammonia-source molecular beam epitaxy (ammonia-MBE). The samples also range from thick GaN and thick AlGaN Schottky diode structures to fully operational AlGaN/GaN HEMTs. The HEMTs were processed by collaborators and only an overview of the processing is given, but the Schottky diode processing developed is explained in great detail including the mask layout for the latest generation mask set and with the full process follower given in Appendix A.

Chapter 4 covers a range of experimental techniques used in the course of research. Generally, these are conventional techniques that were developed elsewhere. Before presenting the derivation of DLTS and deep level optical spectroscopy (DLOS), the physics of carrier emission and capture are presented along with the properties of diodes including derivation of capacitance, steady-state occupancy of deep levels, and the dynamic time-response of deep levels after changes in voltage or light incidence. DLOS and DLTS are quite similar in terms of measurement setup and operation, but the analysis of the transient data is quite different for both techniques. The next section describes the nuances of DLOS and steady-state photocapacitance including models for fitting the optical cross section. The thermal emission rate depends exponentially on temperature unlike the optical emission rate that depends on the incident photon flux and the optical cross-section. This feature of DLTS allows for several automated techniques to extract the defect properties such as thermal cross section, concentration, and energy in the bandgap. The double boxcar, Fourier transform, and Padé-Laplace methods of analysis are derived and compared. The Padé-Laplace method has been applied to other fields of study, mostly for experiments of exponential decay, because it was thought that this method could not handle baseline offsets. It is shown that the Padé-Laplace method can indeed operate with baseline offsets, but a two-step method is needed to ensure that the
baseline offset is on the order of the transient magnitude to achieve the most accurate results. The last section concerning DLTS discusses the ability to vary the fill pulse time to distinguish interacting (e.g. Cottrell cloud, defect clustering, and dislocation core related) from ideal non-interacting defects that would be randomly distributed throughout a film. This is based on Coulombic interaction between defect sites, but if the levels are extremely close they can also interact on the scale of the wavefunction. For defects with overlapping wavefunctions, defect minibands form, which can be distinguished from discrete levels.

The remainder of the chapter focuses on several ancillary techniques including capacitance voltage used to determine the carrier concentration, current-voltage-temperature that reveals details about carrier transport in Schottky diodes, internal photoemission employed to accurately measure the Schottky barrier height, secondary ion mass spectroscopy to determine background impurity concentrations, electron beam induced current used primarily to determine the threading dislocation density, x-ray diffraction to evaluate growth quality, and several scanning probe techniques such as atomic force microscopy. These techniques were primarily applied to Schottky diodes. As eluded to previously, the DLTS and DLOS techniques were designed for diodes, but diodes cannot be used to replicate all of the lateral electric fields, complicated growth structures, interfaces, and other HEMT specific issues. Several additional techniques have been developed to directly measure on HEMTs, but these are presented later in Chapters 6 and 7. Although the ideal method to measure defects specific to HEMTs is by measuring the HEMTs themselves, the number of unknown processes and defects is overwhelming and interpreting these results without a base knowledge of the constituent layers is unwise.

To develop the base knowledge of defects in GaN and AlGaN is daunting as well. Chapter 5 explores the role of growth conditions and methods in the deep level spectrum of $n$-type GaN and AlGaN films. The first section examines the effect of PAMBE or MOCVD on the
defect incorporation where these techniques are quite different in growth temperature, pressure, source material, growth limiting mechanisms, and levels of contamination, but both methods yield high quality GaN films. The largest differences in deep level spectra between the films are the large gallium vacancy (V_{Ga}) concentration in MOCVD films and the absence of the residual substitutional Mg doping in PAMBE films. It is also shown that MOCVD films tend to have higher concentrations of mid-gap and near valence band states and lower concentrations of near conduction band states. In PAMBE growth, the active nitrogen is provided by an rf-plasma, which can emit a range of species including metastable nitrogen, ionic nitrogen and atomic nitrogen. The kinetic energy of the species varies, but in some cases is sufficient to induce point defects or cause GaN decomposition. Therefore, the impact of the nitrogen plasma power’s effect on the defect incorporation was studied in PAMBE Ga-face Ga-rich n-type GaN. It is also possible to grow GaN in an MBE using ammonia as the nitrogen source. From PAMBE to ammonia-MBE, the myriad of nitrogen species are replaced with a hydrogen-rich molecule and the Ga-rich growth environment is replaced with an N-rich one, so differences in the deep level incorporation are expected. Although the ammonia-MBE sample had higher concentrations, which was not unexpected as it has not reached the maturity of PAMBE growth, the defect spectra were very close indicating that ammonia-MBE has great potential. Similar to the nitrogen plasma power study, the ammonia flow rate or Group III/V beam flux ratio of ammonia-MBE growth was also explored to investigate its role in defect incorporation. A relatively systematic increase in defect incorporation was observed with the major exceptions being the gallium vacancy-related level that did not depend on the ammonia flow rate and the suspected nitrogen vacancy-related level that was decreased by a factor of $\sim 6$ at the higher flow rate. Up to now, all of the samples examined were of Ga-face polarity, but N-face GaN offers several advantages
to traditional Ga-polar material. Comparison of $n$-GaN grown by PAMBE in both Ga- and N-face substrates revealed similar near conduction band levels and concentrations for the gallium vacancy-related and substitutional carbon levels, but significantly different spectra of the near conduction and mid-gap states. Finally, the previous studies were all for thick GaN films, but HEMTs consist of both AlGaN and GaN layers. Therefore, a thick AlGaN layer grown by MOCVD was also examined. Using a vacuum referred binding energy model and previous work, the level at $E_C - 3.10$ eV is suspected to be the cation vacancy. As expected, a large $E_C - 3.10$ eV concentration was measured similar to the cation vacancy in MOCVD $n$-GaN. A large concentration of a defect at $E_C - 3.93$ eV was also observed. SIMS did not rule out the possibility of this level being attributed to carbon or magnesium. Similar to MOCVD $n$-GaN, the concentration of levels in the near conduction band region were very low. The only observed level via DLTS was at $E_C - 0.90$ eV and the concentration was $2 \times 10^{13}$ cm$^{-3}$. This survey of the component layers revealed a range of defects and sources that one might expect to see in HEMT devices. Levels that do not correspond to one of these levels might be due to an interface, surface, or induced by the large vertical and lateral electric fields.

Chapters 6 and 7 endeavor to address the shortcomings of correlating HEMT phenomena with bulk deep level measurements by measuring directly on operational HEMTs by adapting the DLTS and DLOS techniques and developing an AFM-based method to locally measure defect concentrations, respectively. Although some work has been done to adapt DLTS to transistors, specifically HEMTs, the ability to do so in a quantitative way that allows comparison between samples and to extract defect concentrations has not been attempted. Varying the gate-source voltage or drain-source voltage to maintain a constant drain current and carefully selecting the transistor operating point was found to provide unique opportunities to isolate defects in the access region and under the gate. In gate control with the bias point
in saturation, changes in access resistance due to deep level emission or capture impact
the voltage at the drain side of the gate edge, but as the device is in saturation this has little
impact on the drain current assuming the current is flat in the saturation regime. In contrast,
any deep level emission or capture under the gate can be treated as a shift in the threshold. In
this case, the change in gate voltage required to keep the drain current constant corresponds
to the shift in threshold voltage. The voltage shift can be correlated with defect concentra-
tions, but to determine the exact concentration knowledge of the defect distribution in the
device is required. Drain-voltage-controlled constant drain-current measurements are also
possible, but clearly this is only possible in the triode regime. In this case, a simple model is
developed separating the parasitic drain-gate access resistance from the intrinsic transistor.
This model reveals that in a constant current mode, the drain voltage of the intrinsic tran-
sistor is constant while the voltage drop across the access region varies with the deep level
occupation. Modeling the lateral depletion region caused by the high lateral fields allows
quantitative defect concentration determination. Beyond constant current measurements, it
is also possible to perform constant transconductance measurements. Similar relationships
and measurement schemes are determined. The second part of the chapter shows several
test runs, which begin with sanity checks to verify the experiment works as expected while
the latter part looks at the role of passivation using constant drain current measurements.

To perform local measurements of defect concentrations, modifications to a commercially
available atomic force microscope, or more generically scanning probe microscope, are
proposed in Chapter 7. This multi-part problem involved identifying lateral non-uniformities
in the access region, performing deep level characterization in these areas and other control
regions, and then watching how the deep level spectra evolves as the device is stressed. This
technique is still under development and many questions still remain, but initial attempts
have been promising. Defective regions could be identified by looking for areas of differing potential or hot spots or light emission during device operation. So far, only potential variation has been tracked. Identifying defects can be performed in multiple ways – looking at a single spot to record transient behavior or scanning the sample under steady-state illumination, for example. Proposed methods of recording the changes include capacitance using scanning capacitance microscopy and potential using scanning Kelvin probe microscopy. A model is developed to correlate changes in potential with corresponding changes in deep level occupation. Finally, some initial results are shown lending credence to this method as a valuable defect identification tool.

Chapter 8 concludes the results of the previous chapters and expresses some ideas for future work. Most notably, the preliminary results of the constant current/transconductance trap spectroscopy and AFM-based optical trap spectroscopy techniques needs continued development in both theory and modeling as well as improved understanding of the experiment such as the best conditions to use to fully exploit the techniques, evaluate the spatial resolution, and the impact of device geometry on measured defect concentrations.
1.4 References


Chapter 2

Background

The \{Al,Ga,In\}N material system is promising for many applications including high frequency, high power HEMTs in addition to solar blind UV detectors, UV LEDs, UV laser diodes, biosensors, pressure sensors, and MEMS.[1–7] Focusing on the properties that give nitrides a huge advantage in HEMT applications, a number of material metrics combine to give GaN-based devices great potential. Table 2.1 summarizes several figures of merit. Johnson's, Baliga's high frequency, Shenai's figures of merit all predict GaN-based devices can achieve higher performance than Si, GaAs, and SiC. These figures of merit are based on the bulk or two-dimensional electron gas (2DEG) mobility in the case of nitrides, critical electric

Table 2.1: Figure of merits from Ref. 8 for various competing material systems. Johnson's figure of merit (JFOM) predicts device performance based on the breakdown field and maximum saturation velocity while Baliga's high frequency figure of merit (BHFFOM) bases device performance solely on switching losses due to the charging and discharging of the input capacitance of power transistors.[8] Shenai et al. developed several figures of merit incorporating thermal effects one of which is QF1.[8]

<table>
<thead>
<tr>
<th>Material</th>
<th>JFOM</th>
<th>BHFFOM</th>
<th>QF1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Si</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>GaAs</td>
<td>11</td>
<td>16</td>
<td>9.4</td>
</tr>
<tr>
<td>4H-SiC</td>
<td>410</td>
<td>34</td>
<td>950</td>
</tr>
<tr>
<td>6H-SiC</td>
<td>260</td>
<td>13</td>
<td>300</td>
</tr>
<tr>
<td>GaN</td>
<td>790</td>
<td>100</td>
<td>910</td>
</tr>
</tbody>
</table>
field, and thermal conductivity. In addition, GaN’s high melting temperature, chemical inertness, and radiation hardness are also quite advantageous in many applications. Phosphides, arsenides, antimonides, and Nitrides are all Group III-V material systems, but unlike the other material systems are only zinc-blende structures are possible, nitrides can be grown in both cubic zinc-blende and hexagonally close packed wurtzite forms. The strong bonding of Group III to N atoms, the lack of crystal symmetry, and the ability to use as many as 3 Group III elements during growth allows for the formation of high mobility, undoped two-dimensional electron gases (2DEG). The implications of this are developed from the crystal structure to polarization and 2DEG formation to device issues.

2.1 Crystal structure

The wurtzite structure is two interpenetrating hexagonally close packed (HCP) sublattices – one sublattice with Group III (e.g. gallium, indium, or aluminum atoms) and the other with Group V nitrogen atoms. One sublattice is offset by 5/8 of the cell height in the c-direction relative to the other sublattice. A diagram of the wurtzite structure is shown in Fig. 2.1. Wurtzite has many common crystal planes some of which are shown in Fig. 2.2. The a-, c-, m-, and r-planes of wurtzite crystal structures exhibit quite different behavior in terms of polarization, mixed versus unmixed surface terminations, and atomic arrangements in the respective planes. Although the other planes would be interesting to study, the research so far has focused only on c-plane nitrides. The polarity and polarization of the c-plane will be explored further.

From Fig. 2.1, it is evident that the wurtzite crystal structure is not symmetric in the [0001] and [000\bar{1}] directions. These orientations are termed Ga- and N-face, respectively, and are shown more clearly in Fig. 2.3. No rotation about the c-direction or addition or removal
Figure 2.1: The dashed lines shows the outline of a unit cell of wurtzite crystal structure. A wurtzite crystal consists of two interdependent hexagonal close packed sublattices displaced by 5/8 on the c(0001) axis. From Ref. 13.

Figure 2.2: Representation of the a-, c-, m-, and r-planes of wurtzite crystals such as GaN, AlN, or InN. From Ref. 12.
of atomic layers will result in identical atomic arrangement for these two directions. This asymmetry impacts the polarization fields and the chemical reactivity of the various faces. The Ga-face surface is very nonreactive in both acids and bases while the N-face surface etches in bases. N-face is a $180^\circ$ rotation from Ga-face, so the direction of spontaneous and piezoelectric polarization is rotated as well. In addition to the crystal structure, Ga, Al, and In form a partially ionic, partially covalent bond to N and together these effects leads to the relatively strong piezoelectric and spontaneous polarization effects, which are a pivotal feature of nitride heterostructure devices.[15]

## 2.2 Polarization

The spontaneous polarization $P_{sp}$, which is the result of a nonzero volume dipole moment in the crystal in the absence of any external influence such as strain or an applied electric field, exists in wurtzite nitrides as a consequence of the lower symmetry of the wurtzite crystal.
The spontaneous polarization occurs in the (0001) direction with a magnitude of -0.029 C/m$^2$.

The noncentrosymmetric nature of both wurtzite and even zinc-blende crystal structures in some directions results in a nonzero piezoelectric moduli. The piezoelectric polarization $P_{pz}$ is the product of the piezoelectric modulus $e_{ij}$ and the stress $\sigma_{ij}$ where both are tensors that can be reduced due to array of the matrices following Voigt’s rules. For the wurtzite structure, the $P_{pz}$ in the [0001] direction is

$$P_{pz} = 2\varepsilon_1 \left( e_{31} - \frac{c_{13}}{c_{33}} e_{33} \right)$$  \hspace{1cm} (2.1)$$

where the $c_{ij}$ are the elastic coefficients for the respective directions, and the strain is applied biaxially (0001) plane for lattice mismatched growth such as AlGaN on GaN. Ref. 17 summarizes the various coefficients needed to calculate the piezoelectric polarization. For the structures studied here, the film is unconstrained in the c-direction and biaxially strained in the c-plane, so the strain is calculated as

$$\varepsilon_1 = \frac{a_{\text{actual}} - a_{\text{unstrained}}}{a_{\text{unstrained}}}$$  \hspace{1cm} (2.2)$$

where $a$ is the in-plane lattice constant (i.e along the $<0010>$ directions).

If spatial variation of the spontaneous or piezoelectric polarization field exists, a polarization-induced electrostatic charge density $\rho_{pol}$ will exist

$$\rho_{pol} = -\nabla \cdot P = -\nabla \cdot \left( P_{sp} + P_{pz} \right)$$  \hspace{1cm} (2.3)$$

In structurally uniform material, the only spatial non-uniformity in $P$ will occur at the surfaces, which is presumably nullified by mobile electrons that migrate to the surface or more likely by charging of surface states.
2.3 2DEG formation

The usefulness of the polarization is more evident in nitride heterostructures. Here, an AlGaN layer grown on a thick GaN layer, both with Ga-face wurtzite crystal structure, is assumed although N-face GaN on AlGaN and indium-based heterostructures are also possible. The thick GaN layer is assumed to be relaxed (i.e. unstrained) so no piezoelectric-induced charge is present in this layer. In this scenario, the polarization-induced interface and sheet charge density $\sigma/q$ would be[15]

$$\sigma/q = \left\{ P_{sp}^{GaN} - P_{sp}^{AlGaN} - 2\varepsilon_1 \left[ e_{31} - \frac{c_{31}}{c_{33}} e_{33} \right] \right\}$$

(2.4)

This simple model yields an upper limit because charge can be screened out by interface or surface charges and because the polarization is found to influence the Schottky barrier height in HEMTs.[17] Calculating the polarizations for Ga-face AlGaN/GaN and N-face GaN/AlGaN structures are shown in Fig. 2.4. Continuing on with the Ga-face AlGaN/GaN heterostructure, the charge at the surface is accounted for by surface charge, mobile electrons, and by charge movement in the Schottky metal. Balancing the charge, the band diagram in Fig. 2.5 can be determined. At the AlGaN/GaN interface, a two-dimensional electron gas (2DEG) forms.
to compensate the polarization charge. The 2DEG has a much higher mobility than the bulk electron mobility and the 2DEG sheet concentration is typically $5 \times 10^{12} - 1.5 \times 10^{13} \text{ cm}^{-2}$ depending on the AlGaN thickness and Al mole fraction. The 2DEG can also be modulated by the Schottky metal contact above it in Fig. 2.5 allowing a high electron mobility transistor to be created by adding source and drain Ohmic contacts on either side of the Schottky gate metal. The advantages of nitride HEMTs are many, but they allow nitride HEMTs to be targeted for high power, high frequency applications.

### 2.4 Defects and HEMT specific problems

As a developing technology, nitride HEMTs have had success in some markets, but they have not yet achieved their full potential a number of reasons. Not only do nitrides need to be grown on foreign substrates due to the lack of a economically viable free-standing GaN substrate, which leads to increased threading dislocation density among other things, but they also have electrically active defects, high fields, and power dissipation in HEMTs.
lead to dispersion, lower power output, reduced max operating frequency, and reliability problems. The two major causes of problems in HEMTs are the result of increased junction temperature and/or defects. The temperature management requires improved package, layer, and substrate design. Electrically active defects can form during growth or during operation in response to high fields, high junction temperature, or strain and can be mitigated with improved growth and/or device design depending on the defect generation source.

A recurring problem with growth of nitrides has been the lack of a native substrate for growth. The common substrates are sapphire, SiC, and Si where the latter two are more commonly used for HEMT devices. These substrates do not even have wurtzitic structure, so growth and optimization on these substrates has been a long standing area of research. In this time, near-native substrates have been achieved through growth of thick GaN films, typically grown by MOCVD or HVPE, on a foreign substrate, which is then removed using laser ablation, removal of a sacrificial layer, or other mechanism. These near-native substrates have low threading dislocation densities in the $\sim 10^6$ cm$^{-2}$, but their cost is such that it is not practical to use them. Arguably, the most common nitride HEMT substrate is SiC with both 4H and 6H polytypes being used. These crystal structures have complicated stacking orders and is 33% hexagonal for the 6H polytype and 50% for 4H.[19] The lattice mismatch of SiC is 3.4% for 6H and 3.77% for 4H polytypes,[20, 21] which is quite large and results in threading dislocation concentrations of $10^9 - 10^{10}$ cm$^{-2}$. Despite the high threading dislocation densities (TDD), the thermal conductivity (4.4-4.9 W/cm/K) is the highest of the common substrates and $\sim$ 3 times larger than GaN ($\sim 1.3$ W/cm/K).[18, 22, 23] Due to the high power density, heat from power dissipation in the HEMT needs to be removed, which is where the high thermal conductivity is very advantageous. Nitronex has built their product line of nitride HEMT on Si (111) substrates. The thermal conductivity is only 1.5 W/cm/K but still slightly higher
than GaN. The dislocation densities as low as $\sim 5 \times 10^9$ cm$^{-2}$ have been reported for MOCVD growth of a 2.5 $\mu$m thick GaN on Si (111) even though the lattice mismatch is 17%.[24, 25] The major advantages of Si substrates are the low cost relative to the other substrates and the extremely high quality, low defect density substrates that have been developed mostly for the Si electronics industry. Finally, sapphire is also a common substrate but not usually for high power electronics due to its low thermal conductivity of $0.3$ – $0.4$ W/cm/K, which is $>10X$ lower than SiC.[22, 23] The advantage of sapphire is that it has the lowest density of threading dislocations of the substrates compared except for freestanding GaN that is currently too expensive to be viable in most markets. GaN templates using direct or patterned (lateral epitaxy overgrowth) growth of GaN on sapphire are commercially available with dislocation densities ranging from $8 \times 10^7$ - $5 \times 10^8$ cm$^{-2}$ from Lumilog. Much research has focused on nucleation conditions and procedures to optimize the material quality, so the GaN templates allow researchers to focus on overgrowth on a stable commercial platform without worrying about optimizing nucleation conditions or variations from nucleation.[12, 26, 27] The majority of the research reported here has been on sapphire as Schottky diodes on thick GaN layers do not need high thermal conductivity substrates and the low TDD ($\sim 10^8$ cm$^{-2}$) often improves leakage current.

Evidence of the leakage caused by dislocations is evident in Fig. 2.6 where the threading dislocation-mediated carrier transport in a heterojunction bipolar transistors (HBT) causes a pronounced increase in leakage current over the window region where the TDD concentration is significantly higher than the wing region.[28] Further evidence of the role of dislocations is found in several scanning probe and transmission electron microscopy experiments. Hsu et al. used scanning Kelvin force, scanning capacitance, and conductive atomic force microscopies to suggest that screw dislocations are the primary source of leakage.
current in Ga-rich MBE-grown films while mixed and pure edge dislocations are negatively charged.[29–31] Cherns et al. determined that edge dislocations were negatively charged in \( n \)-type MOCVD-grown GaN and positively charged in \( p \)-type material using electron holography.[32–34] Charges as high as \( \sim 1 \) electron per unit cell in the c-direction were measured in MOCVD \( n \)-GaN,[35] which brings up the next issue – electrically active defects. For a dislocation density of \( 10^8 \) cm\(^{-2} \), a lattice constant of 5.18 Å in the c-direction, and assuming a charge of 1 electron/unit cell, the volume charge density is \( \sim 2 \times 10^{15} \) cm\(^{-3} \). Previous work has identified some defects potentially decorating dislocations with defect energy levels in the bandgap,[36] which could be responsible for the observed charge. Dislocations, however, are not the only source of defects in GaN as evidenced by the hydrogenation study by Hierro et al.[37]

Several defect-related photoluminescence (PL) bands of yellow, green, and blue wavelengths are also commonly observed in \( n \)-type GaN with yellow luminescence (YL) typically
the most dominant. The suspected source of this YL band at \( \sim 2.2 \) eV is a deep carbon level or gallium vacancy.\(^{[38]}\) Both C and \( V_{\text{Ga}} \) are predicted to form deep levels with energy consistent with the \( 2.2 \sim 2.3 \) eV donor-acceptor pair emission observed in PL.\(^{[39–41]}\) The YL intensity was found to increase with carbon incorporation even when the gallium vacancy concentration is decreasing.\(^{[42, 43]}\) Contradicting these results, positron annihilation spectroscopy results for \( n \)-type GaN have shown that the YL intensity tracks the \( V_{\text{Ga}} \) concentration.\(^{[44]}\) Armstrong \textit{et al.} found a similar defect formation for both C-doped films and \( n \)-type films with \( V_{\text{Ga}} \) where the defect energies were similar but with divergent Frank-Condon energies related to the lattice relaxation related to each defect.\(^{[45]}\) Thus, two mechanisms for YL were identified. In semi-insulating films doped with carbon and \( n \)-type films with gallium vacancies could both lead to 2.2-2.3 eV luminescence.

Carbon, however, is still important as it readily compensates the doping by sitting on a gallium site for \( p \)-type material and nitrogen site for \( n \)-type material.\(^{[41]}\) As carbon is a common background impurity and amphoteric dopant, the \( n \)-type films typically studied exhibit a compensating deep level at \( E_C \sim 3.28 \) eV, which has been identified as the \( C_N \) substitutional defect. Carbon is also suspected to cause the defect at \( E_C \sim 1.35 \) eV due to a split-carbon interstitial defect. Theses defects and many others have been correlated with physical sources, but how the defects behave to degrade different aspects of device performance is an open question. There is, however, no shortage of performance degradations in HEMTs to explore.

Several issues can arise in HEMT devices that can effect device performance and operation. Many of these problems dictate device design considerations such as field plating to mitigate the problems, but this is done at the expense of reduced max operating frequency. These defects can be categorized as DC or RF according to which frequency range where they are primarily seen. Gate and buffer leakage (DC phenomena) reduce the operating
voltage range by limiting the gate voltage for gate leakage and the drain-source voltage for buffer leakage. The gate leakage results from Schottky interface issues or poor insulators in the case of MIS structures. This is a reoccurring problem and is commonly addressed using an insulator layer or surface treatment such as fluorine plasma treatment. The buffer leakage problem is less of a problem now because of the Fe and C doping of the buffer in MOCVD and MBE, respectively. The use of a large metal atom or amphoteric dopant may introduce unwanted deep levels in the device yielding improved buffer leakage but increased RF dispersion, for example. This is avoided by separating the doped semi-insulating from the 2DEG. The more difficult issues to eliminate remain the RF ones.

Several deviations from the expected RF performance of HEMTs exist and are generally termed as dispersion. The dispersive effects can only be caused by defects and self-heating, so although the focus here is on defect-related phenomena self-heating should not be dismissed. It is possible to separate the effects of self-heating and traps in a comparison of the pulsed and DC I-V curves by considering a line of constant power as in Fig. 2.7. The quiescent bias point for the pulsed I-V curves is chosen to overlap a DC operating point. In the absence of trapping, the DC and pulsed I-V curves would all cross at the line of constant power dissipation determined by the quiescent bias point, labeled Q in Fig. 2.7. Further considering the points P, P', R, and R', it is evident that DC and pulsed I-V curves do not intersect at the point of constant power dissipation as expected. One can therefore conclude that self-heating is not the only source of dispersion and trapping must be influencing the I-V measurements.

According to Morkoč, the most common and pronounced effect in HEMTs is gate lag. Gate lag is used to describe the slow transient response of the drain current and typically refers to a change in gate voltage to more positive values with a corresponding slow increase in drain current. Because the drain current increases with time, self-heating cannot be the
source as increased temperature results in increased electron-phonon scattering that results in decreased current flow.\[47\] Therefore variations in the gate potential must cause trapping/detrapping of defects under the gate or in the access regions, but specific defects have yet to be identified.\[47\] Gate lag can reach such levels that the drain current is reduced to 10% of its nominal value.\[48\] An example of the temporal dependence of the drain current on the gate switching is shown in Fig. 2.8. The gate lag exhibits a bias dependence evidenced by the differences in response in Fig. 2.8a and b where it is clear that multiple onsets are visible indicative of multiple traps contributing to the drain current response. The increased gate lag at high biases is consistent with the concept of defects in the access region where at higher biases this “virtual gate” extends further into the access region. Trapped charge modulation near the surface in the gate-drain access region has been observed as far as 0.5 – 1 µm from the gate edge using scanning Kelvin probe microscopy.\[47, 50\] This leakage of electrons from the gate contact to the AlGaN surface is commonly referred to as virtual
Figure 2.8: Experimental drain-current versus time waveform in response to a constant drain voltage of 10 V and gate-source voltage stepped from (a) -7 to 0 V and (b) -12 to -4 V. The HEMT was in series with a 50 Ω resistor attached to the drain terminal. This HEMT exhibits varying degrees of gate-induced lag due to trapped charge. From Ref. 49.
gating or gate extension and occurs due in part to the extremely large electric fields at the gate edge.[51]

Similar to gate lag, drain lag also exists where changes in drain voltage also result in slow transient response of the drain current occurring over seconds in some cases.[48] An example of drain lag is shown in Fig. 2.9 where again multiple onsets in the response indicate more than one defect is involved. Again, the observed effect cannot be due to self-heating as the opposite change in current would be expected. Drain lag is thought to occur due to traps located beneath the channel causing self-backgating.[52] This effect along with the prevalent theory of gate lag caused by gate extension occur in the gate-drain access region, making this area highly scrutinized by researchers.

Another common HEMT problem is knee walkout. When comparing the pulsed I-V or RF voltage swing to the DC I-V curves, the transition between the triode and saturation regions, or knee voltage, can shift to higher drain-source voltages.[53] Knee walkout causes clipping of the RF signal and hence distortion of the output signal along with reduced power output.

Figure 2.9: Evidence of drain-lag phenomenon on a 0.6 mm AlGaN/GaN HEMT. $V_{GS}$ is kept at a fixed value equal to -6 V whereas $V_{DS}$ is pulsed from 20 to 30 V. Self-heating effects are negligible in this case. From Ref. 52.
Figure 2.10: Comparison of the dynamic load-lines between continuous wave RF, pulsed-IV pulsed-RF, and pulsed-I-V measurements at a constant substrate temperature of 30°C. The pulsed I-V curves are pulsed from the $V_{DS}$ and $V_{GS}$ voltages indicated on the graph to $V_{GS} = 0$ V, which was the peak value of the RF $V_{GS}$. Knee walkout is evident by the $V_{DS}$ swing of the RF load lines at the high current end shifting to higher voltages as the $V_{DS}$ bias point increases. It is also evident by the pulsed I-V measurements where the knee voltage shifts to the right at higher $V_{DS}$ values. From Ref. 53.

Figure 2.10 shows pulsed I-V curves pulsed from several different values of $V_{DS}$ in pinch-off. Most evident is a current collapse that occurs for larger $V_{DS}$ values, but the knee voltage also shifts out to larger $V_{DS}$. Because there is minimal power dissipation in pinch-off or in the short time when the pulsed I-V measurement takes place, this effect is trap related. In a related vein, current collapse occurs after a transition from pinch-off to on where the drain current under pulsed conditions fails to reach the DC drain current value because of trapping effects thought to be at the gate edges.[54, 55]

Obviously, the most detrimental impact of defects is device failure where failure can range from catastrophic failure to falling out of device specifications such as a large threshold voltage shift. Premature device failure and resulting unreliability problems can be reduced
Figure 2.11: Arrhenius plot of the mean time to failure (MTTF) for AlGaN/GaN HEMTs on Si showing a 1.7 eV activation energy and an extracted MTTF of $10^7$ hours at a normal operating junction temperature of 150$^\circ$C. From Ref. 57.

through improved device design and defect reduction.[56] Several types of failure and pre-failure phenomena have been observed, but in many cases the preexistent condition that led to the problem cannot be identified. One example of reliability improvement is the effort of Singhal et al. that identified the gate sinking into a Schottky interfacial layer upon device stress.[57] Improving the process by including a gate anneal improved device drift and accelerated lifetime tests showed an expected mean time to failure of $10^7$ hours at normal operating temperatures in Fig. 2.11. Beyond junction temperature, Coffie et al. found a drain-source bias dependence on the power output degradation in HEMTs as well.[58] Others have focused on RF overdrive and found that the gate-drain breakdown voltage and gate burnout could both accelerate failure.[59]

Failure, however, does not occur uniformly across the gate. Researchers have commonly observed “hot spots” or areas along the gate of increased temperature during the off-state or
Figure 2.12 shows electroluminescence (EL) images of a U-shaped HEMT in an off-state with high reverse gate bias and the source and drain shorted to ground. Interestingly, the larger reverse bias does not lead to enlargening of the previous spots but creation of new leakage spots. Similar results were obtained by Inoue et al. [60] The “inverse piezoelectric effect” where the external electric field induces additional piezoelectric-induced strain is one possible source of this effect. [60]

Although large gate bias can result in increased points of leakage, stressing the device in the off-state for extended periods also has an interesting result. Figure 2.13 shows the EL image of an unpassivated device before stressing and after stressing for 5 hr at $V_{DS} = 27$ V and $V_{GS} = -8$ V (pinched-off). It is evident that the EL intensity is dramatically and nearly uniformly decreased after stressing, which was attributed to electron trapping or trap generation on the AlGaN surface and possibly due to hot electron effects. [60]

Hot carrier induced degradation has been indicated as a possible primary cause of degradation in AlGaN/GaN HEMTs by comparing the time dependence of the power degradation. [58] Indeed, strong electron-phonon coupling in nitrides due to their ionic nature leads to efficient phonon generation, which leads to hot phonon generation. Morkoç suggests that these hot phonons with temperatures approaching 2000 K are capable of generating point defects particularly in the presence of existing point defects formed during growth and high local fields present in HEMTs. [47]

The passivation and surface treatments themselves are also a potential source of degradation as SiN$_x$ has been shown to potentially deteriorate with time. [47] Surface passivation has been an integral part of reducing dispersion increasing output power density. [61] but the long term stability is not as well understood. Although SiN$_x$ is probably the most popular passivation material, many other compounds are being evaluated including SiO$_2$, Al$_2$O$_3$,
Figure 2.12: (a) EL image (EL intensity in false colors) of a device biased at $V_G = -30$ V, $V_D = V_S = 0$ V after 100 s at this bias. Only three spots are present. (b) EL image of the same device subsequently biased to $V_G = -35$ V with $V_D$ and $V_S$ again at 0 V. The spot density has increased by ~4X and generally the previously observed spots have increased in intensity. Note that $V_G$ in the figure was $V_D$ in Ref. 60 but is corrected here. After Ref. 60.
Figure 2.13: (a) EL image (EL intensity in false colors) of an unpassivated device biased at $V_{DS} = 20 \text{ V}$, $V_{GS} = 0 \text{ V}$. (b) EL image taken in the same electrical conditions after 5 hr off-state test at $V_{DS} = 27 \text{ V}$, $V_{GS} = -8 \text{ V}$. From Ref. 60.

HfO$_2$, MgO, and Sc$_2$O$_3$. Generally, the application of passivation generally results in a slight increase in the channel sheet charge density, but whether this is due to charge in the passivation or is interface related is not well understood. Surface preparation is also quite important as part of the passivation process and simple solvent cleans, descums, and acid dips can drastically influence RF performance. Additionally, Morkoç raises an interesting issue that the hydrogen used in the passivation deposition process as a carrier gas or in a precursor may diffuse into the GaN or AlGaN to passivate deep levels in the underlying epitaxial layers, but hydrogen, being relatively mobile in many cases, could result in only a temporary improvement of the device performance as elevated operating temperatures would only assist in removing passivating hydrogen.

The source of these failures is an open question with many suspected culprits. Figure 2.14 shows potential mechanisms contributing to HEMT failure and the regions of the device where this is from a more device perspective. From a more materials aspect, the role of point defects, point defect formation, the high dislocation densities, interface and surface-related defects, and the sources of the defects is also important. What is clear though is without
Figure 2.14: Possible mechanisms contributing to HEMT degradation and failure are represented. From Ref. 47.

bridging the gap between materials and device testing the whole picture of what is going on and the causes of it cannot be fully understood.
2.5 References


Chapter 3

Growth and processing of nitride devices

The devices used in this research were grown by MOCVD, plasma-assisted MBE, and ammonia-MBE to investigate the role of growth conditions in defect incorporation. Schottky diodes provided a straightforward platform for conducting well-established capacitance-based measurements while HEMT devices afforded the opportunity to explore HEMT-specific defect mechanisms using constant current and conductance-mode measurements. This chapter first surveys the growth of these devices followed by device processing and typical growth structures used in this research. The HEMT devices were processed at UCSB and MIT using optical and e-beam lithography, but this will not be discussed to the same depth as the Schottky diodes processed at OSU. Because many growth methods and variations were used in this research, the common parameters and a brief introduction to each method or mode is discussed with specific details of growth temperatures, flow rates, and other growth parameters given in the relevant sections.

3.1 MBE growth of nitrides

Molecular beam epitaxy is an epitaxial growth method performed in ultra high vacuum using molecular beams from gas and solid sources to synthesize many semiconductor materials. An effusion (or Knudsen) cell is used for sources with low vapor pressures such as
the Group III elements (Ga, In, and Al) where the sources are heated in crucibles to provide significant fluxes of the elemental species. Reactive nitrogen is provided using a plasma chamber to excite and/or dissociate N₂ molecules or via ammonia (NH₃) that cracks on the substrate surface to provide elemental nitrogen. A typical setup for an MBE chamber is shown in Fig. 3.1. MBE offers several advantages including in situ characterization such as reflection high energy electron diffraction (RHEED), the ability to provide very abrupt interfaces due
Figure 3.2: Growth diagram defining the Ga-flux conditions and substrate temperatures for the Ga-droplet, intermediate, and N-stable growth regimes. From Ref. 4.

to the lower growth temperatures and reduced outgassing from the sidewalls, and reduced background impurity concentrations.[2]

This research has focused primarily on growth of plasma-assisted MBE but also on ammonia-MBE and the less conventional growth of N-face GaN. The plasma-assisted growth of GaN and AlGaN was performed in a Varian GenII MBE system with active nitrogen for growth supplied by an AppliedEpi Unibulb rf-plasma source. Elemental gallium was supplied by a conventional effusion cell. In PAMBE MBE, the Group V/III flux ratio and substrate temperature must be controlled very tightly to achieve high quality growth. Although a high quality N-rich growth regime has been found,[3] conventional PAMBE growth has been done in a Ga-rich flux environment. The conventional growth regime has been extensively studied to determine optimal V/III ratio and substrate temperature. Heying et al. developed a map of these parameters, which is shown in Fig. 3.2.[4] In this regime's N-rich growth, pitted surfaces and low mobilities were observed,[4, 5] which corresponded to large increases in the deep level concentrations.[6, 7] Smooth surfaces and the highest mobilities were found for growths
near the line between the Ga-droplet and intermediate regimes.\cite{4, 5} Gallium droplet formation locally reduces the growth rate as nitrogen cannot easily diffuse through the excess metal so special steps are required to approach the Ga-droplet regime while still minimizing Ga droplet formation. Optimization is a continuing process, but the PAMBE samples used in this research were all grown under the best known conditions of the time, which means near the cross-over point between Ga-droplet and the intermediate Ga-stable regimes and with a temperature between about 710 and 740°C. Specific growth temperatures for each sample are given in the section where each sample is discussed and analyzed.

The bulk of GaN research has been on Ga-face material and unless otherwise mentioned, the samples used in this research were Ga-face. N-face GaN is an interesting alternative to Ga-face nitrides for high frequency HEMTs because the inverted spontaneous and piezoelectric polarization direction allow for improved device layout and potentially higher performance. The N-face sample used in this work was grown in a Varian Gen-II MBE system equipped with standard effusion cells for Ga, Al, and Si along with an Applied-Epi Unibulb RF plasma source to provide active nitrogen. The substrate temperature was measured using a pyrometer calibrated against the melting point of Al. Similar to typical PAMBE growth of Ga-face GaN, the growth was performed in a Ga-rich, N-limited environment using a plasma power of 250 W, \(\text{N}_2\) flow rate of 0.3 cubic centimeter per minute at standard temperature and pressure (sccm), and a Ga beam equivalent pressure of approximately \(5.5 \times 10^{-7}\) Torr measured by an ion gauge mounted in the growth position. This resulted in a growth rate of 220 nm/h. The Ga/N flux ratio was maintained to grow near the transition between the Ga-droplet and intermediate regimes same as Ga-face samples.

Beyond PAMBE growth of nitrides is MBE growth using ammonia as the nitrogen source. Using a Veeco Gen930 MBE system, gallium was provided using a conventional effusion cell
while the ammonia source was a custom made tube with a showerhead design that allowed the ammonia flow rate to vary from 5 to 200 sccm and used 99.9999% pure ammonia further purified using a getter filter. [2] Contrary to conventional PAMBE growth, optimized ammonia-MBE growth was found to take place in an N-rich flux similar to conventional III-V growth. The nitrogen overpressure was found to suppress GaN decomposition and allowed use of growth temperatures above 800°C – well in excess of typical PAMBE growth temperatures. [2] In PAMBE growth, growing near the crossover point from Ga-droplet to intermediate growth regimes results in a very small temperature and flux window where Ga-droplet formation does not occur and inhibit growth while the N-rich enviroment of ammonia-MBE growth window is much larger. These growths resulted in no significant difference in oxygen or hydrogen background impurity levels compared with common MOCVD growth while the carbon concentration was reduced in the ammonia-MBE material. [2] Additionally, the ammonia-MBE growth provides the possibility of achieving higher growth rates, the avoidance of Ga-droplets on the growth surface, overall improved film uniformity, and reduced leakage current along threading disolcations due to the N-rich growth conditions. [2, 8, 9]

### 3.2 MOCVD growth of nitrides

Metal-organic chemical vapor deposition (MOCVD) growth of GaN and AlGaN has been a relatively minor focus of this research. Two MOCVD reactors were used for growth of the samples: one reactor at UCSB used for comparison of thick GaN grown by MOCVD and MBE and another reactor at Sandia National Labs for growth of thick AlGaN. To avoid confusion, the reactor geometry, sources, and other growth conditions are discussed in the appropriate sections. MOCVD is a chemical process where the reactants react at the sample surface to synthesis new material. Intermediate reactions in the gas stream are also possible but
undesirable. Reactants in a carrier gas flow down the reactor to the sample mounted on a susceptor. The flowing gases form a stagnant layer above the sample that the reactants must diffuse through to reach the sample. Hence, MOCVD is a mass-transport limited growth method. MOCVD growth of nitrides is typically performed at much higher temperatures than MBE with typical substrate temperatures of around $\sim 1050^\circ C$.

3.3 Device fabrication

HEMT devices were processed by MIT and UCSB for defect measurements at OSU, but the layer structures and device geometry will be briefly discussed here. The defect measurements of HEMTs are relatively new and as such a typical layer structure and device geometry is not as standardized as the Schottky diode test structures that are processed in-house. Additionally, there are many more options with HEMTs including gate length, access region spacings, and gate geometry (e.g. slant gate or t-gate). One purpose of measuring on HEMTs is to examine defect concentrations in state-of-the-art devices, which inherently implies that device geometry will change as these devices are optimized. This has to be balanced with the need to change as few variables as possible when comparing defect levels in different devices. Here comparisons are made where device geometry is nominally identical except for inclusion of various passivation layers.

The device structures for the UCSB HEMTs is shown in Fig. 3.3. Mimicking the state-of-the-art HEMTs that use SiC substrates due to the large thermal conductivity, the sample used a Si-face 6H-SiC with a nucleation layer and thick GaN buffer to attain sufficient material quality. Despite the thickness, the threading dislocation density is $\sim 10^{10} \text{ cm}^{-2}$. On the same sample, half of the sample surface was passivated with PECVD SiN$_x$ while the other sample remained unpassivated. Due to the order of processing, this resulted in a straight gate walls as
Figure 3.3: UCSB HEMT used for demonstrating the constant drain current and constant transconductance measurements. Several geometries existed on each die. One series of six HEMTs had a gate length varying from 0.5 to 1 $\mu$m with a gate-drain spacing of 2.0 $\mu$m and a gate-source spacing of 0.5 $\mu$m. Another series of six HEMTs had a fixed 0.7 $\mu$m gate length, gate-source spacing of 0.5 $\mu$m, and a gate-drain spacing that varied from 1 to 6 $\mu$m.

The general processing outline followed by Yi Pei at UCSB for these devices was:

1. Acid pretreatment for leakage reduction
2. Ohmic (Ti/Al/Ni/Au)
3. Anneal (870°C)
4. Mesa (160 nm RIE etch)
5. Gate lithography, CF$_4$ treatment, and gate metal evaporation (Ni/Au/Ni)
6. SiN passivation on half of the sample (160 nm)
7. Bond Pads

The lithography masks defined multiple device geometries but the focus was on the HEMTs with a 0.7 $\mu$m gate length, 0.5 $\mu$m gate-source spacing, and 5 $\mu$m gate-drain spacing. This set of samples was used while developing the constant drain current and constant transconductance DLTS and DLOS methods on HEMTs, and during this time the focus was on the
unpassivated sample because a large signal was expected due to large concentrations of unpassivated defects.

Continuing on with the passivation study, a second set of HEMTs from MIT was studied with both varying passivation thickness but also for different passivation materials namely hafnium oxide and alumina. These samples were grown by Nitronex on a Si (111) substrate using the MOCVD and processed by MIT. The Si substrate offers a relatively cheap substrate at the cost of reduced thermal conductivity. The layer structure for the device is shown in Fig. 3.4 where in this case the passivation is under the gate as well. On top of the Si substrate, a proprietary 1 µm transition layer is grown to achieve high-quality growth of subsequent layers. The succeeding layers consisted of a 1 µm uid GaN buffer and a 17 nm AlGaN barrier layer forming the 2DEG just below the AlGaN/GaN interface. MIT processed this sample into
Figure 3.5: Typical Schottky diode test structure used for capacitance-based DLOS, DLTS, I-V-T, and several other diode-based techniques. The low doped top layer allows for accurate detection of defect concentration down to $\sim 5 \times 10^{12}$ cm$^{-3}$ while the highly doped $\sim 3 \times 10^{18}$ cm$^{-3}$ layer provides an efficient means for lateral conduction of electrons between the diode and Ohmic contact. A series resistance between 50-100 $\Omega$ is typical for this structure. The substrate has been one of the major variables and has been changed to study the impact of dislocation density. The semitransparent Schottky diode allowed light to be incident on the depletion region formed by the Schottky contact. Typical Schottky barrier heights would be 1.10-1.13 V with an ideality factor between 1.01 and 1.08 at room temperature. Leakage current varied depending on substrate, growth technique, and crystal polarity, but in all cases the conductance was such that capacitance-based measurements could be reliably performed.

HEMTs with straight gates 3.0 $\mu$m long with 2.0 $\mu$m gate-drain and gate-source spacing. The device performance depended on the passivation. The alumina exhibited little dispersion while the hafnium oxide and unpassivated both indicated significant dispersion based on pulsed I-V measurements. The processing procedure is similar to the UCSB process except the passivation was done before the gate deposition and without the acid pretreatment.

Contrary to the developing HEMT measurements, the Schottky diode test structures are more standardized in our lab. Figure 3.5 shows the typical test structure where the common features are the Ohmic, semitransparent Schottky, low doped top layer, and the highly doped lateral conduction layer used to improve current spreading and the Ohmic contact resistance. The typical top layer thickness and Si doping concentration was 0.5 $\mu$m and $5 \times 10^{16}$ cm$^{-3}$, respectively. The combination of thickness and concentration allows the depletion region
to be contained with this layer entirely over a reasonable range of Schottky diode bias. This pinpoints any change in defect concentration, leakage current, etc. variation to this top layer. Additionally, the low doping is used in the top layer because the detection limits of DLTS and DLOS to defect concentrations is $\sim 10^{-4} N_D$, so this suggests that detection limit of these capacitance-based measurements is $\sim 5 \times 10^{12} \text{ cm}^{-3}$.

This device structure has been used for devices grown via MOCVD and MBE (plasma-assisted and ammonia nitrogen source) and was used to study different growth modes such as Ga-face, N-face, and N-rich as well as studies of threading dislocation density variation. Although there have been a myriad of variations in growth and substrates in order to study the effect of these changes, each growth and processing variable has been carefully controlled to isolate the comparisons between samples to as few variables as possible to allow for unambiguous conclusions. In each study, the exact details of each sample are discussed including substrate, layer structure, dopings, growth modes, and any other relevant information. In addition, all of the Schottky diodes devices were patterned using essentially the same mask set producing the same device area, mesa to Schottky metal distance, etc. in an attempt to keep processing constant and not influence the measurement results. The mask set designed in shown in Fig. 3.6 and is actually the third generation mask set. Although little has changed in terms of the actual Schottky diodes, several improvements have been made including the addition of a numbering scheme and a circular TLM structure for in-process characterization of the Ohmic contacts. Whereas the Ohmic contacts to GaN have not needed an anneal, contacts to thick AlGaN and AlGaN/GaN heterostructures have required annealing. With the TLMs, the contact resistance can be verified before and after the anneal to make sure the Ohmic contact is sufficiently linear and has low enough contact resistance. One unique feature of this mask set is the common Ohmic (“back”) contact, which is the blue layer in
Figure 3.6: Mask set with circular transmission line method structures for Schottky diode formation. The thick front, back, and front contact layers are all shown in inverted contrast while the mesa layer is shown in the correct contrast. For the metal layers, the colored areas represent area where metal is left after lift-off. For the mesa layer, the colored areas represent the tops of the mesas with the clear areas being etched down.

Fig. 3.6. Because only one diode is ever measured at a time, there is no need to isolate the other devices, therefore, the web of “back” contact metal provides a low resistance contact to any diode with one wire bond. The other unique feature of this mask set is the thick front pad. The optical measurements require semitransparent Schottky contacts so the photons can be absorbed in the device depletion region, so the Schottky (“front”) contact is 80 Å thick. Making reliable contact to this thin layer and not damaging it requires an additional pad, which is provided by the thick front pad. This pad occupies less than 7% of the active area so little light is blocked. The pad is square with an edge length of 75 μm so that 1 mil (∼ 25.4 μm) wire can be used to wire bond to the diodes.

The processing of the Schottky diodes used for capacitance-based DLTS and DLOS, I-V-T, IPE, EBIC, and other diode-based measurements was developed at OSU using existing processes at Wright Patterson Air Force Base (WPAFB) and UCSB. The details of this process are
given in Appendix A, but an abbreviated version is given here to convey the general procedure. Here, it is assumed the sample is a standard Ga-face $n$-GaN test structure. Heterostructures and N-face samples require additional steps and a different processing order, which is described in Appendix A. The process always begins with degreasing using an acetone/methanol/isopropyl alcohol sequence. The Schottky diode quality is the most important part of the processing sequence so this is deposited first to put the Schottky metal down on the highest quality, undamaged surface. This works because the sample does not require anneal and the highest temperatures the sample is exposed to is 270°C. Therefore, a bilayer 1813/PMGI resist process is used. The PMGI is exposed in DUV 260 nm light while the 1813 develops in standard g-line (436 nm) light in the aligner. The process requires that the 1813 first be patterned using an aligner then exposed to the DUV light to expose the PMGI. The advantage of this process is the separation of the exposures, PMGI's immunity to acetone, the ease of stripping the 1813 and PMGI in 1165 stripper in less than 2 min., and the ability of PMGI to stand to many chemicals and even dry etching in Cl$_2$-based chemistries. Although an extra step is required to separately develop the PMGI, the flexibility and robustness of the process is ideal for research environments. After developing both the 1813 and PMGI, the sample is barrel ashed for 2 min. to remove any residual organic material (i.e. photoresist), given a 30 s dip in HCl:DI [1:1] to remove any native oxide, and put quickly into the e-beam evaporator to deposit the 80 Å Ni Schottky contact. After deposition, a lift-off process using acetone and a spinner is employed to remove the excess metal. The PMGI remains ensuring that any metal redeposition does not occur in unwanted areas. The PMGI is removed later in a bucket of 1165 stripper at 90°C. The next step of the process is the mesa etch to provide a place for the back contact to reach the lateral conduction layer. PMGI etches approximately twice as fast as GaN, so three layers of PMGI are used to adequately protect the unetched areas as long as
the etch is less than 1 \( \mu m \). The 1813 resist is again used to pattern the PMGI, but this time the mask needs to be aligned to the previous front contact layer. The resist is developed as before except additional DUV exposures are needed to get through all three layers of PMGI. Before the RIE etch, the 1813 is removed because it causes veiling of the sidewalls and pillars to form in the etched region. The etch is purely RIE without any ICP power using a mixture of \( \text{Cl}_2 \) and Ar. This yields an etch rate of 65-70 nm/min., so a 600 nm etch takes approximately 9 minutes and will remove \( \sim 1.2 \, \mu m \) of PMGI. After the etch, the PMGI is removed and the back contact is next. The back and thick front contacts is approximately the same procedure as the Schottky contact. The same bilayer 1813/PMGI resist set is used. The only differences are the use of the back contact mask and a Ti/Al/Ni/Au [100:2000:200:3000 Å] metal deposition for the back contact and thick front mask and Ni/Au [200:2800 Å] for the thick front pads. After the final lift-off and strip, the devices are finished. The first test after the process is finished is an I-V test to check the quality of the Ohmic and Schottky contacts. Typical Schottky barrier heights for this standard process would be 1.11 – 1.13 V with an ideality factor between 1.01 and 1.08. A series resistance between 50-100 \( \Omega \) is commonly measured for a 300 \( \mu m \) diode.

This process has reliably produced high quality Schottky diodes even on unique, non-standard samples. The consistency of processing has been one less variable that could effect the defect measurements. The robustness of the processing has allowed measurements on multiple substrates with a range of threading dislocation density, samples grown by multiple grown modes including MOCVD and MBE, and on both Ga-face and N-face nitride material.
3.4 References


4.1 Introduction

The primary tools used throughout this research to characterize the properties of deep levels in nitrides are deep level transient spectroscopy (DLTS) and deep level optical spectroscopy (DLOS). Unless otherwise specified, this is in reference to the standard capacitance-based measurements. Using a one-sided p-n junction or Schottky diode, the properties of deep levels are determined by analyzing capacitance transients caused by the emission of trapped holes and electrons in these techniques. The first section of this chapter discusses the theory and experimental setup of DLOS. The following sections describe the original boxcar DLTS theory addressing the original transient analysis method, the use of carrier capture kinetics to separate point vs. extended defects, and experimental setup followed by two sections dedicated to Fourier transform DLTS (FTDLTS) and Padé-Laplace DLTS (PLDLTS) as companions to the original boxcar DLTS. It will be shown that these powerful techniques are critically important for the quantification of defects and not only identifies trends in defect formation and incorporation, but also aids in defect identification through careful studies. The understanding the fundamental background of DLTS and DLOS is important on its own,
but it is also the starting point for expanding the theory in Chapter 6 where the theory will be expanded to include current-based transistor measurements.

### 4.2 Schottky and $p$-$n$ diode properties

The simplest semiconductor devices that can be used for capacitance transient spectroscopies are Schottky and $p$-$n$ diodes because their junction capacitance formed by the charge separation caused by the depletion region can be modulated both by applied voltage and by change in trapped charge emission and capture. Capacitive measurements on $p$-$n$ diodes are simplified considerably by assuming an abrupt, one-sided diode – meaning that for the case considered here that the $n$ doping is much less than the $p$ doping. Ideally, the lightly doped side should be doped at least 10 times lower than the other side of the junction. With most of the space charge on one side of the junction, the DLTS experiment is essentially insensitive to the higher doped side. With this, capacitance $C$ is defined as

$$C = \left| \frac{dQ}{dV} \right| = \left| \frac{dQ}{dW} \frac{dW}{dV} \right|$$  \hspace{1cm} (4.1)$$

where $Q$ is the charge, $V$ is the voltage, and $W$ is the depletion width. However, the second term is simply a form of Poisson's equation

$$\frac{dV}{dW} = \frac{W \rho(W)}{\epsilon}$$  \hspace{1cm} (4.2)$$

where $x$ is the distance from the metallurgical junction and $\epsilon$ is the permittivity of the semiconductor. The $dQ/dW$ is simply the charge per unit length at the depletion edge or

$$dQ = A \rho(W) dW$$  \hspace{1cm} (4.3)$$

where $A$ is the device area, and $\rho(W)$ is the net fixed charge density at the depletion edge. This allows one to determine the capacitance for a given depletion depth

$$C = \left| A \rho(W) \right| \left| \frac{\epsilon}{W \rho(W)} \right| = \frac{\epsilon A}{W}$$  \hspace{1cm} (4.4)$$
This equation is the same used for a parallel plate capacitor. It is usually necessary to determine the fixed charge density (i.e. doping) to use during DLOS/DLTS analysis, and it is possible to extract this using the following relation

\[
\frac{dC}{dV} = \frac{dC}{dW} \frac{dW}{dV}
\]  
(4.5)

For the first term, \(dC/dW\) can be determined by differentiating Eq. (4.4) and the second term, \(dW/dV\), was already determined in Eq. (4.2). Filling in these relations one determines

\[
\frac{dC}{dV} = \left[ \frac{-\epsilon A}{W^2} \right] \left[ \frac{\epsilon}{W \rho(W)} \right]
\]  
(4.6)

Rearranging this equation and using Eq. (4.4) one can solve for \(q \rho(W)\), or the net fixed charge density, \(N(W)\),

\[
N(W) = \frac{\rho(W)}{q} = \frac{C^3}{q \epsilon A^2 \left( \frac{dC}{dV} \right)}
\]  
(4.7)

Equations (4.4) and (4.7) are frequently used in C-V measurements to extract the depletion depth and carrier concentration, respectively. A note of caution is required when assuming the value from Eq. (4.7) is the free-carrier concentration because what is actually measured is the fixed charge density near the depletion edge. If the trap concentration is significant and shallow enough to emit their carriers with a time constant less than 1 s, the value measured will be the doping plus or minus the trap concentration depending on the type of trap involved. In fact, Mg doping GaN produces a level 200 meV from the valence band. In the depletion region, the Mg is fully ionized while in the bulk the Mg is 1% ionized. Assuming the fixed charge density equaled the hole concentration would be grossly inaccurate. After calculating the depletion depth and fixed charge density as a function of voltage, it is possible plot \(N(W)\) vs. \(W\). With the capacitance known, it is necessary to delve into trap response in order to develop the theory for capacitance transient spectroscopy-based techniques.
4.3 Physics of carrier emission

Varying trap occupancy within the depletion region changes the fixed charge density so to maintain the charge balance with a fixed bias applied to the sample, the depletion edge has to move slightly. The small change in the depletion edge can be detected as a change in capacitance according to Eq. (4.4). This section will quantify these relations and expand upon them to include all the concepts necessary for the basic understanding of processes involved in DLTS and DLOS. The motivation of this derivation of traditional capacitance-based DLTS and DLOS measurements is discussed in depth to give background for later development of new DLTS analysis techniques and for expanding the theory for current-based measurements on HEMTs in Chapter 6. Much of this derivation and exposition follows the treatment by Blood et al.[1]

The four possible processes can occur for a deep level within the bandgap of a semiconductor to impact the occupancy of the level. They are electron capture, electron emission, hole capture, and hole emission from the trap as depicted in processes (a)-(d) in Fig. 4.1, respectively. To solve for the rate at which the density of occupied traps $n_T$ changes with time, the relationship between all of these possible events needs to be considered. The rate of electron capture is proportional to both the number of electrons in the conduction band $n$ and the number of empty trap states $p_T$. The number of trap states empty of electrons $p_T$ is equal to the total number of defect states $N_T$ minus the number of defect states filled with electrons $n_T$. Since capture always increases the electron concentration of the defect, the rate of change is always positive and given by

$$\frac{\partial n_T}{\partial t}\bigg|_{\text{capture}} = c_n p_T n$$  \hspace{1cm} (4.8)
The constant $c_n$ is known as the *electron capture coefficient* and has units of $\text{cm}^3\text{s}^{-1}$. Now, electron emission from the defect level to the conduction band is considered. Because the conduction band is mostly empty, the rate of emission is independent of the number of empty final states, and the rate is limited by the number of defects occupied by electrons.

Thus,

$$\left. \frac{\partial n_T}{\partial t} \right|_{\text{emission}} = -e_n n_T \quad (4.9)$$

In this case, $e_n$ is the *electron emission rate* and has units of $\text{s}^{-1}$. Following the same argument for the valence band,

$$\left. \frac{\partial n_T}{\partial t} \right|_{\text{capture}} = -c_p n_T p \quad (4.10)$$

and

$$\left. \frac{\partial n_T}{\partial t} \right|_{\text{emission}} = e_p p_T \quad (4.11)$$

where $c_p$ is the *hole capture coefficient* and $e_p$ is the *hole emission rate*. The change in occupancy of the defect considering all recombination-generation (R-G) processes is

$$\frac{\partial n_T}{\partial t} = \left( c_n n_T p_T - e_n n_T \right) \left( c_p p_T n_T - e_p p_T \right) \quad (4.12)$$
Considering the *principle of detailed balance* under equilibrium conditions, the rate of capture and emission to each band must equal zero. This fact can be used to determine a relationship between the capture coefficient and the emission rate for the conduction band

\[ e_n = c_n n \frac{p_T}{n_T} = c_n n \left( \frac{N_T}{n_T} - 1 \right) \]  

(4.13)

However, Fermi-Dirac statistics can be used to establish the relationship between \( n_T \) and \( N_T \) in equilibrium where

\[ \frac{n_T}{N_T} = \frac{1}{1 + \exp\left( \frac{E_F - E_T}{kT} \right)} \]  

(4.14)

Fermi level position also dictates \( n \) and \( n_i \) yielding

\[ n = n_i \exp\left( \frac{E_F - E_i}{kT} \right) \]  

(4.15)

and

\[ n_i = N_C \exp\left( \frac{E_i - E_C}{kT} \right) \]  

(4.16)

So, the relationship between \( c_n \) and \( e_n \) can be written

\[ e_n = c_n N_C \exp\left( \frac{E_T - E_C}{kT} \right) \]  

(4.17)

or for the valence band using similar arguments

\[ e_p = c_p N_V \exp\left( \frac{E_V - E_T}{kT} \right) \]  

(4.18)

Because the capture rate is defined as

\[ c_n = \sigma_n v_{th} \nu_n \]  

(4.19)

where \( \sigma_n \) is the electron capture cross-section and \( v_{th} \) is the thermal velocity. The latter is defined as

\[ v_{th} = \left( \frac{3kT}{m_n} \right)^{\frac{1}{2}} \]  

(4.20)
where $k$ is Boltzmann's constant and $m_n$ is the electron effective mass. Additionally, the effective density of states in the conduction band $N_C$ is

$$N_C = 2 \left( \frac{2\pi m_n kT}{\hbar^2} \right)^{3/2}$$

where $\hbar$ is Planck's constant. This can all be substituted into Eq. (4.17) to determine the electron emission time constant $\tau_n$ for a defect at $E_C - E_T$

$$\tau_n = \frac{\exp((E_C - E_T)/kT)}{\sigma_n \nu_{th} N_C}$$

or

$$\tau_n T^2 = \frac{\exp((E_C - E_T)/kT)}{\gamma n \sigma_n}$$

where $\gamma_n = \left( \frac{\nu_{th}}{T^{1/2}} \right) \left( \frac{N_C}{T^{3/2}} \right)$

where $\tau_n$ is simply the inverse of $e_n$. The $\gamma$ term is generated by dividing out the temperature dependencies from the $\nu_{th}$ and $N_C$ terms and multiplying the left side of the equation by $T^2$ to maintain balance. Simplifying $\gamma$ with all the constants, it becomes $3.25 \times 10^{21} (m_n/m_0)$ cm$^{-2}$s$^{-1}$K$^{-2}$ for an n-type majority carrier defect. Equation (4.23) is a fundamental equation for DLTS and is used to determine both the defect's energy within the bandgap and the thermal cross-section. To determine $\tau_n$ and the defect concentration, more knowledge of how the defect concentration effects the capacitance is needed.

Relaxing the restrictions slightly from steady-state, quasi-steady-state relationships, which are more conducive to real situations, are considered. Under these conditions, the population of carriers is still constant in time, but the requirement that this balance is due to an exact balance between each process and its inverse (e.g. electron capture and electron emission) is relaxed. These conditions are useful for determining steady-state current, recombination-generation minority carrier lifetime, and most importantly for capacitive studies the steady-state occupancy of the deep levels. Rearranging Eq. (4.12) gives

$$\frac{\partial n_T}{\partial t} = (c_n n + e_p) (N_T - n_T) - (c_p p n + e_n) n_T$$

(4.24)
Solving this differential equation results in

\[ n_T(t) = n_T(0) \exp \left( -\frac{t}{\tau} \right) + \frac{(c_n n + e_p) N_T}{c_n n + e_p + c_p p + e_n} \left[ 1 - \exp \left( -\frac{t}{\tau} \right) \right] \]  

(4.25)

where \( n_T(0) \) is initial density of filled states and \( \tau \) is defined as

\[ \tau = \frac{1}{c_n n + e_p + c_p p + e_n} \]  

(4.26)

Based on this it is a simple matter to determine the steady-state occupancy of the deep level by checking the limit as \( t \) approaches \( \infty \)

\[ n_T(t = \infty) = \frac{c_n n + e_p}{c_n n + e_p + c_p p + e_n} N_T \]  

(4.27)

This general solution is difficult to use in practice because there are so many variables that are not known at every moment is space and time. Several reasonable assumptions are possible in the case for a Schottky or p-n diode that allow for great simplification of this result. In the depletion region, \( n \) and \( p \) are negligible, so capture events can be ignored and the equation simplifies to

\[ n_T(t = \infty) = \frac{e_p}{e_p + e_n} N_T \]  

(4.28)

From this point on, the discussion will be limited to the depletion region in an \( n \)-type semiconductor to reduce the number of equations being presented. It is assumed that a n-type Schottky diode or abrupt \( p+ / n \) diode is being measured. This assumption allows for a majority of the change in depletion depth to occur on one-side of the junction. Nonetheless, this theory is just as applicable to \( p \)-type material with simple substitution of variables and change of sign in some cases. Studying an \( n \)-type depletion region with a majority carrier trap above \( E_i \), \( e_p \) is most likely negligible compared to \( e_n \). Therefore, the steady-state condition in this case is \( n_T(t = \infty) = 0 \). It is also important to note at this point that these equations hold
true for both thermal and optical processes and will be important for the theory of DLOS presented later.

Now, considering a defect initially filled with electrons. This could be the case for a DLTS trap that has recently been filled with electrons during the fill pulse or for a DLOS trap deep in the bandgap that in equilibrium is filled with electrons. In such a case, electron and hole capture are negligible within the depletion region and hole emission to the valence band is assumed negligible either because the thermal energy required for this to occur is extremely high or because the energy of the incident photons is insufficient for a transition to occur. In such a case, Eq. (4.12) reduces to

$$\frac{dn_T}{dt} = -en_T$$  \hspace{1cm} (4.29)

The solution to this equation, still assuming $e_p$ is negligible such that $n_T(t = \infty) = 0$, is

$$n_T(t) = n_T(0) \exp(-en_t) \hspace{1cm} (4.30)$$

The emission of electrons as a function of time leads to increased positive fixed charge in the depletion region. Since the net charge remains the same for a given voltage, this leads to decrease in the depletion width and an increase in the capacitance. Using the equation for capacitance and assuming a uniformly-doped n-type Schottky or abrupt, one-sided $p+/n$ junction, the following is educed from Eq. (4.4)

$$C = \frac{\epsilon A}{W} = \epsilon A \left[ \frac{qN}{2\epsilon (V_{bi} - V)} \right]^\frac{1}{2} = \epsilon A \left[ \frac{q (N_D - n_T)}{2\epsilon (V_{bi} - V)} \right]^\frac{1}{2}$$  \hspace{1cm} (4.31)

where $N_D$ is the n-type doping concentration, $V_{bi}$ is the built-in voltage or Schottky barrier height, and $V$ is the applied voltage. This can be rearranged to

$$C = C_\infty \left( 1 - \frac{n_T}{N_D} \right)^\frac{1}{2}$$  \hspace{1cm} (4.32)
where
\[ C_\infty = \varepsilon A \left[ \frac{q N_D}{2 \varepsilon (V_{bi} - V)} \right]^{\frac{1}{2}} \]  

(4.33)

When the trap concentration \( N_T \) is much less than the doping concentration \( N_D \), the equation can be further simplified recalling the binomial approximation (i.e. \((1 + x)^n \approx 1 + nx\), for \( x \) less than \( \approx 0.1 \)) results in
\[ C(t) = C_\infty \left( 1 - \frac{n_T(t)}{2N_D} \right) \]  

(4.34)

Looking at the steady-state change in capacitance from an initial occupancy to a final occupancy, this equation becomes
\[ \frac{C(\infty) - C(0)}{C_\infty} = \frac{n_T(0) - n_T(\infty)}{2N_D} \]  

(4.35)

using Eq. (4.30). Typically one assumes the level is completely filled at \( t = 0 \) such that \( n_T(0) = N_T \) and \( e_p \to 0 \) so \( n_T(\infty) = 0 \). Additionally, the total change in capacitance \( C(\infty) - C(0) \) is typically written as \( \Delta C \). Using these relations and rearranging, the more common form of this equation is arrived at where
\[ N_T = \frac{2N_D \Delta C}{C_\infty} \]  

(4.36)

This is the general form one might expect for both DLTS and DLOS measurements where the difference between the two is the use of the thermal emission rate and optical emission rate, respectively. In the next sections, these relations will be applied to DLOS and DLTS separately to further describe the theory of each.

There is an additional case in DLOS where the deep level energy is \( > E_g/2 \) in a \( n \)-type semiconductor where the deep level is typically filled with electrons in equilibrium. In this situation when the incident photon energy is sufficient to empty the electrons from the deep level to the conduction band \( (h\nu > E_C - E_T) \), there is also sufficient energy for an electron in the valence band to refill the deep level, or a hole emission from the deep level to the valence band.
band. In this case, the capture events are still negligible because it is within the depletion
region, but now both emission processes are possible and the steady-state occupancy of the
trap is given by Eq. (4.28), so using Eq. (4.28) one gets an equation similar to Eq. (4.35)
\[
\frac{\Delta C}{C_{\infty}} = \frac{N_T}{2N_D} \left( 1 - \frac{e_p}{e_n + e_p} \right) = \frac{N_T}{2N_D} \frac{e_n}{e_n + e_p}
\]
(4.37)

As one can see, this is similar to Eq. (4.35) except the change in capacitance is always less
than the \(\Delta C\) of Eq. (4.35) because the last term in Eq. (4.37) is always less than one. This will
be discussed further in the following section.

### 4.4 Introduction to steady-state photocapacitance and deep level optical spectroscopy

Because of the practical limits of temperature and measurement time, DLTS is limited to
detecting energies within \(\sim 1\) eV. Using only DLTS, a large portion of the bandgap can be left
unexplored with the bandgap of the nitride system around 6.03 eV in AlN.[3] Using optical
excitation, it is possible to supplement the DLTS results with defect information from \(\sim 0.5\) eV
(the absorption of the optics) to \(\sim 6\) eV (the absorption of air and the optics) using steady-state
photocapacitance (SSPC) and deep level optical spectroscopy (DLOS). In the depletion region
of a diode, carriers may be trapped at defect sites. If incident light with energy \(h\nu\) greater
than \(E_C - E_T\) for an electron emission or \(E_T - E_V\) for a hole emission, optically stimulated
emission can occur leading to the carrier being swept out of the depletion region by the
electric field and leaving behind fixed charge that impacts the measured capacitance. Both
SSPC and DLOS require special structures that allow light to penetrate uniformly into the
depletion region of a diode. A thin semi-transparent Schottky contact is typically used, which
allows for a very versatile sample that can also be used for DLTS, C-V, I-V-T, IPE, EBIC, and
other measurements.
Before delving into these measurement techniques, it is necessary to understand the emission rate under optical excitation. This treatment will again only consider $n$-type semiconductors, but an analogous derivation is also possible for $p$-type material as well. The change in $n_T$ considering only emission to the conduction band is still given by Eq. (4.29), but now the emission rate is proportional to the incident photon flux $\Phi$ where

$$e_n^o = \Phi (h\nu) \sigma_n^o (h\nu)$$

(4.38)

The $\sigma_n^o$ parameter is the *electron optical cross section* and is related to a deep level’s probability of absorbing a given photon. Using Eq. (4.28), the steady-state occupancy can be simplified to

$$n_T (t = \infty) = \frac{\sigma_p^o (h\nu)}{\sigma_p^o (h\nu) + \sigma_n^o (h\nu)} N_T$$

(4.39)

and the change in capacitance simplifies to

$$\frac{\Delta C}{C_{\infty}} = \frac{N_T}{2N_D} \left[ 1 - \frac{\sigma_p^o (h\nu)}{\sigma_p^o (h\nu) + \sigma_n^o (h\nu)} \right]$$

(4.40)

It is typically assumed that $\sigma_p^o \to 0$ yielding the familiar

$$N_T = \frac{2N_D \Delta C}{C_{\infty}}$$

(4.41)

If the cross-section of the minority band cannot be ignored, the change in capacitance is lower than the ideal case and the extracted trap concentration will be a lower limit on the actual trap concentration.

When $h\nu$ is lower than $E_C - E_T$, the optical cross-section is 0 and the light cannot cause the deep level to empty. Above $E_C - E_T$, the optical cross-section is non-zero and is function of the light energy, so in steady-state the deep level will be completely emptied by the incident photons. SSPC is concerned with the steady-state capacitance value $C_{\infty} (h\nu) - C_{dark}$ as a function of $h\nu$. When $h\nu > E_C - E_T$, the electrons are emitted leaving behind additional fixed
positive charge (i.e. the unpopulated defect sites), which leads to an increase in capacitance. At the onset energy, $h\nu = E_C - E_T$ and hence defines the energy of the trap. Positive onsets (increases in the slope of the SSPC) are indicative of an electron emission to the conduction band (i.e. the trap energy is referenced to the conduction band) while negative onsets indicate hole emission to the valence band (i.e. $h\nu = E_V - E_T$ at the onset). The trap concentration can also be extracted from the SSPC spectrum. The change in capacitance from the onset to the saturation point is $\Delta C$ and is used in Eq. (4.41) to calculate the defect concentration. As noted before, this may be a lower limit on the trap concentration.

The onsets in SSPC can be broad making identification of an exact level difficult. Moreover, the observed energy transition can actually be the classical ionization energy $E_0$ and a lattice relaxation related energy referred to as the Franck-Condon energy $d_{FC}$. In some cases, the change in occupancy for a defect necessitates a change in local bonding configuration and this gives rise to an energy increase needed to emit a carrier. This added energy is the Franck-Condon energy. The theories of Chantre et al. and Passler can be used to extract $d_{FC}$ from DLOS spectra. This, however, requires more information than just the steady-state values.

The change in $n_T$ can be calculated using Eq. (4.29), assuming electron emission is the dominant process, assuming the level is initially completely full, and evaluating at $t = 0$

$$\left. \frac{dn_T(t)}{dt} \right|_{t=0} \equiv -e_n^0 (h\nu) N_T = -\Phi (h\nu) \sigma_n^0 (h\nu) N_T$$

(4.42)

This can be converted to a capacitance derivative using the time derivative of Eq. (4.34) leading to

$$\frac{1}{C_\infty} \frac{dC(t)}{dt} \bigg|_{t=0} \equiv -\frac{1}{2N_d} \left. \frac{dn_T(t)}{dt} \right|_{t=0}$$

(4.43)
Subsequently, the change in capacitance at $t = 0$ is directly proportional to $\sigma_n^0 \Phi$

$$\sigma_n^0 (h\nu) = \frac{2N_d}{\Phi (h\nu) N_T C_\infty} \frac{dC}{dt} \bigg|_{t=0} \quad (4.44)$$

The flux is measured so it can be divided out leaving just the cross section. Reflection and small absorption in the sample cannot be accounted for with the current equipment, so the optical cross section is always given in arbitrary units. It should be noted, however, that the reflection and absorption mainly due to the semi-transparent Schottky contact could also have some impact on the shape of the optical cross section, but the optical cross section varies by several orders of magnitude so any small variations from absorption and reflection will have negligible impact on the shape. Reflection and absorption can, however, impact the overall magnitude of $\sigma_n^0$, so arbitrary units should be stated instead of cm$^2$.

The shape of the optical cross section contains the information for both $E_0$ and $d_{FC}$. One can first assume $d_{FC}$ is negligible and look at the quality of the fit to see if more complex models that account for $d_{FC}$ are necessary. Lucovsky first formulated an equation for the optical cross section by modeling the optical matrix elements for transition from the deep level to the conduction band, accounting for the density of states within the conduction band, and assuming a delta function for the deep level potential.[4] Disregarding the temperature dependence and focusing on the energy dependence, the cross section has the following form[4]

$$\sigma_n^0 (h\nu) \propto \frac{E_i^{1/2} (h\nu - E_i)^{3/2}}{(h\nu)^3} \quad (4.45)$$

for $h\nu > E_i$ and where $E_i$ is $E_C - E_T$ when referencing the conduction band and $E_T - E_V$ referenced to the valence band. Fitting is normally done with $\log(\sigma_n^0)$ vs. $h\nu$ to emphasize the onset of the cross section and not the higher energy area where it is more likely another defect may be responding. If the data is broader than the Lucovsky fit, this is an indication
of unoptimized $\Delta C/\Delta t$ and/or significant broadening due to lattice relaxation. Generally, shorter time windows result in sharper cross section data because long time windows tend to result in a lower slope calculation due to curvature than shorter ones. This needs to be balanced with signal-to-noise ratio and the signal from other traps that may be responding. If non-zero initial times are used to avoid a faster trap that is responding in the same energy range, one should take extra care as the broadening may be due to the choice of time window.

If the local atomic configuration is strongly coupled to the deep level's state, carrier emission and capture will result in lattice reconfiguration further resulting in phonon generation. If phonon-assisted photoionization is occurring during the DLOS measurement, the measured $\sigma_n^o$ will broadened with absorption for $h\nu < E^o$ where $E^o$ is the optical ionization energy. The thermal activation energy is lower than $E^o$ by the Franck-Condon energy $d_{FC}$. Using the model developed by Chantre et al., the effect of $d_{FC}$ on the cross section is given as[5]

$$
\sigma_n^o = \frac{E^o_\text{FC}^2}{h\nu \sqrt{4\pi kT d_{FC}}} \int_1^\infty \sqrt{\frac{x-1}{E^o}} \left[(x-1)/E^o + m\right]^{-2} \exp \left[ \frac{E^o_\text{FC}^2 (h\nu/E^o - x)^2}{4kTd_{FC}} \right] \, dx \quad (4.46)
$$

$m$ is the effective mass for the $\Gamma$ valley and also assumes transitions only to the $\Gamma$ valley because the other valleys are more than 1.2 eV greater in energy.[6] The parameters $E^o$ and $d_{FC}$ are determined using a non-linear least squares fit to the data.

Similarly, Pässler recently developed a model of the cross section specifically for GaN-based HEMTs.[7] This theory specifically accounts for the phonon energy in GaN and finds very large Franck-Condon energies of up to 1.46 eV for a 0.45 eV thermally activated level. The cross section for this model is

$$
\sigma_n^o \propto \frac{1}{h\nu \sqrt{2\pi d_{FC} \hbar \omega \coth (\hbar \omega / 2kT)}}
$$
The thermal excitation energy $E^T$ is the energy for an electron to be promoted from the minimum of the trap $E_1$ to $E_C$ with the assistance of phonons. During an optical excitation $E^O$, the electronic transition is fast compared to the lattice relaxation resulting in a vertical transition since the heavy atoms cannot respond in the same time frame. The extra energy needed $D$ is the Franck-Condon energy. From Ref. 7.

$$\int_0^{\infty} \frac{E_k^{3/2}}{(E_k + E^o - d_{FC})^2} \exp \left[ \frac{-(\hbar \nu - E^o - E_k)^2}{2d_{FC} \hbar \bar{\omega} \coth (\hbar \bar{\omega} / 2kT)} \right] dE_k$$

where $E_k$ is $\hbar \nu - E^o$ and $\hbar \bar{\omega}$ is the phonon energy. This model has not been used in the current research mostly because the Franck-Condon energies have been negligible, but this model could prove useful in the future for the GaN HEMT measurements.

To visualize the Franck-Condon energy and lattice relaxation effects, it is necessary to invoke the configuration coordinate diagram shown in Fig. 4.2. If the change in occupancy of a deep level leads to a change in lattice distortion, optical electron excitation of such a
state occurs quickly relative to the time required for the heavy atom to displace (i.e. the Franck-Condon principle).\[8\] The coordinate axis in Fig. 4.2 represents the effective motion of a defect atom due in this case to the electron occupation change.\[9\] Parabolic dependence on the configuration is assumed approximating electron-phonon interaction.\[9\] For an optical process where phonon generation cannot occur during the transition event, the atomic configuration cannot immediately respond, so the transition must be vertical, which is depicted by the $E^0$ transition in Fig. 4.2. After the transition, the carrier will thermalize down to the $E_C$ valley with the generation of the phonons. For a thermal process where phonons are directly involved in the process, the configuration can reconfigure during the event and only the $E^T$ energy is needed. A more in-depth, theoretical discussion of configuration coordinate diagrams, their construction, and why energies measured by photoluminescence (PL) and DLOS differ is available in Ref. 10.

With the theoretical background for SSPC and DLOS, the measurement system and common setting used throughout the research need to be presented. A schematic of the equipment is given in Fig. 4.3. DLOS is performed by radiating a sample with monochromatic light and recording the capacitance transient. After illumination, the traps are refilled with a fill pulse of higher (more positive) bias. After each transient, the energy of the monochromatic light is increased and the process is repeated. By starting a low energies, defects remain filled until the photon energy is sufficient to excite the level. The DLOS measurement consists of optical, electrical, and control components. The control is handled using a Windows-based PC by running one of several custom Labview programs. The electrical components for capacitance-based DLOS are the probe station, capacitance meter, digitally-controlled capacitance offset box, arbitrary function generator for managing voltage pulses to the sample, computer-based data acquisition card, and optionally an oscilloscope for observing
Figure 4.3: Schematic of DLOS equipment depicting Xe and QTH lamps for spanning a range of energy from 0.5 eV to 6.0 eV, monochromators for selecting single wavelengths of light, mirror-based beam steerer for selecting the light source, fast shutter for accurately starting the optically stimulated transient, temperature controlled stage for measurements at room temperature and anywhere between 4 K and 400 K, and capacitance meter and other equipment to measure and acquire the transient data.

the voltage pulses and capacitance transients. The major components of the optical system are the 1000 W Xe lamp, 1000 W QTH lamp, two monochromators for selecting the desired wavelength from each lamp, a beam steerer to select the lamp, and a fast shutter to enable synchronization of the start of the transient with the onset of light and to block light during the fill pulse.

To perform these measurements, a series of Labview programs were written to automate the measurement process and allow for high speed data acquisition. The DLOS software suite consists of three main programs: Initialization.vi, DLOS.vi, and DLOS Analysis2.vi. Appendix B contains screenshots and a short description. The Initialization.vi program is used to open/close shutters and set up each piece of equipment (e.g. capacitance meter, voltage function generator, monochromators, and capacitance offset box). Because the DLOS
experiments require 12-24 hours in the dark to allow the deep levels to reach equilibrium, the DLOS.vi does not, in general, setup the equipment as this would potentially apply a voltage change requiring additional wait in the dark. Instead, the equipment needed for DLOS is manually set up before placing the sample in the dark when the light is aligned using the Initialization.vi. The DLOS.vi program performs the actual measurement – setting the wavelength of light, applying the fill pulse, recording the capacitance transient, saving the data, and repeating this process until the run is complete. For DLOS analysis, the photon flux needs to be measured. “Photon Flux Measurement.vi” can be used for this purpose. This program uses a thermopile with an aperture of known area to measure the power density. The program automates the data acquisition by controlling the monochromator and measuring the flux at each wavelength. Thermopiles are temperature sensitive but their response is nearly independent of wavelength, so to make reliable measurements a shutter blocks off the light to baseline the thermopile offset at each wavelength before measuring the change in power from the incident light beam. Knowing the power density, measured in mW/cm², and the wavelength of the incident light, the photon flux $\Phi$ can be calculated

$$\Phi = \frac{P}{qAhv} \quad \text{(4.48)}$$

where $P$ is the power measured by the thermopile and $A$ is the aperture area.

Many subroutines exist to change the monochromator filters and gratings to maximize the flux and prevent unwanted multiples of the wavelength generated by grating monochromators from the output, set up and record from the data acquisition card, etc. The transient recorded is somewhat complex. The data acquisition card connected to the capacitance meter output records data at 50 kHz for the entire transient, but data is downsampled using a
simple averaging routine. The following

\[ C(k) = \frac{1}{S} \sum_{i=0}^{S-1} C(Sk + i) \]  

(4.49)

indicates the averaging method used where \( S \) samples of \( i \) are averaged for each \( k \)th element.

A value of 200 is used for \( S \) during the first second and 2000 for the remaining time. The higher data rate at the beginning allows for shorter DLOS time windows during analysis. The average time is calculated for each \( k \)th element and stored in time/capacitance pairs in the data files.

The equipment is typically controlled via general purpose interface bus (GPIB), but the fast shutter, for example, uses serial communication. All of this software and even equipment setup is unique. While this adds a level of complexity to the system, it is also allows much greater flexibility. With a functional understanding of Labview, it is possible to implement new techniques in less than a day in some cases. Obviously, there are many aspects of our system that are complicated subjects on their own, but lack of knowledge in any of these areas can be detrimental to the validity of the results due to changing something incorrectly in the program, not using the correct filters, misaligning a lamp, and the list goes on.

Proper setup of the optical system is necessary to ensure valid results and identify artifacts due to filter or grating changes. A good resource for learning about the optics, monochromators, lamps, and shutters used in our setup from Ref. 11. The single, most subtle mistake someone can make is choosing filters and filter crossover points that are incorrect. If the monochromator is set to 1240 nm, the grating will diffract light at 620 and 310 nm at the same angle as 1240 nm resulting in multiple wavelengths of light instead of monochromatic light. Long-pass filters are used to block light with a wavelength lower than the cut-off wavelength of the filter. The cut-off wavelength is defined as the point where 50% of the light is transmitted, so for the previous example, the cut-off frequency would need to be greater than 620 nm to guarantee no light at 620 or 310 nm would pass. The cut-on wavelength for
determining what light is blocked should be where < 1% of the light is transmitted while the cut-off wavelength for determining if light is passes should be where greater than ~ 80% of the light is transmitted. Additionally, if the filter or grating cross over points are chosen incorrectly, all photons may be blocked over a range of wavelengths. This can impart features or lack there of into the DLOS and SSPC spectra. Therefore, great care should be taken to ensure the optical system imposes as few artifacts as possible and that any artifacts that do exist are identified and removed if possible during the analysis.

With the program briefly described and a few big caveats mentioned, the common operating conditions for most nitride samples are mentioned. Unless otherwise stated, the Xe lamp based measurements were run from 1.2 to 3.6 eV in 0.02 eV increments for GaN samples and up to 4.5 eV for Al$_{0.30}$Ga$_{0.70}$N with the same energy increment. The quartz-tungsten-halogen (QTH) lamp was used for energies between 0.5 and 2.0 eV again with a 0.02 eV energy increment. The flux of the Xe lamp is in the $10^{16}$ photons/cm$^2$ s. All samples were pulsed to 0.00 V for 10 s between each transient to allow the deep level to refill. The transients were measured at -0.50 V (quiescent bias) for times between 300 and 350 s. This time was experimentally determined to guarantee the transients reached steady-state at each incident energy. The voltage pulse of filling and measuring results in a thermally induced transient for the shallow levels, so an experimentally determined delay time typically between 15 and 60 s was used between the fill pulse and shutter opening for measurement to ensure the thermal transient had extinguished. When measuring a new sample, these times should be experimentally determined by running a test run for a few wavelengths and with zero delay after the fill pulse to determine the thermal transient decay time.
4.5 Introduction to deep level transient spectroscopy

Deep level transient spectroscopy (DLTS), developed by D. V. Lang in 1974, is a powerful and well known technique for identifying deep level thermal cross-section, concentration, and energy position in the bandgap.[12] The basic concept of DLTS has changed relatively little since its inception, and most of the advances have resulted from the increased capability of computers to do complex storage and analysis of the transient data. There are a few exceptions to this including applying the technique to conductance and current of , for example.[13–16] Much of developments to the DLTS theory have involved improving the analysis, but improving upon the original boxcar integration in terms of signal-to-noise ratio have been limited. A myriad of variations of DLTS are available including the standard boxcar DLTS, Fourier transform DLTS (FTDLTS), Laplace-transform DLTS, and covariance method of linear predictive modeling (CMLPM) to name a few.[12, 17–20] Several of these variations have been implemented at OSU to aid in defect identification. Each method has its advantages and disadvantages with regard to required signal-to-noise ratio (SNR), ability to separate defect levels closely spaced in energy, accuracy, and ease-of-use. The old Bio-Rad DL4600 DLTS system, which was previously used at OSU, used the original double boxcar method. The following sections will describe the original boxcar method, Fourier transform method, and a new method called the Padé-Laplace method. Before discussing the different analysis methods, it is beneficial to discuss the experimental operation of the DLTS system at OSU to show how it works and how the different analysis methods can be applied to the same data sets.
4.5.1 Operation and experimental setup of deep level transient spectroscopy

The experimental setup of the DLTS system is rather simple. The sample is mounted in a temperature-controlled cryostat and connected to a capacitance meter and voltage function generator. The order of events for DLTS are to apply a fill pulse with bias $V_{FP}$ for time $t_{FP}$, adjust the bias to the quiescent voltage $V_R$ and record the capacitance transient for the necessary amount of time, increment the temperature, and repeat the process. During the fill pulse, a majority carrier trap in n-type material below the Fermi level will fill with electrons as can be seen in Fig. 4.4a. After the fill pulse, the quiescent bias is applied, typically -0.5 V and always less (more negative) than the fill pulse bias, so the depletion region extends deeper into the material (see Fig. 4.4b). Now most of the filled traps in the depletion region are above the quasi-Fermi level and would prefer to be empty, so electron emission to the conduction band occurs according to Eq. (4.34). As the emission occurs, the fixed charge density increases for an electron trap in n-type material, but because the applied bias is constant the total charge is constant the depletion depth must decrease to maintain the total charge. As the depletion region decreases, the capacitance increases, which is shown in Fig. 4.4c and d. As this process is repeated at higher and higher temperatures, the emission rate increases according to Eq. (4.23) and is shown here rearranged for clarity

$$e_n(T) = \sigma_n \nu_{th} N_C \exp \left( -\frac{E_C - E_T}{kT} \right)$$

(4.50)

Again, remember that the emission rate constant $\tau_n$ is simply $1/e_n$ assuming negligible emission to the valence band. The emission rate varies exponentially with temperature, so the challenge is to develop a system that can determine this rate from the capacitance
Figure 4.4: Schematic illustration of the effect of capture and emission of electrons from an electron trap. (a) The sample is biased in the fill pulse regime to allow $E_T$ to fill with electrons. The fill pulse bias is always greater than the measurement bias to allow traps to fill. The bottom part of the figure shows the bias scheme as a function of time while the other graph shows the resulting capacitance. (b) At the start of the measurement, the levels are completely filled resulting in a lower charge density in the depletion region. This dictates the depletion depth and capacitance. (c) As the traps emit, the charge density increases due to the uncovered fixed charge of the traps. In this case, the depletion region shrinks as the same total charge can be uncovered in a smaller volume, and the capacitance increases. This is the standard intuitive process for a deep level exposed to a voltage change. From Ref. 2.
transient, which is given by

\[ C(t) = C_\infty \left( 1 - \frac{n_T}{2N_D} \right) = C_\infty \left[ 1 - \frac{N_T}{2N_D} \exp \left( \frac{-t}{\tau_n} \right) \right] \] (4.51)

which is similar to Eq. (4.34) but \( n_T(0) \) is assumed \( N_T \). With today’s computing power and memory capacity, recording parts or all of the transients is relatively simple. The difficult part is extracting the emission rate/time constant from recorded data especially considering that more than one trap can have similar time constant that would need to be extracted separately and this is where the double boxcar, Fourier transform, Padé-Laplace, and other analysis methods are used. Once the emission rate vs. temperature is extracted, the trap activation energy and capture-cross section can be determined by modifying Eq. (4.23) to

\[ \ln(\tau_n T^2) = (E_C - E_T) \frac{1}{kT} - \ln(\sigma_n \gamma_n) \] (4.52)

Plotting \( 1/kT \) vs. \( \ln(\tau T^2) \) gives trap energy from the slope and the cross-section is extracted from the \( y \)-intercept. This is very straightforward so far, but as alluded to earlier the extraction of the emission rate constants is nontrivial and this has been the basis for the plethora of analysis techniques. The following sections will describe three of these techniques.

**4.5.2 Double boxcar method for DLTS Analysis**

The original DLTS analysis method used a double boxcar integrator to determine the temperature for a given emission rate.\(^{[12]}\) A hardware implementation of the double boxcar integrator multiplied the capacitance transient by the double boxcar waveform shown in Fig. 4.5. Stated more simply, the DLTS signal \( \Delta C_0 \) is simply \( C(t_2) - C(t_1) \) Better signal-to-noise ratio is possible without significant distortion of the signal by averaging the capacitance around \( t_2 \) and \( t_1 \), which is depicted in Fig. 4.5 and implemented in the double boxcar DLTS at OSU. When the emission rate constant \( \tau \) is very small, the transient is extremely quick.
Figure 4.5: Double boxcar signal multiplier digitally applied to capacitance signal. Simplistically, the $\Delta C$ is $C(t_2) - C(t_1)$, but in reality signal-to-noise ratio can be greatly improved by averaging the signal around $t_1$ and $t_2$. Mimicking the old Bio-Rad DL4400 DLTS system, the new system uses $t_2/t_1 = 2.5$ and averages 10% of $t_2$ around each time.

and reaches steady-state before $t_1$ and $\Delta C_0$ is therefore approximately zero. In contrast, the transient with a very large $\tau$ is so slow that the capacitance is essentially constant resulting in a zero $\Delta C_0$ once again. In between these two extremes, there exists a peak in the measured $\Delta C_0$. This phenomenon is depicted in Fig. 4.6 where $\tau$ is small at the top of the figure and long at the bottom. The change in capacitance for a given rate window is given by

$$\Delta C_0 = C(t_2) - C(t_1) = \frac{C_\infty N_T}{2N_D} \left[ \exp\left(\frac{-t_1}{\tau_n}\right) - \exp\left(\frac{-t_2}{\tau_n}\right) \right]$$  \hspace{1cm} (4.53)

Taking the derivative of Eq. 4.53 and setting it to 0 to find the maxima or minima, the peak in the $\Delta C_0$ signal can be found as

$$\tau_n^* = \frac{t_2 - t_1}{\ln\left(t_2/t_1\right)}$$  \hspace{1cm} (4.54)

At this peak, the combination of $t_2$ and $t_1$ will cause $\Delta C_0$ to peak when the capacitance transient emission rate constant $\tau_n$ equals the $\tau_n^*$ from Eq. (4.54). Essentially, the double box
Figure 4.6: Illustration of how a typical DLTS trap spectrum is determined by the various capacitance transients. At low (bottom transient of figure) temperatures the trap emits too slowly to generate a change in capacitance while at high temperatures (top transient of the figure) the trap emits too quickly to be seen at $t_1$ and $t_2$. At intermediate temperatures, the spectrum peaks at $T_1$ where the trap emission rate equals the double boxcar rate window. From Ref. 12.
car method searches the capacitance transients as a function of temperature and peaks for a majority carrier defect when the trap emission rate $\tau_n$ equals the emission rate determined by the boxcar settings $\tau_n^*$. Though not discussed here, a valley for a minority carrier defect can exist in the DLTS spectra and this minimum corresponds to $\tau_p = \tau_p^*$. Thus by changing $t_2$ and $t_1$, the boxcar integrator searches for a different emission rate versus temperature. The inverse of $\tau_n^*$ in Eq. (4.54) is commonly called a rate window. Searching for several emission rate constants using several rate windows, one develops a set of emission rate constants and peak temperatures that can be used in Eq. (4.52) by using a linear regression of the data plotted in a Arrhenius relationship ($1/kT$ vs. $\ln(\tau_n T^2)$) to extract the trap characteristics.

From the Arrhenius plot, one is able to extract the capture cross-section and trap energy, but it is also important to determine the trap concentration, $N_T$. This can be determined in a relatively straightforward manner if one knows the total change in capacitance using Eq. (4.36) repeated here for simplicity

$$N_T = 2N_D \left( \frac{\Delta C}{C_\infty} \right)$$  \hspace{1cm} (4.55)

The only issue is that the total change in capacitance is unknown at this point because the change in capacitance the double boxcar detects is not at $t = 0$ and $\infty$. However there is a relation between the double boxcar $\Delta C_0$ and the total $\Delta C$ where

$$\Delta C = \frac{\beta \beta^1 (\beta^1 - 1)}{\beta - 1} \Delta C_0$$  \hspace{1cm} (4.56)

and $\beta = t_2 / t_1$. The ratio $\beta$ is kept constant at 2.5 for all rate windows to be consistent with the older Bio-Rad DL4400 system and also to keep the peak height constant between rate windows and maintain a uniform correction factor of 3.07. Equation (4.56) then becomes

$$N_T = 6.14N_D \left( \frac{\Delta C_0}{C_\infty} \right)$$  \hspace{1cm} (4.57)
If $\beta$ varied, $\Delta C$ remains constant, which means that $\Delta C_0$ varies proportionally to $r$. This means the DLTS spectra would have varying peak heights. The ratio $r$ is kept constant to maintain the same signal-to-noise ratio. Typically varying peak heights at the different rate windows is indicative of non-exponential transients or a temperature dependent cross-section.[1]

The boxcar method is quite elegant, and is used quite heavily despite the revolution in computing power, memory capacity, and the host of other methods people have developed. The downsides of this analysis technique are the necessity to find the peaks for each trap for each rate window and the broad peaks means that defects with close energy levels can have overlapping peaks that necessitate peak fitting to separate. Most other techniques developed in subsequent years extract $\tau_n$ directly without the need for peak identification and/or fitting.

### 4.5.3 Fourier transform deep level transient spectroscopy

The Fourier transform method of DLTS (FTDLTS) uses the same acquisition system, but relies on the Fourier transform to analyze the capacitance transients. Weiss *et al.* showed that the fast Fourier transform (FFT) of the DLTS transients could be used to directly extract the emission rate at a given temperature thus allowing direct plotting of the Arrhenius plot.[18] The major advantage of this technique is not having to fit individual peaks like boxcar DLTS. This method can only properly interpret single exponential transients meaning only one defect has significant emission at any given temperature. If two or more traps are emitting simultaneously, it is possible to separate them if the traps have sufficiently different activation energies.

For a capacitance transient given as

$$f(t) = \Delta C \exp\left(-\frac{t + t_0}{\tau_n}\right) + C_\infty$$

$$f(t) = \Delta C \exp\left(-\frac{t + t_0}{\tau_n}\right) + C_\infty$$

(4.58)
where $\Delta C$ is the amplitude of the capacitance transient, $C_\infty$ is the steady-state capacitance, and $t_0$ is the end of fill pulse. If one takes the continuous Fourier transform of Eq. (4.58), the following equations for the $n$th real and imaginary coefficients, respectively, are obtained\[18\]

\[
a_n = \frac{2\Delta C}{T_w} \exp\left(-\frac{t_0}{\tau_n}\right) \left[1 - \exp\left(\frac{T_w}{\tau_n}\right)\right]
\]

(4.59)

and

\[
b_n = \frac{2\Delta C}{T_w} \exp\left(-\frac{t_0}{\tau_n}\right) \left[1 - \exp\left(\frac{T_w}{\tau_n}\right)\right]
\]

(4.60)

where

\[
\omega = \frac{2\pi}{T_w}
\]

(4.61)

and $T_w$ is the measuring time. With this in mind, simple algebraic manipulation can be used to derive the following most useful expression for $\tau_n$ (Ref. 18)

\[
\tau_n(a_n, b_n) = \frac{1}{n\omega} \frac{b_n}{a_n}
\]

(4.62)

Because the real measurement requires numerical analysis, the discrete Fourier transform is required. Only Fourier coefficients from 1 to 5 are typically used because higher values amplify the transient noise. With $\tau_n$ calculated directly, an Arrhenius plot, based on Eq. (4.52), be constructed and the trap activation energy and cross-section ascertained. The major advantage of this technique is not having to fit each peak for each rate window analyzed as in boxcar DLTS because the FTDLTS yields the Arrhenius plot directly. Figure 4.7 is a possible Arrhenius plot that may result from two closely spaced energy levels. By choosing different $T_w$ measurement times from the data, one deep level can be enhanced to allow fitting over a larger range. Generally, it was observed that the Fourier transform is most sensitive to a $T_w$ approximately 3–5 times larger than the $ tau_n$ being extracted. A program was developed that automatically used multiple $T_w$’s and also optimized $T_w$ for the highest sensitivity as well.
Figure 4.7: Example Arrhenius plot using FTDLTS analysis from Ref. 18. FTDLTS is only able to detect one deep level at a time. This example shows two traps with close energy levels where depending on the time window chosen for the analysis the transition from detecting one deep level to the other can be shifted.

4.5.4 Padé-Laplace deep level transient spectroscopy

A new digital DLTS system has been built using a computer, data acquisition card, capacitance meter, arbitrary function generator and temperature-controlled stage. This setup also separates the running of the experiment from the data analysis by recording the transients for later analysis. This allows for easy implementation of many other DLTS analysis methods. For example, FTDLTS and standard boxcar DLTS have difficulty separating defect energy levels closely spaced in energy, which is addressed by Laplace DLTS, or high resolution DLTS (HRDLTS).[19] The limitation of this method is the strict requirements on very high signal-to-noise ratio, which requires thousands of iteration to sufficiently average out the noise. This DLTS program can be set to average many transients at a few temperatures and save all the data, so HRDLTS could be implemented. Unfortunately, performing the numerical
inverse Laplace transform on real data is considered an ill-posed problem because there is not a unique solution to the data if any noise exists. Several methods exist to work around this problem, but a better analysis method would combine the resolution of the HRDLTS with the ease of use of FTDLTS and double-boxcar DLTS. For a good review of other DLTS methods with respect to accuracy, convergence robustness, signal-to-noise ratio (SNR), energy resolution, and computing time see Ref. 21.

Here, the Padé-Laplace method is presented as an alternative analysis method that has many of the features of other DLTS methods but with few of the disadvantages. Similar methods such as the Padé-Z-transform method, which can give comparable results, but because of the large number of data points recorded for each transient (~ 1 million points) the Laplace transform is superior as the number of points closely approximates a continuous function. Padé-Laplace is commonly applied to biological applications and has many features useful for both DLTS and DLOS. The Padé-Laplace method was first formalized in 1988 by Claverie et al. for the analysis of decay rates in multi-exponential functions.[22] The theory for this can be found in Refs. 22, 23 and 24 and is reviewed here.

DLTS transients have a general form of

\[ C(t) = \sum_{n=1}^{N} \Delta C_n \exp \left( \frac{-t}{\tau_n} \right) + C_\infty \quad (4.63) \]

where \( C \) is the capacitance, \( N \) is the number of traps in the sample, \( \Delta C_n \) is the change in capacitance due to trap \( n \), \( \tau_n \) is the trap emission constant for trap \( n \), and \( C_\infty \) is the steady-state capacitance. The equation to transform a time domain function \( f(t) \) to the Laplace domain \( F(s) \) is[25]

\[ F(s) = \mathcal{L}\{f(t)\} = \int_{0}^{\infty} f(t) e^{-st} \, dt \quad (4.64) \]
where $s$ is independent variable in the Laplace domain. Additionally, the $n$th derivative of $F(s)$ is

$$F^{(n)}(s) = \int_0^\infty (-t)^n f(t) e^{-st} dt$$  (4.65)

Using the Laplace transform, the terms in Eq. (4.63) can be transformed into the $s$-domain[25]

$$\mathcal{L}\{\exp(-\alpha t)\} = \frac{1}{s + \alpha}$$  (4.66)

and

$$\mathcal{L}\{u(t)\} = \frac{1}{s}$$  (4.67)

where $\mathcal{L}$ is the Laplace transform operator and $u$ is the step function. In the $s$-domain, the capacitance is

$$C(s) = \frac{C_\infty}{s} \prod_{n=1}^{N} \frac{\Delta C_n}{s + \tau_n^{-1}}$$  (4.68)

Now, the form of the solution in the $s$-domain is known, but a method to transform the data is needed. Although analytic solutions for the capacitance transients is possible, the Laplace transforms of digitized data are not directly possible. The solution is to approximate the Laplace transform using a Taylor series expansion and Padé approximant to yield a solution that is a rational fraction as in Eq. (4.68). The Padé approximant $P_L(x) / Q_M(x)$ is given as

$$\frac{P_L(x)}{Q_M(x)} = \sum_{n=0}^{\infty} a_n x^n$$  (4.69)

where $a_n$ is the coefficients of a rational function approximating the Laplace transform and $P_L$ and $Q_M$ are the polynomial functions of the Padé approximant

$$P_L(x) = p_0 + p_1 x + \ldots + p_L x^L$$  (4.70)

and

$$Q_M(x) = q_0 + q_1 x + \ldots + q_M x^M$$  (4.71)
In this case, the order of $P_L$ is $L$ and $Q_M$ is $M$. To match the form of Eq. (4.68), the order of $M$ is $N + 1$ and $L$ is $N$. Assuming the $a_n$ coefficients are known, the Padé coefficients are calculated solving the system of equations from Eq. (4.69) and separating out the terms of equal order of $x$

$$Q_M(x) \sum_{n=0}^{\infty} a_n x^n = P_L(x) \quad (4.72)$$

Now, the poles (roots of $Q_M$) and residues will give the solution to Eq. (4.68), so it is a matter of determining the $a_n$ coefficients from the data.

The Taylor series expansion can be expressed as

$$y(x) = \sum_{n=0}^{\infty} \frac{y^{(n)}(a)}{n!} (x - a)^n \quad (4.73)$$

where the function $f$ is evaluated at $a$ and $x$ is the independent variable. The solution is to use the Taylor series expansion at $s_0$, which is

$$F(s) \approx \sum_{n=0}^{2N} \frac{1}{n!} \int_{0}^{\infty} (-t)^n f(t) e^{-s_0 t} d t (s - s_0)^n \quad (4.74)$$

Combining this with Eq. (4.65) gives

$$F(s) \approx \sum_{n=0}^{2N} \frac{1}{n!} \int_{0}^{\infty} (-t)^n f(t) e^{-s_0 t} d t (s - s_0)^n \quad (4.75)$$

To fit this equation into the form of Eq. (4.69), $x$ is substituted for $s - s_0$ giving

$$F(s) \approx \sum_{n=0}^{2N} \left\{ \frac{1}{n!} \int_{0}^{\infty} (-t)^n f(t) e^{-s_0 t} d t \right\} x^n \quad (4.76)$$

where the $a_n$ can be identified from Eq. (4.69) as

$$a_n = \frac{1}{n!} \int_{0}^{\infty} (-t)^n f(t) e^{-s_0 t} d t \quad (4.77)$$

It is nearly complete, but the integral in the previous equation must be approximated by a numerical integral so it can be performed on the data. This gives

$$a_n \approx \frac{1}{n!} (-t)^n \sum_{i=0}^{R} f(i \Delta t) e^{-s_0 i \Delta t} \Delta t \quad (4.78)$$
Finally, $a_n$ can be calculated, which allows the Padé approximant to be calculated. The roots of the denominator (poles) are the emission rates of the traps and the corresponding residuals are the changes in capacitance due to each trap. Due to the substitution of $x$, the poles need to be offset by $s_0$ to determine the actual values.

So now a new mathematical method has been derived for use with DLTS, but what are the advantages and limitations. The advantages include:

1. $\tau$'s are determined directly without need for peak fitting.

2. $N_T$'s are directly calculated from residuals $(2N_D\Delta C_n/C_\infty)$.

3. It is able to separate $\tau$'s of comparable value (i.e. levels closely spaced in energy). [22, 24] Although rigorous comparison of HRDLTS and PLDLTS has not been done, the tests in Ref. 19 for HRDLTS and Ref. 24 for PLDLTS show very similar resolutions, but PLDLTS is computationally easier and does not suffer from the convergence problems of HRDLTS when SNR is not high.

4. It can operate on large sample sets (e.g. > 1 million data points).

5. One consequence of large sample sets is large internal averaging that allows for convergence even with low signal-to-noise ratio.

6. The number of traps with $\tau$'s able to be detected by the data acquisition system is determined by the analysis.

7. Unlike methods like FTDLTS, PLDLTS is insensitive to capacitance offsets with a caveat. The capacitance offset can be accounted for in the analysis, but the PLDLTS is run in a two-pass mode. The first time, the offset is accurately determined and removed so
that the second pass does not contain a pole at \( s = s_0 \) that contains a very large value compared to the trap residuals. This improves detection ability.

The major disadvantages of the Padé-Laplace method are the choice of \( s_0 \) and \( N \). As noted by Ref. 24, a method for searching for the best \( s_0 \) is not generally agreed upon. To determine the optimal \( s_0 \), the parameters extracted for a given \( s_0 \) were used to reconstruct the transient. The mean-squared error (MSE) was calculated for the data minus the reconstructed transient to evaluate the choice of \( s_0 \). In a brute force method, \( s_0 \) is slowly varied recording the MSE at each point. The point of lowest MSE is assumed to most correct answer. This is computationally intensive, but the entire DLTS spectra can be determined in a few minutes with this method. Additionally, the number of simultaneous transients to extract is nearly trivial on a case by case basis. If too many terms are considered, the Padé approximant residuals of the extra terms are near zero or the poles become complex signifying sinusoidal behavior. It is generally possible to identify the extra terms and rerun the analysis with the correct number of terms, which is done to improve the accuracy of the extracted parameters. If too few terms are considered, all of the Padé approximant residuals and poles are convolutions of all of the components and can by decaying sinusoids. However, automating this process to let the computer choose the most appropriate solution is more difficult. Because the number of simultaneously emitting traps on a time scale of \(~ 100 \mu s\) to \(~ 1 s\) is generally limited due to the exponential relationship between \( \tau \) and \( T \), the analysis can be performed for a range of \( N \) then the best value of \( N \) can be chosen graphically over a temperature range of interest. Looking at an Arrhenius plot (\( \ln(\tau T^2) \) vs. \( 1/kT \)) for \( N = 1, 2, \ldots \) along with a plot of the residuals vs. \( 1/kT \), the smallest number of \( N \) that produces a straight line on the Arrhenius plot with consistent values of the residuals is the best choice. The analysis program for this
method that implements these features is “PLDLTS Analysis.vi” and is further explained in Appendix B.

This method is also useful for DLOS where multiple deep levels can respond at once. Currently, \( \frac{dC}{dt} \) is extracted around \( t = 0 \) where the optical cross section is proportional to \( \frac{dC}{dt} \). The problem is that all of the traps with energy \( E_C - E_T > h\nu \) respond. Often choice of the time range over which the slope is calculated can mostly separate each trap, but this is non-ideal. A better method might be to use the Padé-Laplace method to extract the emission rates, so the cross section for each deep level could be directly calculated from Eq. (4.38).

### 4.5.5 Capture kinetics via DLTS

Under ideal situations, defects are randomly distributed and in low enough concentration such that the no defect occupancy affects any other defect occupancy. If defects are present in very high concentrations, clustered, or decorate dislocations, for example, the defects can be physically close enough that the Coulombic barrier associated with one defect can influence the occupancy of other defects. Interacting defects vs. ideal non-interacting defects can be determined via differences manifested in the DLTS spectra. Studying the capture kinetic behavior allows one distinguish ideal point defects from interacting ones. Another non-ideality that can arise for multiple reasons is a distribution of energy levels for a given defect instead of the ideal discrete energy level. Determination of the energy distribution is also possible using capture kinetic analysis although a few modifications to the ideal theory previously presented is required.

In capture kinetics the fill pulse time is varied to measure the change in the DLTS peak height. The changes that occur here are a result of incomplete filling of the level where in the past it was assumed \( n_T(0) = N_T \), so it is necessary to focus on fill pulse time when capture is
the dominant process. Keeping the discussion restricted to \( n \)-type material, the change in \( n_T \) is

\[
\frac{dn_T}{dt} = c_n n p_T = c_n n (N_T - n_T) \tag{4.79}
\]

assuming the level is completely empty at the start of the fill pulse the solution is

\[
n_T(t) = N_T \left[ 1 - \exp \left( -c_n n t' \right) \right] \tag{4.80}
\]

The \( t' \) is used to signify a difference between the measurement time \( t \) that starts at the end of the fill pulse and \( t' \) that starts at the beginning of the fill pulse. This expression is evaluated at the end of the fill pulse \( t_{FP} \) right before the bias changes and emission becomes the dominant process. The value of \( n_T \left( t' = t_{FP} \right) \) will be called \( n_{T0} \) and corresponds to \( n_T \left( t = 0 \right) \) in previous equations. For a sufficiently long fill pulse time, \( c_n n t_{FP} \) is large and \( n_{T0} \rightarrow N_T \). Conversely, as \( c_n n t_{FP} \rightarrow 0 \), \( n_{T0} \rightarrow 0 \). If the change in capacitance \( \Delta C_F \) with a fill pulse \( t_{FP} \) is compared to the change in capacitance for an infinitely long fill pulse \( \Delta C_\infty \), the ratio is

\[
\frac{\Delta C_F}{\Delta C_\infty} = \frac{n_{T0,F}}{N_T} = 1 - \exp \left( -c_n n t_{FP} \right) \tag{4.81}
\]

This can be rearranged to a more usable form

\[
\ln \left( 1 - \frac{\Delta C_F}{\Delta C_\infty} \right) = -c_n n t_{FP} \tag{4.82}
\]

Because \( c_n n \) is constant, the left side of the above expression should be linear with \( t_{FP} \) for an ideal, non-interacting defect to the point where the time is long enough for the level to completely saturate. This measurement can be performed by finding the saturation point where \( \Delta C_F \) no longer increases with \( t_{FP} \), which defines \( \Delta C_\infty \). Then measuring \( \Delta C_F \) at several shorter fill pulse times, a plot can be generated to determine if the defect is ideal.

When the defects are non-uniformly distributed and closely spaced due to linear arrangements or clustering for example, a Coulombic interaction can occur influencing the
Figure 4.8: Before capture (top), the defect is uncharged and has no Coulombic barrier associated with it. After capture begins (bottom), a time-dependent barrier forms reducing the capture rate of the next electron. In order for the effect to be observed, neighboring defects have to be close enough to be influenced by the resultant Coulombic potential. From Ref. 27.

electron capture of neighboring defects. This Coulombic barrier to capture is modeled as a time-dependent potential $\phi(t)$ that builds up around the defects and reduces the subsequent electron capture as in Fig. 4.8. Assuming Boltzmann statistics, the potential barrier modifies Eq. (4.79) becoming[28]

$$\frac{dn_T}{dt} = c_n n (N_T - n_T) \exp \left[ -\frac{q\phi(t)}{kT} \right] \quad (4.83)$$

The function of the potential barrier is unknown but it must depend on the defect occupancy $n_T$, so $\phi(t)$ is represented by its Taylor series expansion. A general expansion of the potential
becomes

\[ \Phi(t) = a_1 n_T(t) + a_2 n_T^2(t) + \ldots \]  

(4.84)

where \( a_1 \) and \( a_2 \) are coefficients of the expansion whose value depends on the specifics of the defect arrangement. After neglecting the higher order terms, solving Eq. (4.83) with an assumption of \( t \sim 0 \), and developing the equation for capacitance, the solution is[28–31]

\[ \Delta C \approx C_{\infty} \frac{kT}{2N_D q a_1} \ln \left( \frac{t_f}{\tau_E} \right) = C_{\infty} \frac{kT}{2N_D q a_1} \left[ \ln \left( \frac{t_f}{\tau_E} \right) - \ln \left( \tau_E \right) \right] \]  

(4.85)

where \( \tau_E \) is the time constant associated with the formation of a time-dependent Coulombic barrier. The time constant \( \tau_E \) becomes[31]

\[ \tau_E = \frac{\tau_n kT}{n q a_1} \]  

(4.86)

From Eq. (4.85) if a plot of \( \Delta C \) vs. the natural log of the fill pulse is linear, the deep level is exhibiting trapping kinetics consistent with that of a linearly arranged extended defect. An example of an interacting defect is the \( E_C - 0.9 \) eV level commonly observed in \( n \)-type GaN, which is shown in Fig. 4.9.

Further understanding of interacting defects is possible considering the theory of Schröter et al., which allows one to distinguish bandlike and localized states.[32] This model shows that continuous energy bands are likely to result from pure dislocation states due to wavefunction overlap. In contrast, for the case where point defects decorate a dislocation, Schröter showed that such wavefunction overlap would be minimal and a deep level associated with this type of defect formation would manifest as a localized state and not exhibit band-like character.[27, 32–34] Figure 4.10a and b shows simulation and experimentally observed behavior of bandlike levels, respectively. The bandlike level in Fig. 4.10 is distinguished by the peak shifting to lower temperatures for longer fill pulse times. As the fill pulse increases, more of the defect band fills starting furthest from the conduction band then filling available states above this.
Figure 4.9: Capture kinetics study of the $E_C - 0.9$ eV level in $n$-type GaN. Double boxcar DLTS runs were performed with fill pulse times ranging from 32 µs to 100 ms. The inset shows a plot of $\Delta C$ peak heights versus the $\log(t_{FP})$, the test for interacting behavior. The linear relationship fits the model for interacting behavior suggesting these defects are clustering (possibly around dislocations in a Cottrell cloud).

Figure 4.10: (a) Simulated DLTS spectra of interacting levels for a bandlike energy distribution from Ref. [32]. (b) Measured minority carrier bandlike distribution of states in Si$_{0.7}$Ge$_{0.3}$/Si from Ref. [35]. The simulation uses a broader distribution of states, which causes the larger distortion of the DLTS curve at larger fill pulse times. The features of the bandlike distribution are the decrease in peak temperature and increase in peak height at longer fill pulse times and the high temperature side of the curves typically nearly overlap.
The defect band can be considered as an infinite series of discrete levels with their own $E_C - E_T$ and cross-section. Assuming the cross-section is constant, the emission rate will vary exponentially with $E_C - E_T$. For a short fill pulse where only the bottom of the band fills, the DLTS will approximate the peak of a discrete level. As the fill pulse is increased, the bottom part of the band will have the same emission response but additional parts of the band will emit faster (i.e. at lower temperatures) due to a lower $E_C - E_T$. The result of this is three-fold. First, the peak temperature will shift to lower temperatures as the averaged band $E_C - E_T$ decreases for longer fill pulse times. Second, the peak height will increase because more defects are captured and emitted at the longer fill pulse times. Third, the high temperature side of the DLTS peak will nearly line up. This is because of the exponential dependence of the emission rate on temperature meaning that the bottom part of the band dominates this part of the spectra.\textsuperscript{32} On the other hand, a discrete interacting level has the same “effective” $E_C - E_T$ independent of the fill pulse length, by definition, so the emission rate is constant and therefore the peak temperature is also constant.

Because it was already determined that the $E_C - 0.9 \text{ eV}$ level in $n$-type GaN is an interacting defect, it can also be analyzed to identify if it is discrete or bandlike. The capture kinetics study in Fig. 4.9 exhibits a peak temperature shift of $< 1.5 \text{ K}$ from the 31 $\mu$s to 100 ms fill pulse times. This suggests that although the defects are close enough to Coulombically interact, the level remains discrete.

### 4.6 Supporting characterization techniques

Several other ancillary techniques have been used in this research but will not be explained in as great as detail as they are more common. Capacitance-voltage (C-V), current-voltage-temperature (I-V-T), internal photoemission (IPE), secondary ion mass spectroscopy
(SIMS), x-ray diffraction (XRD), and several scanning probe microscopy (SPM) techniques are all briefly presented.

4.6.1 Capacitance-voltage

One technique used to assess the quality and doping concentration of the diodes used for DLTS and DLOS is the capacitance-voltage (C-V) measurement. The data from the C-V curve is used with Eq. (4.7) to determine the fixed charge density (typically taken as the doping concentration) as a function of voltage. Equation (4.4) can be rearranged to determine the depletion depth \( W \) from the capacitance

\[
W = \frac{\epsilon A}{C}
\]

where \( \epsilon \) is the static electric permittivity and \( A \) is the junction area. Together the local doping concentration vs. depletion depth can be plotted. This result is compared with the expected concentration and the layer structure. The spatial resolution of C-V is limited by the Debye length \( L_D \)

\[
L_D = \sqrt{\frac{kT\epsilon_s}{q^2n}}
\]

where \( n \) is the electron concentration. If the doping varies abruptly compared with \( L_D \) the calculated concentration will not accurately replicate the actual doping. In general, the extracted concentration is actually \( N_D^+ - N_A^- + N_T \) where the trap concentration \( N_T \) is assumed to be an electron trap. If the trap concentration is significant, it can cause the extracted \( N \) to differ from the free electron concentration. Additionally, for \( p \)-GaN where Mg is used as a dopant, the free hole concentration is 1% of the Mg concentration at room temperature in the bulk due to magnesium’s large activation energy. However in the depletion region, the band bending leads to complete ionization of the Mg and hence a much smaller depletion depth than expected for the free carrier concentration. To further complicate the issue, the
large activation energy of magnesium means that its occupancy cannot respond quickly to voltage changes and has a frequency dependent response. At high frequencies, \( N \) approaches \( p \). At low frequencies, \( N \) approaches \( N_A \). All of these issues need to be considered when interpreting C-V and N vs. W plots.

C-V measurements are an easy way to quickly assess the quality of a diode for other longer experiments. As a rule of thumb, the capacitance in pF should be 10 times greater than the conductance in \( \mu \)S. This ensures that the capacitance meter can accurately measure the capacitance. If this is not the case or the capacitance scale is much smaller than the capacitance of the sample, the accuracy of the capacitance meter should be tested to determine if it is working properly. This is done by adjusting the capacitance offset by half the full scale range. If the capacitance on the meter changes by 90–110% of the offset change, one can assume the meter is accurately measuring.

4.6.2 Secondary ion mass spectroscopy

When dopant and impurity concentrations are needed, secondary ion mass spectroscopy (SIMS) is often the best choice as it can measure absolute concentrations as a function of depth. SIMS relies on an ion beam to sputter material off the sample then collects via an electric field and analyzes the ejected secondary ions using a mass analyzer. SIMS detection limits can often be in the parts per billion range. Because the matrix material and sputtering ion beam determine the ion yield rate of a species, absolute concentrations of these species is determined using a calibration standard of known concentrations. The signal is scaled by the etch rate, which is determined by measuring the etch depth after the SIMS is done. The depth sensitivity is determined by the flatness of the etching surface. Additionally, SIMS spectra
require careful interpretation as impurity profiles can tail off more slowly from the surface or interfaces than the true profiles.

### 4.6.3 Electron beam induced current

EBIC is a powerful technique to spatially resolve electrically active defects in the depletion region of a diode in plan view. EBIC is performed in an SEM using the SEM’s electron beam to spatially inject excess carriers into the diode’s depletion region. The back and front contacts are connected to a sensitive, high speed picoammeter. If the injected electrons do not recombine in the depletion region, they are swept out of the depletion region and cause an increased measured current. Electrically active defects are excellent sources for recombination and lead to a reduced current surrounding the defect. A grey scale image is then produced where black is the lowest excess current and white is the maximum. EBIC has been used extensively to count dislocations in GaN and other semiconductors.[36] EBIC has primarily been used in this research to confirm threading dislocation densities in the films to correlate with observed variations in defect densities. Figure 4.11 shows a typical EBIC image where the dark spots correlate with threading dislocation density.
Figure 4.11: EBIC image of $n$-GaN with 8 nm-thick Ni Schottky contact with threading dislocation density of $\sim 4 \times 10^7$ cm$^{-2}$.

4.6.4 X-ray diffraction

Another supporting technique is XRD, which is a powerful technique to quantitatively look at film quality. XRD maps the x-ray intensity diffracted near the Bragg angle between crystal planes in the GaN. For a good review of XRD and reciprocal space maps see Ref. 37 or 38. In this case, reciprocal space maps are made for the GaN epitaxial peak, where x-ray signal is detected throughout the sample including the nucleation layer, and both the $\omega$-relative FWHM and $\omega - 2\Theta$ FWHM are compared. Large FWHM’s in the $\omega$-relative direction are indications of mosaic structures. This technique is useful to see if different growth conditions, such as N-plasma power, impact the structural quality of the films, which would indicate unoptimized growth conditions and would have a significantly higher impact than desired variable (such as N-plasma power).
Figure 4.12: Schematic diagram of Park Systems’ atomic force microscope. The Z-scanner controls the cantilever movement and force of the tip on the sample surface while the XY-flexure scanner moves the sample under the tip. The X and Y positions are tracked to record where the Z height data is collected. The force is kept constant using a laser position detection system (not shown) as a feedback and recording the position of the Z scanner required for constant force. From Ref. 39.

4.6.5 Scanning probe microscopy

The use of atomic force microscopy (AFM) and related techniques are just beginning to developed to study GaN and AlGaN/GaN HEMTs in the EMDL group but are already showing great promise. This section briefly describes AFM, scanning Kelvin probe microscopy (SKPM), scanning capacitance microscopy (SCM), and scanning thermal microscopy (SThM) – the four standard techniques that might be the most useful.

AFM uses a cantilever with an atomically sharp tip as shown in Fig. 4.12 to profile the sample surface. The height of the cantilever is controlled via the Z-scanner and the sample is moved under the tip using the XY flexure stage in the Park Systems XE-NSOM system. Not shown in Fig. 4.12 is the laser and position sensitive photodetector system used to monitor the position of the tip. There are several modes of AFM, but contact and “true non-contact”
modes will be discussed here. The force interaction of the tip with the sample surface is attractive at first due to van der Waals force then becomes repulsive as the tip becomes closer due to electrostatic forces and is shown in Fig. 4.13.[39]

In contact mode, the tip is kept in the repulsive regime and the Z-scanner is adjusted to keep the deflection of the cantilever constant. Because the Z-scanner moves to keep the distance to the surface constant, recording the Z-scanner position allows one to directly map the surface. To the relatively high forces involved compared to the attractive regime, AFM tips are not as sharp so they have the strong enough to handle the forces.

“True non-contact mode,” as Park Systems refers to their version non-contact mode where the AFM never hits the surface as in tapping mode AFM. In non-contact mode, the tip is oscillated at the maximum slope on the side of the resonant peak. When the tip approaches the sample surface, it experiences the attractive van der Waals forces, and this causes a change in the effective spring constant, shift of the resonant frequency, and a change in the amplitude of oscillation at the operation frequency. This mode maintains the same amplitude
of oscillation by adjusting the mean sample-tip distance (i.e. constant force). The Z-scanner position is used to track the surface topography. Using non-contact mode, the tip can be atomically sharp because it never experiences the high forces of the Coulombic regime, so atomic lattice resolution is achievable in this mode. The Park AFM has many more options and abilities other than imaging topography. Because potential, capacitance, force, and other parameters vary with tip-sample distance, topography is typically measured in other modes to maintain the tip-sample distance. Scanning Kelvin probe microscopy measures the voltage potential of the sample surface using an AC voltage signal with frequency $\omega_0$ on the tip and DC bias $V_{DC}$ is adjusted to match the surface potential. A lock-in amplifier is used to measure the tip deflection amplitude at $\omega_0$, referred to as $A(\omega_0)$. When the DC voltage difference between the tip and sample approaches zero, $A(\omega_0)$ approaches zero. Therefore by dynamically adjusting $V_{DC}$ to minimize $A(\omega_0)$, the surface potential can be mapped. If the local surface defect ionization changes, a change in surface potential will result.

Another electrical AFM technique is scanning capacitance microscopy. The concept is relatively straightforward where the tip and sample form a capacitor. In theory, the capacitance could directly be measured by applying an AC voltage and measuring the out-of-phase current, but higher sensitivity can be achieved by using an inductor-capacitor (LC) resonator and only measuring a relative $dC/dV$ change. The relative $dC/dV$ is not quantitative, so comparing results between samples is not feasible. A system to measure capacitance while achieving high sensitivity is currently under development at OSU. Implementing this would allow local measurement of the deep level concentration.

Finally, scanning thermal microscopy used in thermal contrast mode allows temperature variations of $\sim 0.1^\circ$C with a spatial resolution of $< 100$ nm.[39] While a technique like this
might not be expected for defect characterization, a method like this can be used to identify regions of increased temperature that might indicate highly defective regions. These regions could indicate where a device could fail, so correlations could be made between failure spots and defect spectra.
4.7 References


Chapter 5

Studies of individual component layers of AlGaN/GaN HEMTs

Understanding the source of electrically active point and extended defects, whether incorporated during growth or generated during device operation, is central to advancement of GaN-based HEMT electronics – whose importance continues to increase. Defects are known to cause performance degradations including dispersion, current collapse, knee walkout, gate and drain lag, and even device failure.[1–4] Electrically active defects in HEMTs can be:

- intrinsic (e.g. \( V_{\text{GA}} \)) or extrinsic (e.g. \( C_N \))
- point or extended (e.g. dislocation related)
- bulk, surface, or interface related
- in the gate or access regions
- buffer, barrier, or passivation related, or
- created or migrate under high stress/high temperature operation

If the defect levels can be tied to a specific layer in the HEMT device and the physical source of the level identified before delving into the actual HEMT devices, the complex structures and myriad of potential and observe HEMT phenomena will be easier to understand. Additionally, defects are known to cause performance and reliability related issues in nitride HEMTs, light emitting diodes, laser diodes, and other nitride electrics, but it is possible to reduce total defect concentrations simply by studying the component layers and feeding back the defect information back to grower to find optimized growth conditions.
Although Chapters 6 and 7 develop methods for direct measurement of defects in HEMTs, this chapter focuses on the component layers through capacitance-based measurements. Nitride HEMTs are commonly grown using MOCVD or plasma-assisted MBE (PAMBE) growth methods but other emerging technologies such as ammonia-MBE are studied. In addition to comparing and contrasting the various growth mechanisms, several growth parameters within each growth method have also been explored. MBE and MOCVD operate on very different principles, so the first section examines the differences between these two growth methods. In the second section, the impact of the rf-plasma power of the nitrogen plasma source PAMBE growth of Ga-rich, Ga-face $n$-GaN is explored while maintaining a constant V/III beam flux ratio. Ammonia-MBE growth shows great promise to yield high quality state-of-the-art HEMTs in the future, so the third section explores a comparison between the more conventional PAMBE with ammonia-MBE growth using Ga-face $n$-GaN. Growth conditions are also very important in ammonia-MBE growth, so the fourth section explores the importance of V/III ratio using a constant Ga beam flux with increasing ammonia flow rates. Having focused on various growth conditions, the fifth section examines the impact of crystal polarity comparing traditional (0001) Ga-face growth to (000$\bar{1}$) N-face material. N-face offers many potential advantages for HEMTs, so this relatively new mode is being actively explored. The exploration of $n$-GaN finishes with a study on the impact of threading dislocations on the defect incorporation in MOCVD-grown GaN. Thick AlGaN layers for capacitance studies have been difficult to grow because of the lattice and thermal expansion coefficient mismatches, but the chapter wraps up with the defect characterization of an MOCVD-grown 30% AlGaN layer to obtain a first look at defects throughout the AlGaN bandgap.
A natural initial comparison is between the defects present in MOCVD and MBE films grown under optimal conditions for the given growth technique. Both MOCVD and MBE-growth produce high quality, state-of-the-art HEMTs, but differ greatly due to large differences in growth temperature, reaction versus kinetic limited growth mechanisms, and very different growth pressures. For these reasons, it is expected that some variation in deep level energies or concentrations will be evident between the two growth methods. The MOCVD sample was grown on sapphire using optimal nucleation conditions to achieve a nominal threading dislocation (TD) density of $6 \times 8 \times 10^8$ cm$^{-2}$. The top layer, shown in Fig. 5.1 and probed using DLOS and DLTS, was grown at $1040^\circ$C, 760 Torr, used ammonia and trimethylgallium as the precursors, and hydrogen as the carrier gas. To match the MOCVD sample as much as possible, the MBE film was grown on a nearly identical template with a nominally identical TD concentration and character to ensure any differences in defect spectra are not attributable to dislocations themselves. The MBE sample was grown at $740^\circ$C in the Ga-stable regime of...
Fig. 3.2 near the line between the intermediate and Ga-droplet regime.[5] More details of the growth chambers and other growth details and processing can be found in Chapter 3.

The near conduction band defects were studied using DLTS as shown in Fig. 5.2. For box-car DLTS, the peaks correspond to a single trap, and by plotting several different rate windows on an Arrhenius plot the activation energy and thermal cross-section can be determined. Positive peaks correspond to majority carrier traps, which are electrons in this case so the energies are referenced to the conduction band. The slope of the Arrhenius plot is the trap activation energy and the cross section depends exponentially on the $y$-intercept. The DLTS levels are strikingly similar as both samples contain levels at $E_C - 0.25$, $E_C - 0.60$, $\sim E_C - 0.90$ eV, but the concentrations and relative concentrations between defects in the same sample are quite different. Although these samples are several years old and display higher values than current state-of-the-art material, the conclusions generally remain true. These samples were chosen because the samples were grown near the same time so as not to bias the comparison. These upper bandgap states also display identical capture carrier kinetics (See Section 4.5.5 for details) in both growth methods indicating that the $E_C - 0.25$ and $E_C - 0.60$ eV levels appear to act as extended defects saturating at long fill pulse times, and the $\sim E_C - 0.90$ eV level acts as an ideal extended defect.[6] This is discussed in more depth in Chapter 4 and Section 5.6. Since no difference in defect configuration between growth methods is observed, it implies the defects are caused by identical sources that could be from common impurities or native defects. In contrast, the SSPC spectra of the MOCVD and MBE films in Fig. 5.3 differ in several respects. Positive onsets in the SSPC are indicative of majority carrier deep levels with onset energies referenced to the conduction band. The step height of the SSPC corresponds to the deep level concentration. The opposite trend to the DLTS is observed here, where the traps observed via SSPC for the MOCVD films tend to of higher concentration than the MBE films.
Figure 5.2: DLTS spectra of the MOCVD and MBE samples using a quiescent bias of $-0.5$ V and $10$ ms fill pulse of $0.0$ V. Both films exhibit the same $E_C - 0.25$, $E_C - 0.60$, and $\sim E_C - 0.90$ eV levels revealed by the Arrhenius plot but in quite different concentrations. The concentrations, determined by the magnitudes of the peaks, reveals that the MBE film has significantly higher concentrations of these near conduction band states.
Figure 5.3: The steady-state photocapacitance results for typical MBE and MOCVD material grown at the same time several years ago reveals three levels in each sample. The DLOS experiments were performed with a −0.5 V quiescent bias and 10 s fill pulse at 0.0 V. The \( E_C - (1.28 - 1.35) \) eV level, thought to be a carbon interstitial, appears in both samples. The gallium-vacancy related level at \( E_C - 2.64 \) eV is in much higher concentration in the MOCVD sample. The main difference in the spectra between the MOCVD and MBE samples is the defect near the valence band edge. In the MOCVD sample, the \( E_C - 3.22 \) eV level is associated with magnesium while the \( E_C - 3.28 \) eV level in MBE corresponds with carbon.
The $E_C - 1.35$ eV level is generally observed via both growth techniques typically in similar concentrations agreeing with the assignment of this level to a carbon interstitial as carbon impurities are common to both growth environments.[7] As opposed to the $E_C - 1.35$ eV level, which has small concentrations and broad onsets that are difficult to identify in the DLOS analysis, the cross sections of the $E_C - 2.64$, $E_C - 3.22$, and $E_C - 3.28$ eV levels can be accurately modeled to determine precise energy levels. The gallium-vacancy related level at $E_C - 2.64$ eV is present in substantially higher concentration for the MOCVD film.[8–10] This level has been attributed to gallium vacancies, so it is reasonable to expected its concentration to depend on growth mode.[8] The $E_C - 3.22$ eV trap in the MOCVD material that we have shown to be residual Mg$_{Ga}$ acceptors within background $n$-GaN, is not present for the MBE material, which does not have the same Mg memory effect.[11] The $E_C - 3.28$ level in the MBE material has been attributed to carbon on nitrogen substitutional acceptors. This level has similar energy to the $E_C - 3.22$ eV level in MOCVD material, but Armstrong et al. demonstrated that this Mg level is distinct from the C level common in MBE material using DLOS analysis.[7]

The general trends of MBE films compared with MOCVD films of similar structure are higher concentrations of near conduction band state, lower concentration of the gallium-vacancy related level, and a carbon deep acceptor as opposed to the magnesium deep acceptor in MOCVD material. The most striking similarity between the growths though is the similarity of the deep level energies. This suggests that these levels are due to impurities common to both MBE and MOCVD, common structural defects, or intrinsic defects. With this brief introduction to deep level spectra in two distinct growth methods, we now focus on MBE growth parameters starting with the N plasma power.
5.2 Study of deep level incorporation in PAMBE $n$-GaN varying the nitrogen source plasma power

Continuing the study of defects in GaN as a subset of the AlGaN/GaN HEMT device, the focus turns to growth conditions in MBE. It is possible to evaluate possible improvements to the growth process including increased growth rate, which is the focus of the this section. Later sections will examine and compare the MBE growth of GaN using an ammonia nitrogen source instead of the typical RF plasma source examined now.

There has been significant progress in the performance of high-frequency, high-power AlGaN/GaN high electron mobility transistor (HEMT) structures grown by molecular beam epitaxy (MBE) in recent years.[12–15] Plasma-assisted MBE (PAMBE) growth of nitride devices in which an RF plasma is used to provide reactive nitrogen during III-nitride epitaxy is currently the dominant MBE method and recently reported HEMT performance with a power added efficiency of 64% with an output power density of 12 W/mm at 48 V in Class AB operation.[15] Despite record performance, the growth rate of PAMBE-grown films remains low (∼0.2 $\mu$m/hr) and increasing the growth rate is highly desirable to compete with other growth methods. The impact of deep level incorporation with increased growth rate is not well studied but significantly important to predicting device performance as deep levels have been associated with the cause of dispersion, knee walkout, and performance degradations.[1–3] PAMBE growth is typically performed with a low V/III near the onset of Ga droplet formation, so the growth rate is dictated by the nitrogen incorporation rate, which is opposite typical III/V growth where the Group III arrival rate determines the growth rate. To achieve higher growth rates in this mode, an increased flux of reactive nitrogen is required, which can be accomplished with increased plasma power. In fact, growth rate is approximately linearly
correlated with rf-plasma power over the range of plasma power studied. However, the nitrogen molecular beam from rf-plasma sources contains a complex mixture nitrogen species that whose ratio and energy changes with the rf plasma power.[16] Previous studies have suggested that some nitrogen species might introduce significant concentrations of point defects,[17] so this study attempts to identify of individual levels and concentrations as a function of nitrogen rf-plasma power in PAMBE-grown Si-doped GaN. A combination of deep level optical spectroscopy (DLOS) and deep level transient spectroscopy (DLTS) is employed to probe deep levels in Schottky diodes and to cover the entire GaN bandgap for changes in the deep level spectrum.

5.2.1 Experimental

In order to characterize the impact of the nitrogen plasma power in rf-plasma-assisted MBE, a series of GaN layers were grown by rf-plasma MBE using three rf plasma power settings of 150, 300, and 400 W. In each case, the Ga beam flux was adjusted to maintain nominally identical V/III flux ratios and were all grown at 740°C to maintain optimum growth conditions near the crossover between the intermediate and Ga droplet growth regime of the GaN growth phase diagram.[18] Under these conditions, the growth rate is determined by the nitrogen arrival rate, and the measured growth rates were 0.21, 0.42, and 0.5 µm/hr as a function of increasing N plasma power for the three nominal setting used, which were determined by RHEED intensity oscillations. Additionally, to minimize the impact of differences in dislocation character and density in interpreting the deep level results, all three samples were grown on identical Si-doped ultra-low dislocation density Lumilog substrates, which were doped to ~1×10^{18} cm^{-3} and possessed a nominal threading dislocation density (TDD) of 6–8×10^{7} cm^{-2}. Growth was initiated directly on the substrate
using the respective plasma power and each test layer was intentionally doped with Si at a concentration of \( \sim 8 \times 10^{16} \text{ cm}^{-3} \) to facilitate high sensitivity trap spectroscopy measurements, and all test layers were \( \sim 0.6 \mu\text{m} \) in thickness. After MBE growth, the samples were processed into Schottky diodes using a mesa structure to form the Ohmic contact to the highly doped substrate. An 80 Å thick Ni layer was used to facilitate a semi-transparent Schottky contact to enable light from the DLOS experiment to penetrate the GaN layer while the unannealed Ti:Al:Ni:Au Ohmic contact was deposited after a Cl-based RIE etch. These standard DLOS structures have been described in detail elsewhere.[8, 19]

Possible correlations between deep levels and the RF plasma power used to supply active nitrogen was monitored using DLOS in both transient and steady state (steady state photocapacitance - SSPC) modes, and by DLTS. Prior work has shown that the combination of thermally stimulated DLTS and optically stimulated DLOS enables quantitative characterization of deep levels throughout the entire GaN bandgap.[20, 21] DLTS was performed using a standard double boxcar method, which for these \( n \)-type GaN Schottky samples can detect traps present within \( \sim 1 \text{ eV} \) of the conduction band edge, over a temperature range from 300 to 400 K. DLOS and SSPC utilize a tunable, monochromatic incident light ranging from 0.5 eV to \( \sim 6.0 \text{ eV} \) to optically stimulate deep level emission throughout the bandgap by illuminating through the semi-transparent Ni Schottky metal. All DLOS and SSPC measurements were performed at 300 K using a \( -0.5 \) V quiescent reverse bias, followed by a 10 s, 0.0 V pulse to fill states. Optical stimulation commenced after a delay of 45 s in order for thermal transients due to carrier emission from shallow states to be completed. Deep level concentrations were determined from the measured step change in the SSPC once a level is stimulated. Precise energy levels are found from detailed analysis of the photocapacitance transient measured for each value of incident light energy, from which optical cross sections can be
obtained. Fitting the optical cross section to known models[22, 23] provides complete details
of the deep level energy (i.e. optical ionization energy and local lattice relaxation effects). Comprehensive reviews of DLOS and SSPC can be found in Chapter 4 and in Refs. 24 and 25. The trap spectroscopy studies were augmented by secondary ion mass spectroscopy (SIMS), triple-axis x-ray diffraction (XRD), and electron beam induced current (EBIC) measurements in order to address the issue of physical sources for observed levels.

5.2.2 Deep and minority carrier traps detected using deep level optical spectroscopy

Figure 5.4a shows the SSPC spectra obtained for each of the GaN samples grown using rf-plasma power levels of 150 W, 300 W and 400 W for the N source. Several features are of immediate note. First, in spite of the variation in the N plasma power, each sample displays very similar spectral features as indicated by the SSPC onset energies, suggesting that no new physical defects are introduced into the region of the bandgap that is characterized by DLOS, at least down to the DLOS detection limit of the upper $10^{12}$ cm$^{-3}$ range for these samples. Second is the strong, monotonic increase in the concentrations of these deep levels as indicated by the increased SSPC step heights at the onset energies at $E_C - 1.35$, $E_C - 2.60$ and $E_C - 3.28$ eV. Figure 5.4 shows the optical cross sections for each of these states obtained from the transient DLOS analysis, which provides precise assignment of the energy levels for each trap by accounting for Franck-Condon shifts due to local lattice relaxation.[23] Prior work has established strong correlations for each of these states with distinct physical sources in MBE-grown GaN. The often-reported $E_C - 2.60$ eV level is associated with gallium vacancy defects in $n$-type GaN that may also partially complex with residual hydrogen.[8–10] The $E_C - 3.28$ eV level has been associated with residual carbon defects that incorporate as substitutional C$_N$ acceptors that have been shown to partially compensate $n$-type GaN grown
Figure 5.4: (a) Steady-state photocapacitance spectra of rf-plasma assisted MBE-grown \textit{n}-GaN Schottky diodes grown with 150 W (blue dashed line), 300 W (black solid line), and 400 W (red dotted line) plasma powers at 300 K using a -0.5 V reverse bias and 10 sec fill pulse of 0.0 V. The two carbon associated levels at $E_C - 1.35$ and $E_C - 3.28$ and gallium vacancy-related level at $E_C - 2.60$ eV show reduced defect concentrations at lower plasma power. The inset shows the same data for the sample grown at 150 W on its own scale for clarity. (b) DLOS spectra corresponding to the same SSPC data for the 150 W (blue ▼), 300 W (solid ■), and 400 W (red ▲). The fits to the data (lines) were generated using Lucovsky’s model and show the three detected deep levels for each sample.
Table 5.1: Table of trap energy levels and concentration in MBE-grown $n$-GaN for varying nitrogen rf-plasma power.

<table>
<thead>
<tr>
<th>N Plasma Power</th>
<th>$E_C - 0.59$ eV</th>
<th>$E_C - 0.86$ eV</th>
<th>$E_C - 1.35$ eV</th>
<th>$E_C - 2.60$ eV</th>
<th>$E_C - 3.28$ eV</th>
</tr>
</thead>
<tbody>
<tr>
<td>150 W</td>
<td>$1.0 \times 10^{13}$</td>
<td>$1.4 \times 10^{13}$</td>
<td>$3.0 \times 10^{13}$</td>
<td>$1.1 \times 10^{14}$</td>
<td>$1.5 \times 10^{14}$</td>
</tr>
<tr>
<td>300 W</td>
<td>$1.6 \times 10^{14}$</td>
<td>$1.0 \times 10^{14}$</td>
<td>$1.4 \times 10^{14}$</td>
<td>$2.67 \times 10^{15}$</td>
<td>$3.16 \times 10^{15}$</td>
</tr>
<tr>
<td>400 W</td>
<td>$6.8 \times 10^{13}$</td>
<td>$\sim 1 \times 10^{13a}$</td>
<td>$7.0 \times 10^{14}$</td>
<td>$7.78 \times 10^{15}$</td>
<td>$7.82 \times 10^{15}$</td>
</tr>
</tbody>
</table>

*a This level is obscured by a higher temperature peak, so only an approximate concentration can be determined.

by rf plasma MBE.[7] The state at $E_C - 1.35$ eV is also often observed and its source is more tentatively correlated with carbon interstitial defects.[7] From these prior assignments, it appears that an effect of the increasing plasma power for the nitrogen source is to impact the formation of gallium vacancies and perhaps the background concentration of carbon impurities, though SIMS results obtained from each of the samples were inconclusive, with a nominally constant carbon concentration being measured of $\sim 4 \times 10^{16}$ cm$^{-3}$. However, it must be realized that the SSPC/DLOS sensitivity is much greater than that of SIMS and as seen from Table 5.1 the carbon related trap concentrations are a factor of 10 to 100 less than the total carbon concentration. Hence, small changes in carbon concentration undetected by SIMS or changes in the way in which carbon becomes incorporated into the growing GaN film may explain the observed strong dependence of carbon related deep levels on the N plasma power.

The assignment of point defects involving gallium vacancies and carbon-related complexes as primary defects that are influenced by N plasma power is further substantiated by other observations. Figure 5.5 shows plan view images obtained from electron beam induced current (EBIC) measurements made on all three samples, confirming that the dislocation density remains unchanged and constant for all three samples at $8 - 9 \times 10^7$ cm$^{-2}$, which is
very close to the dislocation density values of the underlying Lumilog substrates. The lack of variation in dislocation density is further verified by high resolution x-ray diffraction measurements that showed no observable variation in FWHM for both 2θ and ω scan directions. So the question of why gallium vacancies and carbon deep levels are sensitive to the N source plasma power must be asked. The answer is complicated by many factors that are influenced by the plasma source, some of which cannot be easily decoupled. One obvious direction to consider lies with the N beam flux, and its influence on both growth rate of V/III beam flux ratio. As already mentioned above, the epitaxial growth rate increases with an approximately linear dependence on plasma power in the Ga stable growth regime used here. While a higher growth rate itself may well lead to higher gallium vacancy concentration due to reduced adatom mobility on the growth surface, changes in defect concentration due to changes in V/III beam flux ratio is unlikely here since the gallium beam flux was adjusted to maintain a constant MBE growth condition at the crossover point between the Ga droplet and intermediate regimes of the MBE growth phase diagram for all samples.[5] Even if small variations in beam flux ratio did occur, our earlier work has shown only a weak sensitivity of the formation of gallium vacancies and C_N substitional defects on beam flux ratio with a maximum change of ~ 5 times over a fairly large range of beam flux ratios and surface conditions.[20] Here we instead observe variations in concentrations of these same deep levels by as much as 20 – 50 times, so small, inadvertent flux ratio changes cannot be ascribed to the large variations in deep level concentrations observed here. It is likely that the increase in point defect concentration is related to more complex phenomena. Newman demonstrated that not only does plasma-assisted MBE growth of GaN occurs under conditions where the thermodynamically stable phase is not GaN.[26] For typical growth pressures and temperatures, GaN growth is a meta-stable process controlled by the forward and reverse reactions, which are limited by the
Figure 5.5: EBIC performed on the samples grown with (a) 150 W, (b) 300 W, and (c) 400 W nitrogen source rf-plasma powers show nominally identical threading dislocation densities (TDD) of $6-8 \times 10^7 \text{ cm}^{-2}$. The substrate (TDD) is typically $8 \times 10^7 \text{ cm}^{-2}$ and a maximum of $9 \times 10^7 \text{ cm}^{-2}$. The epilayers show no indication of excess TDD suggesting that the nucleation of growth at the various plasma powers did not contribute to any additional threading dislocations. The lines in the images are due to noise from a low signal-to-noise ratio and settings to enhance the poor contrast. EBIC performed by J. Boeckl.
arrival rate of activated nitrogen species and the large kinetic barrier to decomposition.[26] In addition to the competition between growth and decomposition, the incident energy of the constituent nitrogen molecules, atoms, and ions in the molecular beam are well in excess of the kinetic barrier and Gibbs free energy required for formation of GaN.[26] Study of the EPI nitrogen plasma source have indicated that metastable molecular nitrogen is the dominant activated nitrogen species in the molecular beam and that ionic nitrogen atom concentration is \(< 3 \times 10^{-5}\%\).\[16, 27\] As the rf-plasma power is increased, the ionic content of the molecular beam tripled while the concentration of atomic nitrogen flux more than doubled. Both of these constituents have been shown to decrease mobility, growth rate through enhanced decomposition, and suspected of increasing point defect concentrations.[16, 17, 28] The increased trap concentrations of the \(E_C\)−1.35, 2.60, and 3.28 eV deep levels for the higher plasma powers is consistent with the previously suspected increased point defect concentration caused by the increased concentration or energy of atomic/ionic nitrogen. Ptak \textit{et al.} determined that the ionic and atomic nitrogen flux only increased by a factor of 2-3 when the plasma power of the EPI source was increased by a factor of 2. For the \(E_C\)−3.28 eV level, for example, the concentration increased by a factor of \(\sim 21\) from 150 W to 300 W and \(\sim 52\) from 150 W to 400 W, which is a significantly larger rise than the ionic and atomic nitrogen flux increases. If the increase in damaging nitrogen flux was the dominant source of the increased trap concentration, one would expect the a one-to-one correlation between flux and trap concentration increase, but this is not what is observed. It is possible that the increased growth rate could be responsible for the increased defect concentrations due to growth surface chemistry changes such as reduced adatom mobility, but the results of Lee \textit{et al.} suggest that the Ga adatom mobility increases with growth rate.[28] While this cannot completely rule out the effects of growth rate, a more feasible solution exists. If the energy
Figure 5.6: Trap concentrations plotted against rf-plasma power in an Arrhenius relationship exhibits a linear relationship for the $E_C-1.35$ (red ▼), $E_C-2.56$ (green ■), and $E_C-3.28$ (black ▲) eV deep levels between samples grown at 150, 300, and 400 W. The superlinear change in trap concentration of these level suggests that the linear increase in growth rate, potential small variation of V/III ratio, and any change in the atomic and ionic nitrogen atom flux in the nitrogen molecular beam during growth are not responsible for the observed relationship. The connection between the plasma power and deep level concentration suggests that changing energy of the constituents in the molecular beam are likely responsible for the observed behavior.

of the ionic or atomic nitrogen atoms increased with plasma power as one might expect, an energy activated process would be observed. In this case, an Arrhenius-style relationship between trap concentration and rf-plasma power might be expected. The following relation

$$ N_T = A \exp \left( -\frac{P_a}{P} \right) $$

(5.1)

where $A$ and $P_a$ are constants, $P$ is the rf-plasma power, and $N_T$ is the trap concentration was used to fit the data and is shown in Fig. 5.6. This suggests that the dominant source of the
increased point trap concentration of the probable interstitial carbon, gallium vacancy related, and substitutional carbon defects results from increased energy of constituent nitrogen species.

5.2.3 Traps in the upper GaN bandgap detected using DLTS

Since DLOS measurements in \textit{n}-type GaN do not easily detect traps near the conduction band due to deep level thermalization competing with optical emission, conventional DLTS measurements were made on the same samples as a function of the N plasma power used in MBE growth. As was the case for the very deep states, DLTS did not reveal new states as a function of plasma power, and all detected levels ($E_C - 0.59$ eV and $E_C - 0.86$ eV) have been widely reported in earlier works.[19, 20, 29] However, unlike the deep and near-valence band states associated with gallium vacancies and carbon point defects that were detected by DLOS, the DLTS-detected traps at $E_C - 0.59$ and $E_C - 0.86$ eV revealed no systematic dependence on N plasma power. Figure 5.7 shows the DLTS spectra obtained for the three samples where deep levels at $E_C - 0.59$ and $E_C - 0.86$ eV are noted, along with the corresponding DLTS Arrhenius plots that provide the energy levels for these states. The concentrations of these states as well as those detected by DLOS/SSPC are summarized in Table 5.1, for all plasma power values used in this study. The concentration of the $E_C - 0.59$ eV level erratically varied with N plasma power, similar to its behavior in samples grown as a function of well-controlled Ga/N beam flux ratio in earlier work.[20] The $E_C - 0.86$ eV state also displayed no real dependence on plasma power but for this case the change in concentration from sample to sample was virtually negligible (though the appearance of a high temperature shoulder for the 400 W sample that is evident for the shorter rate window in the DLTS experiment shown in Figure 5.5b implies the introduction of another deep level, making it difficult to quantify.
Figure 5.7: DLTS spectra of the rf-plasma assisted MBE-grown $n$-GaN Schottky diodes grown with 150 W (blue ▼ and line), 300 W (black ■ and line), and 400 W (red ▲ and line) plasma powers using a $-0.5$ V reverse bias and 10 ms fill pulse of 0.0 V with (a) 80 s$^{-1}$ and (b) 2 s$^{-1}$ rate windows. For the 400 W plasma power, a high temperature peak masks the peak associated with the $E_{C} - 0.86$ eV level and because a peak is not observed in enough rate windows, an energy was not obtainable in this case. Despite this, both the $E_{C} - 0.86$ and $E_{C} - 0.59$ eV levels exhibit no systematic dependence on plasma power. (c) Arrhenius plot for the different deep levels corresponding to the peaks from each rate window for each sample.
the true concentration of the $E_C - 0.86$ eV state in that sample). These results suggest that whatever the physical sources for these deep levels, they appear to be generally insensitive to the N plasma source power conditions. Earlier work to address their potential sources showed both of these levels to display the logarithmic trapping kinetics behavior that is associated with dislocation-point defect complexes as the physical source for both traps.\cite{6, 29–31}

Though the specific structure of both complexes are not known, what is implied is that the near valence band and midgap traps detected by DLOS, which are associated with simple point defect sources such as gallium vacancies and carbon substitutional impurities, are far more sensitive to the N plasma source conditions for $n$-type GaN than are traps associated with dislocations. This suggests that the simple but deep point defects observed via DLOS in $n$-type GaN all depend strongly on N plasma source power and its effects on the growth surface conditions, whereas dislocations, even if they are decorated by point defects, do not appear to be strongly sensitive to N plasma power.

### 5.2.4 Conclusions

In summary, DLOS and DLTS measurements of $n$-type GaN grown by plasma assisted-MBE revealed strong and differing dependencies of individual deep levels on rf-plasma power. The carbon and gallium vacancy-related point defect levels at $E_C - 3.28$, $E_C - 2.60$, and $E_C - 1.35$ eV displayed the strongest impact of rf-plasma power with an Arrhenius-like incorporation relationship suggestive of plasma damage caused by the change in incident energy of the constituent nitrogen species and/or a change in the ratio of the constituent species. In contrast, the upper bandgap levels at $E_C - 0.59$ and $E_C - 0.86$ eV exhibit no correlation with rf-plasma power for the range of plasma power studied here. Total trap concentrations were increased by a factor of $\sim 20$ and $\sim 50$ for the 300 W and 400 W samples, respectively, compared
to the 150 W sample. Because in the Ga-rich growth regime the growth is limited by the arrival rate of active nitrogen and the increased plasma power results in more active nitrogen, this study shows that with currently optimized growth conditions that higher plasma powers or higher growth rates result in lower quality films in terms of defect concentrations. These results suggest that increased growth rates with low defect densities may possibly be achieved by reducing any ionic nitrogen species and atomic nitrogen species in the higher power plasma output.

5.3 Impact of nitrogen source on deep level spectra in $n$-GaN grown by MBE

Use of an RF plasma source is not the only source of nitrogen in nitride MBE as ammonia, the nitrogen source in MOCVD growth, can also be used in MBE growth. Understanding deep level incorporation is an important metric of the potential to grow high quality GaN using an ammonia nitrogen source as opposed to the more widely-studied technique of RF nitrogen plasma-assisted MBE. Ammonia-based GaN MBE growth is now receiving great interest for device growth applications since it offers numerous potential advantages to RF-plasma-based MBE growth including: higher growth rates, improved film uniformity, no Ga-droplets, and much larger parameter space in which growth conditions may be optimized.[32] However, compared to both plasma-based MBE and MOCVD growth of nitride materials and devices, ammonia based MBE growth is at an early stage. Here we focus our study on the introduction of trap states in ammonia-based MBE-grown GaN layers in order to assess the nature and concentration of electrically active defects in this promising class of GaN epitaxy. Comparing the defect spectra between plasma and ammonia nitrogen sources will allow detection of new defect sources in ammonia-MBE growth as well as changes in concentration of common defects due very different conditions (N-rich for ammonia MBE and Ga-rich for PAMBE),
differences of the growth surface, differences in growth temperature, and excess hydrogen.

Knowledge of the defect spectrum will rapidly assist the development of devices since the role of deep levels as traps and recombination-generation centers is central to the performance of both electronic and optoelectronics. This section focuses on a comparison of deep level incorporation of ammonia and PAMBE MBE growth of GaN.

### 5.3.1 Experimental

To characterize the differences between trapping spectra within $n$-Ga$\text{N}$ grown by ammonia-based and N-plasma based MBE, nominally identical test structures were grown by both methods on identical GaN templates. This allowed us to minimize any effects from differences in threading dislocation character and density, and other substrate effects in this exploratory study. The template was a commercial standard $n$-type Lumilog substrate doped with Si at a concentration of $\sim 3 \times 10^{18}$ cm$^{-3}$ and a threading dislocation density of $\sim 5 \times 10^8$ cm$^{-2}$. For the N-plasma generated samples, the initial 200 nm thick layer was grown at 715°C near the crossover from the intermediate to Ga droplet growth regimes followed by a 500 nm thick $n$-type (Si) doped layer with a nominal concentration of $5 \times 10^{16}$ cm$^{-3}$, to support DLOS and DLTS experiments, following our previously reported growth conditions.[20] The ammonia samples were grown at a substrate temperature of 770°C using two different nominal V/III ratios of 750 and 1500, which were achieved by varying the ammonia flux during growth. This enabled a more in-depth evaluation of factors that may influence trap introduction for the ammonia-based MBE GaN material. It should be noted that the growth conditions used when performing MBE with nitrogen supplied by an ammonia source and a plasma source are quite different to result in optimal film properties, most notably being that the ammonia source enables N-rich growth conditions, more similar to that of conventional III-V MBE.
growth. After growth, samples were processed into DLOS test structures that are also used for DLTS measurements, so that the entire bandgap can be accessed for trap spectroscopy. Full details of the fabrication procedures can be found in Chapter 3.

5.3.2 Results and discussion

The normalized SSPC spectra shown in Fig. 5.8 reveal the presence of similar, multiple deep levels for both the higher V/III ratio ammonia-based MBE and N-plasma MBE samples with onset energies at 1.28, 2.62, and 3.28 eV. All of these onsets are positive indicating that their energy is referenced to the conduction band. Previous studies of plasma MBE-grown GaN have attributed this $E_C−3.28$ eV level to the $C_N$ substitutional defect that acts to compensate $n$-type films.[7] The $E_C−2.62$ eV level has been seen in both MBE and MOCVD films and has been previously correlated to $V_{Ga}$ or complexes of $V_{Ga}$−$H$.\[11, 33\] In MOCVD material, recent positron annihilation spectroscopy work has shown significant concentrations of the $V_{Ga}$−$2H$ complex, which combined with the theoretical work of Neugebauer et al. suggests the $E_C−2.62$ eV level is likely the $V_{Ga}$−$2H$ complex.\[33, 34\] The onset at $E_C−1.28$ eV matches well with known levels with onsets reported at $E_C−1.3−1.35$ eV that have appeared to be linked to carbon as an interstitial defect for both MOCVD and MBE material.\[7\] The similarity between the deep level spectral content for both ammonia and plasma based MBE grown samples demonstrates that the use of an ammonia source to supply nitrogen, even with very different growth conditions, does not appear to introduce new states in the bandgap range probed by SSPC/DLOS, i.e. from the valence band edge up to $E_C−1$ eV, that are above the detection limit of these measurements (around $1−5×10^{12}$ cm$^3$ in these low doped layers). To access the upper $\sim$1eV of the $n$-GaN bandgap, DLTS measurements were performed. Unlike the SSPC/DLOS experiments, DLTS reveals spectral differences between N-plasma and
ammonia-based MBE materials, indicating the presence of different deep levels. Figure 5.9 shows a comparison of DLTS data obtained from the higher ammonia flux MBE sample and from the N-plasma MBE sample (noting the 5x expansion applied to the rf-plasma MBE data in order to better compare spectral features). Besides the obvious differences in trap concentrations, to be discussed below, analysis of the ammonia-based MBE sample data indicates a newly observed level at $E_C - 0.40$ eV, but with a concentration that is near the DLTS detection limit. Further studies are needed to identify potential sources of this level. The shallowest, $E_C - 0.24$, and deepest, $E_C - 0.72$ eV, levels measured via DLTS appear in both samples while the commonly observed $E_C - 0.60$ eV level that has been attributed to point defects that potentially cluster around dislocations also appears in both samples but with a small shift in peak temperature as seen in Fig. 5.9.[20]
Figure 5.9: DLTS spectra of $n$-GaN Schottky diodes grown using ammonia (square) and rf-plasma (triangle) nitrogen sources (note separate axes for the ammonia and rf-plasma spectra where the rf-plasma spectrum is 5x relative to the ammonia). The $E_{C-0.25}$ eV $V_N$-related level is unaffected by the N source in optimized conditions while the $E_{C-0.60}$ eV level shows the largest change with a ~20X reduction in the rf-plasma sample. An unidentified level at $E_{C-0.40}$ eV is also present in the ammonia-MBE sample at low concentration. Inset shows the Arrhenius plot for the different deep levels observed in the ammonia and rf-plasma samples.

The quantitative concentrations of traps detected by DLOS and by DLTS are summarized in Table 5.2. Both the N-plasma MBE results and earlier results obtained for high quality MOCVD (metalorganic chemical vapor deposition) grown $n$-GaN are listed in addition to the ammonia MBE result to provide an appropriate viewpoint regarding the quality of the ammonia-MBE materials. The primary differences between N-plasma and ammonia-based MBE-grown $n$-GaN are manifested in increased concentrations for three distinct levels at $E_{C-3.28}$ eV, $E_{C-2.6}$ eV and $E_{C-0.60}$ eV, for the ammonia-MBE GaN. These are often reported levels and have been attributed to, respectively, background carbon that is incorporated as a $C_N$ acceptor, the gallium vacancy $V_{Ga}$, and an unknown point defect source.[7, 8, 35] These higher concentrations may not be surprising given the comparatively little amount of work done to optimize ammonia-MBE growth to date, but again, to put this in a proper
Table 5.2: Summary of deep levels and concentrations in $n$-GaN films grown using various growth conditions. The MOCVD results were previously reported in Ref. 7.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Deep level Concentration (cm$^{-3}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$E_C$=−3.28</td>
</tr>
<tr>
<td>N-plasma MBE</td>
<td>$3.0 \times 10^{14}$</td>
</tr>
<tr>
<td>Higher NH$_3$/Ga flux</td>
<td>$1.7 \times 10^{15}$</td>
</tr>
<tr>
<td>Lower NH$_3$/Ga flux</td>
<td>$3.1 \times 10^{15}$</td>
</tr>
<tr>
<td>MOCVD$^b$</td>
<td>$2.0 \times 10^{16}$</td>
</tr>
</tbody>
</table>

$^a$ The energy of these levels was slightly higher at 0.62 eV and the cross-section was $\sim 5X$ higher at $2 \times 10^{-14}$ cm$^2$.
$^b$ MOCVD levels at $E_C$=−3.22, $E_C$=−2.5−2.6, $E_C$=−1.35, $E_C$=−0.60 and $E_C$=−0.25 eV, respectively. Note that the $E_C$=−3.22 eV state is due to residual Mg acceptors, and overwhelms the presence of residual carbon acceptors.[7]

perspective one can see that in spite of this relative immaturity that the trap concentrations are measurably lower than very advanced, high quality MOCVD $n$-GaN as shown in Table 5.2. Moreover, high performance electronic devices have been demonstrated using ammonia-MBE despite the relative immaturity of the field and the somewhat defect concentrations compared to plasma-assisted MBE suggests that further optimization that might push device performance far beyond the already promising current levels.

5.3.3 Conclusion

The impact of the nitrogen source in MBE growth for $n$-GaN was explored to better understand the role of defects in ammonia-based MBE growth. The DLOS results show that background carbon and gallium vacancies are the dominant defects in ammonia-based MBE as it is in N-plasma MBE. Similarly, the DLTS results also show two common dominant peaks at $E_C$=−0.24 and $E_C$=−0.60 eV. Varying the V/III ratio in the ammonia-based MBE films showed a strong dependence of the $E_C$=−0.24 eV level on V/III flux ratio, which is explored further in the next section. Identifying all the electrically active defects in the bandgap of materials
is not only useful for later general defect measurements directly on HEMTs, but potentially also for later HEMT comparisons of the ammonia-MBE and PAMBE HEMTs to help identify differences in dispersion.

5.4 Correlation of deep levels in ammonia-MBE $n$-GaN grown with increasing V/III flux ratio

In the previous section, the best optimized ammonia-MBE-grown $n$-GaN was compared with more traditional, optimized PAMBE growth, showing that nearly identical deep levels were observed in both materials. The defect incorporation was also systematically investigated for increasing ammonia flow rates using a fixed Ga beam flux resulting in a series of three samples grown at increasing V/III beam flux ratios. Early in the growth optimization process, it was important to correlate defect introduction with growth parameters to assist in material optimization and facilitate performance improvements in HEMTs. Already though, ammonia-MBE AlGaN/GaN HEMTs have shown excellent power and efficiency performance, reaching an output power density of 11.2 W/mm and recently reported power-added efficiency (PAE) of 70.3% at 10 GHz.\[36–38]\ An output power density of 7.9 W/mm at 30 GHz, PAE of 24\%, and extrinsic $f_{\text{MAX}}$ of 116 GHz were the first millimeter-wave measurements of an ammonia-MBE HEMT though the low reported PAE was partially because of the difficulty of impedance matching because the maximum available load reflection coefficient was reached.\[39, 40]\ Despite these early successes, only limited information on the presence and properties of deep level defects in ammonia-MBE-grown GaN have been reported, and what has been reported is limited to the exploration of traps within $\sim$1 eV from the GaN conduction band edge.\[41, 42]\ Just as was the case for GaN grown by plasma-assisted MBE,\[20]\ knowledge of traps existing anywhere within the 3.4 eV GaN bandgap and how their presence depends on growth parameters is expected to assist the continued development of ammonia-MBE grown
GaN materials and devices. This study focuses on the detection and analysis of GaN bulk trap states and their dependence on systematically-varied NH$_3$/Ga flux ratios in ammonia-based MBE growth through the application of deep level optical spectroscopy (DLOS) and related methods.

GaN samples were grown using ammonia-MBE growth with two different NH$_3$/Ga beam flux ratios to systematically study the impact of V/III beam flux ratio on deep level incorporation. The gallium flux was maintained at constant beam equivalent pressure of $2.0 \times 10^{-7}$ Torr while the ammonia flux was $3.7 \times 10^{-4}$ Torr and $2.0 \times 10^{-4}$ Torr (beam fluxes were measured by an ion gauge in the growth position) for the higher and lower NH$_3$/Ga flux ratio samples, respectively. These growth conditions result in an N-rich growth environment and a constant Ga-limited growth rate of $\sim 300$ nm/hr. All samples were grown on identical Lumilog standard GaN template substrates to minimize ambiguity in this comparative study. The Lumilog template consisted of a sapphire substrate with a $\sim 3$ µm GaN overlayer doped with Si to a concentration of $\sim 3 \times 10^{18}$ cm$^{-3}$ and had a threading dislocation density of $\sim 5 \times 10^8$ cm$^{-2}$. The growth structure consisted of an initial 200 nm thick n$^+$ GaN lateral conduction and back contact layer that was Si-doped to $\sim 3 \times 10^{18}$ cm$^{-3}$, followed by an 0.5 µm thick n-type GaN layer grown at a constant substrate temperature of 770°C, and doped with Si to an electron concentration of $\sim 5 \times 10^{16}$ cm$^{-3}$. To facilitate DLOS and deep level transient spectroscopy (DLTS) measurements of the lightly n-type layers, devices were processed into 300 µm square Schottky diodes using semitransparent 80 Å-thick Ni contacts with mesa etch isolation using Cl$_2$-based reactive ion etching. The Ohmic contacts were formed in the mesa etch trenches to the underlying n$^+$ GaN layer using Ti:Al:Ni:Au (100:2000:200:3000 Å). A sufficiently low series resistance of $\sim 50$ Ω for the diodes was achieved without need to anneal. High Schottky barrier heights of 1.08 eV and 1.03 eV (by internal photoemission
spectroscopy), low room temperature reverse leakage current densities (at –1 V reverse bias) of \( \sim 1 \times 10^{-10} \text{ A/cm}^2 \) and \( \sim 1 \times 10^{-9} \text{ A/cm}^2 \), and forward-biased ideality factors of 1.08 and 1.10 with over 7 decades of linearity were measured for the higher and lower NH\(_3\)/Ga flux ratios, respectively, demonstrating the necessary quality for these diodes to provide accurate comparison via deep level characterization. Moreover, a uniform background doping concentration of \( \sim 5 \times 10^{16} \text{ cm}^{-3} \) was determined via C-V measurements that revealed ideal \( 1/C^2 \) versus V behavior. To investigate defects throughout the GaN bandgap, a combination of DLOS, steady-state photocapacitance (SSPC) and DLTS measurements were used. The DLOS/SSPC was conducted at 300 K using a Xe lamp coupled to a monochromator with 0.02 eV resolution and capable of photon energies between 1.2 and \( \sim 5.5 \) eV. Precise deep level energies were determined by calculating the optical cross section \( \sigma^o \) from the extracted emission rate and measured photon flux and fitting the resulting curve to the theoretical Lucovsky model,[22] which produced excellent fits to the optical cross section without the need to account for lattice relaxation effects via the Franck-Condon shift using the model developed by Chantre et al.[23] To access the near-conduction band region of the bandgap, conventional DLTS was employed. DLOS, SSPC, and DLTS measurements were performed using a quiescent and fill pulse biases of –0.5 V and 0.0 V, respectively. The fill pulse time was 10 s for the DLOS with a 300 s measuring time and 10 ms fill pulse for the DLTS measurements. Full details of the fabrication procedures and measurement setup can be found in Chapter 3.

Figure 5.10 shows the SSPC spectra obtained for both NH\(_3\)/Ga flux ratios, which bear close spectral similarity to prior results for \( n \)-type GaN grown by both plasma-assisted MBE and by metalorganic chemical vapor deposition (MOCVD),[8, 9, 11, 20] indicating the mid-gap to near valence band levels detected by DLOS for the different epitaxial growth methods do not appear to cause the incorporation of different sources of deep levels (whether extrinsic
Figure 5.10: Steady-state photocapacitance spectra of ammonia-based MBE $n$-GaN Schottky diodes for higher (blue dashed line) and lower (red solid line) NH$_3$/Ga flux ratios at 300 K using a −0.5 V reverse bias and 10 sec fill pulse of 0.0 V. The two carbon associated levels at $E_C-1.28$ and $E_C-3.28$ show reduced defect concentrations at the higher NH$_3$/Ga flux ratio while the concentration of the gallium vacancy-related level at $E_C-2.62$ eV is insensitive to NH$_3$/Ga flux ratio. Note the spikes at ∼1.5 eV are due to the Xe lamp source. The inset shows the optical cross section obtained via DLOS in logarithmic format. The fits to the data (lines) were generated using Lucovsky’s model for the three detected deep levels.

or intrinsic). The SSPC onsets at $E_C-3.28$ eV and $E_C-1.28$ eV have been earlier ascribed to carbon impurities for both plasma-assisted and MOCVD GaN where the dominant, near valence band state was associated with C$_N$ defects that act as compensating acceptors in $n$-GaN and the $E_C-1.28$ eV state was correlated with carbon interstitials.[7, 8] Here, a clear dependence of deep level concentration on NH$_3$/Ga flux ratio is observed, with somewhat lower concentrations of both carbon-related traps observed for the sample grown with a higher NH$_3$/Ga flux ratio but with a more pronounced reduction for the C$_N$ substitutional acceptor state. It is not clear from this data whether the decrease in concentration for this level is due to a reduction in carbon concentration or a reduction in vacant nitrogen sites at
Table 5.3: Summary of deep levels and concentrations in ammonia-MBE n-GaN films. Below detection limit is signified by BDL.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Deep level concentration (cm$^{-3}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$E_C$−3.28  $E_C$−2.62  $E_C$−1.28  $E_C$−0.72  $E_C$−0.62  $E_C$−0.40  $E_C$−0.24</td>
</tr>
<tr>
<td>Lower NH$_3$/Ga flux ratio</td>
<td>3.1×10$^{15}$  7×10$^{14}$  9.2×10$^{14}$  BDL  7.7×10$^{14}$  BDL  1.5×10$^{14}$</td>
</tr>
<tr>
<td>Higher NH$_3$/Ga flux ratio</td>
<td>1.7×10$^{15}$  6×10$^{14}$  6.6×10$^{14}$  1.5×10$^{13}$  4.1×10$^{14}$  3×10$^{12}$  2.7×10$^{13}$</td>
</tr>
</tbody>
</table>

the higher NH$_3$/Ga flux ratio that would reduce subsequent carbon substitution. However, the trends observed suggest that defect concentrations may be further reduced by utilizing a higher NH$_3$/Ga flux ratio.

In contrast to these carbon-related deep states, the well known V$_{Ga}$-related defect, which exhibits an SSPC onset near $E_C$−2.6 eV in Fig. 5.10, shows negligible dependence on NH$_3$/Ga flux ratio. The measured concentrations for this deep level of $6\sim7\times10^{14}$ cm$^{-3}$ for both flux ratios. Contrary to state-of-the-art MOCVD n-GaN material where the gallium vacancy levels tends to be a dominant level, the concentration here is a small fraction of the total defect concentration, which is more similar to plasma-assist MBE growth. As growth temperature increase from PAMBE to ammonia-MBE to MOCVD growth, the concentration of this level may loosely depend on the equilibrium vacancy concentration based simply on the growth temperature.

Additional traps closer to the conduction band edge were detected using conventional thermal DLTS measurements. Figure 5.11 shows the DLTS spectra for both samples in semilogarithmic format to accentuate the near detection limit trap at $E_C$−0.40 eV. The associated Arrhenius plots for each deep level are shown in the inset and concentrations are summarized in Table 5.3. The two dominant levels seen by DLTS in the ammonia-based MBE
Figure 5.11: DLTS spectra of ammonia-MBE n-GaN Schottky diodes for the higher (blue ▼) and lower (red ▲) NH$_3$/Ga flux ratios using a −0.5 V reverse bias and 10 ms fill pulse at 0.0 V. The DLTS spectra are shown in logarithmic format to reveal the very low concentration $E_{C-0.40}$ and $E_{C-0.72}$ eV levels. Inset shows the Arrhenius plot for the different deep levels observed in the ammonia samples.
films, $E_C = 0.62$ and $E_C = 0.24$ eV, are frequently detected in $n$-GaN grown by other methods [6, 11, 20, 31, 43] and have been previously observed in ammonia-MBE $n$-GaN.[42] The deep level at $E_C = 0.24$ eV is similar to the D ($E_C = 0.25$ eV) level that has been previously associated to N vacancy-related complexes[31, 43] and shows the highest dependence on NH$_3$/Ga flux ratio compared with all the other DLOS and DLTS levels – being almost six times lower concentration in the higher NH$_3$/Ga flux ratio sample. This behavior is consistent with its assignment as a nitrogen vacancy-related level. The concentration of the often-reported $E_C = 0.62$ eV level, for which the physical source remains controversial,[11, 21] is also reduced for higher NH$_3$/Ga flux ratios, exhibiting a reduction by approximately a factor of two. The fact that these two electron traps, along with the three deeper states observed by DLOS are all commonly observed in high quality $n$-GaN grown by a wide range of methods indicates that the sources of these deep levels are not unique to the ammonia-MBE growth process or sources. The DLTS measurements, however, also reveal the presence of two other electron traps at $E_C = 0.72$ eV and $E_C = 0.40$ eV. In fact these traps, exhibit the opposite dependence on NH$_3$/Ga flux ratio compared to the other deep levels with concentrations that increase from below our detection limit for these samples ($\sim 1 \times 10^{12}$ cm$^{-3}$) for the lower NH$_3$/Ga flux ratio sample to very low but measurable concentrations in the higher NH$_3$/Ga flux ratio sample of $1.5 \times 10^{13}$ and $3 \times 10^{12}$ cm$^{-3}$ for the $E_C = 0.72$ and $E_C = 0.40$ eV levels, respectively. These levels have also been previously observed in both MBE and MOCVD films but with no clear identification of physical source.[21, 41] Their low concentrations suggest that they are unlikely to significantly impact device performance.

In summary, DLOS and DLTS measurements on $n$-type GaN grown by ammonia-MBE revealed strong and differing dependencies of individual deep levels on NH$_3$/Ga beam flux ratio but with a total trap concentration that already compares favorably with much more
mature GaN epitaxy methods. A shallow electron trap at $E_C - 0.25 \text{ eV}$ shows the strongest
dependence, with a factor of six reduction in trap concentration for two-fold NH$_3$/Ga beam
flux ratio increase, demonstrating consistency with its nitrogen vacancy-related association.
In addition to this state, several other commonly reported levels at $E_C - 0.62 \text{ eV}$, $E_C - 1.28$
eV and $E_C - 3.28 \text{ eV}$ also showed reduced concentrations at higher NH$_3$/Ga flux ratios, but
with somewhat lessened effect. In contrast the concentration of the V$_{\text{Ga}}$ state at $E_C - 2.62 \text{ eV}$
was found to be relatively invariant with NH$_3$/Ga flux ratio, but revealed properties more
consistent with V$_{\text{Ga}}$ defects present in MOCVD grown GaN than in plasma-assisted MBE
grown GaN, implying that gallium vacancy incorporation may strongly depend on whether
the GaN growth surface is N-rich or Ga-rich. Finally, the higher NH$_3$/Ga flux ratio sample did
reveal the emergence of two levels, at $E_C - 0.40$ and $E_C - 0.72 \text{ eV}$, their low trap concentrations
coupled with the presence of only common point defects and the marked suppression of
the potential shallow nitrogen vacancy-related donor-like trap at $E_C - 0.25 \text{ eV}$ implies great
promise for continued optimization of ammonia-MBE grown GaN based devices.

5.5 Comparison of defect incorporation in N-face and Ga-face polarity growth of PAMBE $n$-GaN

Previously, the studies of defect incorporation due to changes in growth conditions in-
cluding ammonia flow rate, plasma power, and growth method had been the focus, but
many other changes in growth are possible. Attention is now directed to changes in defect
introduction based on various substrate conditions starting off with GaN crystal polarity.
N-face (000$\bar{1}$) GaN and its alloys offer numerous possible improvements to traditional Ga-
face HEMTS device design resulting from the reversal of the spontaneous and piezoelectric
polarization charges discussed in Sections 2.1-2.3. In AlGaN/GaN heterostructures, the po-
larization charges cause the formation of a 2DEG at the GaN/AlGaN interface, which is the
Figure 5.12: Direction of spontaneous and piezoelectric polarization in Ga- and N-face heterostructures. In both cases it is assumed the AlGaN layer is under the critical thickness and fully strained by the GaN template substrate. In both structures, a 2DEG forms at the AlGaN/GaN interface slightly offset into the GaN layer. After Ref. 44.

Figure 5.13: Ball-and-stick model of wurtzite crystal structure for the Ga-face and N-face polarities of GaN. Wurtzite lacks symmetry in the c-direction evidenced by the different relative atomic positions for the different orientations. After Ref. 45.

basis of the HEMT device. The spontaneous and piezoelectric-induced polarization is shown in Fig. 5.12 for the wurtzite crystal structure shown for Ga- and N-face in Fig. 5.13 from Chapter 2 and repeated here for clarity. With few exceptions,[41] most III-nitride research on polar material has focused on Ga-face, but the reversed polarization direction and hence reversed device structure in a simplistic sense offers many potential advantages to improve HEMT performance and properties.
N-face GaN on AlGaN heterostructures grown by plasma-assisted MBE have demonstrated comparable mobility and 2DEG sheet charge density to Ga-face AlGaN on GaN layers,[46] but the GaN surface of the N-face structure implies Ohmic contacts are made to the 2DEG through a lower bandgap and therefore lower barrier material. This advantage has led to contact resistances < 55 \ Ohm \ \mu{m}.[47] In addition, the AlGaN back barrier provides a better back barrier leading to stronger confinement, low subthreshold current, and reduced short channel effects.[4] To date, however, only limited information on the presence and properties of deep level defects in N-face MBE-grown GaN have been reported, and what has been reported is limited to the exploration of traps within ~1 eV from the GaN conduction band edge.[48] Just as was the case for Ga-face GaN grown by plasma-assisted MBE,[6, 11, 19, 20, 49, 50] knowledge of traps existing anywhere within the 3.4 eV GaN bandgap and how their presence depends on growth parameters is expected to assist the continued development of N-face MBE grown GaN materials and devices. Moreover, it is of great fundamental interest to understand how deep level incorporation may depend on choice of polarization direction and resulting growth surface. This study focuses on the detection and analysis of GaN bulk trap states and their dependence on polarization direction in plasma-assisted MBE growth through the application of DLOS and related methods.

N- and Ga-face n-type GaN samples were grown using PAMBE growth under conditions optimized for each growth direction to systematically study the impact of V/III beam flux ratio on deep level incorporation. To minimize the impact that variation in threading dislocation density (TDD) might have on the deep level incorporation, substrates were chosen with approximately equal TDD. The Ga-face sample was grown on a Lumilog standard ULD template with a TDD of ~8 \times 10^7 \ cm^{-2} while a free-standing N-face GaN substrate with a TDD in the mid to high 10^7 \ cm^{-2} was used for the N-face sample, which is also depicted in Fig. 5.14.
Both growth structures consisted of an initial 200 nm thick $n^+$ GaN lateral conduction and back contact layer that was Si-doped to $\sim 3 \times 10^{18}$ cm$^{-3}$, followed by a 500 nm thick $n$-type GaN layer, and doped with Si to an electron concentration of $\sim 5 \times 10^{16}$ cm$^{-3}$. The optimized growth temperatures for the Ga- and N-face samples was 715°C and 710°C, respectively. These growth conditions result in a Ga-rich growth environment and a constant Ga-limited growth rate of $\sim 210$ nm/hr. The samples were processed using the typical GaN test structures using a mesa etch to deposit the Ohmic contacts on the heavily doped layers. The N-face sample required a unique processing procedure to prevent etching in basic solutions such as photoresist developer, which is described in Appendix A. Precise deep level energies were determined by calculating the optical cross section $\sigma_o$ from the extracted emission rate and measured photon flux and fitting the resulting curve to the theoretical Lucovsky model,\cite{22} which produced excellent fits to the optical cross section without the need to account for lattice relaxation effects via the Franck-Condon shift using the model developed by Chantre \textit{et al.}\cite{23} To access the near-conduction band region of the bandgap, conventional DLTS was employed. DLOS, SSPC, and DLTS measurements were performed using a quiescent and fill
pulse biases of -0.5 V and 0.0 V, respectively. The fill pulse time was 10 s for the DLOS with a
300 s measuring time and 10 ms fill pulse for the DLTS measurements.

The SSPC and DLOS results of the N-face and Ga-face samples are shown in Fig. 5.15 and what is immediately obvious is the very similar spectral features in the near valence
band region. The $E_C - 3.28$ eV level, which has been associated with residual carbon de-
fects that incorporate as substitutional $C_N$ acceptors that partially compensate $n$-type GaN
grown by PAMBE,[7] has a concentration of $2 - 3 \times 10^{14}$ cm$^{-3}$ – determined by measuring the
step height of the onset in Fig. 5.15a. The precise energy level is determined by fitting the
Lucovsky model to the optical cross-section shown in Fig. 5.15b. Here, it was determined
that the Lucovsky model, which assumes lattice relaxation effects are negligible, matched
the data well. Additionally, the other two observed levels, $E_C - 2.62$ and $E_C - 1.28$ eV, also
followed the Lucovsky model. The level at $E_C - 2.62$, most evident in the DLOS spectra of
Fig. 5.15b, also appears in both samples with nearly identical concentration of $8.5 - 9 \times 10^{13}$
$\text{cm}^{-3}$. This often-reported level is associated with a gallium vacancies possibly complexed
with hydrogen.[8–10] The similarity in concentration suggests that the $V_{Ga}$ formation is in-
sensitive to crystal orientation during growth while previous studies[7, 11, 19] have found
that N-rich/H-rich growth conditions found in MOCVD and ammonia-MBE growth strongly
impact $V_{Ga}$ formation. Despite the relative consistency in the near valence band states, the
midgap to conduction band states are quite in contrast. The $E_C - 1.28$ eV level possibly related
to split-interstitial carbon due to its correlation with carbon incorporation and theoreti-
cal prediction of bandgap energy is almost ubiquitously observed in previous $n$-type GaN
studies[7] is noticeably absent in the N-face material but present in the Ga-face sample with
a concentration of $1.0 \times 10^{14}$ $\text{cm}^{-3}$. Since the $C_N$ level is nearly identical in both samples, this
Figure 5.15: (a) Steady-state photocapacitance spectra of rf-plasma assisted MBE-grown N-face (blue dashed line) and Ga-face (red solid line) $n$-GaN Schottky diodes at 300 K using a $-0.5$ V reverse bias and 10 sec fill pulse of 0.0 V. The carbon associated levels at $E_C-1.28$ and $E_C-3.28$ are absent and in the same concentration, respectively, in the N-face sample while gallium vacancy-related level at $E_C-2.62$ eV is not sensitive to the growth face. The inset reveals more detail about the low energy spectrum of the N-face sample where the $E_V+1.4$ eV and $E_C-1.7$ eV were determined from the onsets in the SSPC. (b) DLOS spectra corresponding to the same SSPC data for the N-face (blue △) and Ga-face (red ○) samples. The fits to the data (lines) were generated using Lucovsky’s model.
lends credence to the idea of different surface conditions during growth. Instead, two previously unobserved levels at \( \sim E_C - 1.7 \) and \( \sim E_V + 1.4 \) eV, determined via SSPC onsets, manifest themselves in the N-face sample in relatively low concentrations (\( \sim 5 \times 10^{13} \) cm\(^{-3} \) and \( \sim 5 \times 10^{13} \) cm\(^{-3} \), respectively). Work is ongoing to identify the source of these levels.

To examine the upper bandgap states, DLTS was employed – the results of which are shown in Fig. 5.16a with corresponding Arrhenius plot in Fig. 5.16b. The largest difference in deep level concentration and the only common level between both samples is the probable nitrogen vacancy-related level at \( E_C - 0.25 \) eV,[19] which is \( 3.5 \times 10^{13} \) cm\(^{-3} \) in the Ga-face and in the N-face sample is almost an order of magnitude higher with a concentration of \( 2.3 \times 10^{14} \) cm\(^{-3} \). The emission rate constant of this level would be \( \sim 100 \) ns at room temperature, decrease exponentially at elevated temperatures, and could negatively impact device performance by inducing current collapse or gate lag, for example, depending on the spatial location of the deep levels. The other possibly common level in both samples is the \( E_C - 0.65/0.66 \) eV level, which has a concentration of \( 1.3 \times 10^{13} \) cm\(^{-3} \) in the Ga-face and \( 3.6 \times 10^{13} \) cm\(^{-3} \) in the N-face. The level is thought to be nitrogen interstitial-related based in N implant studies conducted by Haase et al.[51] The increase in these intrinsic defect concentrations for the N-face may be the result of slightly unoptimized growth, but further study is needed to determine the role of growth temperature and flux ratio in this growth mode.

The other DLTS-measured levels in both samples differ in energy, which was determined by fitting with Gaussian peaks. In the Ga-face sample, one additional deep levels are observed at \( E_C - 0.60 \) eV, which is commonly observed in PAMBE,[6] ammonia-MBE,[19] MOCVD,[8] and HVPE[35] grown \( n \)-type GaN films. This almost omnipresent defect is likely related to an intrinsic or common extrinsic defect. Several have suggested that this level may be related to an \( N_{\text{Ga}} \) antisite,[35, 51, 52] but the high formation energy (\( \sim 7 \) eV for \( n \)-type GaN)[53] suggests
Figure 5.16: (a) DLTS spectra of rf-plasma assisted MBE-grown N-face (blue dashed line) and Ga-face (red solid line) $n$-GaN Schottky diodes using a $-0.5$ V reverse bias and 10 ms fill pulse of 0.0 V for the 50 s$^{-1}$ rate window. The significant difference in total concentration between the N-face and Ga-face $n$-GaN sample develops from the $E_C - 0.25$ eV level thought to be nitrogen vacancy related. (b) Arrhenius plot for the different deep levels corresponding to the peaks from each rate window for each sample.
that the $1.9 \times 10^{13} \text{ cm}^{-3}$ concentration observed is many orders of magnitude higher than predicted. This level, however, is ascribed to a point defect-dislocation complex in Section 5.6 using DLTS capture kinetics and correlation with threading dislocation density. In the N-face sample, additional deep levels are observed at $E_C - 0.24$ and $E_C - 0.44$ eV. The $E_C - 0.24$ eV level is distinct from the $E_C - 0.25$ eV level as evidenced by the separate peak in the DLTS spectrum for the N-face sample and the Arrhenius plot. The concentration of this level is near the detection limits with a concentration of $6 \times 10^{12} \text{ cm}^{-3}$. The $E_C - 0.44$ eV level is similar to levels observed in ammonia-MBE $n$-GaN material,[19] but with no clear identifiable physical source for this level.

DLTS and DLOS measurements on N-face and Ga-face $n$-type GaN grown by PAMBE revealed total measured deep level concentrations of $\sim 6 \times 10^{14} \text{ cm}^{-3}$, which is unexpectedly low given that N-face material has been studied much less than the traditional Ga-face material. This suggests that while the growth surface does play a small role in emergence of new traps and suppression of others, the impact of this is much lower than previously growth condition studies such as nitrogen plasma power (see Section 5.2) or previously studied growth modes such as ammonia-MBE.[19]

In summary, the N-face $n$-GaN showed a decrease in carbon acting as deep interstitial levels near $E_C - 1.3$ eV but with similar concentration of $C_N$ ($E_C - 3.28$ eV). Furthermore, the $V_{Ga}$-related state at $E_C - 2.62$ eV was insensitive to N-face versus Ga-face polarization evidenced by the same measured concentration in both samples. The emergence of several new levels, all present in low concentration, at $E_C - 0.24$, $E_C - 0.44$, $E_V + 1.4$, and $E_C - 1.7$ eV warrant further study to evaluate both their physical source and also any potential impact on HEMT device performance. Finally, the most striking contrast between the N-face and Ga-face deep level spectra is the nearly 10X increase in the $E_C - 0.25$ eV level thought to
be nitrogen vacancy related. As a shallow level, this will have a room temperature time constant of \( \sim 4 \mu s \), which could easily contribute to HEMT performance issues. Investigations into this level and others in N-face HEMT is currently in progress, but this is not the only substrate-related study on HEMT planned or in progress.

5.6 Impact of dislocation density on trap concentrations in MOCVD-grown \( n \)-GaN Schottky diodes

Dislocations are suspected to impact device performance and reliability due to increased leakage, or other unidentified mechanisms.\(^{[54–58]}\) Despite speculation, little direct evidence of the role of dislocations has been presented. From a materials perspective, the role of dislocations is more understood as screw dislocations in MBE material are known to act as leakage paths and edge dislocations are thought to be negatively charged, but many questions still remain in this aspect as well. A fundamental way to answer many of these questions is to identify any electrically active defects due to the dislocation core itself or defect segregation, as this information directly reveals trapping information but this information can also be used in future quantitative defect studies of HEMT reliability to monitor these defects for correlations with gate leakage, power output, current slump, or other relevant parameters. Before HEMTs can be analyzed, it is necessary to examine the properties of dislocations in “bulk” material to search for any dislocation related defects.

GaN lacks a viable native substrate at this point, so dislocations in high concentrations compared with other semiconductors are unavoidable. However, the range of growth variables for GaN epitaxy such as nucleation conditions, substrates, selective area growth, and film thickness means that the threading dislocation density (TDD) can vary from \( \sim 10^{10} \text{ cm}^{-2} \) down to \( \sim 10^6 \text{ cm}^{-2} \). Therefore, by maintaining identical growth conditions and initiating
Figure 5.17: Device structures for the MOCVD threading dislocation density study. The samples were grown under nominally identical conditions except for the substrate template threading dislocation density, which was varied from $\sim 4 \times 10^7 \text{ cm}^{-2}$ to $8 \times 10^9 \text{ cm}^{-2}$.

Growth on MOCVD-grown GaN templates where only the threading dislocation is varied, a series of samples can be grown to exclusively study the effects of dislocations.

The $n$-GaN samples used to examine the effects of dislocations were grown by MOCVD on three template substrates: lateral epitaxial overgrowth (LEO), Type “A”, and Type “C”. The LEO template was a commercial $n$-type Lumilog 2S-LEO (two step lateral epitaxially overgrown) substrate with a TD density of $\sim 4 \times 10^7 \text{ cm}^{-2}$ and Si concentration of $\sim 3 \times 10^{18} \text{ cm}^{-3}$. The Type “A” template consisted of a $\sim 3 \mu\text{m}$ thick atmospheric pressure MOCVD grown GaN film with the last $\sim 2 \mu\text{m}$ Si doped to $\sim 1 \times 10^{19} \text{ cm}^{-3}$ deposited on a c-plane sapphire substrate. Grown under optimal conditions, this template had a TD density between $6 - 8 \times 10^8 \text{ cm}^{-2}$. The Type “C” template is nominally the same growth structure as the Type “A” template but due to slightly different nucleation conditions had a TDD of $8 \times 10^9 \text{ cm}^{-2}$. The high doping of the template acted as a lateral conduction layer for low series resistance and low contact resistance Ohmic contact following the procedure in Chapter 3. The device structure is shown in Fig. 5.17. The MOCVD GaN layer grown on the templates and probed by the DLOS and DLTS measurements consisted of a uid $n$-GaN layer with a thickness of 0.8 $\mu\text{m}$ grown
Figure 5.18: I-V curves of three MOCVD $n$-GaN films grown on different GaN template substrates. The threading dislocation density ranged from $4 \times 10^7$ cm$^{-2}$ for the LEO sample with the lowest leakage current to $6 \times 8 \times 10^8$ cm$^{-2}$ for the Type "A" sample, which had degraded Schottky characteristics compared with the LEO sample, and finally to $8 \times 10^9$ cm$^{-2}$ for the Type "C" sample that was the leakiest and exhibited nearly completely Ohmic behavior.

under atmospheric pressure at $1040 \, ^\circ\text{C}$. Subsequent electron-beam-induced current (EBIC) measurements of nearly identical structures with Type "A" and LEO templates confirmed an average TDD value of $\sim 7 \times 10^8$ cm$^{-2}$ for the former (higher TDD sample) and $\sim 4 \times 10^7$ cm$^{-2}$ for the latter (lower TDD sample). This provided a factor of $\sim 17$ difference in TDD between the two sample types – to support the goals of this study.

The I-V curves of the three samples in Fig. 5.18 shows a strong dependence on the Schottky characteristics including leakage and ideality factor on the TD density. Because the Type “C” sample's I-V was basically Ohmic with extremely high leakage current, this sample was unsuitable for capacitance-based measurements and therefore DLTS and DLOS could not be performed on this sample. Even without the Type “C” sample the TD density
Table 5.4: Summary of deep level concentrations as a function of threading dislocation density (TDD) in MOCVD n-GaN. BDL signifies below detection limits.

<table>
<thead>
<tr>
<th>Energy</th>
<th>Deep level concentration (cm$^{-3}$)</th>
<th>Type “A”/LEO change</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E_C - 0.59$</td>
<td>1.4×10$^{13}$</td>
<td>1.7×10$^{14}$</td>
</tr>
<tr>
<td>$E_C - 0.90$</td>
<td>2×10$^{12}$</td>
<td>BDL</td>
</tr>
<tr>
<td>$E_C - 1.35$</td>
<td>3×10$^{13}$</td>
<td>4×10$^{14}$</td>
</tr>
<tr>
<td>$E_C - 2.64$</td>
<td>4×10$^{14}$</td>
<td>2×10$^{15}$</td>
</tr>
<tr>
<td>$E_C - 3.22$</td>
<td>1×10$^{15}$</td>
<td>2×10$^{15}$</td>
</tr>
<tr>
<td>TDD</td>
<td>4×10$^{7}$ cm$^{-2}$</td>
<td>6 – 8×10$^{8}$ cm$^{-2}$</td>
</tr>
</tbody>
</table>

 Variation between the LEO and Type “A” samples are large enough to see clear trends in trap concentration with TD density.

Figure 5.19 shows the DLTS spectra for the two samples where the $E_C - 0.60$ eV level dominates the spectra. The concentrations for these levels is summarized in Table 5.4. The $E_C - 0.60$ concentration increases by ~12X in the higher TD density sample, which is quite similar to the 15 – 20X increase in TD density between the samples. The concentration of the $E_C - 0.90$ eV level evident in the LEO sample is below the detection limit in the Type “A” sample indicating little correlation with threading dislocation density. This result is quite profound since previous studies of these two levels indicated extended behavior possibly due to dislocation related effects.[6, 29] In Section 4.5.5, the concept of capture kinetics using varying fill pulse times in DLTS is discussed. Based on the relationship between the peak height and fill pulse time and variation of peak temperature, the Coulombic interaction of levels can be inferred and some information on arrangement of defects determined. The peak heights of $E_C - 0.90$ eV level exhibited a completely linear relationship with the logarithm of the fill pulse time, shown in Fig. 5.20 – indicative of ideal extended defect behavior. The $E_C -
Figure 5.19: DLTS spectra for the (a) Type “A” and (b) LEO MOCVD $n$-GaN samples using a quiescent bias of $-0.5$ V and a 10 ms fill pulse of 0.0 V. The insets show the $E_C - 0.59$ eV Arrhenius data for the different rate windows. The activation energies and cross-sections were extracted from the Arrhenius plot while the concentration was calculated from the DLTS signal, electron carrier concentration, and equilibrium capacitance value.
Figure 5.20: DLTS experiments performed with different pulse times on the $E_C - 0.90$ eV level. The capture kinetic analysis in the inset for an ideal interacting defect indicates that the defects are close enough to influence the capture of neighboring defects based on Coulombic interaction.

$0.90$ defects are non-uniformly distributed and potentially forming clusters, linear arranged defects, or other arrangement involving closely spaced defects. Previously, it was concluded that the $E_C - 0.90$ eV could be forming a Cotrell cloud around dislocations,[6, 30] but the $E_C - 0.90$ trap’s lack of dependence on TDD strongly suggests this is not the case. Instead, the defect may be forming point defects in clusters. As opposed to the $E_C - 0.90$ eV level, the $E_C - 0.60$ eV level in Fig. 5.21 exhibits more complex behavior. The right inset of Fig. 5.21 should be linear for uniformly distributed defects. Clearly, this is not the case. Nevertheless, the level does not follow the ideal interacting behavior either evident of the non-linear behavior in the left inset of Fig. 5.21. The assumption for in the derivation (Eq. (4.85)) of the interacting defects is that the fill pulse time is short. A possible reason for this non-linearity is due to the violation of this assumption, which results in an eventual saturation of the potential when emission and capture rate both become significant. This non-uniformly distributed
Figure 5.21: DLTS experiments performed with different pulse times on the $E_C - 0.60$ eV level. The capture kinetic analysis in the right inset for an ideal point defect is clearly non-linear indicating the level is absolutely not an ideal point defect. The level, however, does not follow the ideal interacting behavior either. Several assumptions are implicit in the derivation of the interacting behavior that may not be valid at the longer fill pulse times where the defect level may merely be saturating.

and possibly interacting behavior means this level could be associated dislocations. Indeed, the strong dependence of the $E_C - 0.60$ trap concentration on the TD density suggests this is the case. Delving further, the observed peak temperature shift from the $10 \mu s$ to $100$ ms fill pulse times was only $3$ K, which implies little spread of the deep level energy. Following the progression of logic in Section 4.5.5, this further implies that the defects are close enough to interact Coulombically but not close enough for wave function overlap. This suggests that this defect energy is not associated with the dislocation core itself where the defects would be close enough to have overlapping wave functions. With these conclusions, the $E_C - 0.60$ can tentatively be assigned to a point defect/dislocation complex.

With only DLTS, this analysis would be done and association of deeper states with dislocations would not be observed. However, DLOS results clearly reveal a strong dependence
of deep and near valence band states on dislocation density. Figure 5.22 shows the mid-gap and near valence band steady-state photocapacitance spectra for both the LEO and Type “A” samples with the results summarized in Table 5.4. Three levels are evident in the SSPC at $E_C - 1.35$, $E_C - 2.64$, and $E_C - 3.22$ eV. From Table 5.4, it is clear that the level which most strongly correlates with the TDD is the $E_C - 1.35$ eV level. This level has also been tentatively associated with a split carbon interstitial level and is cautiously suggested to be a carbon point defect/dislocation defect complex.\[24\] This relation holds for both MOCVD and MBE films as well and is supported by a plot of the $E_C - 1.35$ concentration versus residual carbon concentration measured with SIMS in Fig. 5.23. The figure clearly shows a dependence of the concentration of this state on both carbon content as well as threading dislocation density, strongly suggesting that the physical source of this defect is a carbon-dislocation complex. It is interesting to extrapolate to the leakage current dependence on dislocation density, since
Figure 5.23: Measured trap concentration of the $E_C - 1.35$ eV level versus the residual carbon concentration measured by secondary ion mass spectroscopy. The concentration of the $E_C - 1.35$ level correlates with both carbon incorporation and threading dislocation density. The squares represent MOCVD growth while the MBE results are circles. After Ref. 24.

The decoration of dislocations by electrically active carbon could indeed create low resistance localized leakage paths to the Schottky contact.

From Table 5.4, the DLOS levels at $E_C - 2.64$ eV (Ga vacancy related) and $E_C - 3.22$ eV (compensating Mg acceptor) exhibit a clear but somewhat weaker dependence on TDD. However it must be noted that at these energies the SSPC concentrations are only lower bounds since there is the possibility of competing hole emission to the valence band for levels present below $E_g/2$. Considering the Ga vacancy state a bit further, we have already discussed at length that this state has long been associated with $V_{Ga}$-related complexes,[6, 33] and is probably related to the $V_{Ga} - 2H$ level based on work of Hautakangas et al. in MOCVD GaN and the theoretical work of Van de Walle et al.[33, 34] Earlier work has actually suggested that the yellow luminescence emission associated gallium vacancies was found to preferentially occur at low angle grain boundaries.[59] If we ignore the possible lack of saturating the
states argument just made for the moment, the weaker dependence observed here might suggest that the formation of gallium vacancy related levels at low angle grain boundaries/dislocations is favorable but since this level also forms in significant concentration in the bulk as well the full order of magnitude increase in concentration to linearly track the change in TDD is not observed. Finally, the $E_C - 3.22$ eV deep level exhibits little dependence on the TD density. This is not surprising given the assignment of Mg acceptor incorporation as the physical source for this level and that magnesium clustering around dislocations has not been reported to be an issue in the literature.

This dislocation study in MOCVD grown GaN has reached several important conclusions about the nature of defects in GaN. The $E_C - 0.90$ eV level exhibits interacting behavior possibly due to clustering of defects but is not dislocation related. The $E_C - 0.60$ and carbon-related $E_C - 1.35$ eV levels exhibit strong correlation to increases in TDD suggesting these levels form as point defect-dislocation complexes and potentially contribute to the enhanced leakage current in higher TDD samples. The $E_C - 2.64$ eV level shows a dependence on TDD that is clearly evident but somewhat weaker than the $E_C - 0.60$ eV and $E_C - 1.35$ eV states, but this could be attributable to a lack of trap saturation leading to a lower bound concentration for this state, or it could be a result of Ga vacancies being distributed both in the non-dislocated field as well as around dislocation cores. Either way, there is a clear impact on $V_{Ga}$-related defects and such $V_{Ga}$-dislocation complexes have been discussed in the past. As expected, the Mg acceptor at $E_C - 3.22$ eV does not show direct TDD dependence. Finally, none of the levels observed appear to be “pure” dislocation core related levels as such a level would not only exhibit interacting behavior but also an energy distribution of the deep level due to Pauli exclusion causing mini-band formation from closely space core defects. This strong correlation between trap concentrations, TD density, point defect configuration,
and degradation of Schottky diode characteristics indicates that electronic devices such as HEMTs can be directly improved through a combination of reduced point defect density, including excess carbon impurities, and reduced TD densities.

5.7 Defect identification in thick, MOCVD-grown n-AlGaN on AlN

High quality AlGaN films are a critical component of the AlGaN/GaN HEMTs and so far we have discussed wide variations in the growth of GaN, which for HEMTs is the buffer. As seen in later chapters, it is the AlGaN/GaN interface that is key and the AlGaN material quality in addition to the GaN buffer that are crucial for high performance devices. Of course AlGaN as a bulk material itself is of great importance for ultraviolet optoelectronics,[60–62] and this section focuses on the first quantitative DLOS/DLTS trap studies of AlGaN. Together with the GaN results, the component layers of an AlGaN/GaN HEMT device structure is fully characterized. In this section, a systematic investigation of deep levels using DLTS and DLOS to probe the entire Al$_{0.30}$Ga$_{0.70}$N bandgap was performed. These results are correlated with SIMS analysis and theoretical density functional theory (DFT) to develop correlations between deep levels and their potential physical sources.

5.7.1 Experimental details

The bulk AlGaN sample used in this study was grown by metal organic chemical vapor deposition (MOCVD) in an EMCORE D-125 reactor with a short jar geometry via an ongoing collaboration with Sandia National Laboratory (A. Allerman). Trimethylgallium, trimethylaluminum, and ammonia sources were used to grow the AlGaN layers and silane was used as the n-type dopant source. It was found that carbon incorporation increased at lower growth pressures but deleterious parasitic, gas-phase reactions caused nanoparticle incorporation at higher growth pressures, so a growth pressure of 75 Torr was chosen for
these layers. The gas-phase reactions were also minimized by using a lower V/III ratio. The sample was grown on an AlN substrate, which allows thick AlGaN films to be grown without cracks because the growing AlGaN film is in compression but results in threading dislocation of $\sim 2 \times 10^9$ cm$^{-2}$. The growth on the AlN substrate was initiated with a 1.5 $\mu$m unintentionally doped (uid) AlN layer followed by several microns of uid Al$_{0.30}$Ga$_{0.70}$N. A highly doped $n$-type Al$_{0.30}$Ga$_{0.70}$N lateral conduction layer 1.4 $\mu$m thick and doped $\sim 1 \times 10^{18}$ cm$^{-3}$ was grown to provide a contact point for the Ohmic metal and series resistance. The top layer, which is probed by the DLOS and DLTS measurements, was 0.58 $\mu$m thick, grown at 1070°C, and Si-doped. The carrier concentration measured by C-V indicated the film was $n$-type with an electron concentration of $2.6 \times 10^{17}$ cm$^{-3}$. A semitransparent Schottky contact (80 Å Ni) was used to form the Schottky diode with an area of 0.0841 mm$^2$ in this study. Mesa etch isolation using Cl$_2$-based reactive ion etching provided trenches to the underlying $n+$ AlGaN layer and Ohmic contacts consisting of Ti:Al:Ni:Au (100:2000:200:3000 Å) were deposited. The contacts to the $n+$ AlGaN had Schottky character before annealing but upon annealing at 850°C for 30 s in a nitrogen ambient the contacts were Ohmic with a contact resistance of $8 \times 10^{-6}$ Ωcm$^2$, measured by the transfer length method (TLM), and a $\sim 45$ Ω series resistance was achieved for the Schottky diode structures. To investigate defects throughout the AlGaN bandgap, a combination of DLOS, steady-state photocapacitance (SSPC) and DLTS measurements were used.

5.7.2 Results and discussion

Standard double boxcar DLTS measurements, shown in Fig. 5.24, revealed only one deep level above detection limits ($\sim 1 \times 10^{13}$ cm$^{-3}$) with a concentration of $2 \times 10^{13}$ cm$^{-3}$. An activation energy of $E_C - 0.87$ eV and capture cross-section of $\sim 1 \times 10^{-13}$ cm$^2$ was determined analyzing
Figure 5.24: DLTS spectrum for the \( n \)-AlGaN Schottky diode using a \(-0.5\) V quiescent bias and 10 ms, 0.0 V filling pulse. Only one peak is evident at high temperature with a concentration of \( \sim 2 \times 10^{13} \) cm\(^{-3} \).

Unlike many GaN deep levels where they are observed in films grown by multiple growth methods, the \( E_C - 0.87 \) eV level observed here appears to be very distinct from other levels implying that the \( E_C - 0.87 \) eV level may more likely be related to an extrinsic defect than an intrinsic one. This level is also very similar in energy level to the \( E_C - 0.9 \) eV level in \( n \)-type GaN, but comparison of the Arrhenius plot in Fig. 5.26 makes it clear that their are very different characteristics (capture cross section) for each level suggesting that these levels are the result of different physical sources. This implies the \( E_C - 0.87 \) eV level is not a hydrogenic level that forms with the same energy in GaN and AlGaN. Continuing the comparison with MOCVD GaN, the concentrations of the upper bandgap (near conduction band) deep levels for both MOCVD GaN and AlGaN are quite low as opposed to MBE GaN films where these
Figure 5.25: An Arrhenius plot of the DLTS data from Fig. 5.24 indicates a straight line with $E_C - 0.87$ eV energy.

Figure 5.26: An Arrhenius plot of the typical $n$-GaN DLTS levels are plotted for MOCVD (red triangles) and MBE (black squares) samples with the $E_C - 0.87$ eV $n$-AlGaN DLTS level (green circles). It is clearly evident that the $E_C - 0.87$ level in AlGaN is very distinct from the $E_C - 0.9$ eV level in GaN.
Figure 5.27: Steady-state photocapacitance spectrum of the thick MOCVD-grown $n$-AlGaN Schottky diode performed at a $-0.5$ V quiescent bias and 10 s fill pulse at 0.0 V. The spectrum is dominated by the $E_C-3.10$ and $E_C-3.93$ eV deep levels. The inset shows the the SSPC spectrum of a quartz-tungsten-halogen lamp where an onset is evident at $E_C-0.9$ eV and possibly at $E_C-1.5$ based solely on the change in slope.

levels have significantly higher concentrations where Section 5.1 discussed this in more depth for GaN.

The rest of the AlGaN bandgap was explored using optical stimulation of deep levels. The steady-state photocapacitance (SSPC) is shown in Fig. 5.27 where the AlGaN bandgap is evident from the peak at $\sim E_C-4.08$ and several levels are identified by their onsets and fitting to their optical cross section. The $E_C-0.9$ and $E_C-1.5$ eV deep levels evident in the inset of Fig. 5.27 could only be resolved from the SSPC onset and not from the DLOS optical cross section fitting due to low signal-to-noise ratio from the quartz-tungsten-halogen lamp-based transients. Analysis of the onsets in the SSPC is valid but does not provide the accuracy of the DLOS analysis hence the accuracy is only given to $\sim 0.1$ eV, which does not allow us to extract specific information about possible local lattice relaxation. The concentration of the $E_C-0.9$ and $E_C-1.5$ eV levels would normally be determined by the step height in the SSPC.
Figure 5.28: Because the $\Delta C/C_0$ from the SSPC spectrum exceeds 0.1, the deep level concentrations extracted are not accurate due to assumptions being violated. To extract accurate concentrations, lighted-capacitance-voltage measurements were performed. C-V curves are recorded under monochromatic light bias. From these curves, the fixed charge density is extracted as a function of depth. The difference in fixed charge densities at a given depth is due to the change in defect occupation. The light energies were chosen based on the onset energies and bandgap to extract the defect concentrations of individual traps.

spectrum, but due to the large $\Delta C/C$ the binomial expansion assumption in the derivation of DLOS in Chapter 4 is violated and has to be determined via other means, notably LCV (lighted C-V).[24, 64] The LCV method uses the change in fixed charge density extracted from the C-V curves as a function changing the deep level occupancy via sub-gap illumination (i.e. emptying traps at $E_C - E_T$ filled with electrons or filling states at $E_V + E_T$ that are devoid of electrons). This changes the space charge concentration by the density of unoccupied traps, so the difference in extracted fixed charge density is the deep level density. Figure 5.28 shows the LCV results where the optical energies were chosen to isolate individual traps determined by the SSPC and DLOS results. The defect concentration for the $E_C - 0.9$ eV trap was determined by examining the difference between the $E_C - 1.5$ and $E_C - 0.9$ LCV concentrations, which was very approximately $1 \times 10^{13}$ cm$^{-3}$ based on averaging the fixed
Figure 5.29: DLOS spectrum of the thick MOCVD-grown $n$-AlGaN Schottky diode performed at a $-0.5$ V quiescent bias and 10 s fill pulse at 0.0 V. The lines represent fits of the cross section using the Lucovsky model presented in Chapter 4 to extract an accurate defect energy level.

charge densities over the measured depletion depth in an attempt to increase signal-to-noise ratio to detect the low concentration. The similar energy and concentration with the $E_C-0.87$ DLTS level suggests these levels are most likely the same level observed via thermal and optical processes. While optical cross section analysis was not possible, the Franck-Condon energy can be estimated from the difference in optical and thermal energy, which suggests this lattice relaxation energy is $\sim 0.07$ eV based on the precise onset in the SSPC and DLTS energies. The concentration of the $E_C-1.5$ eV level was measured subtracting the $E_C-3.10$ C-V result from $E_C-1.5$ leading to a trap concentration of $2 \times 10^{16}$ cm$^{-3}$, much higher that the $E_C-0.9$ eV trap density.

The dominant level in the SSPC spectrum of Fig. 5.27 is the $E_C-3.10$ eV level, which was determined using the Lucovsky model[22] to fit the DLOS spectrum shown in Fig. 5.29. Again using LCV results in Fig. 5.28 to extract the correct defect concentrations, a concentration of
$2 - 3 \times 10^{17} \text{ cm}^{-3}$ was determined for the $E_C - 3.10 \text{ eV}$ level. The likely source of this level is a cation-vacancy related level because of several factors: its energy level, predicted formation energy, and similarity to GaN. Here we discuss each of these considerations. First principles calculations suggest that in $n$-type material the cation vacancy formation is favorable in GaN consistent with previous results and formation is predicted to be even more favorable in AlN.\cite{10, 65, 66} Based on the energy level predictions in GaN and AlN, the $E_C - 3.10 \text{ eV}$ level agrees well with predicted values assuming Vegard’s law to estimate the energy level in 30% AlGaN and assuming (for now) that the energy level follows the vacuum referred binding energy (VRBE) model first developed by Zunger et al.\cite{67, 68} In this model, the deep level energy referenced to the vacuum level $E_{\text{vac}} - E_T$ is constant irrespective of the composition (i.e. ratio of Al to Ga in this case, bandgap and electron affinity i.e $E_C$). It was predicted and observed that deep, antibonding, cation-site impurities are expected to track the vacuum level and not a band edge as the ratio of Group III/V is varied,\cite{68} which is the case here. If this model is assumed, $E_{\text{vac}} - E_T$ should be constant across the AlGaN material system, as shown in Fig. 5.30. Hence, we can write an expression for the deep level energy $E_C - E_T$ for Al$_{0.30}$Ga$_{0.70}$N as

$$
(E_C - E_T)_{\text{Al}_{0.30}\text{Ga}_{0.70}N} = \chi_{\text{GaN}} + (E_C - E_T)_{\text{GaN}} - \chi_{\text{Al}_{0.30}\text{Ga}_{0.70}N} \tag{5.2}
$$

where $\chi$ is the electron affinity $E_{\text{vac}} - E_T$. Now, while the electron affinity values are in dispute in the literature, we are here only concerned with differences and so we chose values from Wu et al. who have reported an electron affinity of $3.5 \pm 0.1 \text{ eV}$ for GaN\cite{69} and $1.9 \pm 0.1 \text{ eV}$ for AlN.\cite{70, 71} Using these values, the electron affinity extrapolated Al$_{0.30}$Ga$_{0.70}$N is approximately 3.0 eV.

Plugging in 3.0 eV and 3.5 eV for the Al$_{0.30}$Ga$_{0.70}$N and GaN electron affinities, respectively, and our measured GaN vacancy energy of $E_C - E_T = 2.6 \text{ eV}$, this model predicts a cation
Figure 5.30: Band diagram showing the GaN bandgap on the left and AlN bandgap on the right. Using the vacuum referred binding energy model, the cation vacancy (commonly referred to as the gallium vacancy in GaN) is proposed to have a fixed $E_{\text{vac}} - E_T$ energy. In 30% AlGaN, the $E_C - 2.6$ eV cation vacancy related level that is observed for GaN with a 3.39 eV bandgap would be expected to be at $E_C - 3.10$ eV.

vacancy energy of $E_C - E_T = 3.1$ eV for Al$_{0.30}$Ga$_{0.70}$N, which matches our DLOS measurements shown earlier! This attribution is consistent with the work of Nam et al., who tracked the cation vacancy related levels for varying concentrations of Al in $n$-type AlGaN relative to the GaN vacancy band edge and found with some assumptions that the defect energy was constant.[72] Their results with their assumptions closely match the calculations here lending further to the premise that the $E_C - 3.10$ eV level is cation vacancy related and vacuum referred.

Finally, the $E_C - 3.10$ eV concentration in MOCVD-grown $n$-type Al$_{0.30}$Ga$_{0.70}$N is higher than the typical $V_{\text{Ga}}$-related deep level ($E_C - 2.60$ eV) concentration in MOCVD-grown $n$-type GaN ($\sim 6 \times 10^{15}$ cm$^{-3}$,[7] which agrees with the theoretical work that suggests the formation energy of cation vacancies are decreased as one moves along the AlGaN ternary composition range from GaN to AlN.[65, 66] Therefore, there is irrefutable data to suggest that the source for
the $E_C - 3.10$ eV state in $\text{Al}_{0.30}\text{Ga}_{0.70}\text{N}$ is a native cation (group III sublattice) vacancy, closely related to the well established Ga vacancy in pure GaN.

The SSPC spectrum in Fig. 5.27 also revealed a level at $E_C - 3.93$ eV – determined by fitting DLOS spectrum to the Lucovsky model. LCV measurements from Fig. 5.28 revealed a measured trap concentration of $5 \times 10^{16} \pm 2 \times 10^{16}$ cm$^{-3}$ for the $E_C - 3.93$ eV level. This level referenced to the valence band is $\sim E_V + 0.15$ eV and is strong evidence of a shallow acceptor level in $n$-$\text{AlGaN}$. In GaN, both carbon and magnesium can form shallow acceptor levels,[7, 8] and one of these is likely the source of the $E_C - 3.93$ eV level observed here. The formation energies of Mg and C are predicted to remain low in $n$-type $\text{AlGaN}$, and the ionization energy is also predicted to be slightly deeper than the respective level GaN reaching $\sim E_C - 0.3 - 0.5$ and $\sim 0.4$ eV in AlN for C and Mg, respectively.[73–76] Experimental results of Mg activation in near 30% $\text{AlGaN}$ suggests the activation energy ranges between 0.16 and 0.31 eV, which is slightly lower than the theoretically predicted values.[77, 78] SIMS analysis was performed using calibrated standards for $\text{AlGaN}$ to potentially differentiate the the physical source, and this revealed concentrations of $\sim 2 \times 10^{18}$ cm$^{-3}$ for carbon and less than or equal to $2 \times 10^{17}$ cm$^{-3}$ for magnesium. The carbon concentration is high because low pressure (LP) MOCVD growth of nitrides allows carbon to readily incorporate as a shallow acceptor.[7, 60] Unfortunately with respect to differentiating the physical source of the level, both the Mg and C concentrations are higher than the observed trap concentration $\sim 5 \times 10^{16}$ cm$^{-3}$. The high SIMS concentrations, however, does give further support to the theoretical and experimental data of C and Mg that one of these is the source of the $E_C - 3.93$ eV level. Finally, fitting of the optical cross section yielded a good fit using only the Lucovsky model where no Franck-Condon energy is present. It is predicted that the Franck-Condon energy of the shallow substitutional carbon on nitrogen acceptor will increase more than 500 meV in AlN from

168
GaN.[79] Because no Franck-Condon energy was observed for the $E_C−3.93$ eV level and $C_N$ is predicted to have a significant Franck-Condon energy even for 30% AlGaN, this level is tentatively assigned to Mg$_N$ substitutional acceptor level. We know from our prior experience with MOCVD growth of $n$-type GaN that residual Mg acceptors will be present in abundance on the reactor's quartz walls, and that the Mg desorbs at growth temperature. For GaN, the Mg state dominates the C acceptor for MOCVD films, and we expect this to be the case here as well.

### 5.7.3 Conclusion

The first defect characterization of thick $n$-AlGaN by DLTS and DLOS was presented, which completes the component layer analysis of AlGaN/GaN HEMTs. Four deep levels were observed at $E_C−0.87/0.9$, $E_C−1.5$, $E_C−3.10$, and $E_C−3.93$. The DLTS $E_C−0.87$ level is likely related to the DLOS $E_C−0.9$ level suggesting a lattice relaxation energy of $∼0.07$ eV and is not related to the $E_C−0.9$ eV level in GaN due to divergent cross sections. The largest concentration level, $E_C−3.10$ eV, with a concentration of $∼5×10^{17}$ is probably related to a cation vacancy based on predicted energy levels, the vacuum referred binding energy model, and is consistent with the work of Nam et al. Also, a shallow acceptor level at $E_C−3.93/∼E_V+0.15$ was observed and is possibly due to magnesium although carbon cannot be conclusively ruled out. Finally, although defect concentrations in AlGaN are currently moderately high, identification of deep levels should assist in rapidly improving film quality and also completes the picture of bulk defect spectra in AlGaN/GaN layers allowing for characterization of multi-layer, multi-dimensional structures to proceed.
5.8 Summary

This chapter has focused on the component layers of AlGaN/GaN HEMT characterizing not only the defects in these materials but also observing the effects of growth and substrate conditions on the deep level incorporation. Though this may not expose all of the defect related effects in HEMTs, it does provide a relative easy way to explore new growth regimes and fully characterize each layer. In the next chapters, this information is applied to measurements on HEMTs where the current-based measurements rely on this fundamental understanding of defects in GaN and AlGaN to interpret the complex defect signatures from the surface, AlGaN, AlGaN/GaN interface, and GaN buffer.

This chapter has shown that material quality depends on growth conditions and changes to the substrate, and these studies were carefully planned to not only identify the role of each condition on defect incorporation but also find potential physical sources of these levels. The first comparison of growth conditions was MOCVD to plasma-assisted MBE $n$-GaN where generally the same levels were observed at $E_C - 0.25$, $E_C - 0.60$, $E_C - 0.90$, $E_C - (1.28 - 1.35)$, and $E_C - 2.6$ indicating these levels are either intrinsic or due to common impurities such as carbon or oxygen. The main differences between MOCVD and MBE is the dominant acceptor level, which is Mg ($E_C - 3.22$ eV) in MOCVD and C ($E_C - 3.28$) in MBE. MOCVD material tends to have very low concentrations of the DLTS levels (within 1 eV) of the conduction band edge as well.

MOCVD GaN can be grown at higher rates than typical MBE nitrides, so this prompted a study of the nitrogen plasma power in MBE material. At higher growth rates, the concentration of the $E_C - 3.28$, $E_C - 2.56$, and $E_C - 1.35$ eV levels all increased exponentially with plasma power following an Arrhenius-like relationship. Atomic and ionic nitrogen are known to cause increased point defect incorporation and the Arrhenius-like relationship suggests
that the increased carbon and gallium vacancy incorporation may result from increased energy of constituent nitrogen species in the molecular beam. Plasma nitrogen is not the only source of nitrogen in MBE though it is by far the most common. Comparing plasma-assisted and ammonia-MBE growth to examine the effects of the different sources and growth conditions, four defects (\(E_C - 0.62\), \(E_C - 1.28\), \(E_C - 2.62\), and \(E_C - 3.28\)) increased between 5 – 20X in the ammonia-MBE materials. As ammonia-MBE has not been researched to nearly the same degree as plasma-assisted MBE, these differences may decrease dramatically as the ammonia-MBE GaN is further optimized. Carbon, the gallium-vacancy related level, and the point defect-dislocation complex are all areas of improvement for ammonia-MBE growth. This comparison results from the best ammonia-MBE growth, which was studied by characterizing samples with increasing V/III flux ratios. It was found that higher ammonia flow rate for a fixed gallium flux resulted in lower defect concentrations with the \(E_C - 0.25\) eV level having the highest sensitivity. This concluded the study of growth conditions on deep level incorporation.

Also studied was the effect of various substrate conditions looking first at the effect of crystal polarity: Ga-face (0001) and N-face (000\(\overline{1}\)). As with other changes, the main defects remain the same and generally in similar concentration between the two orientations except for the \(E_C - 0.25\) and \(E_C - 1.28\) eV traps where the \(E_C - 0.25\) level thought to be nitrogen vacancy related is increased 10X in the N-face sample while the \(E_C - 1.28\) is absent. This is a good sign for N-face HEMTs because the total defect concentration remains about the same but the reversed direction of polarization offers many advantages over the typical AlGaN on GaN Ga-face structure. Crystal polarity is not the only substrate variable that influences defect incorporation as threading dislocations were found to significantly influence defect
incorporation. The impact of dislocations was quite clear with the reduced TD density resulting in improved Schottky diode characteristics and reduced defect concentrations of the $E_C-0.60$ and $E_C-1.35$ levels. The capture kinetics of the $E_C-0.60$ eV level gave evidence that this level could be dislocation related and the clear correlation with TD density gave even more evidence that this level was a point defect-dislocation complex. The $E_C-1.35$ not only correlated strongly with dislocation density but also with carbon concentration leading to the conclusion that this level is likely a carbon point defect-dislocation complex. Finally, the $E_C-0.90$ level thought to be dislocation related based on its capture kinetics was instead revealed to be independent of TD density suggesting instead that this level may form due to point defect clusters existing in the non-dislocated field regions.

Whereas the bulk of the measurements have been on $n$-GaN, some initial measurements have been performed on MOCVD $n$-AlGaN as well. Not surprisingly, the energy levels in the AlGaN are different than in GaN, but many of the same sources are suspected to cause the observed levels in AlGaN. The dominant features of the AlGaN spectrum, the $E_C-3.10$ trap is probably cation-vacancy related (i.e. gallium vacancy in GaN) and the near conduction band acceptor at $E_C-3.93$ is possibly a magnesium substitutional level. The range of defects in AlGaN and GaN will result in a range of time constants associate with emission from these deep levels and this will manifest itself as time and bias-dependent transients in resistance, current, threshold voltage, and other HEMT parameters.

Although much of this work was used to characterize the component layers and aid in material optimization and defect source identification, this knowledge of defects in nitride materials can now be leveraged to understand the defect sources and vertical location of the levels in HEMTs. This knowledge can also be used in HEMT measurements to separate surface related defects from the bulk defects measured by DLOS and DLTS. The next two chapters
develop methods and measurements of HEMTs to extract defect energies and concentrations then compare initial results with the defects determined here.
5.9 References


180


Chapter 6

Deep level characterization of HEMT transistors

HEMT transistors are composed of the same materials discussed in the previous chapter, but HEMTs also have high lateral fields in normal operation because of the high drain-source voltages that cannot be replicated in Schottky diodes structures in addition to surface traps and hot carriers effects that either cannot be observed or easily separated from other effects. It is highly desirable to measure on actual HEMTs, so all of the trap-related phenomena can be measured and characterized. However, the mere fact that all of these types of defects can be observed makes it difficult to isolate and identify them. In addition, the DLOS and DLTS theory formalized 30-40 years ago for diodes cannot be applied to 3-terminal transistors without losing all the potential benefits of measuring on a transistor. In the two-terminal measurement, the transistor has its source and drain shorted and is then treated as a diode with the capacitance being generated by the gate contact, so virtual gating effects and other bias specific phenomena cannot be observed. Another downfall of the two-terminal transistor measurement is that a sufficient capacitance is needed to measure with standard equipment, which is typically a larger than the gate area of an optimized HEMT. So now a three-terminal, non-capacitance method is needed to measure all the trap-related phenomena along with a method to differentiate defect levels both vertically and laterally in the HEMT structure. This research will leverage the knowledge of the trap spectra in the “bulk” measurements in the
previous chapter to understand the defect sources in the complex HEMT structures. Other
groups have used drain current to characterize HEMTs, but often without any formalized
theory.[1–5] Others have used theories developed for MESFETs and incorrectly applied the
measurement and analysis to HEMTs.[6–8] While the non-quantitative drain current mea-
surements are somewhat useful for identifying levels in a particular device because everyone
assumes and probably agrees that an onset in the steady-state drain current or a change of
slope in the current transients correlates to a defect level, the lack of formalization meant
that defect concentrations could not be determined. This has an unwanted consequence
of not allowing comparison of defect concentrations or relative defect concentrations be-
tween samples. Much of the power of defect spectroscopies comes from carefully designing
experiments where a single variable (e.g. growth condition, substrate, or device geometry)
is adjusted to observe any changes between the growths. Without being able to compare
defect densities, this is not possible. The purpose of this chapter is to both formalize a theory
and measurement setup that enables both determination of defect energy level and defect
concentration and does so with vertical and lateral resolution to spatially locate the levels
using drain current and transconductance measurements. In addition, it is proposed that this
theory has the added benefit of providing a direct connection between device relevant param-
eters such as change in drain-gate access region resistance and change in threshold voltage
on the presence and concentration of specific deep level defects. The equipment necessary
to perform these measurements was not available, so a custom printed circuit board (PCB),
field programmable gate array (FPGA), and Labview programming was used to implement
the feedback circuit and measurement apparatus. The uniqueness of the circuit design has
led us to submit an invention disclosure that is currently pending as a patent application.
Several test runs were performed to verify the expected role of defects in the drain-gate access
region and under the gate. Finally, some initial measurements on an unpassivated HEMT device are shown.

6.1 Theory of constant drain-current and constant conductance

DLOS/DLTS

Dispersion is used to denote a difference between the dynamic and from the static characteristics of a device.\cite{9, 10} The two sources of dispersion are self-heating caused by the power dissipation in the device and changes in trapped charge.\cite{11} The effects of self-heating need to be considered when designing an experiment and choosing operating points for the experiments. In most cases, the thermal transient is either allowed to reach steady state before measuring or in the case of temperature-based scans is avoided by measuring as close to an off-state as possible. By measuring a voltage transient while keeping the drain current constant, several simple relations can be developed so that trap concentrations can be extracted from HEMTs under near normal operation. This not only replicates real world conditions where very high electric fields are present, but also as an intermediate step measures variation in device parameters (threshold voltage and gate-drain access resistance) that can directly be used to predict device performance. This section develops the theory of operation for HEMTs with defects in the gate-drain access region and under the gate and measurement conditions necessary to measure the desired parameters.

6.1.1 Theory for gate-controlled measurements

Developing a method to measure the effects of traps on threshold voltage, resistance, and conductance in HEMTs first requires equations relating the various parameters. The basic procedure followed here is to first develop the equations assuming there are only traps directly under the gate in the AlGaN, at the AlGaN/GaN interface, or in the GaN. Then, the conditions
necessary for this approximation will be presented and tested to show that other defects can be ignored while operating in the saturation regime. Much like the conventional DLTS theory, the hope is to develop a first-order model, which can be used to estimate trap concentrations and identify defect signatures (energy and cross section). Later, once a working model is obtained, the theory can be further refined to include effects of electric field, tunneling, or other parameters much like the original theory for DLTS.[12]

We first consider an ideal HEMT with defects located only under the gate. The following equations are borrowed from Ref. 13 and adapted to include the effects of trapping. In the saturation regime, the drain current $I_{DS,sat}$ can be written as[13]

$$I_{DS,sat} = \frac{2 (V_{GS} - V_T)^2 W \varepsilon_2 \mu}{L d} \left[ 1 + \sqrt{1 + \xi_d} \right]^{-2}$$

(6.1)

where $W$ is the channel width, $L$ the channel length, $V_T$ the threshold voltage, $d$ the gate-2DEG distance, $\varepsilon_2$ is the permittivity of AlGaN, $\mu$ is the 2DEG mobility, and $v_{sat}$ is the saturation velocity. The parameter $\xi$ is a measure of whether the transport is mobility limited ($\xi \approx 0$), saturation velocity limited ($\xi >> 0$), or in between and is given as

$$\xi_d = \frac{2 \mu (V_{GS} - V_T)}{v_{sat} L}$$

(6.2)

The transconductance $g_m$ in saturation is given as[13]

$$g_{m,sat} = \frac{\partial I_{DS,sat}}{\partial V_{GS}}_{V_{DS}=\text{const}} = \frac{2 W \varepsilon_2 \mu (V_{GS} - V_T)}{d L} \left\{ (1 + \xi_d) + \sqrt{(1 + \xi_d)} \right\}^{-1}$$

(6.3)

In the triode regime, the drain current is given as[13]

$$I_{DS,\text{triode}} = \frac{1}{1 + \frac{\mu V_{DS}}{V_{sat} L}} \left\{ \frac{W \varepsilon_2 \mu}{L d} \left[ (V_{GS} - V_T) V_{DS} - \frac{V_{DS}^2}{2} \right] \right\}$$

(6.4)

The 2DEG sheet charge density can be expressed as[13]

$$n_s = \frac{\varepsilon_2}{q d} (V_{GS} - V_T)$$

(6.5)
where \( V_T \) accounting for trapping effects is

\[
V_T = \Phi_b + \frac{1}{q} E_{fi} - \frac{1}{q} \Delta E_C - \frac{q}{\varepsilon_2} \int \int N_{T,2}^{bulk} (t, y) dy dy + \frac{qd}{\varepsilon_2} N_{T,1}^{bulk} (t) \tag{6.6}
\]

where \( N_{T,2}^{bulk} \) is the volume concentration of ionized traps with arbitrary defect distribution in AlGaN, \( N_{T,1}^{bulk} \) is an assumed uniform ionized trap volume concentration in GaN, \( E_{fi} \) is the Fermi level with respect to the conduction band edge in the channel layer, \( \Delta E_C \) is the conduction band discontinuity at the AlGaN/GaN interface, \( y \) is the distance from the Schottky contact into the sample, and \( t \) is time. These parameters are depicted in Fig. 6.1 as well. In general, the traps in the GaN will be filled in accumulation (\( V_{GS} > V_T \)) because the Fermi level is near the conduction band throughout the GaN buffer. In depletion (\( V_{GS} < V_T \)), the occupancy of traps will depend on the location of the Fermi level allowing for both thermal transients as well as optical excitation. The trap occupation of levels in the AlGaN cannot be generalized and will always depend on the location of the Fermi level. The integral for the traps in the AlGaN will generally be integrated over the entire AlGaN thickness \( d \) because this
layer is always depleted while the integral for the traps in the GaN layer will be integrated over the depletion region, which only exists when $V_{GS}$ is less than or around $V_T$. Outside the depletion region, a wealth of electrons are available to recombine with any ionized levels. This equation can be simplified if one assumes various deep level dependencies on depth $y$. In this case,

$$V_T = \Phi_b + \frac{1}{q}E_f - \frac{1}{q}\Delta E_C - \frac{qd}{K_2\varepsilon_2}N_{T,2}^{bulk}(y, t) + \frac{qd}{\varepsilon_2}N_{T,1}^{bulk}(t)$$  \hspace{1cm} (6.7)$$

where $K_2$ is a constant correcting for the deep level profile in $y$. The rest of the derivation assumes a deep level in the AlGaN layer that is uniformly distributed, so $K_2$ is 2. To work with sheet charge densities is simpler, so sheet trap concentrations are defined as

$$N_{T,1} = N_{T,1}^{bulk} x_d \quad \text{ and } \quad N_{T,2} = N_{T,2}^{bulk} d$$  \hspace{1cm} (6.8)$$

where $N_{T,1}$ and $N_{T,2}$ are the total sheet charge densities for the GaN and AlGaN, respectively, $x_d$ is the depletion region in the GaN, and $d$ is the thickness of the AlGaN barrier layer. The rest of this derivation follows the normal capacitance DLTS/DLOS derivation in Chapter 4 making the same assumptions in many cases and developing similar relations. Assuming biasing conditions exist such that the trap is initially completely filled and at some quiescent bias the trap will eventually completely empty, the following relations for the ionized sheet charge density $n_T$ are possible[12]

$$\frac{dn_T}{dt} = -e_n n_T$$  \hspace{1cm} (6.9)$$

and

$$n_T(t) = n_{T0} \left[ 1 - \exp \left( \frac{-t}{\tau_n} \right) \right]$$  \hspace{1cm} (6.10)$$

where $e_n$ is the electron emission rate and can be for either an optical or thermal process, $\tau$ is the electron emission time constant and is $1/e_n$, $n_T$ is the ionized sheet charge density for either the AlGaN or GaN trap, and $n_{T0}$ is the total trap sheet concentration for either the
AlGaN or GaN trap. All of this can be substituted back into Eq. (6.7) to obtain

\[ V_T(t) = \Phi_b + \frac{1}{q} E_{fi} - \frac{1}{q} \Delta E_C \]

\[ - \frac{q d}{\varepsilon_2} \left[ \frac{1}{2} n_{T0,2} \left\{ 1 - \exp \left( -\frac{t}{\tau_{n,2}} \right) \right\} + n_{T0,1} \left\{ 1 - \exp \left( -\frac{t}{\tau_{n,1}} \right) \right\} \right] \]

where \( \tau_{n,1} \) and \( \tau_{n,2} \) are the GaN and AlGaN electron emission time constants, respectively, and depend on temperature, activation energy and cross section following Eq. (4.22) from Chapter 4. Now, a relationship between defects under the gate and the threshold voltage has been obtained. This equation can be used with Eq. (6.1), (6.3) or (6.4) to examine how the drain current and transconductance vary with changes in trap occupancy.

Because a relation now exists between the threshold voltage and trap occupation, the total change in threshold voltage \( \Delta V_{T0} \) for a defect uniformly distributed in the AlGaN layer would be

\[ \Delta V_{T0} = V_T(\infty) - V_T(0) = -\frac{q d N_{T2}}{2\varepsilon_2} \]

(6.12)

For simplicity, assume double boxcar DLTS is used to measure this threshold voltage change. In this case, DLTS measures the threshold voltage at times \( t_1 \) and \( t_2 \). The change in threshold voltage is

\[ \Delta V_T = V_T(t_2) - V_T(t_1) = \frac{q d N_{T2}}{2\varepsilon_2} \left[ \exp \left( -\frac{t_2}{\tau_{thh}} \right) - \exp \left( -\frac{t_1}{\tau_{thh}} \right) \right] \]

(6.13)

where \( \tau_{thh} \) is a generic electron emission thermal time constant for a GaN or AlGaN defect level,

\[ \Delta V_T = \frac{\Delta V_{T0}}{r} \]

(6.14)

and \( 1/r \) is the fraction of the transient magnitude observed. This is given by[12]

\[ r = \frac{\beta / \beta - 1}{\beta - 1} \]

(6.15)
where

$$\beta = \frac{t_2}{t_1}$$  \hspace{1cm} (6.16)

Measuring with a fixed $\beta$ and scanning the temperature, the DLTS signal $\Delta V_T$ will have a maximum (peak) determined by

$$\frac{\partial \Delta V_T}{\partial t} = -\frac{t_2}{\tau_{th}} \exp\left(-\frac{t_2}{\tau_{th}}\right) + \frac{t_1}{\tau_{th}} \exp\left(-\frac{t_1}{\tau_{th}}\right) = 0$$  \hspace{1cm} (6.17)

Solving this yields

$$\tau_{th} = \frac{t_2 - t_1}{\ln\left(\frac{t_2}{t_1}\right)}$$  \hspace{1cm} (6.18)

Thus, a given choice of $t_2$ and $t_1$ determines the electron emission time constant or electron emission rate ($e_n^{th}$), which is the same relation as in traditional capacitance DLTS.[12] The unique combination of $t_1$ and $t_2$ is commonly referred to as a rate window and is referred to by the emission rate that can be calculated from Eq. (6.18). As in standard DLTS theory, the emission rate has an exponential dependence on temperature given by Eq. (4.22) and repeated here for clarity[12]

$$e_n^{th} = \frac{1}{\tau_{th}} = c_n N_c \exp\left(\frac{E_T - E_C}{kT}\right)$$  \hspace{1cm} (6.19)

which can be rearranged to

$$\ln\left(\tau_n^{th} T^2\right) = (E_C - E_T) \frac{1}{kT} - \ln\left(\gamma_n \sigma_n^{th}\right)$$  \hspace{1cm} (6.20)

where

$$\gamma_n = \left(\frac{\nu_{th}}{T^{1/2}}\right) \left(\frac{N_C}{T^{3/2}}\right)$$  \hspace{1cm} (6.21)

and

$$c_n = \sigma_n^{th} \nu_{th}$$  \hspace{1cm} (6.22)
The thermal velocity is \( v_{th} \), \( c_n \) is the electron capture coefficient, \( k \) is Boltzmann’s constant, \( T \) is the temperature, \( \sigma_{th}^n \) is the thermal cross-section for electrons, and \( N_C \) is the conduction band edge density of states. Performing temperature scans with several \( t_2, t_1 \) combinations, a set of temperatures corresponding to a set of emission rates can be plotted on Arrhenius plot of \( 1/kT \) vs. \( \ln \left( \tau_{th}^n T^2 \right) \). Using Eq. (6.20), a straight line can be fit to the data to yield both the trap energy and thermal cross-section.

Before discussing how to measure the change in threshold voltage, the relationships for DLOS are first derived. The correlation between threshold voltage and trap concentration has already been derived, so the trap concentration can already be calculated from steady-state photo-induced threshold voltage shifts using Eq. 6.12

\[
n_{T0} = -\frac{2\varepsilon_2}{qd} \Delta V_{T0} \quad (6.23)
\]

Determining the electron optical cross section \( \sigma_n^o \) requires further work. Following the derivation of Ref. 12, the optical electron emission rate \( e_n^o \) can be calculated using

\[
\left. \frac{dn_T}{dt} \right|_{t=0} = -e_n^o n_{T0} = -\Phi(h\nu)\sigma_n^o(h\nu) n_{T0} \quad (6.24)
\]

where \( \Phi(h\nu) \) is the photon flux of light and \( h\nu \) is the energy of the incident light. The relation between \( n_T \) and \( V_T \) is known so the time derivative of

\[
V_T(t) = \Phi_b + \frac{1}{q} E_{fi} - \frac{1}{q} \Delta E_C - \frac{qd n_{T0,2}}{2\varepsilon_2} \left[ 1 - \frac{n_{T2}(t)}{n_{T0,2}} \right] \quad (6.25)
\]

is simply

\[
\frac{dn_{T2}(t)}{dt} = \frac{2qd}{\varepsilon_2} \frac{dV_{T0}(t)}{dt} \quad (6.26)
\]

Substituting Eq. (6.26) into Eq.(6.24), the optical cross-section can be calculated as

\[
\sigma_n^o(h\nu) = -\frac{\left. \frac{dn_{T2}}{dt} \right|_{t=0}}{\Phi(h\nu) n_{T0,2}} = -\frac{2qd}{\varepsilon_2 \Phi(h\nu) n_{T0}} \left. \frac{dV_T}{dt} \right|_{t=0} \quad (6.27)
\]
Hence, the optical cross-section can be determined from the derivative of the threshold voltage at \( t = 0 \), which is very similar in principle to the original capacitance DLOS.[12] Nearly identical relations are possible for bulk GaN-related levels as well where the only differences is the factor of 2 is removed assuming they are uniformly distributed.

**Gate-controlled constant drain current deep level transient/optical spectroscopy (CI_D-DLTS/DLOS) theory**

So the question is how to determine the change in threshold voltage, so both DLTS and DLOS can be performed. If \( I_{DS} \) is measured while \( V_{GS} \) and \( V_{DS} \) are held constant, the change in \( V_T \) due to trap emission has a complicated relationship with \( I_{DS} \), which given with explicit time dependence is

\[
I_{DS,sat}(t) = \frac{2We_{2}\mu}{Ld\left(1 + \sqrt{1 + \xi_d(t)}\right)^2} [V_{GS} - V_T(t)]^2
\]

This is a difficult relation to determine how a change in current relates to a change in threshold voltage and maybe even more importantly all of the time invariant parameters need to be known to determine the current-to-trap-concentration relation. If the experiment is setup differently (i.e. the drain current is kept constant using the gate voltage as a feedback), the relation to gate voltage becomes very simple

\[
V_{GS}(t) = \sqrt{\frac{I_{DS,sat}Ld\left(1 + \sqrt{1 + \xi_d}\right)^2}{2We_{2}\mu}} + V_T(t)
\]

The time dependence of \( \xi_d \) is lost because \( V_{GS} - V_T \) is constant. The voltages \( V_{GS} \) and \( V_T \) are offset from each other by all the constants and although the constants may vary with temperature they are slowly varying compared with the measurement time. Additionally, only the change in voltage is necessary, which means

\[
\Delta V_{GS}(t) = \Delta V_T(t)
\]
This relation also holds true in the triode regime, but the usefulness of this relation in this regime is greatly reduced because $I_{DS}$ also depends on $V_{DS}$. The greatness of this relation is that none of the other constants ($\mu$, absolute $V_T$, $L$, $W$, or $d$) need to be explicitly measured now. Before moving on to a method for measuring defect concentrations in the access region, it is important to note that the lack of dependence of $V_{DS}$ has positive implications on measurements of defects under the gate in the saturation regime. Any defect occupation change in the access region due to virtual gating, hot carriers, etc. impact the 2DEG density in the access region and can be thought of as change in resistance or variable resistor as shown in Fig. 6.2. In saturation, any change in $V_{DS,int}$ caused by $V_{DS} - R_D(t)I_{DS}$ will not impact $I_{DS}$ and hence $V_{GS}$ will not have to change to keep $I_{DS}$ constant. In theory, measurements in the saturation region with $V_{GS}$ feedback will only detect defects directly under the gate. This premise is tested later in Sec. 6.3 but is a very important conclusion.
Gate-controlled constant conductance deep level transient/optical spectroscopy (CG-DLTS/DLOS) theory

Constant conductance measurements were the inspiration for the new measurement techniques following a theory developed by Deen and Kolev for DLTS measurements on Si MOSFETs and later for Si JFETs.[14, 15] Because the equations are slightly different for HEMTs, the theory can be rederived with these equations to see if the relations still hold. Using the equation for the transconductance in saturation, it is possible to relate variations in the threshold voltage to the transconductance. In this case, the gate voltage is adjusted to maintain a constant transconductance between the source and drain. This allows Eq. (6.3) to be rearranged lumping all of the constants together to give

\[ V_{GS}(t) = V_T(t) + \frac{g_{m,sat} d L}{2 W \varepsilon_2 \mu} \left\{ (1 + \xi_d) + \sqrt{1 + \xi_d} \right\} \]  

(6.31)

Again, the measurement is conducted such that \( V_{GS} - V_T \) is constant, which is needed to maintain a constant transconductance so the terms involving \( \xi_d \) are constant. In this case though, there is a caveat because in the saturation velocity limited regime where \( \xi_d \) is large, Eq. (6.3) becomes

\[ g_{m,sat} = \frac{W \varepsilon_2 v_{sat} d}{d} \]  

(6.32)

Therefore, when \( V_{GS} \) is much larger than \( V_T \), the gate length is small, and/or the mobility is high, the transconductance can become constant and no longer controlled by \( V_{GS} \). Therefore, this velocity saturation regime needs to be avoided in this mode. This is possible but may be more difficult in highly-scaled HEMTs.

The transconductance is fixed, so the HEMT is basically a resistor with resistance \( 1/g_{m,sat} \). These equations were developed for the intrinsic transistor without parasitics. If the gate-drain access resistance is again considered as in Fig. 6.2, the resistance adds to the \( 1/g_{m,sat} \) resistance. Therefore, changes in access resistance would precipitate a change in \( V_{GS} \) to
maintain a constant resistance or transconductance. This seems like a problem, but in reality this is just another limit. If the transconductance is small, the resistance is large and the access region effects will be negligible assuming the intrinsic transistor resistance is large compared with the change in resistance of the access region. This limitation is similar to the $\xi_d$ limitation and basically means that the transconductance needs to be small (i.e. operating just above pinch-off).

Finally, with the caveats of operating such that $1/g_{m,sat}$ is large compared with $\Delta R_D$ and the transistor is not saturation velocity limited, the same relation is developed as in CI$_D$-DLOS where $\Delta V_{GS} = \Delta V_T$.

So far, the theory would suggest that the same information would be detectable via either method, but because of different biasing conditions during the experiments, it may be possible that one method (i.e. the conductance mode) would be more bulk sensitive, but this has not been well tested to confirm this possibility. Conductance mode has not been tested as well and current mode as it is still a work in progress. Until recently, these measurements could not be performed because the equipment needed to vary the voltage did not exist. The experimental setup and custom electronics are discussed in Section 6.2. Additionally, the isolation of under the gate defects from access region defects has not been tested in this mode at this point and relies solely on the theory to develop the separation of the two regions.

### 6.1.2 Drain-controlled CI$_D$-DLOS theory

Defects in the access region are suspected to play a large role in the observed dispersive effects. The gate controlled CI$_D$-DLOS (GC CI$_D$-DLOS) is sensitive to defects under the gate but not to the access region if performed in saturation. It is proposed that GC CI$_D$-DLOS cannot be used to assess access region defects and that drain voltage-controlled measurements are
the answer. Before embarking on this, it is necessary to understand how defects in the access region affect the access region. Consider the charge distribution and resulting band diagram for an AlGaN/GaN heterostructure in the access region where no metal Schottky contact is atop the AlGaN in Fig. 6.3. The definitions for the symbols is given in the figure caption, but it is assumed a surface trap is contributing to the charge distribution of the structure. The net polarization $\Delta P$ induces a charge distribution on either side of the AlGaN. This causes a 2DEG with nominal concentration $n_s$ to form at the AlGaN/GaN interface. This charge balances with charge from the surface $\sigma_s$ to yield charge balancing. A surface trap is introduced that induces a reduction in the 2DEG density to $n_s - n_T$. As $n_T$ becomes larger due to electron emission, the 2DEG concentration is locally decreased. While one could consider this as a
threshold voltage variation that depends position, time, and bias history, the solution to such a problem is not simple. Instead, the resistance can be considered. Given the situation in Fig. 6.3 and assuming this is uniform throughout the access region, the gate-drain access resistance \( R_D \) would be

\[
R_D = \frac{L_{DG}}{qW\mu(n_s - n_T)}
\]  

(6.33)

where \( L_{DG} \) is the gate-drain separation. As \( n_T \) increases, the access resistance increases. This is the simplest case where everything is uniform, but in the most general case

\[
R_D = \int \frac{1}{qW\mu(n_s - n_T)} \, dx
\]  

(6.34)

Unfortunately, the instantaneous values above are not known for every position between the gate and drain so additional simplifications are needed based on the known characteristics of nitride HEMTs. In pinch-off, the virtual gate effect results in leakage of electrons into the access region, which depletes the 2DEG in this region and results in most of \( V_{DG} \) being dropped across this depleted region.[16] This realization prompted a simplified model of the access resistance to be developed where the access resistance consists of two resistances: one from the virtual gate and one from the normal resistivity of the access region channel. This is depicted in Fig. 6.4. In these regions, the resistance is considered approximately uniform. The total access resistance becomes

\[
R_D = \frac{L_{dep}(t)}{qW\mu[n_s - n_T(t)]} + \frac{L_{DG} - L_{dep}(t)}{qW\mu n_s}
\]  

(6.35)

Both \( L_{dep} \) and \( n_T \) have time dependence but at first the change in resistance is considered only in the extremes where the time dependence of \( n_T \) and \( L_{dep} \) need not be known. Comparing the change in resistance between the state when the virtual gate is approximately gone at low bias conditions (\( t = \infty \)) and when the gate extension is maximized at large pinch-off bias conditions (\( t = 0 \), the change in resistance caused by \( n_T \) can be calculated. The
Figure 6.4: Schematic depiction of a cross section of a HEMT with the gate and drain shown. A 2DEG is formed at the AlGaN/GaN interface, but after large lateral electric fields in the access region during pinch-off electrons can leak into the access region from the gate contact and be captured by surface or AlGaN states. This time-dependent phenomenon leads to a variation in access region resistance that is represented by a simple model of a variable resistor due to the gate extension and a fixed resistivity for the normal channel resistance.

The total trap sheet concentration of $n_T$ is $n_{T0}$ and the maximum $L_{dep}$ is $L_{max}$. The change in resistance is

$$\Delta R_D = R_D(\infty) - R_D(0) = \frac{1}{qW\mu} \left[ \frac{L_{max}}{n_s} - \frac{L_{max}}{n_s - n_{T0}} \right]$$

(6.36)

where $\Delta R_D$ is always less than zero for an electron trap and from this the total trap sheet concentration can be calculated

$$n_{T0} = \frac{n_s^2}{n_s + \frac{L_{max}}{qW\mu(-\Delta R_D)}}$$

(6.37)

where no further assumptions have been made. Similar to capacitance mode DLTS/DLOS, the assumption that the trap concentration is much smaller than the 2DEG concentration will be made later. This allows DLOS and DLTS expressions to be derived, but the implications and alternatives are discussed later.
The trap concentration and gate-drain access resistance have been correlated using a relatively simple model, which may not be applicable for all of the observed defects but should be considered more as a starting point to estimate the defect concentrations. Later, the AFM-based measurements can be used to determine distributions of defects between gate and drain contacts and along the access region to verify or refine this model.

The approximation needed to develop the time dependence of \( R_D \) is that \( L_{\text{dep}} \) is constant in time. The general thought process is that when the bias is switched from pinch-off to near zero bias conditions, the traps in the virtual gate will uniformly emit their electrons so \( L_{\text{dep}} \) is constant while \( n_s - n_T(t) \) is increasing as \( n_T \) uniformly approaches zero at every position in the depleted region. The time-dependent resistance in this case is

\[
R_D = \frac{1}{qW\mu} \left[ \frac{L_{\text{max}}}{n_s - n_T(t)} - \frac{L_{DG} - L_{\text{max}}}{n_s} \right]
\]

(6.38)

where \( n_T(t) = n_{T0}\exp(-t/\tau) \). After some algebraic manipulation then assuming \( n_{T0}\exp(-t/\tau) << n_s \), one gets that

\[
R_D = \frac{1}{qW\mu} \left\{ \frac{L_{DG}}{n_s} + \frac{L_{\text{max}} n_T(t)}{n_s [n_s - n_T(t)]} \right\} = \frac{1}{qW\mu n_s} \left\{ L_{DG} + L_{\text{max}} \frac{n_T(t)}{n_s} \right\}
\]

(6.39)

The assumption \( n_T(t) << n_s \) is a crux of the capacitance DLOS/DLTS derivation but may not be true in all cases. In capacitance mode DLOS, the solution is to use lighted capacitance-voltage (LCV) measurements, but in drain-controlled CI\(_D\)-DLOS a similar technique is not available. In this case, Eq. 6.37 is needed, which means that whatever trap causes the \( n_T << n_s \) approximation to fail will need to use the full equation and \( dR_D/dt \) and DLTS will be invalid. The derivation continues assuming this approximation, but when the data is presented a case will arise when the approximation is not valid.
In the case of boxcar DLTS, the equation for the change in resistance \( \Delta R_D \) at times \( t_2 \) and \( t_1 \) becomes

\[
\Delta R_D = \frac{L_{\text{max}}}{q W \mu n_s^2} \left[ n_{T0} \exp \left( \frac{-t_2}{\tau} \right) - n_{T0} \exp \left( \frac{-t_1}{\tau} \right) \right]
\]  

(6.40)

In this case, the trap is assumed to be full at \( t = 0 \) and is empty at \( t = \infty \). A similar relation was developed for the capacitance and CG Cl\text{D}-DLTS theories. From this point, the temperature dependence of \( \tau \) can be added showing that boxcar DLTS is possible in drain-control mode.

As \( t_2 \to \infty \) and \( t_1 \to 0 \), the resistance \( \Delta R_{DG,0} \) becomes

\[
\Delta R_{DG,0} = -\frac{L_{\text{max}} n_{T0}}{q W \mu n_s^2}
\]  

(6.41)

showing that the total change in resistance is proportional to the total trap concentration, which is useful for steady-state photo-induced changes in resistance.

It is also possible to perform DLOS and steady-state measurements in this low concentration regime. To calculate the cross-section, the derivative of the resistance at \( t = 0 \) is needed

\[
\frac{dR_D}{dt} \bigg|_{t=0} = \frac{L_{\text{max}}}{q W \mu n_s^2} \frac{dn_T(t)}{dt} \bigg|_{t=0}
\]  

(6.42)

Using Eq. (4.42) the optical cross-section \( \sigma_n^o \) can be calculated

\[
\sigma_n^o (h\nu) = -\frac{q W \mu n_s^2}{\Phi(h\nu) n_{T0} L_{\text{max}}} \frac{dR_D}{dt} \bigg|_{t=0}
\]  

(6.43)

With the cross section, the models from Section 4.4 can be applied to precisely determine the trap energy. The steady-state change in resistance \( \Delta R_{DG,0} \) can also be used to determine the defect concentrations from specific levels as in capacitance DLOS using Eq. 6.41.

With all of these equations derived, one might ask why gate-controlled measurements cannot be used or how do you measure this change in resistance? First, some voltage definitions are needed. The voltage at \( D' \) in Fig 6.4 is considered in the intrinsic transistor drain.
voltage or $V_{DS}^\text{int}$. The voltage applied to the whole structure (i.e. at point D in Fig 6.4) remains $V_{DS}$. GC ClD-DLOS/cannot be used in saturation because changes in $V_{DS}^\text{int}$ have no effect on the drain current of the saturated transistor. Therefore no gate voltage response is observed to maintain the constant current since the current does not change. In the triode region, GC ClD-DLOS is also sensitive to defects in the access region because defects in this region will cause a change in $n_s$ and access resistance. For a constant current in the triode regime, the voltage drop across this access region $V_{DS} - V_{DS}^\text{int}$ will vary necessitating a change in the gate voltage to maintain the constant $I_{DS}$. Unfortunately, it is difficult to relate the $V_{GS}$ shift to a change in resistance in the access region because there are too many experimental unknowns and equations that may not completely accurately represent the transistor under test. However, if $V_{DS}$ is adjusted to regulate $I_{DS}$, the situation is much better. To keep the current constant and be insensitive to defects under the gate, $V_{DS}^\text{int}$ must be constant. Said another way, the effects of $\Delta V_T$ must induce small/negligible variation in $\Delta V_{DS}$. It is possible to estimate any changes in the intrinsic transistor (i.e. the HEMT not including the access resistances) based on the operating condition and the $\Delta V_T$ observed. Using data from a passivated HEMT, a curve of $V_{GS} - V_T$ vs. $V_{DS}$ for constant $I_{DS}$ in the triode regime is shown in Fig. 6.5. Using a constant $V_{GS}$ in the experiment, variations in $V_T$ due to defects under the gate will cause a shift in $V_{GS} - V_T$. The variation in $V_{DS}$ minimized by finding the largest slope in Fig. 6.5. At a typical operating point of $V_{GS} = 0$ V, the slope is approximately $-10$ V/V meaning that a 100 mV $\Delta V_T$ would only result in a $\sim 10 - 15$ mV shift in $V_{DS}$. This demonstrates moderate rejection of $V_T$ shifts when $V_{GS}$ is near or above 0 V for this transistor. This is derived more formally in Appendix C.

If $V_{GS}$ and $V_{DS}^\text{int}$ are kept constant during the measurement period of the experiment where these voltages are defined in Fig. 6.6, the drain current through intrinsic transistor
Figure 6.5: Plot of $V_{DS}$ vs. $V_{GS} - V_T$ to maintain a fixed drain current. The marginal change in $V_{DS}$ is decreased as the magnitude of $V_{GS} - V_T$ is increased. At $V_{GS} = 0$ V for this sample, a 100 mV shift in $V_T$ results in a $\sim 10$ mV shift in $V_{DS}$.

will be approximately constant. The applied voltage $V_{DS}$ will need to be varied to keep $V_{DS}^{int}$ constant because the gate-drain access resistance and resulting voltage drop will vary with time. A circuit diagram for this measurement is shown in Fig. 6.6 showing the constant $V_{GS}$ and constant $I_{DS}$. Restated, the constant current in a HEMT in the triode regime with a constant $V_{GS}$ implies that the $V_{DS}^{int}$ is constant because the intrinsic transistor is maintaining it same state. The $\Delta R_D$ is

$$
\Delta R_D = \frac{[V_{DS}(t_2) - V_{DS}^{int}(t_2)] - [V_{DS}(t_1) - V_{DS}^{int}(t_1)]}{I_{DS}} = \frac{V_{DS}(t_2) - V_{DS}(t_1)}{I_{DS}} = \frac{\Delta V_{DS}}{I_{DS}} \tag{6.44}
$$

The measured voltage is now proportional to the gate-drain access resistance, which can easily be calculated and applied to the DLOS and DLTS theories derived previously.
Figure 6.6: Considering the HEMT to consist of a gate-drain access resistance and intrinsic transistor without this parasitic loss, the circuit shows the biased HEMT with constant $I_{DS}$ and constant $V_{GS}$. The drain voltage of the intrinsic transistor is represented as $V_{DS^{int}}$.

**Estimation of the gate extension**

Almost everything is now known or measurable, but the maximum depletion length for a given pinch-off bias has not been determined. Therefore, a model is presented to estimate $L_{max}$ and is based entirely on unpublished work from Rajan et al.\[16] An analytic model was developed to explain the breakdown behavior in AlGaN/GaH HEMTs where the voltage drop in the gate drain access region is considered in terms of a lateral depletion region near the gate contact and the rest of the access region that has little voltage drop across it.

Using the previous diagram, shown in Fig. 6.4, the depletion region near the gate edge and the schematic view of the electron concentration in the 2DEG is shown. Now shown is the uniform, positive polarization-induced charge $+\sigma_p$ at the AlGaN/GaN interface. In the lateral depletion region, this positive charge is uncompensated resulting in both perpendicular and parallel electric fields. Modeling the charge as a series of infinite line charges that are
oriented along the width of the depleted region, the electric field in the parallel direction (in the direction from the gate to the drain) is given by

\[
E_{||,0}(x) = \int_0^{L_{dep}} \frac{\sigma_p}{2\pi \varepsilon} \frac{t-x}{r^2 + (t-x)^2} \, dt
\]

(6.45)

where \(x\) is the distance from the gate edge \((x = 0)\) to \(L_{dep}\) edge, \(r\) is the distance between the point of the electric field calculation and the AlGaN/GaN interface, and \(t\) is a constant of integration to sum the contributions from each infinitesimal line charge. Because the charge distribution is uniform, it can be recognized that the electric field at each edge has equal and opposite values or

\[
E_{||,0}(0) = -E_{||,0}(L_{dep})
\]

(6.46)

Additionally, there is an electric due to the applied voltage between the gate and drain and assuming that all of this electric field is dropped across the depleted region this additional electric field \(E_{||,1}\) is

\[
E_{||,1}(x) = -\frac{V_{DG}}{L_{dep}}
\]

(6.47)

The total electric field is simply the superposition of both electric fields

\[
E_{||}(x) = E_{||,0}(x) + E_{||,1}(x) = \int_0^{L_{dep}} \frac{\sigma_p}{2\pi \varepsilon} \frac{t-x}{r^2 + (t-x)^2} \, dt - \frac{V_{DG}}{L_{dep}}
\]

(6.48)

Because it is assumed the voltage is dropped entirely across the depletion length, the electric field at \(L_{dep}\) is zero. This forms a transcendental equation

\[
E_{||}(L_{dep}) = \int_0^{L_{dep}} \frac{\sigma_p}{2\pi \varepsilon} \frac{t-L_{dep}}{r^2 + (t-L_{dep})^2} \, dt - \frac{V_{DG}}{L_{dep}} = 0
\]

(6.49)

that can be solved to determine \(L_{dep}\). To determine the maximum depletion length, the DLTS and DLOS experiments need to be considered. For the drain-control measurements, the biasing scheme is to first pinch-off the device and apply large \(V_{DS}\). After sufficient time to
allow the depletion region to reach equilibrium at these conditions, the bias with switched
to the triode regime \(V_{GS} = 0, V_{DS} \sim 1.5\) V. The depletion length at the end of the pinch-off
time determines \(L_{\text{max}}\).

This theory was developed by Rajan et al. to interpret and model the breakdown voltage,
which occurs when the peak electric field at the gate edge reaches a critical value when
impact ionization and avalanche breakdown occurs. The numbers produced by this model
have led to reasonable total trap concentrations, but the accuracy of the model has not been
verified at this point. A combination of scanning Kelvin probe microscopy measurements
and modeling with simulation software such as ATLAS will hopefully elucidate any deviations
from the actual lateral depletion.

### 6.1.3 Measurement procedure

Before discussing issues related to trap emission and capture on an experimental level,
the measurement procedure used for the various DLOS/DLTS methods will be reviewed. In
general, the procedure for DLTS measurements is to first fill the traps using the “fill pulse” then
return the bias to a quiescent point where the traps emit as an exponential decay process.
For the double boxcar analysis used throughout this chapter, the gate or drain voltage is
recorded at two times for a given rate window to determine the change in voltage at a given
temperature. The device is pulsed again to refill the traps and the process is repeated. The
temperature is swept and the change in voltage recorded at each temperature. The procedure
for DLOS is quite similar. A fill pulse is applied to fill the traps then the sample is returned to
the quiescent point. The shallow levels begin to thermally emit and after an experimentally
determined time after the thermal emission has basically ceased a shutter is opened to
allow monochromatic light to shine on the sample. This causes deep levels with energy
\( E_C - E_T < h\nu \), where \( h\nu \) is the photon energy, to emit their electrons to the conduction band. Holes can also be excited with energy \( E_T + E_V < h\nu \) to emit a hole to the valence band, which would result in a decrease 2DEG density opposite the behavior for the electron trap. The entire transient is recorded once the light is incident so that the slope at the beginning of the transient and the steady-state value at the end of the transient can be extracted. The photon energy is increased and the process is repeated. This is summarized in Fig. 6.7 for the gate- and drain-controlled ClD-DLOS experiments where the top part shows the biasing conditions as a function of time and the bottom shows the operating points on the transistor I-V curve. These are just schematic representations, and the precise operating point and I-V curves is given later with the other data. With this in mind, the specifics of gate- and drain-control measurements are presented.

The existing equipment in the lab could not adjust the gate voltage to keep the drain current constant, so new equipment was needed to fulfill this roll, and the specifics of this are discussed in more detail in the next section. The equivalent circuit for this measuring in this mode is shown in Fig. 6.8 where the current is sensed across \( R_{\text{sense}} \) and the gate voltage is adjusted using a proportional-integral controller to maintain a constant drain current. The voltage drop across \( R_{\text{sense}} \) is accounted for using a higher voltage at \( V_{DD} \) to achieve the correct voltage at \( V_{DS} \). Because the current is constant, the voltage on \( R_{\text{sense}} \) is constant and the correction factor is constant as well. The \( V_{GS} \) transient needed to maintain the constant \( I_{DS} \) is recorded and is the source of analysis for the DLOS and DLTS measurements.

The equivalent circuit used to maintain a constant \( I_{DS} \) in drain control mode using a fixed gate voltage is shown in Fig. 6.9. The drain current is measured using the sense resistor and the voltage \( V_{DD} \) adjusted by a proportional-integral controller to maintain a constant \( V_{\text{sense}} \) or \( I_{DS} \). The PI controller setpoint is simply the voltage corresponding to the desired
Figure 6.7: Depiction of timing diagram and biasing schemes for the gate- and drain-controlled DLTS/DLOS measurements.
Figure 6.8: Circuit diagram for gate-controlled $I_D$-DLTS/DLOS measurements. The voltage drop across the resistor $R_{\text{sense}}$ is used to determine $I_{DS}$ and the proportional-integral controller varies $V_{GS}$ to achieve the desired drain current input by the $V_{\text{setpoint}}$. The $V_{GS}$ transient resulting from the light or bias change is recorded.

Figure 6.9: Circuit diagram for drain-controlled $I_D$-DLTS/DLOS measurements. The voltage drop across the resistor $R_{\text{sense}}$ is used to determine $I_{DS}$ and the proportional-integral controller varies $V_{DS}$ to achieve the desired drain current input by the $V_{\text{setpoint}}$. The $V_{DS}$ transient resulting from the light or bias change is recorded.
drain current times the value of the sense resistor. The $V_{DS}$ transient is recorded allowing for later DLTS or DLOS analysis.

The theories and procedures for these measurements have been described, but a separate issue is how to bias the devices to transition from completely full trap conditions to completely empty conditions. This can be difficult issue in capacitance mode measurements where it is just a Schottky diode but theories have been developed to account for where the deep level crosses the Fermi level. In a transistor structure, this issue is magnified several fold due to the complex vertical and horizontal structure of the device. This has not been completely resolved at this point. In drain-control, the virtual gate forms during pinch-off at high $V_{DS}$, so the filling and emptying is relatively well understood. In gate-control, intuitively one would expect to fill the deep levels with zero or positive $V_{GS}$ as the Fermi level is highest in the bandgap, but the observed behavior suggests that this is not the whole story. If the refilling is enhanced in pinch-off, it might suggest that a mechanism such as trap assisted tunneling is occurring, which leads to a higher average occupancy than the low bias conditions where the emission would out compete the tunneling. Initial reasoning for measuring assuming the latter case is given in Fig. 6.10 where the gate bias before the shutter opening relative to the dark equilibrium gate bias is plotted. If the gate bias drifted downward and had relatively the same shape as the steady-state gate voltage, the deep levels would not be effectively refilled during the “fill pulse”. This same experiment has been repeated on a different sample with the opposite idea of filling and measurement, but was not able to refill as well evidenced by a drift of the gate voltage during the DLOS run. In addition to the refilling problem, it was discovered that gate-control measurements performed at significant drain current were sensitive to both the defects under the gate as well as the source-gate access region. To eliminate the source-gate effects, the measurement must be performed under low
Figure 6.10: Gate bias before the shutter is opened relative to the initial dark gate bias is plotted. If this drifts in the same direction of the $\Delta V_{GS}$ signal, it means that the traps are not refilled by the fill pulse. Here, the maximum $\Delta V_{GS}$ was approximately $-90\,\text{mV}$, but the dark gate bias shifts very little but only slightly negatively (ignoring the first part explained next) and only a small fraction (~ 4%) of the total $\Delta V_{GS}$ indicating that the traps are refilled quite well. The positive onset at the beginning of the run is due to the sample not being pulsed beforehand, so the onset of pulsing causes a slow but positive shift as the sample reaches a new equilibrium. This has been eliminated by pulsing the sample in the dark several hours before starting the measurement with light.
drain-current conditions. The impact of the source-gate region is discussed in Appendix C. Some initial results were performed where the refilling was improved but with unwittingly including source-gate effects. In all cases, the conclusions are still valid.

6.2 Experimental Setup

These new constant current/conductance methods required the design and construction of new equipment to provide a high speed feedback mechanism to adjust the gate or drain voltage and control the drain current or conductance. This equipment is part commercial-off-the-shelf components and part custom-designed circuitry and circuit boards. To control and communicate the transient information to a computer, a series of Labview programs were written to enable these techniques to be performed.

The system being described is actually the second generation design. The first design was only able to use the gate voltage as the feedback and could only perform constant drain-current. Design of this circuit helped immensely in understanding noise levels, time to reach stable control with negligible following error, and only implementing a single measurement in the suite of techniques simplified the design process. The most difficult part of the design turned out to be designing the circuit to turn off the feedback during the fill pulse time. The other shortcomings of the first generation circuit besides having only gate control and constant current were several millivolts of noise, a few millisecond settling time, limited maximum VDS, difficulty tuning the constant current setting, and having to feed the gate voltage to a data acquisition card to digitize the signal. These non-idealities became the focus of the second iteration.

The general design specifications were:

1. Minimal noise (<1 mV)
2. Software configurable constant current, conductance, transconductance, and capacitance capability
3. Proportional-integral (PI) controller with a 1 MHz loop closure rate
4. $0 \leq I_{DS} \leq 150$ mA with 0.1 mA control
5. $0 \leq V_{DS} \leq 50$ V with 0.1 mV resolution
6. $-10 \leq V_{GS} \leq 0$ V with 0.1 mV resolution
7. Settling time $\leq 200 \ \mu$s
8. Set the PI values, voltages, current, transconductance, etc. all via a computer interface

No general purpose computer can perform a 1 MHz loop with accurate timing, which is necessary for accurate control. Using a microprocessor, FPGA, and other high speed digital options were time prohibitive and would require many iterations to get everything to work correctly. Instead, National Instruments (NI) offered a series of products that could meet part of the necessary requirements. The NI PCI-7813R FPGA-based card offered digital output rate up to 40 MHz, 3M gate reconfigurable I/O, 160 digital I/O pins, an embedded ability to communicate with the host computer, and the ability to program it via Labview. Other FPGA cards were available with analog-to-digital inputs and digital-to-analog outputs onboard, but the maximum conversion rate was only 200 kHz. We instead designed our own schematic and circuit board to interface the analog inputs and outputs to the FPGA for further processing. The circuit board consisted of three primary sections identified in the signal conditioning board section of Fig. 6.11 and were the power conditioning to provide the necessary voltages to power the various integrated circuits (IC) (double line components), analog out for generation of gate and drain voltages (components with gray background), and analog in for measuring the drain current directly or capacitance/conductance through the Boonton 7200 capacitance meter (single line components with white background). By building a few test circuits, it was confirmed that the ICs performed and methodology worked as planned allowing the layout of the full circuit. The layout and soldering of the components was handled by Drew Malonis,
Figure 6.11: Block diagram of FPGA-based constant current, conductance, or capacitance measurement system. The power conditioning (double line block components) provides low noise voltages to the other components. The digital-to-analog (gray, single line components) provide the $V_{DS}$ and $V_{GS}$ voltage outputs to the transistor while the analog-to-digital (white, single line components) are used to read the capacitance or conductance from the external input or the $I_{DS}$ current by measuring the voltage across a sense resistor.
who also worked on much of the circuit design. When implemented the drain bias range was 0 - 50 V, the gate bias range is -10 - 0 V, and the maximum drain current is 150 mA. The use of two 16-bit digital-to-analog converters (DAC) with different scaling factor allows for 0.07 mV resolution on the drain-source voltage ($V_{DS}$) while the 16-bit DAC for the gate voltage provides 0.14 mV resolution. Finally, the current is sensed with a resistor above the drain terminal and the 16-bit analog-to-digital converter (ADC) imparts a 2.2 $\mu$A resolution. This meant that the circuit board fulfilled Goals 4-6. In reality, the circuit board's only functionality was conversion of digital and analog signals but was an essential part in achieving our goals.

All the "brains" of the system were programmed into the FPGA. This is where the proportional-integral (PI) controller, memory management, state machine to sequence from the idle to fill pulse to measurement phases, and computer interface were all implemented. Communication with the ADCs was done via a 16-bit parallel bus with a conversion start line to initiate reading of the analog voltage input. DACs rated to 1 MHz were easily available but not with a 16-bit parallel bus, so a serial peripheral interface (SPI) was used instead. The DACs were capable of 25 MHz SPI, so a SPI controller was implemented in the FPGA and consisted of a state machine to cycle through each bit synchronously with the 25 MHz clock signal needed to keep communication coordinated. With the communication to the board determined, a PI controller with a range of feedbacks (current vs. conductance) and various outputs ($V_{GS}$ vs. $V_{DS}$) needed to be implemented. This proved difficult because the code needed to be very compact to fit into the FPGA and the FPGA does not natively support floating point numbers. This led to many cases of integer rollover and as an example multiplication requires a output buffer twice as large as the input buffers but arbitrary increases in buffer sizes results in extremely large, slow code. Therefore, the bits need to be shifted before multiplication resulting in smaller input buffers. Otherwise, the most significant bits can be
lost resulting in a pseudo-random number instead of the desired result. This wreaks havoc on the control algorithm, and as such, much time was needed to eliminate this problem at every multiplication or addition step. This is clearly not an in depth examination of each programming step but a gross overview of the pitfalls. Next, the program layout and usage is explained. Because this program was part of the hardware and experimental technique development, the program usage and development is explained here.

Figure 6.12 shows the front panel of the program that runs on the FPGA (FPGA PID Control.vi). As with other Labview subroutines, this program is called and the variables are passed to and from the controls and indicators on the front panel. So all of the voltages, times, proportional (P) and integral (I) gains, type of voltage control, etc. are all variables other Labview programs use to control the programming running on the FPGA. In addition to the variables on the front panel, a shared memory buffer is employed to pass the current, conductance, $V_{DS}$, or $V_{GS}$ transients, which occur in real-time unlike the other variables, which are set before the fill pulse is initiated, to the computer. In general, this program should not be used directly as all of the variables are in integer format and not scaled, there is no way to recover the transient data if only running this program, and it is difficult to know which variables need to be used for a given set of conditions. Instead, other programs such as FPGA Initialization.vi, DLTS.vi, and DLOS.vi should be used. The FPGA-based experiments have been fully integrated into the existing DLOS and DLTS programs and supplementary programs have been written to assist in tuning and setting up the FPGA to perform the correct measurement.

The main tuning program is FPGA Initialization.vi, which not only sets the times, voltages, and measurement type, but also is used to run short tests with the HEMT connected to determine the ideal PI settings. It also provides a method to set the current $V_{GS}$ and $V_{DS}$
Figure 6.12: This is the front panel of the FPGA PID Control.vi program. Other Labview programs use this to adjust timing and voltages. Not evident is the shared memory the FPGA opens to pass the large arrays of data during the transient acquisition. Labview converts this program to code the FPGA can run and provides the interface for other Labview programs to access the variables on this screen.
voltages. It is extremely important that these voltage be set when the circuit board is plugged in and before the device under test (DUT) is connected as the gate and drain voltages go to random values when the signal conditioning board is plugged in. The FPGA initializes the board to 0 V when it starts, but if the FPGA starts before the circuit board is plugged in or numerous other scenarios, the voltages can be hazardous to the DUT. As a simple rule, it is important for the user keep the DUT unconnected before the voltages have been manually set to reasonable values (0 V). No autotuning is available, so these must be manually tuned which can be done using the following simple algorithm:

- Set the I gain to 0 and the P gain low (0.0001, for example)

- Run a measurement with time between $1 \times 10^{-3}$ and 2 s using the fill pulse, voltage control, and measurement voltages intended for the DLOS or DLTS run later. NOTE: Use conservative settings on the U (upper) and L (lower) limits to prevent damage to the HEMT in case of instability.

- Slowly turn up the I gain running transients between adjustments. Look for improved settling time, increased overshoot, or added noise.

- At the point of increasing noise or significant overshoot, back off the I gain 20%, and start increasing the P gain. This usually ends up being one-tenth of the I gain – approximately. Again look for improved settling time. Typically, a settling time between 50 and 200 $\mu$s can be achieved.

- Once you’ve reached this point, the tuning should be pretty good. Adjust the time scale to 0 to $1 \times 10^{-4}$ s and adjust both the P and I gains slightly to see if the noise, overshoot, and settling time can be improved.
Figure 6.13: This is the FPGA Initialization.vi program used to setup experiments by setting voltages, times, and PI settings. This program provides graphical feedback via a graph to assist in tuning the PI settings for optimal performance – typically with a settling time between 100 and 250 µs.

Once the FPGA is tuned for a given set of conditions, the DLOS or DLTS experiment can be run. A screenshot of the FPGA Initialization program is shown in Fig. 6.13.

The DLTS.vi and DLOS.vi programs are designed to be used in capacitance, constant drain current, constant conductance, and constant capacitance modes. In the cases where the FPGA is used (i.e. anytime a feedback is needed to maintain a constant feedback value), the voltages, measurement setup variables, and everything except the measurement time is read from the FPGA and stored in the data file. Once the values are set using the FPGA
Initialization.vi program, the values do not need to be repopulated in the DLOS.vi or DLTS.vi programs. The operation of the DLTS.vi and DLOS.vi programs is discussed in Appendix B.

6.3 Experimental verification

In theory, any gate controlled measurement that stays within the saturation region and is performed on a device with a perfectly flat $I_{DS}$ in this regime (i.e. $I_{DS}$ does not depend on $V_{DS}$) will not detect any change in the gate-drain access resistance because the voltage drop across the access region is changing the voltage across the intrinsic transistor but not affecting the drain current. Hence, the gate voltage does not have to change to keep the drain current constant. Figure 6.14a shows a gate controlled CI$_D$-DLOS measurement performed in the saturation regime on an unpassivated PAMBE grown HEMT grown on SiC (022808A) with the same measurement point ($V_{DS} = 5$ V, $I_{DS} = 27$ mA) but different fill pulse conditions. The first run (red) fill pulse was $V_{DS} = 5$ V and $V_{GS} = -8$ V while the second run (blue) fill pulse was $V_{DS} = 10$ V and $V_{GS} = -8$ V. The drain voltage was limited to 10 V at the time of these experiments, which is why the measurements were performed with only a 5 V difference in “fill pulse” bias. The steady-state change in gate voltage spectra shown in Fig. 6.14b are nearly identical for both cases. This is an unpassivated device, so if any gate-drain access region-related effects were contributing to the signal, one would expect to observe them in this situation. The lack of any significant difference lends credence to the idea that gate-controlled measurements are insensitive to access region defects if the measurement is performed in saturation. In this case, there are some gate-source access resistance effects because the measurement was performed at a larger drain current. The $\Delta V_{GS}$ is the same for both cases though, so the effects of the source-gate should be the same for both cases as they indeed appear to be.
Figure 6.14: Two gate controlled Cl₂-DLOS measurements performed in the saturation regime on an unpassivated PAMBE grown HEMT grown on SiC (022808A). The first run (red) was performed at $V_{DS} = 5 \text{ V}$ and $I_{DS} = 27 \text{ mA}$ during the measurement phase with a fill pulse of $V_{DS} = 5 \text{ V}$ and $V_{GS} = -8 \text{ V}$. The second (blue) had the same conditions in the measurement phase ($V_{DS} = 5 \text{ V}$ and $I_{DS} = 27 \text{ mA}$), but the fill pulse was $V_{DS} = 10 \text{ V}$ and $V_{GS} = -8 \text{ V}$. The operating points are shown in (a) and the resulting steady-state gate voltage shift is shown in (b). Because the drain voltage is the only difference between the runs, the only expected change is in the deep level occupation in the gate-drain access region. The close agreement of both steady-state spectra suggests that this measurement was insensitive to the changes in the access region as predicted.
Figure 6.15: Gate-controlled DLOS experiment performed on an AlGaN/GaN HEMT with and without passivation. The passivation should in theory not affect the defects detected under the gate. The near overlap of the DLOS spectra suggest that the gate-control is indeed only sensitive to defects under the gate.

For evidence of complete access region insensitivity, some initial measurements of the same HEMT with and without passivation measured in gate-control with a drain current of 1 mA revealed nearly identical behavior despite having a dramatically different response in drain current where this passivation study is currently ongoing and the being performed with Drew Malonis. Swinging the gate voltage from 0 V during the fill pulse to $\sim -5$ V during the measurement, the DLOS spectra for the passivated and unpassivated samples were nearly identical evidenced by Fig. 6.15. The differences in spectra between Figs. 6.15 and 6.14 is due to the source-gate access resistance.

In the initial experimentation phase, gate-controlled measurements in the triode were also acquired. A comparison of the results in saturation and triode regimes are shown in Fig. 6.16. The $\Delta V_{GS}$ signal is greatly increased for the measurement in the triode regime. Because the change in $V_{DS}$ does not effect the $\Delta V_{GS}$ magnitude as shown in Fig. 6.14, the
Figure 6.16: Two gate controlled ClD-DLOS measurement performed on an unpassivated PAMBE grown HEMT grown on SiC (022808A). The first run (green) was performed at $V_{DS} = 2$ V and $I_{DS} = 27$ mA (triode regime) during the measurement phase with a fill pulse of $V_{DS} = 10$ V and $V_{GS} = -8$ V. The second (blue) had the same conditions in the fill pulse phase ($V_{DS} = 10$ V and $V_{GS} = -8$ V), but the measurement phase in this case was $V_{DS} = 5$ V and $I_{DS} = 27$ mA (saturation regime). The operating points are shown in (a) and the resulting steady-state gate voltage shift is shown in (b). The gate bias change is the same for both measurements, so the defects under the gate should have the same response. However, the experiment in the triode region now has a dependence on $V_{DS}$ and the response is much larger than the measurement in saturation due to the access region defects which are impacting the access resistance and voltage on the drain. The measurement in the triode regime is sensitive to both defects under the gate and access regions but is dominated by the access region defects. In this case, it is reasonable to assume that the drain controlled measurements are dominated by access region defects.
difference must be due to the $V_{DS}$ dependence on the defects in the access region causing the access resistance to vary. As the extraction of the access region effects from the difference in $\Delta V_{GS}$ was inherently flawed due to the need to a strong dependence on the absolute $V_T$ value that would cause the equations not to converge or vary greatly with small shifts in the estimated value of $V_T$. The solution of course was the drain-controlled measurements.

For the drain control, the necessary requirement was that $V_{DS}^{int}$ be constant so that the change in resistance can be calculated. It was shown in Fig. 6.5 that the shift of $V_T$ would only cause $V_{DS}$ to vary by $\sim 1/10$th of the $V_T$ shift. In Fig. 6.14, the threshold voltage shifted by $\sim 90$ mV, which predicts an approximate $V_{DS}$ shift of 9 mV. Figure 6.17 shows that the typical $\Delta V_{DS}$ shift for a drain-controlled DLOS measurement of an unpassivated sample is much larger than 9 mV shift caused by $V_T$ shift on the same HEMT suggesting that the small shift in $V_T$ can be ignored. If this condition is not met, it should be possible to measure the $V_{GS} - V_T$ vs. $V_{DS}$ for the measurement current used and use the $\Delta V_{GS}$ vs. $h\nu$ from the gate-controlled measurements to subtract its contribution from the steady-state change in $V_{DS}$.

Finally, though not discussed in the previous derivation to maintain some level of simplicity, the impact of varying source-gate access resistance $R_S$ was also considered. To avoid any deviations due to $R_S$, several further restrictions are placed on the measurement as discussed in Appendix C. For gate-control, the requirement is that $I_{DS}\Delta R_S(t)$ is small compared with $\Delta V_{GS}$, which can be accomplished by having a sample with small total $R_S$ or using low $I_{DS}$ conditions. In the case of drain-control, the situation is more complicated and measurement of $\Delta R_S(t)$ cannot be avoided. To ensure that this term adds to $R_D$ and does not cause second order effects, the measurement should be conducted at large $V_{GS} - V_T$ values, which is already a requirement to avoid coupling in $V_T$ shifts. Generally, the time variation of $R_S$ is considered to be much smaller than the time variation of $R_D$ because the voltage swings on the drain
Figure 6.17: Initial drain-controlled CI\textsubscript{D}-DLOS result of an unpassivated PAMBE-grown HEMT grown on SiC (022808A). The experiment was performed at $V_{GS} = 0$ V and $I_{DS} = 27$ mA (triode regime) during the measurement phase with a fill pulse of $V_{DS} = 10$ V and $V_{GS} = -8$ V. The result here shows that a 140 mV change in $V_{DS}$ was observed in drain control. This can later be converted to a change in access resistance.

side are much larger than those on the source side and it is probably reasonable to ignore its contribution. Some initial measurements in gate-control mode using different measurement current indicate that this is not a problem at least in this mode, but a more systematic study is needed verify this result. With these checks completed, some initial results are given next.

### 6.4 Initial results

This new suite of HEMT characterization techniques can now be applied to an unpassivated AlGaN/GaN HEMT transistor to evaluate the new techniques, examine the defect spectra in an operational HEMT, and compare with previous “bulk” deep level spectra where appropriate. This sample was grown and processed by our collaborators at the University of California, Santa Barbara and more information about the processing can be found in Chapter 3. Because the concentration and energy of defects can vary depending on growth
chamber and because the majority of the “bulk”, Schottky diode samples were grown at UCSB, a better comparison with the HEMT deep level spectra should be possible. Although many of our samples are grown on sapphire because of the low dislocation density that helps with leakage on the large area contacts, state-of-the-art nitride HEMTs are grown on SiC substrates, which is the major difference between the growth. The layer structure for the HEMT, shown in Fig. 6.18, was grown using plasma-assisted MBE on Si-face 6H-SiC and using the optimized growth conditions necessary to prevent Si outdiffusion from the SiC substrate. The access region spacings were 0.5 \( \mu \text{m} \) and 5 \( \mu \text{m} \) for the source-gate and gate-drain spacings, respectively, and had 0.7 \( \mu \text{m} \) gate length. This sample is part of a passivation study and because the passivation was deposited after processing the gate the gate walls are vertical with no field plating or slanted gates. This unpassivated sample, however, never received the passivation layer.

The Si-face of the substrate results in Ga-face (0001) growth, and following the ideal growth conditions mentioned previously for most PAMBE-growth nitrides, the V/III ratio was maintained in Ga-rich conditions growing near the crossover between the intermediate
The mismatch between GaN and SiC is large, so direct growth of nitrides on SiC results in high threading dislocation densities of $\sim 1 \times 10^{10}$ cm$^{-2}$, but high quality, low leakage HEMTs are still attainable. The use of Ga-face growth and the 30 nm 30% Al mole fraction AlGaN on GaN results in formation of a 2DEG and a equilibrium, zero bias band diagram under the gate given in Fig. 6.19. To date, a vast resource on the effects of traps on HEMT device performance exists,[10, 18–21] but little work has attempted to assign specific traps to device performance problems and does so by measuring quantitative deep level concentrations. Combined with the nano-scale resolution from the AFM-based DLOS techniques developed in Chapter 7, these techniques promise to bridge the gap between fundamental material science studies and high-level device measurements.

The most well-developed techniques so far are the gate-controlled mode experiments that focus on defects under the gate. Figure 6.20 shows the gate-controlled CI$_D$-DLTS results of the unpassivated HEMT. The transistor I-V curves for the HEMT are shown in Fig. 6.20a.
Figure 6.20: (a) Pulsing/measuring conditions for gate-controlled DLTS. The highest current corresponds to $V_{GS} = 0$ and the other curves decrease by 1 V. (b) The gate-controlled DLTS results show a shoulder at low temperature that could be an indication of a broad trap or possibly due to a self-heating effect. A mid-temperature peak was evident with a calculated concentration of $\sim 6 \times 10^{10}$ cm$^{-2}$ based on Eq. (6.23). (c) The Arrhenius plot of the DLTS rate window peaks was linear indicating an activation energy $E_C - E_T$ of 0.59 eV. This energy is coincident with a well-known GaN level indicating that this may be a buffer related defect.
with the fill pulse conditions of $V_{GS} = -6 \, \text{V}$, $V_{DS} = 3 \, \text{V}$ and measurement conditions of $V_{DS} = 3 \, \text{V}$, $I_{DS} = 20 \, \text{mA}$. The measurement occurs in saturation, so the shift in $V_{GS}$ corresponds to the shift in $V_T$. Several DLTS rate windows (20–1000 s$^{-1}$) can be seen in Fig. 6.20b where a low temperature shoulder and peak at mid-temperature are evident. In the case of Cl$_D$-DLOS, the DLTS spectra are inverted compared with standard capacitance DLTS. In boxcar DLTS, the times chosen to measure the change in $V_{GS}$ determine the emission rate (rate window) that will result in a maximum change (peak). The peak temperatures are used with the corresponding rate windows in the Arrhenius plot to determine the activation energy and thermal cross section. In this case, downward peaks (valleys) are indications of electron traps and the energy is referenced to the conduction band. With this background, the low temperature shoulder could be a broad electron trap or a self-heating related issue causing a somewhat constant change in $V_{GS}$. The Arrhenius plot for the mid-temperature peak is shown in Fig. 6.20c. The clearly linear trend indicates a deep level energy of $E_C - E_T$ of 0.6 eV and a large thermal cross section. The $E_C - 0.59$ eV energy signature is very similar to the nearly omnipresent $E_C - 0.59 - 0.62$ eV in bulk GaN suggesting that the gate-control may be sensing this level in the GaN buffer. This level at room temperature would have time constant of 50 µs and cause a 19 mV shift in $V_T$ as the device was pulsed between the off- and on-states. The concentration calculated from Eq. (6.23) indicates a trap sheet concentration of $\sim 6 \times 10^{10}$ cm$^{-2}$.

Thermal emission of deep levels is limited to $\sim 1$ eV from a band edge over reasonable ranges of measurement temperatures, so Cl$_D$-DLOS is used to detect deeper levels. The steady-state gate voltage (SSVG) change is plotted in Fig. 6.21a with the bias conditions and transistor I-V shown in Fig. 6.21b. The SSVG spectrum has been flipped as indicated by the inverted scale on the right $\Delta V_T$ axis. Corresponding to a steady-state photocapacitance
Figure 6.21: Gate-controlled DLOS spectra and pulsing/measuring conditions on the transistor I-V curve for an unpassivated AlGaN/GaN HEMT.
measurement, positive onsets (becoming more negative in voltage in this case) indicate electron emission to the conduction band. The trap concentration for individual levels is determined by the step height of the corresponding level. Several onsets are evident in at $E_C-2.3 \ (\sim 1 \times 10^{11} \ \text{cm}^{-2})$, $E_C-3.3 \ (\sim 5 \times 10^{10} \ \text{cm}^{-2})$, and $E_C-3.7 \ (\sim 1 \times 10^{11} \ \text{cm}^{-2}) \ \text{eV}$ with one or more additional defects below $E_C-2.0$ evident by the slope at the beginning of the run. The onset at $E_C-3.3 \ \text{eV}$ suggests that this may be due to the carbon substitutional defect in the GaN buffer located at $E_C-3.28$.\[22,23\] This again suggests that the gate-control method is able to detect deep levels located in the buffer. The $E_C-3.7 \ \text{eV}$ level is higher than the GaN bandgap, so this level must be located in the AlGaN layer. In the only thick AlGaN studied so far (Section 5.7), the closest level to the $E_C-3.7 \ \text{eV}$ deep level was at $E_C-3.93 \ \text{eV}$ and thought to be due to carbon or magnesium. Even after accounting for any small variation in composition, these level are unlikely to be the same, but at the same time it is not unreasonable that MOCVD- and PAMBE-grown films may have different defect spectra. Several runs show a slight onset around $\sim E_C-2.3 \ \text{eV}$, but this level has not previously been observed in $n$-GaN or $n$-AlGaN. Finally, the level or levels with onsets below $E_C-2.0 \ \text{eV}$ have not been resolved at this point although the quartz-tungsten-halogen lamp would allow trap detect down to 0.5 eV. In this case the measurement has source-gate access region sensitivity due to the high drain current (see Appendix C for details) and does not allow a specific conclusion that all of the deep levels are restricted to the access region. The solution to this is perform the same experiment at low drain current. Figure 6.22 was run under conditions to eliminate the effects of source-gate access resistance (i.e. low drain current). Here the total change in $V_T$ is only $\sim 30 \ \text{mV}$ and the AlGaN related level ($E_C-3.7$) is nearly absent. Even the $E_C-3.3 \ \text{eV}$ level is much reduced. The dominant level in the spectra is $\sim E_C-2.0 \ \text{eV}$ evident by the onset in the SSVG.
Figure 6.22: Gate-controlled DLOS experiment (SSVG) performed on an AlGaN/GaN HEMT without passivation. The fill pulse conditions were $V_{GS} = 0$ V and $V_{DS} = 0.5$ V while the measurement conditions were $I_{DS} = 2$ mA ($V_{GS} \sim -5$ V) and $V_{DS} = 2$ V.

Summarizing these measurement capabilities of gate-control, measurements at significant drain current observe both gate and source-gate access region variation while low drain current measurements sense only the region under the gate. These initial results are quite promising as it might indicate that much of the device degradation is to access region effects as the source-gate access region contributed an $\sim 60$ mV shift in the gate voltage, which was twice as large as the $\sim 30$ mV threshold voltage shift.

Having examined the defects under the gate, drain control measurements were performed to examine defects in the gate-drain access region. In drain control, the measurement is performed in the triode regime so that the intrinsic transistor depends on $V_{DS}$ and the fill pulse is designed to apply a large lateral electric field between the gate and drain contacts in pinch-off to evoke the virtual gate extension. Figure 6.23a shows the transistor I-V curve of the same unpassivated HEMT but is a different transistor on the same sample along with
the biasing conditions for the drain-controlled CI_D-DLTS experiment. The DLTS results in Fig. 6.23b with rate windows ranging from 20 to 2000 s\(^{-1}\) where negative peaks indicate electron traps exhibit a single deep level. The Arrhenius plot in Fig. 6.23c of this deep level reveals a good linear fit to the Arrhenius data with an energy of \(E_C - 0.47\) eV and cross section of \(\sim 3 \times 10^{-15}\) cm\(^2\). This energy does not coincide with any PAMBE \(n\)-GaN suggesting that this is not from the GaN buffer but instead from the AlGaN layer or AlGaN interface. From the biasing conditions used during pinch-off (\(V_{GS} = 10\) V), the maximum lateral depletion \(L_{max}\) was estimated at 0.118 \(\mu m\). At the measurement conditions, the -15 mV change in \(\Delta V_{DS}\) corresponds to a -1.5 \(\Omega\) change in gate-drain access resistance. This unpassivated sample has a large enough response that the \(n_T << n_s\) approximation is not quite valid because \(n_T\) is only one third of \(n_s\), so the full equation without approximations, Eq. (6.37), was needed to determine the trap concentration. Using this equation, the concentration calculated was \(\sim 3.7 \times 10^{12}\) cm\(^2\), but due to the estimation of the gate extension and approximations in the model the accuracy of this number is probably within a factor of 2 or 3. A relatively large defect concentration is needed before the resistance is significantly affected.

Drain-controlled CI_D-DLOS measurements were performed to examine the rest of the defect distribution with energy > 1 eV. The drain control measurements were again performed in the triode regime with a fill pulse into pinch-off with a \(V_{DS}\) of 10 V according to Fig. 6.24a. The results of the DLOS run are shown in Fig. 6.24b where a much larger 4.6 \(\Omega\) change in resistance or a total \(\Delta V_{DS}\) of 140 mV was observed for the total change compared with the DLTS results. Three levels appear in the steady-state change in resistance at \(E_C - 1.3\), \(\sim E_C - 2.3\), and \(E_C - 3.7\) eV. In this run, the extracted defect concentrations were of comparable value to \(n_s\), so the equations derived for DLOS analysis are not relevant for this sample. For this reason, the defect energy levels are extracted from the SSVG onsets and not from the DLOS
Figure 6.23: (a) Pulsing/measuring conditions for drain-controlled DLTS. (b) Drain-controlled DLTS results indicate a single distinguishable peak around 250 K with a rough concentration of $\sim 3 \times 10^{12}$ cm$^{-2}$. Because the calculated concentration is on the order of the 2DEG concentration, Eq. (6.37) was used to calculate the trap concentration accounting for all the DLOS and DLTS observed trap simultaneously as the linear relationship breaks down not allowing for traps to be considered individually. (c) The Arrhenius plot of the DLTS rate window peaks was linear indicating an activation energy $E_C - E_T$ of 0.47 eV and a thermal cross-section of $\sim 3 \times 10^{-15}$ cm$^2$. 
Figure 6.24: Drain-controlled DLOS spectra and pulsing/measuring conditions on the transistor I-V curve.
analysis. Because the Fermi level is pinned under the gate at the Ni/AlGaN interface/surface, the defect occupancy cannot be modulated very near the this interface. This means that surface related effects should not be seen in gate control, but this limitation does not exist for drain control. This information can be used to examine energy levels common between the two areas to evaluate potential vertical locations of defects. This works in pre-stressed devices, but where devices have been RF or DC stressed the defect migration or generation may give rise to differences between the two regions. The absence of the $E_C-3.3$ eV level indicates that the drain control is most likely not sensing the GaN layer as there is little to no indication of the GaN band edge or near valence band level omnipresent in GaN. The levels at $E_C-2.3$ and $E_C-3.7$ are again present as they were in the gate-control mode suggesting that even if these levels are at the surface that they penetrate into the AlGaN layer to some extent. Otherwise, these levels should not be sensed in the gate-control mode that should be insensitive to defects that are exactly at the surface. The presence of a level at $E_C-1.3$ eV may be an indication of a GaN level, but this level usually occurs in small concentrations relative to the $E_C-3.28$ eV level so this level may be new and in the AlGaN potentially in the GaN but less likely. It does leave an open question as to how sensitive the drain-control mode is to the AlGaN and GaN layer, but additional biasing scheme may make this question moot as well. Currently, the pinch-off $V_{DS}$ is much smaller than the operating voltage of state-of-the-art HEMTs. The increased bias or only semi-pinching off the channel may result in higher concentrations of hot electrons and these carriers are more likely to affect both the AlGaN and GaN layers.

Because the change in resistance is larger than DLTS results, the linear approximation is still not appropriate. Using Eq. (6.37) the total trap concentration is $\sim 8 \times 10^{12}$ cm$^{-2}$, which is again given with the accuracy level that the concentration could be 2 or 3 times lower
based on the accuracy of the approximations at this point. There is some question to how these deep levels are actually distributed in the channel as well. The deepest levels may not be modulated by the applied field because the Fermi level may not be swept through these level. Instead, these very deep levels may exist throughout the gate-drain access region. The change in resistance is a quantitative number, but to more accurately estimate defect concentrations a survey of the defect distribution is needed. This is currently in progress as part of the AFM-DLOS experiment presented in the next chapter.

These constant current measurements, both gate and drain control, are the first successful demonstration of such measurements determining defect energy and concentration of specific traps in a HEMT. Additionally, these defects have clearly isolated the effect of defects in the access region and under the gate by demonstrating the impact of defects on the gate-drain access resistance and threshold voltage shift, respectively. The high defect concentrations in the access region and large change in access resistance suggest that this may contribute to a number of dispersion effects such as knee walkout. Continuing work will allow correlation with specific dispersion mechanisms with additional measurements such as pulsed I-V and RF measurements.

6.5 Conclusion

Expanding the defect characterization abilities from capacitance-based Schottky diode measurements, several new constant drain-current or constant conductance methods have bee developed to characterize defects in HEMT transistors. Theory was presented to show that these methods can relate changes in operating voltage to maintain the current/conductance with changes in defect occupancy. Similar to the previous capacitance measurements, the theories were derived for both thermal (DLTS) and optical (DLOS) excitation of levels allowing
for measurement of defect energies throughout the bandgap of GaN and AlGaN. Additionally, two modes of operation were proposed to isolate defects under the gate and in the gate-drain access region by using the gate voltage and drain voltage, respectively, as the feedbacks and measuring in the appropriate conditions. For gate-control mode, measuring in saturation with a transistor that has a reasonable independent drain current dependence on drain voltage is necessary for isolating the defect measurement to under the gate. For the case of the conductance feedback in gate control, there are additional requirements that the conductance of the intrinsic transistor be much smaller than the conductance of the access region, so that the intrinsic transistor dominates the conductance and the effects of the access region will be negligible. Additionally, the transistor must be operated in the saturation regime with an operating point that is not completely saturation velocity limited otherwise the transconductance will not depend on the gate voltage. On the other hand, drain-control mode is operated in the triode regime where a simple model is presented separating the intrinsic transistor without the parasitic gate-drain access resistance and modeling this separately as a variable resistor that is composed of the normal channel resistance and an additional term that accounts for the virtual gate or gate extension effect. The requirements for this mode is that $V_{GS}$ be at 0 V or above to minimize the effects of threshold voltage shifts in this mode. This requirement is device dependent and should be verified with a plot of $V_{DS}$ vs. $V_{GS}$ to determine the rejection ratio at the drain current for the drain-controlled test conditions. The other requirements for all of these modes is minimizing self-heating effects as much as possible for DLTS measurements and not measuring in conditions that cause HEMT degradation during the measurement.

Not presented was a conductance-mode drain-control measurement. In this case, the transistor with transconductance $g_m$ has a resistance of $1/g_m$, which is in series with the
gate-drain access resistance. In a constant conductance mode in saturation or triode with a new definition of the intrinsic transistor resistance, the variations from threshold voltage shifts cannot be separated from the access resistance variation because $R_D + 1/g_m$ needs to be constant but both are of comparable value and both are varying. Contrary to the other techniques where the concept has been to keep the intrinsic transistor in a constant state, this method would vary the transconductance to match the total resistance. Therefore, this method seems infeasible for isolating and quantifying defects in the gate-drain access region.

With the theory presented, some initial measurements on an unpassivated AlGaN/GaN HEMT were introduced. The gate control measurements revealed a 105 mV shift in the threshold voltage corresponding to a $\sim 3.5 \times 10^{11}$ cm$^{-2}$ observed trap concentration. The $E_C - 0.59$ and $E_C - 3.3$ levels are possibly due to the GaN buffer as these levels are quite similar to defects measured in thick $n$-GaN films. The $\sim E_C - 2.3$ and $E_C - 3.7$ eV levels were observed in both the gate- and drain-control methods indicating that levels are not contained solely at the surface as the gate metal should prevent detection of surface traps. Additionally, the large energy of the $E_C - 3.7$ eV reveals that this level must be in the AlGaN layer. Finally, the drain-control measurements reveal a DLTS level at $E_C - 0.47$, which appears to be distinctly different than the $E_C - 0.59$ eV level under the gate and in bulk GaN. This suggests that this level may be surface related as this sample has not been intentionally stressed to cause defect generation or migration.

In evaluating these results in both gate- and drain-control modes, it is important to compare the total trap concentrations to existing work to determine if these results are on par with existing measurements. Meneghesso et al. reported near 100% gate lag in similar HEMTs indicating trap densities nearly equal to $n_s$. As these were gate-pulsed measurements and due to virtual gating, the initial trap estimation of $\sim 8 \times 10^{12}$ cm$^{-2}$ in the gate extension would agree
Koley et al. used scanning Kelvin probe microscopy with UV light excitation to observe an \( \sim 80\% \) reduction in \( n_s \), which in this case would be \( \sim 8 \times 10^{12} \text{ cm}^{-2} \) in excellent agreement with the observed variation. The main advantage of the constant current/conductance measurements over these existing measurements is the ability to directly resolve energies levels, separate variation due to threshold shifts from parasitic resistance changes, and spatially resolve the deep levels with both laterally and vertically by correlating with a wealth of knowledge of bulk defects.

Although the initial results are promising, the measurements continue to evolve and improve as the understanding from the measurements feeds back to the theory. Nevertheless, the measurements have reached a stage where they can now be applied to a whole new series of HEMTs including two passivation studies, a study on the influence of RF stressing on deep level incorporation, the effect of dislocation density, the impact of crystal polarity, and finally comparing N-rich PAMBE with existing Ga-face PAMBE technology.
6.6 References


Chapter 7

Next generation of quantitative nano-scale defect characterization in AlGaN/GaN HEMTs

It has been shown that quantitative defect analysis of high electron mobility transistors (HEMT) is possible using a variation of deep level transient and optical spectroscopies, but as macroscopic methods they cannot identify spatial variations in defect concentrations on the nanometer scale. Again, the solution to this problem is development of new measurement techniques. In collaboration with Prof. Pelz’ group in the Physics’ Department, three methods have been devised to attack this problem although all of them involve use of atomic force microscopy (AFM) discussed in Chapter 4. The AFM offers many advantages including atomic resolution and the ability to measure in many modes such as topography, surface potential, capacitance, temperature, current, and electric force. In addition, the AFM can be combined with a near-field scanning optical microscope (NSOM), which allows for light emission or detection on a scale smaller than the wavelength of the light by using evanescent waves of light. Using DLOS as the model, the fundamental principle used in developing these new methods has been to rely on optical stimulation of deep levels to affect the measurement variable initially using surface potential, capacitance, or current although other useful scanning probe techniques may be identified later. These three planned modes of operation are briefly discussed along with some promising first results.
7.1 Near-field scanning optical microscopy

NSOM can be operated in an illumination, collection, reflection or combined illumination-collection mode. NSOM uses a tapered fiber optic cable to bring light exit point from the fiber very close to the sample surface. In the case of the Park Systems XE-NSOM, a cantilever with a 20 – 100 nm hole in the tip is used as the aperture and the typical AFM topography feedback is used to maintain contact or constant separation non-contact mode. The NSOM tip can be scanned across the sample surface while simultaneously recording the NSOM feedback variable, which could be light intensity, current, wavelength, etc. NSOMs are typically designed to work at a fixed wavelength or a few discrete wavelengths, so significant technological challenges exist in modifying an existing commercial NSOM system to allow operation over a continuous range of wavelengths from ~1.2 to ~6 eV. The advantage of NSOM though is the ability to illuminate a nm-scale region of the HEMT limiting defect emission to a small region. By monitoring the drain current, it might be possible to detect variations in drain voltage that would correspond to defect emission using a feedback mechanism similar to the constant-drain-current DLOS. Because the signal-to-noise ratio would be very small, a specially processed HEMT with a small gate width would probably be required so that the equilibrium-current-to-defect-related-current-change ratio would be small. This method can be considered a drain-current DLOS experiment discussed in Chapter 6 being scaled down to the nm-scale. Another use of the NSOM may be in the collection mode for imaging luminescence from an operating HEMT. Meneghesso et al. and others have identified electroluminescence (EL) associated with defective regions in the access regions and correlation of EL to defect spectra using the NSOM emission mode or other localized AFM-DLOS measurements would help answer several open questions about these defects.[1, 2]
Several issues including a dichroic beamsplitter in the NSOM optical column that transmits part of the light spectrum to a coaxial camera and other part to the NSOM light collection/illumination system, a single mode fiber connection that cannot be connected to the fiber bundle from the monochromator output have slowed the progress if this development. Because the NSOM aperture is much smaller than the wavelength of light, the light forms an evanescent wave that not only decays in intensity exponentially from the aperture but also has a very low emission/collection efficiency. A solution to this is to use larger apertures, which decreases spatial resolution but increases light transmission, but there is still the concern that still may not be efficient enough. At this point, several issues have been identified and systems designed to accommodate future NSOM capabilities, and work is currently ongoing with Park Systems to identify improvements to the optical column to the point where testing can begin in earnest.

7.2 Lighted capacitance microscopy

Scanning capacitance microscopy (SCM) is an increasingly common mode of AFM where an AC bias relative to the sample is applied to a metalized cantilever and tip. During this time, the AFM tip is mechanically oscillated at a resonant frequency (50 – 400 kHz, typically) to maintain a fixed sample-tip separation. The AC bias with a frequency range between 890 MHz and 1050 MHz induces a displacement current in the capacitor formed between the tip and sample, which is fed to a lock-in amplifier through a current preamplifier. Measuring the capacitance directly is difficult, so an additional AC bias is applied between 17 – 18 kHz typically with a large amplitude. The variation in capacitance from the peak and valley of the low frequency waveform is measured with the ~1 GHz waveform, so what really is measured
is $dC/dV$. Figure 7.1 shows the typical high-frequency capacitance curve for a $p$-type metal-insulator-semiconductor structure along with the capacitance derivative, $dC/dV$. Although $dC/dV$ measurements are very sensitive, it is difficult to determine an absolute capacitance that can be used to estimate local defect concentrations. Currently being pursued is a solution to measure the capacitance directly using a Boonton 7200 capacitance meter or a homebuilt capacitance meter, but this requires careful design and testing to probe the limits of absolute capacitance measurements. The Boonton 7200 is a lock-in amplifier-based solution with a fixed detection frequency of 1 MHz. Because the Boonton is specially designed to provide high speed, low noise capacitance measurement and is currently used for the large area DLTS and DLOS measurement, it is difficult to scale this from 300 $\mu$m dimensions to 0.1 $\mu$m as this requires a huge increase in sensitivity to achieve similar resolution at the nano-scale. This extremely challenging task is being approached by both engineering creative methods.
to achieve the highest resolution possible while simultaneously setting up a testbed for systematically testing progress in achieving this goal.

Currently being pursued is a homebuilt capacitance bridge circuit small enough to be placed in the AFM head where the current SCM module sits. The advantage of the capacitance bridge is the ability to measure extremely small capacitances (0.1 aF, for example). Building an extremely small homemade capacitance-based Whetstone bridge only capable of measuring extremely small capacitances and mounting it very near the cantilever would minimize stray capacitance allowing for improved SNR. Additionally, it should be possible to use the null voltage from the Whetstone bridge to adjust the voltage on the tip to keep the capacitance of the tip-sample fixed. Such a setup may allow for extremely high resolution. This would be a constant capacitance DLTS/DLOS setup and would not suffer from the normal DLOS/DLTS requirement of $N_T < 0.1 N_D$. For AlGaN/GaN heterostructures, $dC/dV$ is quite high near the threshold voltage, so a constant capacitance experiment would have decreased voltage noise compared with the capacitance noise allowing for further scaling of the device dimensions. Using the feedback circuit designed for the constant drain-current measurements in Chapter 6 and the Boonton, this concept can be tested.

The second thrust to find the scaling limits of capacitance by experimentally testing the response of equipment to smaller and smaller capacitances. To experimentally identify these scaling issues, a new mask has been designed with Schottky diodes ranging in size from 125 nm x 125 nm to 256 µm x 256 µm with each size scaling by a factor of 2. The mask set, shown in Fig. 7.2, is designed to have a series of diodes of each size. At the smallest sizes, a 2D array was fabricated to aid in their search as the dimensions are smaller than the wavelength of visible light and well below the resolution of our microscope. The Schottky contacts are also too small to reliably probe in our LakeShore TTP4 cryogenic stages where the scaling test
Figure 7.2: Mask set designed to explore the minimum capacitance measurable for future absolute capacitance measurements in the AFM. The green layer is the Ohmic contact layer (filled in regions represent where metal is deposited) that forms a web throughout the sample so only one “back” contact is needed. The blue layer defines the mesa etch mask where the filled in areas represent the areas that are not etched. Finally, the red areas and squares define where the semitransparent Ni Schottky contact is deposited. Circular transfer length method structures are shown so the in process analysis is possible to identify the contact resistance and sheet resistivity of the film. This mask set can be used with the typical structures, which are thick films of GaN, AlGaN, InGaN, etc. or heterostructure samples such as AlGaN/GaN. Because the features in the Schottky mask are so small, this mask needs to be written with e-beam lithography while the other masks are just optical masks.
of the capacitance is being performed, so Prof. Pelz’ group has designed a nanopositioner manipulator to accurately contact these small contacts. The smallest contacts may only be usable later in the AFM where the topography maps can first be used to identify the 80 Å-thick contacts before measuring them in capacitance mode. It is expected that the nanopositioner will be usable at least to the 1 µm size diodes. Measurements on this sample and others will allow the Boonton, any custom capacitance meters, and the constant capacitance concept to be tested.

The tip-sample interaction in SCM is usually treated as a parallel plate and with this approximation the theory from Chapter 4 for C-V, DLTS, and DLOS directly applies. Similar to the constant drain current measurements, the change in required the capacitance constant would equal the change in threshold voltage \( V_T \) and the trap density could be calculated from (6.23) shown here again

\[
N_T = \frac{-2e_{AlGaN}}{qd} \Delta V
\]

(7.1)

where the traps are assumed to be uniformly distributed in the AlGaN layer. In this manner, DLOS measurements could be performed at chosen spots in the access regions following nearly the same procedure as normal capacitance DLOS on the micron scale. It would also be possible to scan the tip at fixed incident photon energy to record the voltage shift as function of position allowing for spatial discrimination of deep level concentration. Because the threshold voltage may naturally vary as a function of position due to spatial nonuniformities in AlGaN composition, doping, relaxation, or other piezoelectric induced variation, the deep level concentration variation would need to be plotted by subtracting the image under illumination from the dark illumination producing a \( \Delta V_T \) image.

The full potential of the SCM measurements for DLOS, or even some DLTS if a temperature dependent stage is acquired, has yet to be determined. The initial work shows that the theory
is basically identical to the typical macroscopic DLOS measurements, but higher signal-to-noise ratio may be possible if the experiment is run in a constant-capacitance mode. In this case, the analysis is slightly different but easily derived following much of the gate-controlled constant-drain-current DLOS theory where the change in threshold voltage is measured. This and other increases in signal-to-noise ratio will be explored in the process of scaling down using the new mask set with a range of contact sizes.

7.3 Lighted Kelvin probe microscopy

Scanning Kelvin probe microscopy is (SKPM) is an established scanning probe technique and as such requires the least adaptation to achieve the goals of this work, which is to identify and quantify concentrations of defects with spatial resolution on the nanometer scale. SKPM was first proposed in the 1980’s and became commercially available in the 1990’s, and was used principally to investigate work function-induced differences in highly conductive materials.[4, 5] SKPM offers the ability to track the relative surface potential (RSP), which is basically the difference in work functions between the tip material and sample surface, by minimizing the electric field between the tip and sample surface, which will be shown is an effective means of measuring defect concentration and energy level. The relative surface potential can be monitored as a function of time, position, wavelength of incident light, and biasing conditions giving a large window to experiment in and develop effective means of characterizing materials.

The basic principle of SKPM is to keep the tip voltage of a conductive AFM tip at a voltage such that no electric field exists in the air gap between the tip and sample surface. Qualitatively, this is accomplished by applying an AC voltage \( V_{AC} \) to the tip to monitor the force induced by any electric field between the tip and sample. A DC bias \( V_{DC} \) is adjusted
until no net electric field is present and therefore no force exists on the tip due to the contact potential difference between the tip and sample. In reality, it is more complicated. The electrostatic force $F$ on the tip due to an electric field $E$ neglecting any Coulombic force is:

$$F = -\frac{1}{2} \frac{dC}{dz} V(t)^2$$

(7.2)

where the voltage between the tip and sample is given as

$$V(t) = V_{DC} - V_{con} + V_{AC} \sin(2\pi f t)$$

(7.3)

and $C$ is the capacitance, $V_{con}$ is the difference in work functions between the tip and surface, $f$ is frequency of the applied AC voltage, and $z$ is the distance between the tip and sample.

Figure 7.3b shows the band diagram of an AlGaN/GaN sample with a voltage applied to the SKPM tip such that $V_{DC} = V_{con}$. The band diagram was created using a charge distribution represented in Fig. 7.3a. The time dependence of the force becomes

$$F(t) = -\frac{1}{2} \frac{dC}{dz} \left[ (V_{DC} - V_{con})^2 + 2(V_{DC} - V_{con}) V_{AC} \sin(2\pi f t) + V_{AC}^2 \sin^2(2\pi f t) \right]$$

(7.4)

Using the trig identity that $\sin^2(\Omega) = \frac{1 - \cos(2\Omega)}{2}$, Eq. (7.4) becomes

$$F(t) = \frac{dC}{dz} \left[ -\frac{(V_{DC} - V_{con})^2}{2} - \frac{V_{AC}^2}{4} - (V_{DC} - V_{con}) V_{AC} \sin(2\pi f t) + \frac{V_{AC}^2}{4} \cos(4\pi f t) \right]$$

(7.5)

The terms can be grouped according to their time dependence giving several DC terms, a sine term that has frequency $f$, and a cosine term that has frequency $2f$. Using a lock-in amplifier to measure the force magnitude at frequency $f$, the amplitude of the sine term can be monitored, and set equal to zero if $V_{DC} = V_{con}$, which give zero electric field in the tip-sample gap as previously mentioned. Koley et al. derived the equations for SKPM assuming a fixed charge at the AlGaN surface and found that an offset in $V_{DC}$ would exist but would reduce as $\sim 1/z$.[6] Therefore, tests of the feedback voltage $V_{DC}$ versus tip/sample separation
Figure 7.3: (a) Schematic charge distribution of an AlGaN layer grown on a thick GaN layer. \( \Delta P \) is the net polarization-induced charge, \( \sigma_s \) is the positive surface charge that compensates the 2DEG electron sheet concentration \( n_s \), and \( n_T \) is a trapped electron sheet density at the surface that is reflected in the 2DEG density. (b) Schematic band diagram of a typical AlGaN/GaN heterostructure used HEMTs with an SKPM tip. No electrostatic-induced force exists between the tip and sample because the SKPM tip has an applied voltage \( V_{con} \) such that no electric field exists between the tip and AlGaN surface. The AlGaN/GaN band structure, shown to scale complete with the vacuum energy, has a 30 nm 30% AlGaN on a thick relaxed uid \( (n \sim 5 \times 10^{16} \text{ cm}^{-3}) \) GaN. The metal and AlGaN work functions are designated \( q\Phi_m \) and \( q\Phi_{AlGaN} \), respectively. The electron affinities of the AlGaN \( q\chi_{AlGaN} \) and GaN \( q\chi_{GaN} \) are also shown. A surface trap is represented at the AlGaN/air interface that could be populated or depopulated by light or bias conditions. The green curve represents the electron distribution of the 2DEG with sheet charge density \( n_s \).
would yield a plateau where the Coulombic interaction with the surface was negligible and
SKPM $V_{DC}$ applied to the tip to minimize the force at frequency $f$ would correctly equal $V_{con}$.

This contact potential $V_{con}$ is a measure of the change in work functions between the AFM
tip and the semiconductor surface and can be affected by changes in traps in a depletion
region or insulator, lateral external electric fields, composition, local strain relaxation, or
other piezo-induced variation. Assuming the measurement can be performed to isolate the
variation solely due to trapping effects, the measured change in barrier height can be quanti-
tatively related to the change in trap occupancy and hence trap concentration assuming one
is able to fully modulate all of the traps of a given level. Given the charge distribution shown
in Fig. 7.3a, the charge distribution $\rho$ is

$$
\rho (t) = \begin{cases} 
    n_s - \Delta P - n_T (t), & 0 < x < 0^+ \\
    0, & 0^+ < x < d^- \\
    \Delta P + n_T (t) - n_s, & d^- < x < d 
\end{cases} 
$$

(7.6)

where $\Delta P$ is the total polarization-induced charge density, $n_s$ is the 2DEG sheet charge
density, and $n_T (t)$ is the trap sheet concentration as a function of time $t$. A constant electric
field $E$ can be calculated from the charge density knowing that $\nabla \cdot E = \rho / \varepsilon$ and yielding

$$
E (t) = \begin{cases} 
    0, & x < 0 \\
    \frac{1}{\varepsilon} [n_s - P - n_T (t)], & 0^+ < x < d^- \\
    0, & x > d 
\end{cases} 
$$

(7.7)

Finally, the potential $V_2$ can be calculated using $\nabla V_2 = -E$ to calculate

$$
V_2 (t) = \frac{q [\Delta P - n_s + n_T (t)]}{\varepsilon_{AlGaN}} d 
$$

(7.8)

The surface barrier height $\Phi_B$ is $V_2 + \Delta E_C - E_{Fi}$. If the change in potential across the AlGaN
layer $\Delta V_{2,0}$ is defined as the potential when the trap is fully occupied minus the potential
when the trap is fully empty, the total trap concentration can be calculated. Additionally,
$\Delta V_{2,0}$ is approximately the change in barrier height $\Delta \Phi_{B0}$ while the 2DEG is in place because
$E_{Fi}$ will change very little. At which point, the total trap sheet concentration $n_{T0}$ is

$$
n_{T0} \approx \frac{K \varepsilon_{AlGaN} \Delta \Phi_{B0}}{d} 
$$

(7.9)
where $K$ is 1 for the surface trap derived here or 2 for a trap concentration uniformly distributed throughout the AlGaN layer. If it is known that the level is uniformly distributed, the volume charge density $N_T$ can be calculated from the sheet charge density $n_T$ as $N_T = n_T / d$.

The voltage applied to the SKPM tip to neutralize any electric field between the tip and sample $V_{con}$ is often referred to as the relative surface potential (RSP). Assuming no compositional variation, the relation between barrier lowering and RSP is simply

$$RSP(t) = -\Delta \Phi_B(t)$$

(7.10)

If measurements are performed in a single spot, there can be no variation in composition or polarization properties, so quantitative trap information can be measured with fewer artifacts. However if scans across the surface are performed, variations in RSP can arise from the lateral nonuniformities in the barrier height. Nevertheless, comparing two SKPM maps of the same area, it is possible to extract quantitative information on defect densities because both maps will have the same variations in everything but trap concentration and most artifacts should cancel out.

Koley and Spencer et al. also used above bandgap UV laser illumination to explore the effects of trapping and detrapping processes.[7] They found that illumination caused nonuniform variations in relative surface potential, changes in the drain current in an AlGaN/GaN heterostructure, and that these changes could persist for minutes to hours.[7] Silicon nitride passivation was also studied and shown to reduce the change in drain current and surface barrier height,[7] but this method did not allow for specific energy levels to be measured and the use of above bandgap light generated large numbers of electron-hole pairs that could result in unexpected effects from the free carriers or large surface photovoltage that is generated. This promising method may be adapted to a DLOS-type measurement on the nanometer-scale using a monochromatic light source instead of a fixed-wavelength UV laser.

253
The equipment setup is quite similar to the normal DLOS setups where an optics table has a 1000 W Xe lamp, monochromator, lenses, shutters, a fiber optic cable to bring light in to the acoustic enclosure to shine the monochromatic light on the sample surface in the AFM. The AFM head and cantilever block a significant solid angle of access to the sample surface right by the AFM tip, so the light must be brought in at an oblique angle.

In Chapter 4, the equations of carrier emission were shown. Previously, this emission was measured via a change in capacitance, drain current, or transconductance, but now the RSP is the measured parameter. Because the change in barrier height is a physical parameter, $\Phi_B$ is considered and can later be related to RSP using Eq. (7.10). The barrier height can be written as

$$\Phi_B(t) = \Phi_0 + \frac{q n_T(t) d}{\epsilon_{AlGaN}}$$

(7.11)

where the barrier height without any traps $\Phi_0$ is

$$\Phi_0 = \frac{q (\Delta P - n_s) d}{\epsilon_{AlGaN}} + \Delta E_C - E_{Fi}$$

(7.12)

Using the time relations for trap emission from Chapter 4, the barrier height with thermal or optical emission becomes

$$\Phi_B(t) = \Phi_0 + \frac{q d}{\epsilon_{AlGaN}} n_{T0} \exp(-e_n t)$$

(7.13)

where

$$n_T(t) = n_{T0} \exp(-e_n t)$$

(7.14)

and $e_n$ is the emission rate for an electron trap whether optical or thermal emission. Additionally, this assumes an electron trap is fully occupied at $t = 0$ and fully empty at $t = \infty$. This direct correlation between $\Phi_B$ and $n_T$ means that DLTS is possible but it is also easy to show that DLOS-type measurements are also possible given the time derivative of $\Phi_B$

$$\left. \frac{d\Phi_B}{dt} \right|_{t=0} = -\frac{q d}{\epsilon_{AlGaN}} \left. \frac{dn_T}{dt} \right|_{t=0}$$

(7.15)
Combining with Eq. (4.42) yields

\[
\sigma_n^0(h\nu) = \frac{\varepsilon_{AlGaN}}{q d\Phi(h\nu) n_{T0}} \frac{d\Phi_B}{dt} \bigg|_{t=0}
\]  

(7.16)

This relation means that the Lucovsky and Chantre models of optical cross-section presented in Chapter 4 can be fit to \(d\Phi_B/dt\) and accurately identify defect energy levels. The equivalent of steady-state photocapacitance (SSPC) is also possible. Using Eq. (7.9), the steady-state photopotential (SSPP) can be used in the same manner as SSPC to determine the defect concentrations.

It has been shown that DLOS looking at \(d\Phi_B/dt\), SSPP, and DLTS using \(\Delta\Phi_B\) are possible. These measurements do not have the concept of scanning, common in AFM, so adaptations are needed in some cases to take full advantage of all the possible variables. First, the position is kept constant allowing the contact potential to be recorded as a function of time analogous to standard DLOS. Measurements on the gate-drain access region of an AlGaN/GaN HEMT were made after first measuring topography shown in Fig. 7.4. The AFM reveals a trench between the gate contact on the bottom and the drain contact at the top, which is the gate-drain access region. The AFM revealed some non-uniformities in the height along the gate-drain access region, but because the sample was repeatedly handled, used throughout the setup of the AFM equipment, and potentially exposed to surface contaminants the source of these variations could be intrinsic to the growth and processing or have resulted from the later handling. Therefore, few conclusions are being drawn about these anomalies in topography or potential in these locations until a more controlled sample can be analyzed for similar “bumps”. Placing the AFM tip at the black cross in Fig. 7.4, the transistor was first biased into pinch-off \((V_{GS} = -4 \text{ V})\) with a reasonably large drain-source voltage of 10 V to leak electrons from the gate into the gate-drain access region forming the virtual gate. The biases were then turned off and most of the thermal transient allowed to subside. Finally, the shutter was
opened to allow monochromatic light to hit the sample while the RSP was recorded. Doing this in the dark and at several wavelengths should allow both determination of trap concentrations from the SSPP (change in barrier height vs. incident photon energy) and trap energies from DLOS ($d\Phi_B / dt$ at $t = 0$ vs. incident photon energy). This is shown schematically in Fig. 7.5 where the bias time was $\sim 45$ s, the time between the bias change and shutter opening was $\sim 15$ s. Because a program has not yet been written to acquire the data automatically, the transients were only gathered at a few wavelengths at this point and are shown in Fig. 7.6. The AlGaN bandgap was $\sim 4.0$ eV and the GaN 3.38 eV. The dark and 2.70 eV light were both less than the bandgaps of the GaN and AlGaN. No indication of surface photovoltage was observed as expected since the below bandgap illumination would not generate electron-hole pairs (EHP). The observed transients have similar character with some thermally induced
Figure 7.5: Biasing scheme for measuring surface Schottky barrier lowering in the access region of nitride HEMTs. In pinch-off ($V_{GS} = -4$ V in this case) and large drain bias ($V_{DS} = 10$ V was the largest stress attempted so far), electrons leak from the gate filling states on the AlGaN surface. In addition, defect state below the Fermi level can also be filled in equilibrium. The bias was removed and the transient from levels able to emit carriers thermally was allowed to subside before opening a shutter to allow monochromatic light of various wavelengths to impinge upon the sample. The resulting optically-induced transient was recorded similar to DLOS experiments allowing for quantitative defect energy and concentration determination.
Figure 7.6: The measured change in barrier height at the black cross in Fig. 7.5. After a fill pulse in pinch-off at $V_{GS} = -4 \text{ V}$, $V_{DS} = 10 \text{ V}$, and a 15 sec. delay, the shutter was opened and the resulting transient recorded. The change in barrier height is measured relative to the dark, equilibrium relative surface potential (or contact potential). This was done for the dark and at 2.70, 3.40, and 4.06 eV. The first three transients show some existing thermal transient, but also the illuminated transients exhibit some difference in $\Delta \Phi_B$ presumably due to trap emission. The 4.06 eV light is above the AlGaN bandgap and produces a strong surface photovoltage not evident in the below AlGaN bandgap transients.
The measured change in barrier height in the dark with no delay after a fill pulse in pinch-off at $V_{GS} = -4$ V, $V_{DS} = 10$ V at the black cross in Fig. 7.5. The change in barrier height was determined from the measured relative surface potential subtracted from the contact potential, which was estimated from the dark, equilibrium relative surface potential.

emission from the bias change 15 s beforehand. In fact, the thermal transient for this sample was quite large as seen in Fig. 7.7. The 2.70 eV transient has slightly greater magnitude presumably because of additional emission of trapped electrons. The 3.40 eV light will excite EHPs in the GaN layer but with a photon flux of $\sim 1 \times 10^{16}$ photons/cm²/s, an assumed excess carrier lifetime of $\sim 1 \times 10^{-9}$ s, and assumed complete conversion of every photon to an EHP, only $10^7$ electrons/cm² would be maintained, which is much smaller than the typical 2DEG concentration ($5 \times 10^{12} - 2 \times 10^{13}$ cm⁻²). Furthermore, the valence band offset would also act to repel excited holes and any excited holes would have a high probability of annihilating in the 2DEG preventing them from entering the AlGaN to be captured. Finally, above AlGaN bandgap illumination at 4.06 eV was used to show the effect of surface photovoltage generation. Here, the initial slope of the transient was limited by the feedback speed of the external...
Figure 7.8: Schematic band diagram of a 30 nm, 30% AlGaN/GaN heterojunction. The top (red) band diagram is shown assuming the barrier height is pinned at 1.7 V in the dark and under equilibrium conditions. The lower (green) band diagram is shown after illumination with surface states being emptied by the exposure to light. It is assumed here that the light energy is sufficient to cause electrons trapped in surface states to be excited to the AlGaN conduction band where they are swept away by the electric field into the 2DEG. It is important to note that the electrons generally cannot be excited all the way up to the vacuum level where they could escape into the air as the electron affinity is in the range of 3-4 eV for low percentage AlGaN films and this energy is needed in addition to the trap energy $E_C - E_T$.

circuitry. After this, the transient is flat at approximately $-1.64$ V indicating a large number of trapped positive charge due to the creation of EHPs in the AlGaN layer. The below AlGaN bandgap illumination yielded offsets in the approximate steady-state barrier lowering. The initial model for this difference is shown in Fig. 7.8 where photons with energy greater than $h\nu > E_C - E_T$ can cause the surface traps to emit an electron to the conduction band where it can be swept away by the electric field in the AlGaN and reside in the 2DEG. This emission of surface traps will lower the barrier height shown by the $q\Delta\Phi_B$ in Fig. 7.8, which corresponds directly to the number of carriers that are emitted according to Eq. (7.9). An estimate of the trap concentration of defects with energy between $\sim E_C - 1.0$ and $E_C - 2.7$ eV can be determined where the lower energy limit would be because low energy traps ($E_C - E_T < \sim 1.0$ eV) would thermalize quickly before the measurement begins (i.e. waiting for the thermal transient to die off) and the upper limit is determined by the incident light energy, which

260
Figure 7.9: Typical (a) surface barrier and (b) current transients measured for an AlGaN (35%, 202 Å)/GaN heterostructure sample before, during, and after illumination. The laser first induces a large positive charge density on the surface that increases the 2DEG concentration and leads to excess current. After the laser is turned off, the current slowly decreases as the bare surface barrier height increases. A shift in the barrier height and drain current remains after 400 s after illumination suggesting that some mechanism is in place to prevent electrons from being recaptured or holes from emitting from these defect locations. From Ref. 7.

is 2.70 eV in this case. Calculating the trap concentration from the end of the 2.70 eV light and the dark transient yields an approximate defect concentration of \( \sim 5 \times 10^{11} \text{ cm}^{-2} \). The same thing can be done between 2.70 eV and 3.40 eV transients yielding \( \sim 4 \times 10^{11} \text{ cm}^{-2} \) for defects with energy between \( E_C - 2.70 \) and \( E_C - 3.40 \) eV. Finally, looking at the magnitude of the dark transient yields defect concentrations for defects that can thermalize on this time scale. In this case, \( \sim 2.5 \times 10^{12} \text{ cm}^{-2} \) defects were detected with energy \( E_C - E_T < \sim 1.0 \) eV.

Although previous nano-scale measurements have observed variations in trap occupancy and even the sign of the charge, the quantitative energy resolution has not really existed.[7] Figure 7.9 shows the relative barrier height and drain current before, during, and after UV laser illumination. The laser illumination causes electrons to be emitted from surface traps evident from the large decrease in surface barrier height, which leads to an increased 2DEG density and higher drain current. The UV light will induce EHPs that would enable higher
current densities as well, but because the photon flux is not given it is not possible to estimate the current from this effect. When the light is turned off, the some levels are quickly filled evident by the increase in barrier height. After 400 s, the barrier height is still far from the initial value suggests that some surface defects cannot easily refill or thermally emit a hole to regain the equilibrium occupancy. Unfortunately without a tunable light source, it is not possible to individually identify deep levels contributing to the drain current and surface barrier height variations. This is one easy way for the new AFM-DLOS system to excel because of its tunable, monochromatic light source with resolution of $\sim 0.02$ eV.

The next step for this measurement is to automate the transient acquisition process so that plots of $\Delta \Phi_B$ vs. incident photon energy and $d\Phi_B/dt$ vs. incident photon energy are possible allowing for trap concentration determination and fitting to find trap energy levels, respectively. This spot check can be performed in “clean” areas and areas of possible or observed degradation along the access region to systematically look for defects that may be responsible for the degradation processes. However, a system is needed to survey the access region to identify new areas of degradation and attempt to determine future areas of degradation by hopefully identifying some tell (regions of naturally higher or lower surface potential, for example).

Several methods should be employed to search for latent defects that may not become an issue for device performance until some amount of RF or high-field DC stressing. Areas of high temperature in the access regions have been found using scanning thermal microscopy (SThm), cathodolumenscence, electroluminescence, and micro-Raman spectroscopy.[1, 8–10] The Park Systems XE-NSOM AFM system can be upgraded to perform SThM, Raman spectroscopy, and possibly electroluminescence, but initial SKPM mapping measurements suggest that barrier height maps may potentially be a feasible method to identify area of
future failure, too. This is very theoretical at this point, but hopefully SThM with the Park AFM or CL measurements with Prof. Brillson in the near future will be able to correlate the SKPM maps to others’ observed lateral inhomogeneity. Lin et al. has already shown some correlation between CL and SKPM, so this method is promising.[9]

As part of the investigation of SKPM to explore defects in HEMTs, surface maps were taken under steady-state illumination at various wavelengths. Shown in Fig. 7.10, the AFM topography is shown with corresponding SKPM maps in the dark and at 2.70, 3.40, and 3.80 eV illumination. The SKPM images map the contact potential variation across the sample. The images were taken after biasing the sample into pinch-off, waiting for thermal equilibrium to occur, and then starting the imaging process. In this case, the white regions indicate higher relative surface potential or low barrier heights assuming the other effects are negligible while the dark brown regions would be areas of higher barrier height. The images in Fig. 7.10 show an evolution from dark to near bandgap photon energy illumination. The variation is position and wavelength dependent – potentially indicative of laterally inhomogeneous distributions of defects in the gate-drain access region. As the measurement infrastructure is improved, these measurements will be correlated with the drain-controlled CI_D−DLOS experiment discussed in Chapter 6. It is important to note however that individual SKPM images cannot be used solely to determine variation in defect concentrations as a given images is a conglomeration of all defects with \( E_C - E_T < h\nu \) or \( E_T + E_V < h\nu \) emitting in addition to the lateral variation due to polarization or composition variation. For this reason, the difference between images under given illumination might more quantitatively portray the defect concentration and lateral variation of concentration of a specific defect selected by the wavelengths of incident light for the two SKPM images being compared. While it is important to note that the sample has been used throughout the AFM setup process and
Figure 7.10: AFM topography (top) and SKPM images (below) of the same gate-drain access region. The gate contact is on the bottom and the drain contact is on the top of the images. The SKPM images were taken in the dark and under steady-state illumination at 2.70, 3.40, and 3.80 eV incident photon energy as indicated in the figure. The higher surface potential at the same spatial location for different illumination indicates a reduced barrier height or more net positive charge at the surface. At different spatial locations, the variation can be due to variations in barrier height or other effects that would cause the contact potential to vary.
required soldering to the package so some or all of these variations could be caused by previous device bias stressing, surface contamination, or other unknown factor, it seems compelling that the majority of the higher surface potential anomalies in the gate-drain access region (light colored regions) seem to originate from the gate contact. This condition might be expected under DC stressing conditions where the highest fields form at the gate edge on the drain side and this could lead to nonuniform leakage of electrons into the gate-drain access region from the gate contact. These regions seem to spread under higher energy light, which might be expected as surface defects could reach the higher conducting anomalous surface potential regions.

It is also possible to explore line scans of the SKPM images, shown in Fig. 7.11, to obtain some quantitative results. The horizontal scans (dashed lines in Fig. 7.11) show significant variation along the access region, which again could be caused by trap density/occupancy variation, variations in electron affinity from compositional variation, or sheet charge variation from strain nonuniformities. Generally, the curves from the dark and 3.80 eV steady-state illumination appear similar although separated in RSP. There are two notable exceptions at 12.5 µm and ~25 µm. At 12.5 µm, the peak in the dark curve appear to be mostly pinned as the 3.80 eV exhibits a dip at the same point while the surrounding area and increased more than 700 mV. At ~25 µm, a peak is evident in the dark scan but nearly indistinguishable in the 3.80 eV scan. These anomalies are intriguing if not completely understood at this point. If the point at 12.5 µm is highly conducting due to surface contamination or some degradation process, it may communicate well with the gate or drain contact helping it maintain a more constant surface potential. Looking at the vertical scans, the obvious increase in peak height under illumination is expected as it corresponds to an increase in positive surface charge, which is consistent with other measurements. However, the sloped sides of the potential
Figure 7.11: Line scans of the dark and 3.80 eV SKPM images in Fig. 7.10 were taken to examine the spatial variation of the potential under the two extreme cases (dark and near bandgap illumination). The higher potential (green) lines are from the 3.80 eV image and the dashed lines correspond to the horizontal line scan while the solid lines correspond to the vertical line scans.
may be real or due simply to the resolution limit for this structure where the metal contact protrude from the sample surface. Further testing is needed to find the resolution limits with the raised metal contacts.

Interestingly, some of these anomalous regions in Figs. 7.10 and 7.11 correspond to small bumps in the AFM topography. This could just imply the surface is contaminated or that some feature during growth is contributing to inhomogeneity. It is unlikely that these features were formed during device operation as it is possible to remove material near the gate edge on the drain side after high field stressing (possibly due to vacancy migration) but this is additional material and the scale is much larger than the removal process.[11] Many questions remain to the importance, repeatability, and evolution of these inhomogeneities observed. The first step is to repeat these measurements on another sample to determine if the problem was caused post-processing or is intrinsic to these kinds of devices. Then, high DC field stressing will be used to see if these features appear or grow in size trying to repeat some of the work of Meneghesso et al. performed with electroluminescence measurements.[1] Hopefully, an initial indication of degradation points will be visible at early stages of device stressing that will allow DLOS-type measurements at the various stages of defect development. An initial suspect in defect formation is actually gallium vacancy migration as high DC stress have suggested an accumulation of vacancies at the gate edge that migrate due to a relatively low migration barrier and large numbers of hot electrons and hot phonons.[12]

7.4 Summary

The AFM equipment is nearly completely set up, but a large amount of experiment design, programming, equipment optimization, and equipment design and fabrication is still needed before the full potential of this equipment will be realized. Designing the NSOM and SCM
aspects of the AFM-based measurements is still underway with the NSOM-based measure-
ments possibly hindered by restrictions in the optical column of the Park XE-NSOM AFM
while the SCM measurements will be limited by the minimum noise that can be achieved.
Several improvements are suggested such as using a homebuilt capacitance bridge and per-
forming constant capacitance measurements to improve the measurement signal-to-noise
ratio.

With the current system setup, some initial SKPM measurements were performed pri-
marily to evaluate the AFM/SKPM with monochromatic, tunable light source system but also
to get some initial results from an AlGaN/GaN HEMT. The transient measurements at a single
spot revealed charge transport consistent with the observed virtual gate phenomenon. In
pinch-off, negative charge leaks into the access region leading to increased surface barrier
height. When the bias is removed, the negative charge dissipates causing the surface barrier
height to decrease. It was shown that light with energy above the AlGaN bandgap led to a
large surface photovoltage, but this was avoided using below bandgap illumination. The
below bandgap illumination is more useful because it will allow for systematic probing of
the bandgap to identify deep levels in the AlGaN, at the AlGaN surface, or in the passivation
(in the case of passivation). The initial dark and below AlGaN bandgap illumination tran-
sients revealed a total trap sheet concentration of $\sim 3.4 \times 10^{12}$ cm$^{-2}$ for defects above $E_C - 3.40$
eV in the bandgap. These numbers are consistent with defect concentrations determined
by drain-controlled constant-drain-current DLOS measurements and the high frequency
photo-assisted C-V measurements by Swenson et al. and previous SKPM measurements
lending credence to both results.[7, 13] The next step in measurements in this mode are to
automate the transient measurements to acquire them over a series of wavelengths and save
each transient. This will allow cross-section analysis, which was shown to be valid but not
performed here due to the lack of complete data, and concentration calculation for individual traps. The trap concentrations were calculated but not for individual traps again due to the lack of complete data for the whole range of wavelengths.

Beyond localized, single point measurement of defect spectra, lighted SKPM mapping has been employed to image the surface potential in the access region. This technique offers the possibility of measuring the defect concentration as a function of distance in a steady-state illumination by comparing images at various wavelengths of incident light. By looking at differences, any contact potential differences not due to variations in defect occupancy would cancel out. The disadvantage of this mode of measurement is the requirement for time-invariance. Unfortunately the measurement takes several minutes to acquire, so anything varying with time would be convoluted with the imaging process.

Now that some initial experiments have been performed, a wide range of possible experiments are possible or will soon be possible. With these new experiments, it is expected that localized, quantitative defect concentrations and energies will be determined. These experiments will allow for studies of DC and RF bias stressing to watch the evolution of defect spectra at various parts of the access regions. Additionally, measurements have been performed while the device has been off, but measurements during pinch-off may yield additional information such as gate extension length. Adding a further realm of possibilities, these techniques can be applied to thick GaN or AlGaN layers in addition to other material systems to allow localized defect characterization that may relate specific traps with dislocations or other structural defects. This measurement system is in no way limited to HEMTs or the nitride material system. Overall, the current results portend great promise to these localized trap spectroscopy techniques.
7.5 References


Chapter 8

Conclusions and future work

Nitride electronics and specifically HEMTs are maturing quite rapidly, in part, because of intensive global research efforts. The main contributions of this work are expanding the understanding of electrically active defects in thick GaN and AlGaN films, the formulation of new constant current-based techniques to quantitatively determine defect concentrations and energies in operating HEMTs, development of new techniques to quantify defects on a nanometer scale, and provide the first combined defect energy spectrum and concentration measurements of AlGaN/GaN HEMTs. In total, seven new techniques have been developed to apply variations of the standard defect transient spectroscopy techniques to three-terminal HEMTs. The added complexity of two voltages to control, complex device structures (vertical), non-uniform horizontal structures (gate, drain, source, and access regions), large vertical and lateral electric field-related phenomena, and resulting overabundance of possible mechanisms and defect sites make interpretations difficult without a base knowledge of at least the bulk-related defects in the component layers (GaN and AlGaN, in this case). For this reason, the first part of this research focused on studying the role of growth conditions and growth modes in defect incorporation in thick $n$-type GaN and AlGaN layers. Studies of MOCVD vs. PAMBE, N-face vs. Ga-face, NH$_3$-MBE vs. PAMBE, the role of ammonia flow rate and nitrogen plasma power their respective growths, the impact of dislocations, and a first look at
the defect spectrum of thick $n$-type AlGaN have provided a great fundamental knowledge that builds upon years of additional research by our groups and several others.[1–7] With this understanding, the knowledge could be applied to the analysis of the defects studies in HEMTs to begin to separate bulk effects from surface and interface effects not observable in the bulk capacitance defect spectroscopies. The following sections summarize the findings of each of these research thrusts.

### 8.1 Continuing studies of component materials in nitride HEMTs

It is possible to improve HEMT performance and reliability simply by optimizing the material quality of component layers. Although this may not expose all of the degradation mechanisms, it becomes a relatively simple way to make large improvements without the added complexity related to measurements directly on HEMTs. Nitrides can be grown by many different methods starting with high level differences such as MOCVD versus MBE but even within MBE different nitrogen sources can be used and other variables include dislocation density and crystal polarity. Each of these variables requires separate optimization and these measurements are also performed simply to baseline levels of their respective defects in each type of growth.

Summarizing the different effects of growth conditions first, typical MOCVD and PAMBE grown $n$-GaN films were first compared and revealed that several common levels exist, but MOCVD material routinely has higher concentrations of the gallium vacancy-related level ($E_C-2.60 \text{ eV}$) and high residual concentrations of Mg ($E_C-3.22 \text{ eV}$) while the largest near valence band level ($E_C-3.28 \text{ eV}$) in PAMBE material is due to the substitutional carbon on nitrogen acceptor.[8–10] It was also shown that MOCVD films tend to have higher concentrations of mid-gap and near valence band states and lower concentrations of near conduction
band states ($E_C-0.25$, $E_C-0.60$, and $E_C-0.90$). In MBE growth, several growth variables have been studied including the nitrogen plasma power keeping the V/III ratio constant in PAMBE and increasing the V/III ratio by adjusting the ammonia flow rate keeping the Ga flux constant in NH$_3$-MBE. The increasing nitrogen plasma power results in significantly higher concentrations of electrically active defects, in general, following an Arrhenius-style relation suggestive of damage from the high energy nitrogen species during growth. In contrast, the increase in ammonia flow rate and V/III flux ratio led to improved surface morphology and average $\sim$2X defect reduction except for the suspected nitrogen vacancy-related level at $E_C-0.25$ eV that exhibited a sixfold reduction consistent with the increased nitrogen flux during growth.

The differences between crystal polarity and the role of threading dislocations was also studied in “bulk” $n$-GaN samples. Although the Ga-face and N-face samples were both grown under Ga-rich conditions with nearly identical substrate temperature, the defect incorporation was sensitive to the reversal of crystal polarity and resultant change in bonding and growth surface interactions. The carbon substitutional and gallium vacancy-related concentrations stayed approximately constant, but the near conduction band states were quite varied. Not only did the N-face sample exhibit new traps not seen before in Ga-face samples grown by MBE and MOCVD, but also the suspected N vacancy-related level at $E_C-0.25$ eV increased the most with nearly an order of magnitude in the N-face sample. Despite the large increase in the $E_C-0.25$ eV level, the total trap concentration remained approximately constant due to the absence of the possible carbon split interstitial ($E_C-1.28$ eV) level in the N-face sample.[2, 11] N-face growth is in its infancy compared with Ga-face growth and defects concentrations will likely decrease in N-face samples as the material is further optimized.
The role of threading dislocations (TD) was also studied in MOCVD-grown $n$-GaN, and several very important conclusions can be drawn from the results. The $E_C - 0.90$ eV level thought previously to be dislocation related\cite{12,13} showed no correlation with TD density strongly suggesting that defect may form in clusters or other relationship but not related to dislocations. The ubiquitous $E_C - 0.60$ eV and suspected carbon interstitial $E_C - 1.35$ eV levels tracked the TD density closely suggesting that these levels are point defect-dislocation complexes and potentially contribute to the enhanced leakage current in higher TDD samples. The $E_C - 2.64$ eV level showed a weaker dependence on TD density, which could be due to a lack of trap saturation being over $E_g / 2$ or a result of Ga vacancy complexes being distributed both in the nondislocated field as well as around dislocation cores. In either case, the $V_{Ga}$-related defects show a clear correlation with TD density. Finally, the Mg acceptor at $E_C - 3.22$ does not show direct dependence TDD.

Although much of the work so far has focused on the GaN buffer due in part to the difficulty of growing crack-free, thick AlGaN layers, work as begun to examine AlGaN as a component material in HEMTs using thick AlGaN films to isolate the defect spectrum exclusively to AlGaN. Similar to MOCVD $n$-GaN, MOCVD $n$-AlGaN exhibits large near valence band defect concentrations and low near conduction band defect concentrations. The $E_C - 3.10$ eV level identification as a cation vacancy-related level is consistent with the vacuum referred binding energy model common for deep, non-hydrogenic defects where a defect energy level is constant relative to the vacuum energy as the aluminum mole fraction is varied. A large concentration of the $E_C - 3.93$, or $E_V + 0.15$, eV level is probably the result of substitutional carbon or magnesium, but SIMS results revealed high concentrations of both. Therefore, carbon, magnesium, or a combination of both may be responsible for this level. The other observed levels at $E_C - 0.9$ and $E_C - 1.5$ eV were observed with concentrations of
$\sim 1 \times 10^{13}$ and $2.4 \times 10^{16}$ cm$^{-3}$, respectively, where the $E_C - 0.9$ eV level common energy with a DLTS level at $E_C - 0.90$ suggests that these levels may be the same defect and because little energy difference exists between the DLOS and DLTS energies near zero Franck-Condon energy or lattice relaxation would exist. These deep levels do not correlate with known GaN levels, so their physical source remains under investigation.

Reducing the HEMT structures down to their component layers, defects, that are all potentially sources of degradation of electrical performance, have successfully been identified in GaN and AlGaN films. Not only can the individual layers be optimized to reduce defect concentrations, which is applicable to all nitride electronics, but the defect signatures can also be used later in studies on operational HEMTs to identify the layer containing the defect and the source responsible for loss of performance.

### 8.2 New frontiers in defect characterization of HEMTs

In Chapter 6, the concept that constant drain current for measurement of defects in HEMTs was introduced. Initially developed as a method where the gate voltages was regulated to maintain a fixed drain current (gate-controlled $C_{ID}$-DLOS/DLTS and $C_{GD}$-DLOS/DLTS), it was found the quantitative trap energies and sheet concentrations could be determined from defects located under the gate contact if the measurement was performed in saturation. Not only can defect concentrations be determined but the threshold voltage shift caused by change in occupation of individual defects can also be measured. Both thermal and optical excitation are possible, which allows for DLTS and DLOS-type measurements and hence defect identification throughout the GaN and AlGaN bandgaps. Initial measurements of an unpassivated AlGaN/GaN HEMT revealed defects at $E_C - 0.59$, $E_C - 2.0$, $E_C - 3.3$, and $E_C - 3.7$ eV via gate-controlled $C_{ID}$-DLTS/DLOS. These levels caused a 90 mV variation in
the threshold voltage with time constants ranging down to ~1 ms for the shallowest level at room temperature. The $E_C - 0.59$ eV level matches favorably with the well-known $E_C - 0.60$ eV in GaN suggesting that this level may be related to the GaN buffer layer. The $E_C - 3.3$ eV, or $E_V + 0.1$, level is also quite similar to the carbon substitutional acceptor level in $n$-GaN again suggesting a role of the buffer in device performance. The $E_C - 3.7$ eV level is above the bandgap of GaN and can only be located in the AlGaN layer, but is shifted from the $E_C - 3.93$ eV level observed in MOCVD-grown $n$-AlGaN. If the same holds true for GaN, the $E_C - 3.93$ eV level may be Mg related as this is a common contaminant in MOCVD growth while the $E_C - 3.7$ eV level may be carbon related, but additional work is needed to confirm or refute this observation. In total, the trap concentration of defects located under the gate was found to be $\sim 3 \times 10^{11}$ cm$^{-3}$, which is quite reasonable.

Gate-controlled measurements were also performed in the triode regime where much larger variations in gate voltage were observed and found to be due to access region related defects, but attempts to quantify the levels using this method were fraught with difficulties and less-than-ideal assumptions. This led to the concept of the drain voltage controlled method where now the gate voltage is fixed and the drain voltage is varied to keep the drain current constant. It was shown that near or above $V_{GS} = 0$, the highest insensitivity to defects under the gate is found and under these conditions the gate-drain access resistance variation could be calculated. This device metric could prove valuable to device modelers and device process/growth researchers because quantitative trap specific impact on gate-drain access resistance can be measured along with bias dependence and temperature-dependent time constants in some cases. So far, a simple model has been developed assuming the variation in gate-drain access resistance is entirely due to the gate extension phenomenon in AlGaN/GaN HEMTs. Using a simple analytic equation to estimate the gate extension, the trap
concentration can be estimated. In drain-control, defects at $E_C - 0.49$, $E_C - 1.3$, $E_C - 1.8$, and $E_C - 3.7$ eV were present in measurements of the same unpassivated AlGaN/GaN HEMT. The $E_C - 3.7$ eV AlGaN-related level and $E_C - 1.3$ eV levels were again present, but the other “under gate” levels have been replaced by new access region related defects that appear to be different than typical GaN or AlGaN bulk defects based on initial measurements. This suggests that although the bulk measurements are useful, other factors impact the access region probably related to the AlGaN surface based on recent passivation studies. Early passivation study results show that passivation strongly impacts the total change in resistance measured and trap concentration calculated. The total change in resistance was $\sim 4.5 \, \Omega$, which corresponded to a total gate extension-related trap concentration of $\sim 8 \times 10^{12} \, \text{cm}^{-2}$. This number has some variability at this point as modeling is needed to confirm or improve the estimation of gate extension and the AFM-DLOS measurements will confirm if all of these levels are virtual gate related or if some of the levels are spread throughout the access region.

8.3 Providing quantitative defect characterization of semiconductor devices with nano-scale resolution

Although great strides have been made to extend defect transient spectroscopies to HEMTs, the can never achieve nano-scale resolution along the width of a transistor in the access regions. To resolve this problem and provide a nano-scale suite of experiments that can be applied to nitride HEMTs as well as other semiconductor material systems, several AFM-based methods have been explored or are in process of being set up to measure quantitative variation in defect concentrations with high spatial resolution. NSOM, SCM, and SKPM were proposed as potentially good candidates to achieve this goal. While the SCM apparatus is being scaled down to find the minimum resolution, the SKPM equipment has been assembled to the point where nano-scale measurements can be performed. Analysis of AlGaN/GaN band
diagrams and charge distributions reveal that quantitative defect concentrations and energy can be determined by measuring the change in the bare surface barrier height. Using light or temperature similar to DLOS and DLTS, the occupation of surface and AlGaN defect levels can be controlled and using subbandgap illumination the generation of electron hole pairs and subsequent surface voltage effect can be avoided according to initial measurements. SKPM, designed to measure the contact potential difference between a tip and oxide-free sample, can be applied to nitride heterojunctions where the AlGaN layer appears as an insulator if the tip is backed off sufficiently to avoid Coulombic forces between surface charges and the tip. The contact potential difference depends on the work function difference between the tip and sample surface and can be affected by variations in composition, trap occupancy, and polarization-induced charge variation. SKPM images can contain artifacts from these non-trap related phenomena, but by comparing images before and after illumination it should be possible to eliminate the artifacts. The SKPM can also be performed at single locations allowing for the time dependence of the trap emission to be recorded at the various wavelengths. In this case, the variation in contact potential should again directly relate to the variation in barrier height. Analyzing the initial time in the transients and the steady-state variation in barrier height at each wavelength, DLOS and SSPC/style measurements can be performed allowing for defect energy and concentration determination. Total defect concentrations in the gate-access region for the tests so far are $\sim 3.5 \times 10^{12}$ cm$^{-2}$, consistent with the initial estimates from the drain-control CI$_D$-DLOS/DLTS measurements on the unpassivated HEMT.
8.4 Future work

Great strides have been made to understand defect incorporation in AlGaN/GaN HEMTs through the use of traditional capacitance DLOS/DLTS measurements and the development of several new methods that have allowed for defect identification on operational HEMTs and do so with nano-scale resolution in the case of the AFM-based measurements. This has opened up a whole new realm of possibilities not only for nitrides HEMT but also for many wide bandgap material systems including non-HEMT devices. But not to be forgotten, the research of new growth methods and material optimization will continue to be performed using conventional diode structures where analysis is more straightforward, not having to worry as much about interface, surface, or electric field effects, and sample sets can be carefully designed to minimize the differences to a single variable. With this in mind, several practicable extensions of the existing work are identified.

In the realm of bulk measurements, the research has focused almost entirely on GaN. Although thick AlGaN layers have been touched upon, the whole range of experiments performed on GaN are available with the added ability to adjust the aluminum mole fraction as well. The obvious study is a series of samples intentionally co-doped with carbon or magnesium to examine the $E_C - 3.93$ and $E_C - 3.7$ eV levels previously identified to unambiguously determine their sources. Although the defects responsible for the trap levels in $n$-AlGaN have been proposed, a study of the Al mole fraction would allow verification of the vacuum referred binding energy model and cation vacancy designation. However, this is not to say that all research has been completed in GaN as the newly identified N-rich growth regime where high-quality GaN films can be attained is another area of study along with N-rich, N-face growth as well.[14, 15] Because N-face material is already showing great promise and because a regime common to typical V/III MBE growth where the Group V-rich flux ratios (N-rich) are
prevalent has been found to provide low surface roughness and high electron mobilities, this may quickly become a competitive growth mode. In recent years the focus of this research has been for nitride HEMT applications, but in parallel the optoelectronics camp is developing nitrides with slightly different goals. In their case, Nakamura et al. have led the development of \(a\)-plane and \(m\)-plane non-polar nitride growth for optoelectronic applications.[16, 17]

This information on “bulk” material is not only of value to understanding the role of defects in nitride HEMTs but also on its own by developing fundamental understanding of defects in nitrides, which can be applied to other electronic applications such as LEDs, laser diodes, and high power rectifiers, for example.[18–23]

The first quantitative energy and concentration measurements of defects in the gate-drain access region and under the gate have been demonstrated. Although the results from the new techniques agree with commonly reported total trap concentrations and the techniques are being successfully applied to studies of Si\(N_x\) passivation, more work is needed to identify any potential artifacts of the measurements, which may be specific to issues related to nitride HEMTs such as electric field dependence, passivation bandgap lower than the GaN or AlGaN bandgap issues, and refilling of levels in depleted films such as the passivation and possibly AlGaN. In theory, the thin passivation layers could absorb a fraction of the light if their bandgap is in the range of the experiment, but for thin layers the absorption would still allow light to penetrate the other layers of the transistor. The larger issue is the effect the absorption may have on the transient slope or DLOS analysis and, of course, the impact of the electron-hole-pair generation in the passivation. The other issues can all be described by bias dependent effects. As defects are not accounted for in typical transistor equations, it is not well-known how the bias or electric field will effect the occupation (filling or emptying) including possible Poole-Frenkel effects. Although additional modeling may be helpful to
understand the complex band diagram in HEMTs that cannot be treated in one dimension is the case for Schottky diodes, the advancement in understanding will inexorably come from continued measurements and analysis. With so few measurements being performed so far, a multitude of possible experiments are possible. Of these, the most exciting study will be examining the effects of device aging/stressing, whether RF or DC, on defect evolution. As discussed in Chapter 2, numerous degradation and failure mechanisms exist where most if not all of these are defect-related phenomena.

Although threading dislocations are suspected to play a role in both device performance and failure,[24, 25] the relationship, if any, has not been clearly identified. Using samples ranging from bulk GaN substrates to direct GaN on SiC, the dislocation density could be varied by orders of magnitude, potentially allowing for direct correlation of defects to device non-idealities. These devices could also be stressed to once again evaluate the aging process. Recently suggested, del Alamo et al. have identified a critical voltage required for permanent device degradation to occur including increased source and drain resistance, increased gate leakage, and reduced drain current and is thought to occur via defect formation at the gate edge.[26] Given that del Alamo et al. suspect a defect hopping model, dislocations may be a likely suspect with defect migration or formation occurring or gettered by the dislocations. Using these new techniques, the impact of such defects could be quantitatively studied in terms of increased access resistance, threshold voltage shifts, and defect concentrations with 0.02 eV resolution of the defect energies as well.

Passivation is known to reduce the virtual gating effects and other forms of dispersion, so this provides a great means of identifying particularly troublesome defects that impact gate lag, knee walkout, and current collapse, for example.[27] Various forms of passivation are being tried including silicon nitride, aluminum silicon nitride, hafnium oxide, alumina,
and magnesium oxide to name a few. Drain-controlled \( \text{Cl}_{\text{D}} \)-DLOS/DLTS, which has been shown to be gate-drain access region sensitive, can be performed to study the quality and stability of the passivation layers. The defect spectroscopy measurements should be done in tandem with pulsed I-V measurements at various temperatures, which has not been well exploited using the in-house DIVA pulsed I-V system at this point, to correlate the effects of defects and dispersion phenomena. It is also likely that the various passivations will depend on deposition mode and conditions as well as the growth method of the HEMT layers as well given that MOCVD, MBE, and such have different defect spectra. This would provide a wealth of studies, so these studies may need to be limited to the highest performance passivations to allow for adequate studies.

As previously pointed out, a simple analytic equation is being used to estimate the lateral depletion in the gate-drain access region due to the applied voltages in pinch-off. The real lateral depletion will depend on the electrically active defects as they have comparable concentration to the polarization-induced surface charge and hence will depend on the passivation as well. Experimental evidence of the lateral depletion will be the best source of information for revising the analytic model or determining the lateral depletion width on a case-by-case basis, but computer modeling could be used to potentially iteratively solve for the depletion width and defect concentrations based on the measured \( \Delta R \), access region 2DEG concentration, and mobility. The best potential way to measure the lateral depletion depth will be the AFM-based measurements currently in development.

The recently purchased and assembled AFM system with variable monochromatic light source is just beginning to give results, which are quite promising. The three areas of thrust for the AFM-based nano-scale measurements are modified versions of near-field scanning optical microscopy (NSOM), scanning capacitance microscopy (SCM), and scanning Kelvin
probe microscopy (SKPM). Each of these areas is in various stages of development with SKPM being the nearest to completion. It has been shown that SCM and SKPM can achieve quantitative energy and concentration information with nanometer scale resolution, which is definitely achievable for the SKPM with some question to how far the absolute capacitance measurement will scale down from the existing micron scale.

The AFM measurements have direct application to many of the studies already mentioned but they can also be used more generally on other bare surface materials including nitrides but also phosphides, arsenides, and other semiconductor material systems. For example, SKPM could be used to measure the variation bare surface barrier height correlating with surface features or structural defects. As another example, the current-AFM capabilities of the system could be used to explore the role of dislocations in GaP-on-Si solar cells similar to studies in GaN.[28, 29] The possibilities are so endless that the remaining discussion will be limited to nitrides. Meneghesso et al. have used electroluminescence to identify localized defect regions that evolve under different biasing conditions but without the resolution to determine what defects cause the increased luminescence.[30] These experiments can be repeated with the added ability to determine local trap concentrations and energies, which is really the ultimate goal as defects would be directly and unambiguously correlated to device degradation. These mechanisms may be different with passivation or growth modes, so other samples will need to be tested as well. Further speculation is possible, but because little expected results exist it means that future results will probably dictate much of the continued research.

The more immediate need for the AFM measurements is continued infrastructure improvements. The SKPM measurements require additional programming to automate the transient capture process in addition to some image processing programming to subtract
one image from another and to perform line scans at the exact same locations. The existing Park Systems software falls short in these respects but much of it is beyond the scope of the commercial-off-the-shelf system as well. The scaling of the capacitance to find the resolution limit is underway, but the testing cannot begin until the small capacitance structures can be probed in the TTP4 stages and the piezopositioner probes are currently being fabricated to allow this test to begin. Once the test has begun, additional electronics will need to be tested to determine if they can outperform the Boonton 7200 capacitance meter. Once this is complete, the absolute capacitance measuring device can be transitioned to the AFM where SCM measurements can begin. Because the SCM measurements will be nearly identical to the SKPM measurements except for the measured feedback, little programming should be needed to enable localized SCM-based DLOS measurements. Finally, the NSOM is the area in most need of work because not only is the wavelength range of the the optics column limited but how to do this form of defect spectroscopy measurement quantitatively is also yet to be determined.

In the end, the ultimate goal is correlate defects both to their source and to the impact they have electrically. The quantitative ability of these measurements has been stressed but a somewhat more subtle purpose of this is to allow comparison between samples. In this way, measurements on one sample will not be scaled by threshold voltage variation, drain current variation, work function variation, and the numerous other device parameters in diodes, transistors, or bare surfaces. These new experiments should be seen as a framework to allow comparisons of samples with complex geometries, layer structures, and defect-induced phenomena.
8.5 References


Appendix A

Nitride Schottky diode processing

A.1 Introduction

The processing of Schottky diodes for this research was conducted at Wright Patterson Air Force Base (WPAFB). The process developed built upon a standard processes at WPAFB and UCSB. The process varies depending on the sample being processed. The most common change in processing steps occurs between GaN and Al containing structures. For think AlGaN layers and AlGaN/GaN heterostructures, the Ohmic contact needs to be annealed. In this case the processing steps are

1. Wafer clean
2. Mesa (if needed)
3. Ohmic
4. Ohmic anneal
5. Semitransparent Schottky
6. Thick front pad

Our Ohmic contacts are typically annealed at 850°C for 30 s in a nitrogen ambient. For example, the Ohmic contact to the thick $n$-$\text{Al}_{0.30}\text{Ga}_{0.70}\text{N}$ went from completely rectifying ($> 1 \text{ MΩ}$) before the anneal to a contact resistivity of $8 \times 10^{-6} \text{ Ωcm}^2$ and a sheet resistance of 261 $\text{Ω}/\square$ after the anneal.
For GaN samples, an anneal is not necessary for the Ohmic contact so this is typically the third masking step for this process. Although the TLM structures have rectifying behavior, the back contact area is large enough that reasonable Ohmic character presumably by defects or regions of high surface charge where tunneling can readily occur. Typical series resistance for a Schottky diode on GaN would be $50 - 100 \, \Omega$, which is quite sufficient for capacitance-based spectroscopy. In this case the processing steps are

1. Wafer clean
2. Semitransparent Schottky
3. Mesa
4. Ohmic without anneal
5. Thick front pad

Other unique sample structures have required further customization of the processing steps, but these are determined on a per sample basis. For processes that have significant risk with the modified process, test samples are used to reduce the chance of destroying the real sample. One such case is processing of N-face GaN. N-face nitrides etch in bases such as photoresist developer, so this requires either a hard mask process (e.g. SiO$_2$) or a change in processing steps. For N-face GaN that does not require annealing of the Ohmic contact, the Schottky contact can be a blanket deposit with a wet etch to remove the Ni in wanted areas. Transene TFB was previously used to etch the Ni layer, but the etch was uneven and took $\sim 30$ min to remove 80 Å of Ni. However, the process did work for our process. The processing steps for this process are shown in Fig. A.1.

The following sections describe each processing step where the processing order is determined by sample requirements previously mentioned.
Figure A.1: N-face process flow to fabricate Schottky contacts on thick $n$-GaN using the typical growth structure where the top layer is lightly doped that the Schottky contact is deposited on and a heavily-doped lateral conduction layer that is contacted for the Ohmic contacts.
A.2 Wafer Clean

- Secure wafer on spinner and spin at 500 rpm
- 15 sec. acetone spray gun
- 15 sec. acetone spray bottle
- 15 sec. isopropyl spray bottle
- Dry with N₂ on clean Texwipes
A.3  Ni Schottky Contact

Dehydration Bake

☐ 1 min. 110°C hot plate bake

PMGI SF11 Coat

☐ Flood wafer with PMGI SF11
☐ 30 sec spin at 4000 rpm, ramp at 200 (verify before coating)
☐ clean backside with EBR (edge bead removal) and swab
☐ 2 min. 200°C hot plate bake
☐ Cool wafer on cooling plate

Flood DUV Exposure

☐ 25 sec Deep UV Exposure

1813 Coat

☐ Flood wafer with 1813 photoresist
☐ 30 sec spin at 4000 rpm, ramp at 200 (verify before coating)
☐ clean backside with acetone and swab
☐ 75 sec. 110°C hot plate bake
☐ Cool wafer on cooling plate

Align and Expose Schottky Mask

☐ Roughly align mask with edge of wafer
☐ 15 sec. expose on MJB-3 mask aligner

351 Develop

☐ 30 sec. spin develop with 351:DI [1:5] developer at 500 rpm
☐ 30 sec. spin rinse with DI at 500 rpm
☐ Dry with N₂ on clean Texwipes

Inspect Lithography

☐ Inspect photoresist with microscope making sure features are open, clean, and sharp

DUV Expose

☐ 200 sec. deep UV exposure

101 Develop
☐ 2 min. bucket develop with 101 developer
☐ 6X DI rinse in automatic rinse tank
☐ Dry with N₂ on clean Texwipes

**Inspect Lithography**

☐ Inspect photoresist and PMGI features with microscope making sure features are open, clean, and sharp. PMGI undercut should be obvious.

**O₂ Plasma Asher**

☐ 4 min. LFE barrel asher (verify settings are 200 W, 500 sccm O₂, 2.5 – 3 Torr)

**Measure Photoresist Height**

☐ Measure resist thickness with profilometer and record ___________ µm

**Pre-metal Dip**

☐ 10 sec. DI rinse
☐ 30 sec. HCl:DI [1:1] etch
☐ 4X DI rinse in automatic rinse tank
☐ Dry with N₂ on clean Texwipes

**Metal Deposition**

☐ Submit Metal Request (80 Å Ni)

**Metal Lift-off**

☐ Heat 1165 Stripper to approximately 90°C
☐ Secure wafer on spinner and spin at 500 rpm
☐ 15 sec. acetone spray gun
☐ 15 sec. acetone spray bottle
☐ 15 sec. isopropyl spray bottle
☐ Dry with N₂ on clean Texwipes
☐ Check back side, clean with acetone and isopropyl. Re-clean front side in case of acetone on front

**PMGI Strip**

☐ 2 min. soak in 1165 stripper at 90°C
☐ 6X DI rinse in automatic rinse tank
☐ Dry with N₂ on clean Texwipes

**Inspect Schottky Metal**

☐ Inspect Schottky metal with microscope.
A.4 Mesa Etch

Dehydration Bake

☐ 1 min. 110°C hot plate bake

PMGI SF11 Coat

☐ Flood wafer with PMGI SF11
☐ 30 sec spin at 4000 rpm, ramp at 200 (verify before coating)
☐ clean backside with EBR (edge bead removal) and swab
☐ 5 min. 270°C hot plate bake
☐ Cool wafer on cooling plate

PMGI SF11 Coat 2 (up to ∼0.6 µm etch)

☐ Flood wafer with PMGI SF11
☐ 30 sec spin at 4000 rpm, ramp at 200 (verify before coating)
☐ clean backside with EBR (edge bead removal) and swab
☐ 5 min. 270°C hot plate bake
☐ Cool wafer on cooling plate

PMGI SF11 Coat 3 (up to ∼1 µm etch)

☐ Flood wafer with PMGI SF11
☐ 30 sec spin at 4000 rpm, ramp at 200 (verify before coating)
☐ clean backside with EBR (edge bead removal) and swab
☐ 5 min. 270°C hot plate bake
☐ Cool wafer on cooling plate

1813 Coat

☐ Flood wafer with 1813 photoresist
☐ 30 sec spin at 4000 rpm, ramp at 200 (verify before coating)
☐ clean backside with acetone and swab
☐ 75 sec. 110°C hot plate bake
☐ Cool wafer on cooling plate

Align and Expose Mesa Mask

☐ Align mask with Schottky metal layer alignment marks
☐ 15 sec. expose on MJB-3 mask aligner

351 Develop
☐ 30 sec. spin develop with 351:DI [1:5] developer at 500 rpm
☐ 30 sec. spin rinse with DI at 500 rpm
☐ Dry with N₂ on clean Texwipes

**Inspect Lithography**

☐ Inspect photoresist with microscope making sure features are open, clean, and sharp

**DUV Expose**

☐ 200 sec. deep UV exposure

**101 Develop**

☐ 1 min. bucket develop with fresh 101 developer
☐ 6X DI rinse in automatic rinse tank
☐ Dry with N₂ on clean Texwipes

**DUV Expose 2**

☐ 200 sec. deep UV exposure

**101 Develop 2**

☐ 1 min. bucket develop with fresh 101 developer
☐ 6X DI rinse in automatic rinse tank
☐ Dry with N₂ on clean Texwipes

**DUV Expose 3**

☐ 100 sec. deep UV exposure

**101 Develop 3**

☐ 1 min. bucket develop with fresh 101 developer
☐ 6X DI rinse in automatic rinse tank
☐ Dry with N₂ on clean Texwipes

**Inspect Lithography**

☐ Inspect photoresist and PMGI features with microscope making sure features are open, clean, and sharp. PMGI undercut should be obvious. **Remove 1813**
☐ 15 sec. acetone spray gun
☐ 15 sec. acetone spray bottle
☐ 15 sec. isopropyl spray bottle
☐ 15 sec. DI spray bottle
Dry with N\textsubscript{2} on clean Texwipes
Swab back side with acetone if necessary

**Measure Photoresist Height**

- Measure resist thickness with profilometer and record before etch thickness = 
  \_______________ \(\mu\text{m}\)

**RIE Etch**

- Record target etch depth, \_______________ \(\mu\text{m}\), and etch time (target/[0.07 m/min]) _min.
- Recipe: 20 sccm Cl\textsubscript{2}, 5 sccm Ar, 5 mTorr, 40 W RIE, 145 V DC Bias, No ICP, 10°C wafer temp., 20°C wall temp., He 5 mTorr/10 sccm. Gives an etch rate of 70nm/min.

**Measure Photoresist Height after RIE Etch**

- Measure resist thickness with profilometer and record after etch thickness = 
  \_______________ \(\mu\text{m}\)

**PMGI Strip**

- 2 min. soak in 1165 stripper at 10°C
- 6X DI rinse in automatic rinse tank
- Dry with N\textsubscript{2} on clean Texwipes

**Inspect wafer**

- Inspect wafer with microscope making sure PMGI is removed.

**Measure Photoresist Height after strip**

- Measure resist thickness with profilometer and record III-N etch depth = 
  \_______________ \(\mu\text{m}\)
- Calculate PMGI etch depth (before etch - after etch + III-N etch depth) = 
  \_______________ \(\mu\text{m}\)
- PMGI etch rate (PMGI etch depth/etch time) = \_______________ \(\mu\text{m}/\text{min}\)
- III-N etch rate (III-N etch depth/etch time) = \_______________ \(\mu\text{m}/\text{min}\)
A.5  Ohmic Contact

Dehydration Bake

☐ 1 min. 110°C hot plate bake

PMGI SF11 Coat

☐ Flood wafer with PMGI SF11
☐ 30 sec spin at 4000 rpm, ramp at 200 (verify before coating)
☐ clean backside with EBR (edge bead removal) and swab
☐ 2 min. 200°C hot plate bake
☐ Cool wafer on cooling plate

Flood DUV Exposure

☐ 25 sec Deep UV Exposure

1813 Coat

☐ Flood wafer with 1813 photoresist
☐ 30 sec spin at 4000 rpm, ramp at 200 (verify before coating)
☐ clean backside with acetone and swab
☐ 75 sec. 110°C hot plate bake
☐ Cool wafer on cooling plate

Align and Expose Back Contact Mask

☐ Align mask with Mesa layer alignment marks
☐ 10 sec. expose on MJB-3 mask aligner

351 Develop

☐ 30 sec. spin develop with 351:DI [1:5] developer at 500 rpm
☐ 30 sec. spin rinse with DI at 500 rpm
☐ Dry with N₂ on clean Texwipes

Inspect Lithography

☐ Inspect photoresist with microscope making sure features are open, clean, and sharp

DUV Expose

☐ 200 sec. deep UV exposure

101 Develop
☐ 2 min. bucket develop with 101 developer
☐ 6X DI rinse in automatic rinse tank
☐ Dry with N₂ on clean Texwipes

**Inspect Lithography**

☐ Inspect photoresist and PMGI features with microscope making sure features are open, clean, and sharp. PMGI undercut should be obvious.

**O₂ Plasma Asher**

☐ 2 min. LFE barrel asher (verify settings are 200 W, 500 sccm O₂, 2.5 – 3 Torr)

**Measure Photoresist Height**

☐ Measure resist thickness with profilometer and record _____________ μm

**Pre-metal Dip**

☐ 10 sec. DI rinse
☐ 30 sec. HCl:DI [1:1] etch
☐ 4X DI rinse in automatic rinse tank
☐ Dry with N₂ on clean Texwipes

**Metal Deposition**

☐ Submit Metal Request (Ti/Al/Ni/Au 100:2000:200:3000 Å)

**Metal Lift-off**

☐ Heat 1165 Stripper to approximately 90°C
☐ Secure wafer on spinner and spin at 500 rpm
☐ 15 sec. acetone spray gun
☐ 15 sec. acetone spray bottle
☐ 15 sec. isopropyl spray bottle
☐ Dry with N₂ on clean Texwipes
☐ Check back side, clean with acetone and isopropyl. Re-clean front side in case of acetone on front

**PMGI Strip**

☐ 2 min. soak in 1165 stripper at 90°C
☐ 6X DI rinse in automatic rinse tank
☐ Dry with N₂ on clean Texwipes

**Inspect Schottky Metal**

☐ Inspect Schottky metal with microscope.
A.6 Thick Front Contact

Dehydration Bake

☐ 1 min. 110°C hot plate bake

PMGI SF11 Coat

☐ Flood wafer with PMGI SF11
☐ 30 sec spin at 4000 rpm, ramp at 200 (verify before coating)
☐ Clean backside with EBR (edge bead removal) and swab
☐ 2 min. 200°C hot plate bake
☐ Cool wafer on cooling plate

Flood DUV Exposure

☐ 25 sec Deep UV Exposure

1813 Coat

☐ Flood wafer with 1813 photoresist
☐ 30 sec spin at 4000 rpm, ramp at 200 (verify before coating)
☐ clean backside with acetone and swab
☐ 75 sec. 110°C hot plate bake
☐ Cool wafer on cooling plate

Align and Expose Thick Front Contact Mask

☐ Align mask with Mesa layer alignment marks
☐ 15 sec. expose on MJB-3 mask aligner

351 Develop

☐ 30 sec. spin develop with 351:DI [1:5] developer at 500 rpm
☐ 30 sec. spin rinse with DI at 500 rpm
☐ Dry with N₂ on clean Texwipes

Inspect Lithography

☐ Inspect photoresist with microscope making sure features are open, clean, and sharp

DUV Expose

☐ 200 sec. deep UV exposure

101 Develop
☐ 2 min. bucket develop with 101 developer
☐ 6X DI rinse in automatic rinse tank
☐ Dry with N$_2$ on clean Texwipes

**Inspect Lithography**

☐ Inspect photoresist and PMGI features with microscope making sure features are open, clean, and sharp. PMGI undercut should be obvious.

**O$_2$ Plasma Asher**

☐ 1 min. LFE barrel asher (verify settings are 200 W, 500 sccm O$_2$, 2.5 – 3 Torr)

**Measure Photoresist Height**

☐ Measure resist thickness with profilometer and record _______________ µm

**Pre-metal Dip**

☐ 10 sec. DI rinse
☐ 30 sec. HCl:DI [1:1] etch
☐ 4X DI rinse in automatic rinse tank
☐ Dry with N$_2$ on clean Texwipes

**Metal Deposition**

☐ Submit Metal Request (Ni/Au 200:2800 Å, Standard Schottky)

**Metal Lift-off**

☐ Heat 1165 Stripper to approximately 90°C
☐ Secure wafer on spinner and spin at 500 rpm
☐ 15 sec. acetone spray gun
☐ 15 sec. acetone spray bottle
☐ 15 sec. isopropyl spray bottle
☐ Dry with N$_2$ on clean Texwipes
☐ Check back side, clean with acetone and isopropyl. Re-clean front side in case of acetone on front

**PMGI Strip**

☐ 2 min. soak in 1165 stripper at 90°C
☐ 6X DI rinse in automatic rinse tank
☐ Dry with N$_2$ on clean Texwipes

**Inspect Schottky Metal**

☐ Inspect Schottky metal with microscope.

302
Appendix B

The appendix of Labview programs written to support the research

An entire suite of programs was written to support the research presented. Simple programs from current-voltage, internal photoemission, and capacitance-voltage to more complicated programs for DLTS, DLOS, current-voltage-temperature, FPGA initialization, AFM-DLOS, and others have been developed to support the mostly custom equipment assembled to not only research nitrides but also many other III-V semiconductors. Several analysis programs have also been written to extract information from the data including a doping calculator, Arrhenius plot generator based on DLTS peaks, $dC/dt$ and steady-state capacitance calculator for DLOS data, and I-V analysis program to extract ideality factors, series resistance, and barrier height. Of the five DLOS systems, the software is currently used on four of them.

Before delving into the programs, the equipment is briefly described. Below is a list of equipment and brief description:

- Boonton 7200 capacitance meter
- A heavily modified LakeShore TTP4 cryogenic vacuum stage to support the wide range of samples and measurements performed
- Newport-Oriel light sources, monochromators, lenses, filters, and mirrors
- Agilent DSO6014 digital oscilloscope
- Keithley 6514 electrometer
- Agilent 33220A arbitrary function generator
B.1 Stage operation

B.1.1 Vacuum procedures

The most important requirement that must be met for pump down or venting is that the probes be out of contact and not near each other. This is because the probes move ≈ 1 mm from atmosphere to vacuum and vice versa. To pump down when the stage is at atmosphere, open the Speedivalve and turn on the pump by pressing the Start/Stop button. When the turbo spins up to 75 krpm, the cold cathode gauge can be turned on by pressing the Emis button. This should only be turned on when the pressure in the stage is less than 0.01 Torr. To vent the stage or turn off the active pumping, first turn off the cold cathode gauge by pressing the Emis button, close the Speedivalve, and then press the Start/Stop button. If the stage is already under vacuum when one wants to begin active pumping again, the procedure is slightly different. In this case, the vacuum pump should be turned on first. After the turbo is mostly spun up, open the Speedivalve slowly so the turbo is not loaded down too much. When the valve is completely open and the pressure low enough, the cold cathode gauge can be turned on. This procedure is designed so that air from the foreline is never vented into stage, which could suck in oil and other contaminants were this to happen. A high likelihood of damage to the stage will occur if the stage is vented while at low temperatures. This is because the stage will cryopump the air freezing it on the stage parts. This can cause large thermal gradients that can warp and damage seals and other parts of the stage.
B.1.2 Temperature procedures

Refer to the LakeShore TTP4 manual for directions for temperature pump down procedures, but a brief explanation is given here. When the stage has reached $\sim 2 \times 10^{-5}$ Torr, the temperature can be lowered. To do this, the transfer line is first put into a dewar and sealed to prevent loss of “air” pressure. Then, the transfer line should very carefully be put into the TTP4 stage without bumping the stage, which can cause loss of contact or scratching of contacts. There is a seal on the TTP4 stage that needs to be tightened to prevent pressurized LN2 or nitrogen gas from escaping before going through the cold finger. Once the transfer line is installed, the foot valve on the transfer line can be opened to allow LN2 to flow through the stage. The rate of flow is controlled both by the foot valve and flow regulator on the exit. The rate of temperature decrease should be limited to 0.06 K per update of the temperature controller screen. If the flow is too high, the heater will not be able to override the cooling capacity and the temperature will not be controlled. In this case, the flow needs to be decreased to allow the heater to stay in the 50-75% range.

If the stage does not reach 78 K or the transfer lines have condensation on them, the transfer lines will need to be vacuumed out again. There are fittings needed to pump these out and typically they pump down to $\sim 2 \times 10^{-6}$ Torr. This is typically needed every 3 months or so.

Above room temperature, the transfer line needs to be carefully removed. The radiation shield must be kept below 325 K, so a nitrogen backflow system is in place. In this case, nitrogen gas first flows through the radiation shield and then through the cold finger cooling the radiation shield. A valve on the exit switches between the flow regulator and nitrogen backflow. Turn the valve to the nitrogen cylinder and turn the nitrogen cylinder on at 10 psi
or less to activate the process. The hissing noise of nitrogen going through the stage should be evident. The maximum temperature of the stages is 400 K.

B.2 Current-voltage measurements

Before measuring anything else on a new sample or diode, an I-V measurement is necessary to verify that the sample is acceptable for further measurements (i.e. rectifying behavior, low leakage current, correctly connected, etc.) The I-V measurements can also be much more than just a cursory glance at the sample. In this case, the resistance, ideality factor, and Schottky barrier height can be determined assuming thermionic emission. To go even further, the transport mechanisms can be studied carefully using numerous different theories to better understand the Schottky interface and semiconductor surface. Many things can be determined such as lateral Schottky barrier height variation, tunneling, and thermally assisted tunneling to name a few. Complete analysis typically requires knowing the temperature dependence, which is possible using the I-V-T program, which is briefly discussed at the end of this section.

B.2.1 General I-V theory

To determine the series resistance two techniques are employed. First, to get a rough idea of the series resistance, the x-axis values of voltage can be replaced with \( V - IR_s \). By varying \( R_s \) and observing when the log \( I \) vs. \( V \) curve is linear, the series resistance can be estimated. The series resistance can also be extracted using the method developed by Cheung and is

\[
\frac{dV}{d(\ln I)} = R_s I + n \frac{kT}{q} \tag{B.1}
\]

An example of this kind of plot is given in Fig. B.1.
B.2.2 Schottky I-V-T theory

Current-voltage-temperature (I-V-T) measurements are performed to determine many details including barrier height, ideality factor, and carrier transport mechanisms. The standard diode equation used to extract the I-V parameters assuming thermionic emission is

$$\ln(I) = \frac{qV}{nkT} + \ln(I_s) \text{ for } qV > 3kT$$  \hspace{1cm} (B.2)

where $q$ is the elementary charge of an electron, $n$ is the ideality factor, $k$ is Boltzmann's constant, and $I_s$ is the saturation current. Fitting the data to the previous equation, we can extract $n$ and $I_s$. The saturation current for a Schottky diode is expressed as[2]

$$I_s = A_e A^* T^2 \exp \left( \frac{-q\Phi_{b0}}{kT} \right)$$  \hspace{1cm} (B.3)
where $A_e$ is the Schottky diode area, $A^*$ is Richardson’s constant, and $\Phi_{b0}$ is the zero-bias barrier height. After rearranging, $A^*$ and $\Phi_{b0}$ can be obtained from

$$\ln\left(\frac{I_s}{A_e T^2}\right) = \ln A^* - q\Phi_{b0} \frac{1}{kT} \quad (B.4)$$

Plotting $1/kT$ vs $\ln (I_s/A_e T^2)$ is referred to as a Richardson plot, where the slope is the barrier height and the $y$-intercept is the $\ln (A^*)$. The theoretical Richardson constant with units of $A/K^2/cm^2$ is given by[2]

$$A^* = \frac{4\pi q m^* k^2}{h^3} = 120 \frac{m^*_r}{m_0} \quad (B.5)$$

By assuming a value for the electron effective mass in GaN of 0.222$m_o$, the theoretical Richardson constant for n-type GaN is found to be 26.64 $A/K^2/cm^2$.[3] This theoretical Richardson constant can be compared with the Richardson constant extracted from the Richardson plot to help verify the validity of this approach.

The previous derivation is correct assuming thermionic emission (TE), but this is often not the case. Relatively high doping and lateral Schottky barrier height variation, for example, can cause the I-V-T results to vary significantly from the thermionic emission model. Determining the correct current transport method can be difficult as many models exist and in some cases multiple models will produce the equally good fits. If possible, it is best to consider the forward bias current as well as the reverse leakage current to help decipher the transport method. Even if one transport model fits over a given temperature rage, it is possible for the transport mechanism to depend on both temperature and applied voltage, which can greatly complicate analysis. Several models are briefly discussed here to overview the transport mechanisms.

Thermionic-field emission (TFE) is a common transport mechanism in higher doped films. To determine if TFE should be a dominant transport mechanism the thermal energy
$kT$ is compared with the characteristic energy $E_{00}$

$$E_{00} = \frac{h}{4\pi} \left( \frac{N_D}{m_e^* \varepsilon_s} \right)^{\frac{1}{2}} = 1.85664 \times 10^{-11} \text{V} \cdot \text{cm}^{1.5} \left( \frac{N_D}{m_e^* \varepsilon_s} \right)^{\frac{1}{2}}$$  \hspace{1cm} (B.6)

where $m_e^*$ is the electron effective mass, $m_r^*$ is the electron effective mass relative to the electron rest mass, $\varepsilon_s$ is the permittivity, $\varepsilon_r$ is the permittivity relative to the vacuum permittivity, $h$ is Planck’s constant, and $N_D$ is the donor doping concentration. If $E_{00} << kT$, TE should be the dominant transport. If $E_{00} \sim kT$, thermally assisted tunneling through the top of the Schottky barrier or TFE is dominant. Padovani et al. developed equations for TFE where the current still has an exponential dependence on applied voltage $V$ \cite{4, 5}

$$I = I_{fwd TFE} \left[ \exp \left( \frac{qV}{n_F kT} \right) - 1 \right]$$  \hspace{1cm} (B.7)

where the forward-bias ideality factor $n_F$ is

$$n_F = \frac{E_{00}}{kT} \coth \left( \frac{E_{00}}{kT} \right)$$  \hspace{1cm} (B.8)

and the saturation current $I_{fwd TFE}$ is

$$I_{fwd TFE} = \frac{AA^* T^2 \sqrt{\pi q E_{00} (\Phi_B - V + V_n)}}{kT \cosh (E_{00}/kT)} \exp \left[ \frac{qV_n}{kT} - \frac{qkT (\Phi_B + V_n)}{E_{00} \coth (E_{00}/kT)} \right]$$  \hspace{1cm} (B.9)

for forward bias where $A$ is the area, $A^*$ is Richardson’s constant, $T$ is the temperature, $\Phi_B$ is the barrier height, and $V_n$ is $(E_C - E_F)/q$. In reverse bias, another set of equations is needed for reverse bias \cite{4, 5}

$$I = I_{rev TFE} \exp \left( \frac{qV}{n_R kT} - 1 \right)$$  \hspace{1cm} (B.10)

where the reverse-bias ideality factor $n_R$ is

$$n_R = \frac{E_{00}}{kT} \left( \frac{E_{00}}{kT} - \tanh \left( \frac{E_{00}}{kT} \right) \right)^{-1}$$  \hspace{1cm} (B.11)

and the reverse saturation current $I_{rev TFE}$ is

$$I_{rev TFE} = \frac{AA^* m_e^* T^2 \sqrt{\pi q E_{00} (\Phi_B - V \cosh^2 (E_{00}/kT))}}{kT \cosh (E_{00}/kT)} \exp \left[ \frac{-q\Phi_B}{kT} \right]$$  \hspace{1cm} (B.12)
Figure B.2: Thin surface barrier model band diagram from Ref. 7. Depicted is the conduction band of an n-type semiconductor with Schottky barrier $q\Phi_B$ with voltage $qV_a$ applied. Higher fixed charge density near the surface leads to thinner barrier and enhanced thermionic-field emission at $q\Phi_m$ with a Gaussian distribution.

where $A^*_m$ is the Richardson constant of the metal. Additionally,

$$n_F^{-1} + n_R^{-1} = 1$$  \hspace{1cm} (B.13)$$

so as the forward ideality factor increases, the reverse ideality factor decreases. At $n_F = 2$, $n_R$ is also 2 and the device is completely symmetric, which is normally very undesirable. This model often fits GaN data quite well, but typically the doping calculated from $E_{00}$ is generally higher than the measured bulk doping. Hasegawa et al. explained this assuming a thin surface barrier (TSB) where the ionized charge was in excess of the bulk charge. Figure B.2 shows how current could more easily tunnel through the thin surface barrier. Kotani et al. developed analytic equations for the current assuming one-dimensional transport and the derivation with mostly correct equations is available in Ref 5. The TFE and TSB models still have temperature dependent reverse leakage current, but if temperature independent reverse
leakage current is observed direct tunneling or field emission may be the dominant transport mechanism.

Field emission (FE) is typically observed in highly doped semiconductors and moderately doped semiconductors at low temperature. Padovani and Stratton also developed the analytic equations for field emission for both reverse and forward bias.[4] In forward-bias, the field emission current is

\[ I = I_{wdFE} \exp \left( \frac{qV}{E_{00}} \right) \]  

(B.14)

where the forward saturation current \( I_{wdFE} \) is

\[ I_{wdFE} = \frac{2\pi A A^* T^2 E_{00}}{kT \log(2(\Phi_B - V)/V_n)} \frac{\exp \left( -\frac{q\Phi_B}{kT} \right)}{\sin \left( \pi kT \log(2(\Phi_B - V)/V_n) / (2E_{00}) \right)} \]  

(B.15)

With FE transport in forward bias, the voltage exponential term depends on \( E_{00} \) only instead of \( kT \) as in TFE and TE transport. This results in a weaker temperature dependence in forward bias. In reverse bias, the current is given as

\[ I = A A^* E_{00}^2 \left( \frac{\Phi_B - V}{\Phi_B} \right) \exp \left( -\frac{2(q\Phi_B)^{3/2}}{3E_{00} \sqrt{q\Phi_B - qV}} \right) \]  

(B.16)

In this case, the reverse leakage current has no temperature dependence.

The TE, TFE, TSB, and FE models all assume the Schottky barrier height (SBH) is spatially uniform, but several theories have been developed for inhomogeneous barriers as well. Models by Tung and Werner \textit{et al.} address lateral barrier height variation on two different length scales. Werner and Güttler proposed a model of a Gaussian distribution of Schottky barrier heights with the variations occurring on a lateral scale much less than the thickness of the depletion depth.[8] In this model, only the lowest SBH regions contribute to the current at low temperature resulting in a lower effective barrier height. As the temperature increases, more and more of the SBHs contribute to the current. Although the mean SBH \( \overline{\Phi_b} \) and the
standard deviation of the barrier height $\sigma_s$ remains the same, the extracted barrier heights
with temperature show a temperature dependence. A significant advantage of this model
is that temperature independent fitting parameters give rise to the temperature dependent
behavior of the ideality factor and barrier height. The following relations can be developed
assuming the SBH has a Gaussian distribution and $\Phi_b$ and $\sigma_s^2$ vary linearly with applied
voltage[8]

$$\Phi^j_b = \Phi_b - \frac{q\sigma_s^2}{2kT}$$ (B.17)

where $\Phi^j_b$ is the measured SBH at a given temperature.

$$\Phi_b = \Phi_{b0} + \rho_2 V$$ (B.18)

and

$$\sigma_s^2 = \sigma_{s0}^2 + \rho_3 V$$ (B.19)

where $\rho_2$ describes how the barrier height varies with applied bias and $\rho_3$ is the change in
$\Delta\sigma_2$ with applied bias. With this, the equation for current is

$$I = I_0 \exp \left( \frac{q\Phi^j_b}{kT} \right) \exp \left( \frac{qV}{kT} \right)$$ (B.20)

which can be rewritten as

$$I = AA^* T^2 \exp \left\{ -\frac{q}{kT} \left[ \Phi_{Bn} + \rho_2 V - \frac{q}{2kT} \left( \sigma_s^2 + \rho_3 V \right) \right] \right\} \exp \left( \frac{qV}{kT} \right)$$ (B.21)

where the traditional saturation current is determined by grouping together all of the terms
that do not depend on the applied voltage

$$I_0 = AA^* T^2 \exp \left( -\frac{q}{kT} \Phi_{Bn} + \frac{1}{2} \left( \frac{q}{kT} \sigma_s^2 \right) \right)$$ (B.22)

and the ideality factor is determined by looking at the voltage terms

$$n = \left[ 1 - \rho_2 + \frac{q}{kT} \rho_3 \right]^{-1}$$ (B.23)
These equations can be used in the normal current equation, \( I = I_0 \exp \left( \frac{qV}{nkT} \right) \). To determine \( \rho_2 \) and \( \rho_3 \), a plot of \( n^{-1} - 1 \) vs. \( 1/kT \) should yield a straight line with the intercept and slope being related the voltage parameters \( \rho_2 \) and \( \rho_3 \) according to Eq. (B.23). In similar fashion to the derivation of the Richardson plot a relation can be developed to extract the Richardson constant, zero-bias barrier height, and zero-bias standard deviation

\[
\ln \left( \frac{I_0}{AT^2} \right) = \ln \left( A^* \right) - \Phi_{bn} \frac{q}{kT} + \frac{\sigma_1^2}{2} \left( \frac{q}{kT} \right)^2
\]  

(B.24)

Using \( q/kT \) as the independent variable and \( \ln \left( I_0/AT^2 \right) \) as the dependent variable, the data can be fit with a general quadratic function. A problem with this model is that the current transport is dominated by relatively small regions of low barrier height, which can result in mean barrier heights much above the internal photoemission and C-V extracted barrier heights. Although the model often fits well, it is important to check the fit parameters to assess how realistic they may be.

Another model of lateral Schottky height variation was considered by Tung. Tung considered several different scenarios of lateral SBH variation on a longer range scale where the diameter of the patches is on the order of the Debye length.[9] One scenario Tung considered was a one-half Gaussian distribution of low barrier height patches with an areal density of patches \( N(\gamma) \)

\[
N(\gamma) = \begin{cases} 
\frac{\rho_p}{\sqrt{2\pi}\sigma_1} \exp \left( -\frac{\gamma^2}{2\sigma_1^2} \right) , & \gamma > 0 \\
0, & \gamma \leq 0
\end{cases}
\]  

(B.25)

where \( \rho_p \) is the areal density of circular patches of differing barrier height, \( \gamma \) is the region parameter defined by Tung (i.e. a product of the patch area and deviation of the barrier height in the patch from the uniform barrier height outside the patches), and \( \sigma_1 \) is the standard deviation of the parameter \( \gamma \). Based on this Tung developed an equation for the current[9, 10]

\[
I = AA^* T^2 \exp \left( -\frac{\Phi_{bn}}{kT} \right) \left[ \exp \left( \frac{qV}{kT} \right) - 1 \right]
\]
\[
1 + \frac{8\pi \rho_p \sigma_1^2 \eta^{1/3}}{9(\Phi_{b0} - V_n - V)^{1/3}} \exp \left( \frac{q^2 \sigma_1^2 (\Phi_{b0} - V_n - V)^{2/3}}{2(kT)^2 \eta^{2/3}} \right) \quad \text{(B.26)}
\]

where \( \eta = \epsilon / qN_D, \) \( V_n = (E_C - E_F) / q, \) and \( \Phi_{b0} \) is the uniform barrier height outside the patches. This equation can be simultaneously fit over a temperature range in forward bias over the linear region where \( A^*, \Phi_{b0}, \rho_p, \) and \( \sigma_y \) are allowed to vary. The standard deviation of the barrier height can be calculated by considering the equation developed by Tung[11]

\[
\Phi^j_b = \Phi_{b0} - \frac{q}{kT} \frac{\sigma_1^2}{2\eta^{2/3}} (\Phi_{b0} - V_n - V)^{2/3}
\]

which can be rewritten as

\[
\Phi^j_b = \Phi_{b0} - \frac{q\sigma_s^2}{2kT}
\]

where \( \sigma_s \) is the standard deviation of the barrier height as in the case of Werner et al. and

\[
\sigma_s = \sigma_1 \left( \frac{\Phi_{b0} - V_n - V}{\eta} \right)^{1/3}
\]

Using this model, several physical parameters can be extracted including the standard deviation of the SBH, patch density, and even the patch area. But being a macroscopic measurement, it is difficult to definitively ascertain that this model accurately represents the microscopic SBH variation and correlate the low SBH regions with physical phenomena.

A very brief review of I-V-T models have been presented to show the range of physical phenomena that have been modeled. In I-V-T, the measurements are the easy parts. The difficult part is the modeling to understand what is causing the observed leakage, SBH, ideality factor, and temperature dependence. Not only is simultaneous fitting often required, but different models can sometimes fit equally well making it more difficult to definitively identify the source of the anomalous I-V-T. Sometimes it is necessary to analyze the current transport to identify processing issues. Other times when working with new materials, identification of the current transport is one part of fully characterizing the new material.
photoemission as a separate method to determine SBH and further the understanding and feasibility of each I-V-T model as well.

**B.2.3 I-V circuit diagram**

To connect a diode for I-V measurement, the following procedure is needed:

- □ Connect the voltage source to the p-type contact (or Schottky contact for a n-type Schottky diode)
- □ Connect the other contact to the ammeter/electrometer

The connection diagram is shown in Fig. B.3. Typically the measurement is performed under vacuum because the probe tips do not wobble as much, but the pump down procedure should be performed before making contact as this would cause severe damage to the sample and probe tips as the vacuum caused the tips to move inward and downward.

**B.2.4 I-V program**

The I-V program is relatively straightforward. The program, as with every other Labview program, is started by clicking the run button (button with white arrow pointing to the right). At this point, the current can be checked by using the quick check to change the voltage and monitor the current – shown in Fig. B.4. The quick check should be used to slowly find the starting reverse bias that will not exceed the current limitations of the ammeter. Failure to do this can result in diode failure or even destroying the electrometer or picoammeter.
Figure B.4: Screenshot of the I-V program used for the DLOS systems.

Once the parameters are all entered, the green go button should be pressed. This initiates the measurement and asks for sample information and filename.

**B.2.5 I-V analysis**

The I-V Analysis program is shown in Fig. B.5 and is used to analyze the data obtained from the I-V measurement. This program can be launched directly after an I-V is completed by turning the flip switch to yes under the "Run Analysis?" label in the I-V program (Fig. B.4), or this program can be run later and the user will be prompted to import the data. The graph on the left of Fig. B.5 allows for the extraction of the $R_s$ though use of the Cheung method described in Subsection B.2.1. The graph on the right of Fig. B.5 is used to calculate the saturation current, barrier height, and ideality factor of a Schottky diode assuming thermionic...
emission. After the linear area of the plot is selected, the calculations are automatically completed.

### B.2.6 Other I-V programs

In addition to this I-V program, there is also an I-V program that uses the Keithley 2400 SourceMeter for high current or high voltage applications and another program that uses the HP 4145B parametric analyzer to perform current-voltage-temperature measurements. These are not described in detail but follow the same basic form as the I-V program described here. Screenshots of two other I-V programs are shown in Figs. B.6 and B.7.
Figure B.6: Screenshot of I-V program for use with the Keithley 2400 SourceMeter.
Figure B.7: Screenshot of I-V-T program for use with the HP 4145B parametric analyzer.
B.3 Capacitance-voltage

The theory for the expected behavior of a C-V curve and the method to extract the doping from the curve is given in Section 4.2. The C-V program can also be used to perform lighted C-V and the only difference is that the sample is exposed to the desired energy of monochromatic light when the C-V is performed. The necessary circuit and program usage is described next.

B.3.1 C-V circuit diagram

In order to physically set up a C-V experiment, the diode needs to be attached following the circuit diagram in Fig. B.8. As with the I-V measurement, the $p$-type contact should be attached to the DUT High. The correct connection can usually be confirmed by applying a small reverse bias to the sample, which should result in a lowering of the capacitance. The connections always need to be confirmed before a large forward bias is accidentally applied to the sample resulting in diode or Boonton failure.
Figure B.8: C-V circuit diagram showing the Boonton 7200 capacitance meter and DLD2005 digital capacitance offset box.

B.3.2 C-V program

The C-V program screenshot is shown in Fig. B.9. This program is designed to be used with the Boonton 7200 capacitance meter or Stanford Research SR831 lock-in amplifier. The use of the lock-in is described in the DLTS-DLOS Equipment Manual and Daewon Kwon’s manual and will not be described further here. The use of the Boonton is straightforward where the connection diagram is given in Fig. B.8, the bias is applied internally, and the computer controls the bias, capacitance offset, and capacitance recording. The maximum bias applied should have a current flow less than 1 µA generally although some samples have worked with currents as high as 100 µA. This is confirmed by always performing an I-V measurement before the C-V measurement. The Full Scale is the range the Boonton uses for the measurement. Small ranges have the highest signal-to-noise ratio, but obviously a limited range. This is overcome by the program automatically changing the capacitance offset to keep the capacitance in range of the capacitance meter, but too many offsets are bad as they
Figure B.9: Screenshot of the C-V program.
often introduce small step offsets in the C-V data and each one has to be removed separately. The C-V program is started like all the other program by pressing the green go button with Labview already running (right arrow button pressed). When the C-V finishes, the doping of the film is calculated as a function of depletion depth following the theory presented in Section 4.2.

B.3.3 Doping calculator program

The doping calculator calculates the depletion depth as a function of voltage using Eq. (4.4) and the doping, or more accurately fixed charge density, using Eq. (4.7). The program (screenshot in Fig. B.10) has few inputs and displays and saves the doping/depletion depth information in ASCII format that can easily be imported into Origin graphing software. The information needed to calculate the data is the permittivity of the semiconductor, area of the diode, and how many points to average over (a point spacing of 3 averages 3 point to the right and 3 points to the left for a total of 7 points). The $y$-axis of the graph is the fixed charge density in cm$^{-3}$ and the $x$-axis is depletion depth in $\mu$m.
Figure B.10: Screenshot of the doping calculator program.
B.4 Deep level transient spectroscopy

The DLTS experiment was a critical part of the research in this thesis and is described in depth in Chapter 4. The general procedure of DLTS is to

- apply a fill pulse (0 V, 10 ms) to fill the deep levels
- return to a reverse bias (–0.5 V) and record the capacitance transient as the carriers emit
- increase the temperature (0.1 K) and repeat the process until the end temperature is reached

The emission rate depends exponentially, so the double boxcar method is able to exploit this behavior to determine the emission rate as a function of temperature to extract the trap energy and cross section.

B.4.1 DLTS circuit diagram

Connecting the equipment for DLTS nontrivial, so a circuit diagram is provided in Fig. B.11. The AFG310/Agilent 33220A function generators provide the fill pulsing while the Boonton 7200 and DLD2005 provide the capacitance measurement and offset capabilities. The Agilent DSO6016A oscilloscope is used to monitor the capacitance transients and is not used by the computer for any analysis or decision making. The fill pulse is sent to the computer via the National Instruments (NI) BNC2210 breakout box to determine the starting point of the capacitance transients that is recorded on AO0. The computer contains an NI M-Series data acquisition card capable of recording 1.5 MSa/sec. These measurements are started from low temperature, so proper vacuum and cooling procedures should be followed.
B.4.2 DLTS program

The DLTS program performs a great number of functions from capacitance-, constant capacitance-, constant drain current-, and constant conductance-mode measurements. The program acquires the data, processes the data, saves the data, controls the temperature, and capacitance offset if applicable. The DLTS program is first discussed in terms of normal capacitance mode then all of the constant mode measurements. The DLTS program is shown in Fig. B.12. The graph on the left displays the double boxcar results of the DLTS experiment. The settings for the double boxcar are the fraction to average, t2/t1 ratio, and # of BC avg., which control the amount of time averaged around the two boxcar times, how much larger t2 is than t1, and how many points are averaged together for the boxcar data, respectively.
Figure B.12: DLTS program screenshot.
scan rate is the acquisition rate for the data acquisition card and # of transients is only used to
determine how many full transients are averaged together to form a single saved transient.
Transient saving is selected or deselected with the Record entire transient control and the
number of points saved is below. The transient files are automatically saved in binary format
and saved with the average temperature in the filename. Several programs exist to convert
the binary data to ASCII format, but a program can easily be written to perform this function
based on the DLTS code if need arises. These recorded transients can be used to perform
different analysis methods including the Fourier and Padé-Laplace methods discussed in
Chapter 4 using the appropriate Labview programs.

A long discussion of the Desired e's and # to avg is needed to explain the operation of
the DLTS measurement. Because low rate windows have large numbers of points to average
within a single transient while large rate windows do not, more short transients are necessary
to achieve the same number of points to average. To do this, the program can handle running
transients with different time lengths. Both controls consist of a series of numbers. It is easiest
to start from the bottom with the 2 from Desired e's and 3 from # of avgs. This means run 3
transients long enough to capture a rate window of 2 s\(^{-1}\). These 3 transients can also be used
to calculate all higher rate windows (shorter transients), so the 4, 10, 20, etc. rate windows
are all calculated from these transients. To supplement these transients, 2 transients are also
recorded for the 4 s\(^{-1}\) rate window. The 10, 20, 50, etc. rate windows can all be calculated from
these transients. In the end, this reads average the 2 s\(^{-1}\) rate window from 3 transients, the 4
s\(^{-1}\) rate window from 5 transients (2+3), the 10 s\(^{-1}\) from 5 transients, the 20 s\(^{-1}\) rate window
from 15 transients (10+2+3), and the 200-6000 s\(^{-1}\) rate windows from 50 transients. These
numbers were empirically determined to give approximately the same signal-to-noise ratio
for the different rate windows. The high rate windows do not have specific averaging because
the measurement time becomes much less than the fill pulse time, so it is almost the same time to run the longer transients. This is performed to yield further averaging of the longer rate windows. This is by far the most complicated part of the program, but thankfully does not need to be adjusted often. When the DLTS experiment is running, the rate windows selected in desired e's are plotted in the graph on the left where the x-axis is temperature and the y-axis is $C(t_2) - C(t_1)$ with units of pF.

The more common controls that need adjusting are in red. The ring control sets whether the run is capacitance, constant drain current, etc. The $V_{rev}$ (V), $V_{fill}$ (V), $t_{fill}$ (ms), $T_{stop}$ (K), and Full Scale (pF) do exactly as their name implies. The duty cycle is typically set to 50%, but the AFG310's cannot be triggered again until the whole waveform has been played so for a 10 ms fill pulse a measurement time less than 10 ms will result in the next fill pulse not being triggered. The program will timeout waiting for the trigger and increment the numTimeouts indicator. Normally, this should be zero, so if it is slowly increasing the likely culprit is the duty cycle, which would need to be increased. The heating rate describes how much to increase the temperature after each set of transients is recorded at a given temperature and the initial capacitance is the capacitance used when the capacitance dialog box pops up to set the capacitance offset.

During a test, the capacitance would normally be automatically adjusted to keep it in range of the capacitance meter, but if for some reason one desires to override this behavior, the Manual Offset can be activated. To write a new capacitance to the offset box, the Write Cap can be used with the Cap to Write control. This takes several seconds to process so just press the button and wait for it to be accepted and pop back out. The experiment can also be paused using the Pause button, which allows the contacts to be reseated or other necessary maintenance.
The last remaining control related to capacitance DLTS is the Update Graph button. If selected, this plots the longest transient in the graph on the right. Unfortunately, Labview has a hard time plotting all of the points on the graph (typically ~2 million). Therefore, if this option is used it should quickly be turned off and one should also right click on the graph and clear the graph to allow Labview to work more efficiently.

As mentioned, the DLTS can also be used with the new constant drain current DLTS (CID-DLOS) experiments. In this case, the values of the PID, Meas, Fill, CID specific, Measurement, DLTS Type, and Voltage Control indicators are all read from the FPGA and set via the FPGA initialization program. This is necessary because the proportion-integral (PI) settings must be tuned beforehand. After changing the Ring setting to the appropriate measurement, several settings that are no longer used are disabled and greyed out. These include the fill pulse, quiescent bias, Boonton full scale, and data acquisition rate (the FPGA operates at a fixed 1 MHz). Other than this, the operation is basically identical to the capacitance DLTS operating procedure described previously.

### B.4.3 DLTS analysis

The Ea Calculator program is a helper program that makes quick analysis of DLTS spectra quite painless. The program can be operated in two modes: load the boxcar data and extract the peaks directly from each peak or entering the previously extracted temperature/emission rate pairs to automatically linear fit the data and extract the trap energy and cross section. Operation of the program is quite easy. First, the data can be loaded on the left side of the screen in Fig. B.13 using the Load DLTS button. Individual DLTS curves can be selected with the Show One DLTS button and the DLTS to display controls. Zooming is available using the standard Labview zooming functionality. The color coordinated cursors should be moved to
Figure B.13: DLTS analysis program.
the peak of the trap under study and the cursors can be found by right clicking on the cursor legend to move the cursor to the center of the screen if it is off-screen. Once all the peaks are defined with the cursors, Update Arrhenius plot can be pressed to calculate the Arrhenius data on the right side of the screen.

On the right side, the program automatically fits the data with a linear regression to all of the points displayed. If some rate windows were not used in the fit, they can be deleted by selecting them in the list below the graph and clicking remove. Points can also be added or modified using the emission rate and Tpeak controls and choosing the appropriate button (Update or Add Point). Once one is happy with the desired fit, Done can be selected to save the rate window, peak temperature, $1/kT$, and $\ln(T^2)$ data for each point. The fit information, trap energy, effective mass, and cross section are also saved in the header in a format easily imported by Origin graphing software. The effective mass control in the fit information should not be neglected as this value impacts the extracted cross-section.

It is important to note that this analysis is only effective for traps that are not overlapping as this case requires multi-peak fitting to extract the true peak temperatures. This has not typically been necessary for $n$-GaN, but several cases have required use of PeakFit to extract multiple overlapping peaks.

### B.5 Capacitance-temperature

The purpose of C-T is to determine the capacitance $C_0$ at the DLTS peak temperatures. This scan must be performed at the same bias as the DLTS experiment, so the values can be used in the $2N_D \Delta C/C_0$ calculation of trap concentration. There is no bias pulsing of the sample in this experiment and the only other requirement is that the experiment be performed slow enough that it remains in thermal equilibrium.
B.5.1 C-T circuit

The C-T circuit connections require only the Boonton capacitance meter and digital capacitance offset box to be connected, which is shown in Fig. B.14. The voltage applied to the sample is from the internal Boonton supply.

B.5.2 C-T program

The C-T program plots the measured capacitance and conductance as a function of temperature on the graph in Fig. B.15. The C-T program is a carryover from the older stages, so the T_heat parameter has to be specifically set for the new LakeShore TTP4 stages to a degree below the starting temperature of the measurement. This temperature determines when the heater is used control the stage temperature, but with the new stages the temperature is controlled throughout the experiment. In this case, the loop time not heating (s) control is not used. The loop time heating (s) control determines how quickly the temperature is incremented and the T_incr (K) control determines how much the temperature is incremented. The T stop (K) is simply the desired temperature to finish the measurement at.
The full scale ring control should be set to a value where the capacitance is still accurately read (i.e. the conductance is not too high) and where the capacitance will not be automatically offset too many times throughout the run to keep the capacitance in range of the capacitance meter. Finally, the $V_{\text{rev}}$ (V) control determines the voltage applied to the sample, so this should be set to the save voltage as the quiescent bias of the DLTS experiments. Once all these parameters are entered and the sample is at the desired starting temperature, the green go button can be pressed to begin the experiment. The program asks for a sample description and filename and pops up a dialog box to allow the initial capacitance to be set. Once this data is entered, the experiment begins. Some information is presented below the graph indicating the current status of the measurement and sample.

Figure B.15: Capacitance-temperature program screenshot.
B.6 Deep level optical spectroscopy

B.6.1 DLOS setup

The DLOS circuit is the nearly the same as the DLTS circuit, but is shown here in Fig. B.16 with the slight modifications. The shutter opening triggers the start of acquisition for the data acquisition card. As in the DLTS, the oscilloscope is only used for the user to observe the transients as they occur and has no other purpose.

A schematic view of the DLOS optical setup is shown in Fig. B.17. The experiment can be run with a quartz-tungsten-halogen (QTH) or xenon lamp depending on the wavelength of interest. Each monochromator has optimized filters and gratings for the usable range of each lamp. The lamps are high power and generate a significant amount of heat that must be properly vented. Even a few minutes after the lamp is turned off, the ventilation should
be maintained to allow for proper cooling of the bulb and housing. The Xe lamp also uses a water filter to cut out the infrared wavelengths. If the cooling water to the water filter is not on, the water will boil out, which can allow the optics after the filter to be subject to additional heating resulting in discoloration or loss of reflectivity for mirrors. The water level should be periodically check to ensure proper filtering.

**B.6.2 DLOS program**

DLOS measurements take a long time because the run must be setup, the sample allowed to reach equilibrium in the dark, and the experiment run. Much like the Cl\textsubscript{2}-DLOS, the DLOS is broken into 2 parts: initialization and experiment. The initialization program allows the equipment to be setup for the upcoming experiment and is shown in Fig. B.18. Once the sample is in contact, the appropriate lamp can be turned on with the monochromator set to a visible wavelength (2.0 eV typically) to allow the optical beam to be aligned to the sample. The shutter buttons, mirror, and write energy buttons all facilitate alignment of the
Figure B.18: Initialization.vi screenshot.
optical beam. Once the beam is focused on the sample, the capacitance meter and function generator can be set for the given measurement. The biasing is usually chosen to match other experiments but the choice of full range of the capacitance meter is chosen based on the sample conductance and the change in capacitance from dark to bandgap illumination. This often requires a short test experiment to determine the change in capacitance, the range of transient emission rates, and the time needed to allow any thermal transient to extinguish before shining light on the sample. These short tests can be run with the DLOS program as it acquire the data and display it on the screen very effectively.

Once the optics are aligned and the test parameters chosen, the sample needs to be left in the dark to allow it to reach equilibrium (deep states to fill) for 12 hours or more. This is critical for the QTH lamp where the changes in capacitance are typically quite small and any remaining transient from trying to reach equilibrium will skew the results. A similar problem can also occur if the sample sits in the dark without pulsing, then the run begins and sample begins seeing a periodic change in voltage. The small change from no pulsing to pulsing can lead to a small transient. This can now be avoided by running the DLOS program in the Hold mode so that the sample is pulsed under the measurement conditions while reaching equilibrium in the dark so the sample can reach complete equilibrium and avoid any initial non-optically induced transients and slope in the SSPC spectrum. Even if the hold mode is not used, it is important that the quiescent bias is applied to the sample during the wait time so that the deep level equilibriate for the measurement bias.

The DLOS program is shown in Fig. B.19. As with the DLTS program, the DLOS program can be used for a myriad of experimental variations including capacitance, constant capacitance, constant drain current, and constant conductance measurements. The first thing to notice is the cheat sheet of common settings in the upper left corner. The screen also has 3
Figure B.19: DLOS program screenshot.
graphs: upper left is no longer used but there is the option to use it if the 2 oscilloscope option is chosen in Source, the lower left graph displays the transient of the last completed measurement, and the right graph displays an SSPC spectrum based on the variation in steady-state capacitance measured so far. This SSPC is automatically generated but not saved.

Much like the DLTS, the CI-DLOS experiments are quite similar to the capacitance-mode DLOS measurements, but the former require additional setup using the FPGA Initialization program to tune the PID settings and set all the voltages in the fill pulse and measurement states. With this briefly discussed, the rest of the setup is independent of measurement mode. Some further explanation of the FPGA set up is described in Chapter 6.

The DLOS fill pulse can be optical, electrical, or nothing at all. With no fill pulse, the \( \frac{dC}{dt} \) calculation is probably less accurate because the calculation of optical cross section depends on the level being completely filled at the onset of the shutter opening, which would not be true and alter the results. The optical stimulation can be used to empty a particular level and refill it later during the measurement, which is referred to as double beam DLOS. This is discussed in Adrian Hierro’s dissertation.[12] Here, the focus is on electrical pulsing.

The DLOS also can be run over different wavelength ranges 0.5 – 2.0 eV for the QTH lamp and 1.2 – 6 eV for the Xe lamp, which are run through different monochromators. The lamp to be used is selected via the ring control below DLOS Type. The settings for the wavelength range (Initial Energy, Energy Increment, and Final Energy) should be based on the available range of the given lamp/monochromator combination. The measuring time and delay after filling pulse, determined previously by the test run, are entered as well. The Filling Pulse Time control is only used for the constant drain current/conductance measurements where the fill pulse time cannot be read from the arbitrary function generator. The times used in the FPGA Initialization are for the tuning and not for the actual test conditions. The Minimum
Scale ring control is nearly useless as the DLOS program no longer uses the oscilloscope for deciding when to adjust the offset capacitance, and the program needed to make sure the DLOS transient fit in the oscilloscope range. Finally, there is a Bias control that allows the bias to be turned off, left on, or turned off for 2 hr. and then turned back on.

When the DLOS program is started, it normally asks where to save the data, sample information, and base filename then starts acquiring data. This process can be interrupted through the use of the Hold button that is above the Common Parameters box. If the Hold button is depressed when the program is run, the emergency shutter will not be opened, the photon energy will not be incremented, and the transients will not be saved. However, the fill pulse and measurement bias will be applied to sample following the timing of the measurement repetitively until the Hold button is turned off. At that point, it will finish the current biasing sequence then begin the actual measurement. While the program is holding, the SSPC graph on the right side continues to plot the steady-state photocapacitance as a function of time instead of photon energy. This allows one to determine when the sample has reached equilibrium. The time and photon energy axes are separate so when switching between hold and normal operation two different sets of data are plotted.

### B.6.3 DLOS analysis

As discussed in Chapter 4, the transients from the DLOS experiments can be analyzed in several ways. To calculate the cross section, the slope of the transient $dC/dt$ needs to be calculated at each wavelength and for the steady state photocapacitance one needs to average the capacitance at the end of each transient. A program was written to facilitate this analysis and a screen shot is shown in Fig. B.20. In the program, 5 analysis options exist though there are really just three categories. The Co and Css analyses both average the capacitance (or
Figure B.20: Screenshot of the DLOS analysis program.
$V_{GS}, V_{DS}, G$, etc. DLOS signals) from the value of Start to Stop for each wavelength. The dC/dt analysis performs a linear regression for each wavelength from Start to Stop. Finally, the Exp Fit and Exp Fit(2) analyses use the $C_{ss}, A_0, e_0, A_1, e_1$ to fit an exponential equation to each transient. This method allows the emission rate to be calculated from a fit instead of $dC/dt$. This is subject to the quality of the fit at each wavelength so the fit needs to be examined for each transient to ensure proper analysis. The double exponential, Exp Fit(2), assumes the transient consists of two traps and allows 2 emission rates to be extracted. Because the transients evolve quite significantly in terms of $\Delta C$ or $A_1/A_2$ in this case, there is also an option to use the fit from the previous exponential or use the same initial guess for each transient. In the end, the fitting is not extremely useful because of the uncertainty of the fitting and is not often used.

The program can perform multiple analyses simultaneously as evidenced by the Selected Analysis listbox. To add an analysis, the desired range is entered into the Start and Stop edit boxes and the desired analysis is selected. The order of the analyses listed in the Selected Analysis listbox is how the data is written to the file so the order can be manipulated with the Move Up and Move Down buttons. They can also be removed with the Remove button. Once everything is in order, the Done button can be selected. This asks for a place to save the information and if any $dC/dt$ analysis is requested asks for the photon flux to calculated $dC/dt$/photon flux, which is proportional to the optical cross section. If the photon flux file dialog box is canceled, the program still runs but the calculation involving the photon flux is not performed.

The best way to choose the appropriate times for the $dC/dt$ is to look at some initial $dC/dt$/photon flux spectra and the SSPC to determine where the traps most likely are taking into account filter and grating changes. Then, by looking at transients below and above
the energy of interest, the time scale with the largest change in $dC/dt$ can be empirically determined with a selection of potential $dC/dt$ times. When the sharpest (least spread out) cross section is obtained, it can be fit to the Lucovsky or Chantre models presented in Chapter 4 to determine the precise deep level energy.

**B.7 Internal photoemission**

**B.7.1 IPE theory**

Given a Schottky diode with a thin Schottky metal, monochromatic light incident on the metal can be absorbed creating an electron-hole pair. If the energy is sufficient to surmount the Schottky barrier, there is a probability that the electron will travel towards the interface and a probability that the electron will make it over the barrier. If it does make it over the barrier, the electron will be swept away by the built-in electric field in the semiconductor. The separation of hole and electron creates the photocurrent that is measured, which is demonstrated in Fig. B.21.

![Figure B.21: Band diagram showing the IPE process.](image-url)
This method was originally developed by Fowler in 1931 and is widely considered the most accurate way to determine barrier height. The measurement is straightforward because all you measure is the photocurrent and photon flux each wavelength. The photoemission yield \( Y \) is defined as\[13\]

\[
Y = \frac{\text{Number of electrons/sec/cm}^2}{\text{Number of photons/sec/cm}^2}
\] (B.30)

When the incident photon energy \( h\nu \) is several kT greater than the barrier height \( \Phi_{b0} \) the photoemission yield is proportional to\[13\]

\[
Y = C (h\nu - \Phi_{b0})^2
\] (B.31)

By plotting the incident photon energy versus the square root of the photoemission yield, a straight line should be achieved and the \( x \)-intercept where \( h\nu \) equals \( \Phi_{b0} \) signifies the barrier height.

An example of IPE results for an \( n \)-type AlGaN Schottky barrier are shown in Fig. B.22. The IPE is basically always performed at zero bias, but small biases can be applied to zero the current more accurately. This is typically in the \( \mu V \) range.
Figure B.22: IPE results of an \textit{n}-type AlGaN sample with semitransparent 80 Å Ni Schottky contact.

**B.7.2 IPE setup**

The optical and electrical setups are shown in Fig. B.23 and B.24, respectively.
Figure B.23: Schematic representation of optical illumination and circuit used for IPE measurements.

Figure B.24: Electrical circuit for IPE measurements.
B.7.3 IPE program

The IPE program is relatively simply as it only increments the wavelength and records the current. The IPE Labview program is shown in Fig. B.25. This measurement can be performed with the Xe or QTH lamps, so this must be chosen in the Common Parameters area. The Initial Energy, Energy Increment, and Final Energy must match the available range for the chosen lamp/monochromator combination as discussed in the DLOS section. The measurement is best performed with an electrometer, but if a large photocurrent is obtained it may be possible to use a picoammeter. These options are available under the Source control. The measurement may take a few seconds to reach an equilibrium current, so the Wait time after shutter control can be used to delay the measurement after the shutter opens. The # to avg. control allows for several measurements of the current at each wavelength. Typically 16 or 32 averages works well.

The light beam needs to be aligned to the sample, so the alignment procedure discussed in the DLOS section should be used. For the photoyield calculation, the photon flux is needed as well. For best results, the photon flux should be measured immediately after the measurement.

As with the other measurements, the measurement is initiated with the green go button. This prompts the user for filename and sample information then proceeds to measure the photocurrent. The file extension will be .pc for these measurements.
Figure B.25: Screenshot of the IPE program.
B.8 Photon flux measurement

The photon flux is needed for several measurements including DLOS and IPE. The flux is measured with a thermopile because it has a very flat response from 200 nm to 5 μm. The thermopile is typically put on top of the stage and the last lens is adjusted to focus the beam onto the 0.081 inch diameter thermopile insert. This results in the correct shape of the photon flux, but not a quantitative value of the flux. The shape of the spectrum is typically what is important and the data can be displayed in arbitrary units. Additionally, it is not known how much of the light is absorbed in the metal or reflected at an interface. Therefore, determining a quantitative value with the thermopile is not meaningful anyways. The fast shutter is now critical to the flux measurement. The program measures the thermopile power with the shutter closed and with the shutter open. The thermopile drifts slightly as the measurement progresses because the thermopile's temperature will typically rise slightly. This drift can strongly impact the calculated photon flux if not properly accounted for.

The photon flux $\Phi_{flux}$ is calculated using the following equation

$$\Phi_{flux} = \frac{P}{qAh\nu}$$  \hspace{1cm} (B.32)

where $P$ is the measured thermopile power, $A$ is the area of the aperture, and $h\nu$ is the photon energy.

B.8.1 Photon flux circuit

To perform photon flux, the thermopile is put in the optical beam under the shutter and on top of the stage typically. The focusing lens is adjusted until the light is focused on the thermopile sensor to achieve nearly the same photon flux as the beam focused on the sample plane. The connections needed for this measurement are shown in Fig. B.26.
B.8.2 Photon flux program

The photon flux program, shown in Fig. B.27, needs several parameters to accurately calculate the flux. The thermopile cannot be directly queried to an analog output proportional to the detected power is output to an oscilloscope where it can be measured and averaged. The thermopile is quite sensitive to temperature variation so to compensate for this the average power is recorded with the shutter closed and opened. The difference in power is an accurate representation of the number of photons hitting the surface. The thermopile also has a large active area but the spot size is much smaller than the active area. To calculate power density instead of total power, a small aperture is place in front of the thermopile with a 0.081 in diameter opening. The thermopile needs to be manually configured to the 2 mW scale and manually offset to 0 though the offsetting is not very necessary. On the other hand, the oscilloscope is automatically configured by the program.
Figure B.27: Photon flux program screenshot.
The graph plots the photon flux in log scale versus wavelength. Most of the settings in this screen are common to other programs but this program does have the unique ability to override the normally automatic filter and grating selection. There is no reason to override the filter selection as this needs to be chosen based on the filter curves and not any measurement, but the gratings can be separately chosen to map out the grating crossover points. The thermopile range should match the range selected on the thermopile itself while the Scope Range should be set to the lowest value. The oscilloscope range is selected such that the analog output of the thermopile fits onto the oscilloscope screen. A file of the photon flux data versus incident photon flux energy is save to the location of the users’ choosing with a .flux extension.

B.9 References


Appendix C

Impact of gate-source access resistance variation on CI_D-DLOS measurements

The derivation in Chapter 6 did not include the gate-source access resistance $R_S$ in the derivation of the CI_D-DLOS/DLTS theories. In gate-control, the equation for the drain current is given as:

$$I_{DS,sat} = \frac{2(V_{GS} - V_T)^2 W\varepsilon_2\mu}{L_d} \left[1 + \sqrt{1 + \xi_d}\right]^{-2}$$  \hspace{1cm} (C.1)

If $R_S$ is explicitly added and the equation rearranged, the following equation is possible:

$$\sqrt{\frac{L_d I_{DS,sat}}{2 W \varepsilon_2 \mu}} \left[1 + \sqrt{1 + \xi_d}\right] = \frac{V_{GS}(t) - I_{DS,sat} R_S(t) - V_T(t)}{\text{Term1} - \text{Term2} + \text{Term3}}$$  \hspace{1cm} (C.2)

The left side is independent of time assuming self-heating is not sufficient to cause a large change in mobility. The terms on the right side all have time dependence. Ideally, any change in Term 1 would directly equal an change in Term 3. In order of this to be true, the change in Term 2 needs to be small. The derivative with respect to time becomes:

$$\frac{d}{dt} V_{GS}(t) = \frac{d}{dt} V_T(t) + I_{DS,sat} \frac{d}{dt} R_S(t)$$  \hspace{1cm} (C.3)

where the drain current is constant because of the experimental conditions. For the equation to simplify down to the assumption in Chapter 6, which is:

$$\frac{d}{dt} V_{GS}(t) = \frac{d}{dt} V_T(t)$$  \hspace{1cm} (C.4)
either $I_{DS,sat}$ must be small or $dR_S/dt$ must be small. The magnitude of $dR_S/dt$ is a property of the HEMT and cannot be varied so the only choice is small $I_{DS,sat}$. In gate control, low drain current avoids the effects of $R_S$ and isolates the observed variation of $V_{GS}$ to $V_T$ and therefore to defects under the gate.

Interestingly, the differences between $\Delta V_{GS}$ at different drain currents might be used to characterize the $\Delta R_S$. A measurement at low drain current would measure $\Delta V_T$. Another measurement at high drain current would have effects of both $V_T$ and $R_S$, so the $V_T$ effects could in theory be subtracted from the high drain current measurement to isolate the effects of $R_S$ as well. In fact, a series of measurements at different drain currents could be performed to look for linear increases in $V_{GS}$ that would be indicative of $R_S$ effects. Using this method, it may be possible to separate the effects of gate-source, gate-drain, and under the gate and determine trap energies and concentrations in each area.

To determine the impact of $R_S$ on the drain-controlled measurements, both access resistances need to be added to the triode drain current equation given as[1]

$$I_{DS,triode} = \frac{1}{1 + \frac{\mu V_{DS}}{V_{sat} L}} \left\{ \frac{W \varepsilon_2 \mu}{L d} \left[ (V_{GS} - V_T) V_{DS} - \frac{V_{DS}^2}{2} \right] \right\}$$  \hspace{1cm} (C.5)

This equation can be rearranged separating the time dependent terms to yield

$$I_{DS,triode} L d \left[ 1 + \frac{\mu V_{DS}}{V_{sat} L} \right] =$$

$$W \varepsilon_2 \mu$$

$$\left[ V_{GS} - I_{DS,triode} (R_S + R_D) - V_T \right] \left[ V_{DS} - I_{DS,triode} (R_S + R_D) \right]$$

$$- \frac{1}{2} \left[ V_{DS} - I_{DS,triode} (R_S + R_D) \right]^2$$  \hspace{1cm} (C.6)

Ideally, the time derivative of this equation would simplify to

$$\frac{d}{dt} V_{DS}(t) = I_{DS,triode} \frac{d}{dt} \left[ R_S(t) + R_D(t) \right]$$  \hspace{1cm} (C.7)

356
then all changes in the drain voltage would correspond to changes in the access resistances.

The derivative of Eq. C.6 becomes

$$0 = \frac{d}{dt}\left(\left[V_{GS} - I_{DS,triode}R_S(t) - V_T(t)\right]\left[V_{DS}(t) - I_{DS,triode}[R_S(t) + R_D(t)]\right]\right) - \frac{1}{2} \frac{d}{dt}\left\{V_{DS}(t) - I_{DS,triode}[R_S(t) + R_D(t)]\right\}^2$$

(C.8)

If $V_{GS} - V_T - I_{DS,triode}[R_S + R_D]$ is assumed to be constant (i.e. the change of $V_T(t)$ and $I_{DS,triode}[R_S(t) + R_D(t)]$ is small compared with the magnitude of $V_{GS} - V_T$), the equation can be simplified to

$$0 = \left[V_{GS} - I_{DS,triode}R_S - V_T\right] \frac{d}{dt}\left[V_{DS}(t) - I_{DS,triode}[R_S(t) + R_D(t)]\right] - \frac{1}{2} \frac{d}{dt}\left\{V_{DS}(t) - I_{DS,triode}[R_S(t) + R_D(t)]\right\}^2$$

(C.9)

This can be expanded taking the time derivative of the last term

$$\left[V_{GS} - I_{DS,triode}R_S - V_T\right] \frac{d}{dt}V_{DS}(t) - I_{DS,triode} \frac{d}{dt}[R_S(t) + R_D(t)] \equiv \left\{V_{DS}(t) - I_{DS,triode}[R_S(t) + R_D(t)]\right\} \frac{d}{dt}V_{DS}(t) - I_{DS,triode} \frac{d}{dt}[R_S(t) + R_D(t)]$$

(C.10)

Because

$$V_{GS} - I_{DS,triode}R_S - V_T \neq V_{DS} - I_{DS,triode}[R_S + R_D]$$

(C.11)

the only solution is

$$\frac{d}{dt}V_{DS}(t) \equiv I_{DS,triode} \frac{d}{dt}[R_S(t) + R_D(t)]$$

(C.12)

which is the desired relationship (i.e. Eq. (C.7)) such that the change in $V_{DS}$ is proportional only to the change in $(R_S + R_D)$. This equation is derived with the assumption that

$$V_{GS} - I_{DS,triode}R_{GS,0} - V_{T0} >> I_{DS,triode}\Delta R_S - \Delta V_T$$

(C.13)

in developing Eq. (C.9) where $R_{GS,0}$ and $V_{T0}$ are the average values and $\Delta R_S$ and $\Delta V_T$ is the magnitude of variation of the respective terms. As with the gate control, the trap induced changes in $\Delta R_S$ and $\Delta V_T$ are properties of the HEMT and cannot be adjusted. The use a large gate voltage ($V_{GS} \geq 0$ V) can mitigate the problem in most cases.
This result modifies the previous derivation in Section 6.1.2. There the measured change in resistance as simply $R_D$ because $R_S$ was ignored as it was assumed to be much smaller than $R_D$, which is probably true in most cases. Now, in the case where $R_S$ is comparable or greater than $R_D$, it can be seen that measured change in resistance is actually $\Delta R_D + \Delta R_S$.

C.1 References


360


370


376


377


