PARISH, CHAD M. Electron and ion-beam characterization of nitride semiconductor devices. (Under the direction of Professor P. E. Russell).

Gallium nitride (GaN) and its alloys are used to manufacture green-to-ultraviolet range light emitting diodes (LEDs) for the solid-state lighting industry. However, heteroepitaxial growth on substrates such as 6H-SiC or $\alpha$-Al$_2$O$_3$ results in LEDs with large densities of crystal defects. Cathodoluminescence (CL) and electron-beam-induced current (EBIC) are SEM-based techniques that are used to probe the optoelectronic behavior of GaN LEDs and defects at the sub-micrometer scale.

This work examines the optoelectronic properties of defects in GaN-based LED devices. First, computer modeling of the polarization fields in quantum wells was performed, and quantitative predictions of cathodoluminescence peak shifts during electron injection, under varying conditions of electrical bias, were made. Results indicate that both polarization and InN-GaN immiscibility strongly influence the device properties, and that polarization fields of $\approx1.33\pm0.15$ MV/cm are present in the quantum wells. Experimental conditions and mathematical treatments for accurate cross-sectional EBIC quantification of the minority carrier diffusion length in GaN LEDs were developed and refined, which allowed quantification of hole and electron diffusion lengths of $L_h\approx92\pm15$ and $L_e\approx42\pm6$ nm, respectively; these short values of $L$ help explain the anomalously high quantum efficiencies of GaN layers despite their high dislocation densities. Combined CL and EBIC techniques were developed for the study of defect populations in GaN LEDs, and results show that large densities of threading defects are present in these devices. Additionally, the effects of focused-ion-beam (FIB) milling as a cross-sectional sample preparation technique for GaN were studied by CL and EBIC. It was found that preparation of GaN devices for CL or EBIC microscopy by FIB causes significant damage and modifies the CL and EBIC response of the devices. By using SEM-CL/EBIC to pinpoint defects in LED devices, site-specific FIB microsampling has been used to prepare samples of defected areas for transmission electron microscopy (TEM).
Analyses of these samples have shown how the identity of crystal defects within the devices directly relates to the optoelectronic behavior observed in CL and EBIC. Densities of defects measured in CL or EBIC correspond with dislocation densities measured via TEM; this indicates that the dislocations are optoelectronically active and influence CL or EBIC behavior. These optoelectronic measurements, in conjunction with SEM and TEM microscopy, indicate a conjunction of high defect density and short diffusion length contribute to anomalously high light emission efficiency, and that the techniques developed and refined in this work can be used to study device performance and optimization.
Electron and ion-beam characterization of 
nitride semiconductor devices 

by 
Chad Michael Parish 

A dissertation submitted to the Graduate Faculty of 
North Carolina State University 
In partial fulfillment of the 
Requirements for the degree of 
Doctor of Philosophy 

Material Science and Engineering 

Raleigh, NC 
2006 

Approved by: 

Prof. P. E. Russell  
(Chair of advisory committee)  

Prof. D. P. Griffis  

Prof. G. J. Duscher  

Prof. M. J. Escuti
DEDICATION

Dedicated to my family, both immediate and extended, who encouraged me through my formative years to pursue my dreams and education.

Dedicated to my wife, Genie, who helped me through the dark days and long nights I spent working towards this day.

With love and gratitude to you all; this dissertation is a testament to your love and support. It would not exist without all of you.
BIOGRAPHY

Chad Parish was born in Las Vegas, Nevada, in 1977, and lived in almost one dozen states while growing up. He graduated from Charlotte Catholic High School, Charlotte, NC, in 1996, and earned a Bachelors of Science, *summa cum laude*, in Material Science and Engineering from North Carolina State University in 2000. He then earned a Masters of Science in Material Science and Engineering from the University of Pittsburgh in 2003, for work performed on phase transformations in high-strength sheet steel. He then earned a Doctorate of Philosophy in Material Science and Engineering from North Carolina State University in 2006 for work performed in the electron microscopy of III-nitride semiconductors, and is now a post-doctoral researcher at the Department of Energy's Sandia National Laboratory, Materials Characterization Department, Electron Optics Group, in Albuquerque, New Mexico.
ACKNOWLEDGMENTS

No work of this magnitude is completed in isolation; I have a large number of people to thank:

- Cree, Inc., and Hitachi High Technologies America for financial support.
- Professor Russell, for taking me into his group, and for guidance and advice over the years.
- My Ph.D. committee members: Prof. Griffis, for teaching me the art of electronics; Prof. Duscher, for teaching me the art of TEM; and Prof. Escuti, for teaching me the art of optics.
- The staff at AIF -- Dale Batchelor, Fred Stevie, Roberto Garcia, and Chuck Mooney -- for helping me solve more problems than I could hope to count here.
- Kristin Bunker, for getting me started in this project on the right track.
- All the AIF graduate students -- Mike Salmon, Jitty Gu, Anthony Garetto, Curt Progl, Chris Penley, and Wingo Wong -- for friendship and support.
- Curt Progl, especially, for working so closely with me on this project. With me leaving, you're now the short-timer in the lab; good luck!
- The NCSU libraries and librarians, who are mostly responsible for my rather excessive reference list.
- Edna Deas, without whom no Materials student could ever hope to graduate.

Lastly, but especially, I acknowledge my wife, Genie, for seeing me through the last three and a half years of toil, labor, and frustration with nothing but love, support, and humor.
# Table of Contents

List of Figures ....................................................................................................................... ix

List of Tables ......................................................................................................................... xviii

1. Introduction .......................................................................................................................... 1

2. Literature Review ................................................................................................................. 3

  2.1 Overview of gallium nitride devices ............................................................................. 3
    2.1.1 Gallium nitride properties and growth ................................................................. 3
    2.1.2 Gallium nitride devices ......................................................................................... 8

  2.2 Dislocations .................................................................................................................... 10
    2.2.1 Overview of dislocations ....................................................................................... 10
    2.2.2 Electronic activity of defects ................................................................................. 11
    2.2.3 Electronic activity of dislocations in GaN: Theory .............................................. 12
    2.2.4 Electronic activity of dislocations in GaN: Experiment ........................................ 14
    2.2.5 Dislocations: summary of literature .................................................................... 29

  2.3 Other III-nitride defects ................................................................................................. 29
    2.3.1 The yellow-band defect ......................................................................................... 29
    2.3.2 Stacking faults .................................................................................................... 31

  2.4 Polarization and segregation effects ............................................................................. 32
    2.4.1 Polarization: theory .............................................................................................. 32
    2.4.2 Segregation: theory .............................................................................................. 41
    2.4.3 Polarization vs. segregation: experiment .............................................................. 45
    2.4.4 Polarization vs. segregation: summary ................................................................. 61

  2.5 Summary of literature .................................................................................................... 62

3. Experimental Techniques ................................................................................................... 63

  3.1 Scanning electron microscopy ....................................................................................... 63
3.2 Beam-injection techniques ................................................................. 67

3.2.1 Beam-injection overview ................................................................. 67
3.2.1.1 Generation of electron-hole pairs .................................................. 67
3.2.1.2 The low-injection condition ............................................................ 77
3.2.1.3 Diffusion and drift ....................................................................... 79
3.2.1.4 Recombination statistics ............................................................... 83

3.2.2 Introduction to electron beam induced current ................................... 84

3.2.3 Mathematical models for cross-sectional EBIC .................................. 89
3.2.3.1 Introduction .................................................................................. 89
3.2.3.2 The exponential model ................................................................. 90
3.2.3.3 Donolato’s technique ................................................................... 91
3.2.3.4 Luke’s technique ......................................................................... 93
3.2.3.5 Bonard and Ganière’s technique ................................................. 95
3.2.3.6 Method of moments .................................................................... 97
3.2.3.7 Other techniques ........................................................................ 98

3.2.4 Theoretical comparison of cross-sectional EBIC models .................. 99
3.2.4.1 The Donolato-Luke methods ....................................................... 99
3.2.4.2 The Bonard & Ganière method .................................................... 104
3.2.4.3 The Ong et al. method ................................................................. 108
3.2.4.4 Theoretical analysis of methods: summary ................................... 109

3.2.5 Cathodoluminescence .................................................................... 109

3.2.6 CL and EBIC spatial resolution issues ............................................. 123

3.3 Transmission electron microscopy ....................................................... 127

3.4 Focused ion beam ............................................................................ 130

3.5 Analytical techniques summary ........................................................ 136

4. Motivation and goals ......................................................................... 137

4.1 Motivation ........................................................................................ 137

4.2 Modeling of the polarization fields in GaN ....................................... 137
4.3 Development of experimental apparatus and techniques ........................................... 138
4.4 Cross-sectional EBIC .................................................................................................. 138
4.5 CL and EBIC analysis of defects ............................................................................... 138
4.6 Analysis of FIB-induced damage ............................................................................... 139

5. Experiments, results, and discussion ........................................................................... 140

5.1 Modeling of polarization fields .................................................................................. 140
  5.1.1 Experimental results .............................................................................................. 140
  5.1.2 Computer model .................................................................................................... 141
  5.1.3 Results of model & comparison of model to experimental data ......................... 159
  5.1.4 Higher-resolution CL injection studies ................................................................. 164
  5.1.5 Discussion of computer model .............................................................................. 168
  5.1.6 Electroluminescence: analytical model and experiment ...................................... 171
  5.1.7 Summary of polarization fields study .................................................................... 175

5.2 Cross-sectional EBIC .................................................................................................. 176
  5.2.1 Sample preparation and experimental techniques .................................................. 176
  5.2.2 Determination of input parameters for Equation (3-25) ....................................... 187
  5.2.3 Cross-sectional EBIC: results ................................................................................ 192
    5.2.3.1 Exponential method ....................................................................................... 197
    5.2.3.2 Donolato-Luke methods ............................................................................... 199
    5.2.3.3 The Bonard and Ganière method ................................................................... 206
    5.2.3.4 Method of moments ...................................................................................... 215
    5.2.3.5 Ong et al. method ......................................................................................... 216
    5.2.3.6 Comparison of techniques ............................................................................ 220
    5.2.3.7 X-EBIC with thermionic SEM ...................................................................... 224
  5.2.4 Cross-sectional EBIC: conclusions ...................................................................... 227

5.3 Cathodoluminescence and plan-view EBIC ................................................................. 229
  5.3.1 Sample preparation and experimental techniques .................................................. 229
5.3.2 CL and Plan-view EBIC: results and discussion ........................................ 240
5.3.2.1 General features observed in LEDs ................................................ 240
5.3.2.2 Spectral data from LEDs and wafers............................................. 244
5.3.2.3 Quantitative analysis of emission mechanisms ............................... 253
5.3.3 CL and EBIC: Conclusions ................................................................. 259

5.4 CL and EBIC of samples cross-sectioned by FIB ...................................... 260
5.4.1 CL of FIB polished areas ................................................................. 260
5.4.2 X-EBIC of FIB polished areas ............................................................ 272
5.4.3 Interpretation of CL and EBIC data of FIBed areas ............................. 281

5.5 TEM analysis of defects .......................................................................... 284

6. Discussion .................................................................................................. 290

7. Conclusions .............................................................................................. 295

Appendices ..................................................................................................... 316

A.2.1 Appendix 2.1: MathCAD 12 code for Bonard and Ganiere curve generation .................................................................................................................. 320
A.2.2 Appendix 2.2: Maple 9.5 code for Bonard and Ganiere curve generation 324
A.3 Appendix 3: C++ code for polarization fields model .............................. 329
A.4 Appendix 4: Imaging circuit description .................................................. 370
A.6 Appendix 6: MathCAD 12 code for Bonard and Ganiere curve fitting .... 375
A.7 Appendix 7: Maple 9.5 code for Method of Moments solver ................. 381
List of Figures

Figure 1: Wurtzite (A) and zincblende (B) crystal structures. Dark gray is Ga, light gray N..............................................................3

Figure 2: Comparison of III-nitride bandgaps and lattice parameters. Data from Table I and DenBaars....................................................5

Figure 3: Misorientation of subgrains and formation of defect-heavy boundary region. The c axis is vertical...........................................6

Figure 4: Illustration of faceted growth behavior; adapted from Hiramatsu et al......7

Figure 5: Schematic of QW structure.................................................................9

Figure 6: Illustration of defect states within a semiconductor. E represents energy, CB the conduction band, and VB the valence band..............11

Figure 7: Effect of temperature and crystallite size on PL & CL spectra. (a) GaN sample with small crystallites, 2K; (b) Large crystallites, 2K; (c) Small crystallites, 300K.................................................................16

Figure 8: CL images of large crystallites. (left) SEM image; (center) Band-edge CL; (right) Yellow-band CL.................................................17

Figure 9: Correlation of AFM and SEM-CL data regarding dislocations.............19

Figure 10: (S)TEM-CL image of GaN-based LED, showing dislocation activities. (A) CL spectrum of thin plan-view TEM specimen of a MQW GaN-based LED. (B) TEM image showing dislocations. (C) Free-exciton (FX) CL image showing dislocation quenching. (D) Yellow-band (YB) CL image, showing no dislocation contrast.........................................................21

Figure 11: Three possible transitions observable in VEELS................................23

Figure 12: Possible model of the YB emission related states................................30

Figure 13: CL and TEM data showing stacking fault luminescence. (A) TEM image (B)-(E) CL data showing emission at the marked wavelength.........................................................................................32

Figure 14: Spontaneous polarization vs. lattice parameter for III-nitrides. Data from Ref. ........................................................................36

Figure 15: Schematic illustration of the Quantum-Confined Stark Effect; E₁>E₂.....38

Figure 16: Wavefunctions in flatband and QCSE quantum wells........................39
Figure 38: Schematic of the four main EBIC modes. $I_b$ denotes the electron beam current; $I_{cc}$ is the EBIC current.

Figure 39: Calculated EBIC curves for $R=200\text{nm}$, $L=1\mu\text{m}$. The $s$ values are labeled on the curves. Calculated using the Donolato-Luke equation, with $d_b=0\text{nm}$.  

Figure 40: Comparison of curves for long and short diffusion length, under constant $R=200\text{nm}$ and $d_b=0\text{nm}$. Dashed curves are for $L=50\text{nm}$; solid lines are for $L=1\mu\text{m}$. Calculated using the Donolato-Luke equation.

Figure 41: Luke-technique effective diffusion length analysis. Here, $L=50\text{nm}$, $s=10^6\mu\text{m}^{-1}$, $d_b=0\text{nm}$. The labels on the curves are the electron ranges used to calculate the curves, and approximate beam energies corresponding to those ranges.

Figure 42: Effect of $s$ and $L$ on effective diffusion length curves $L_s(x)/L$. Solid lines are for $L=1\mu\text{m}$, and dashed lines $R=50\text{nm}$. $R=200\text{nm}$ and $d_b=0\text{nm}$ for all curves.

Figure 43: Effect of increasing spot size $d_b$ on Donolato-Luke curve generation. Conditions $R=200\text{nm}$, $L=50\text{nm}$, $s=10^6\mu\text{m}^{-1}$.

Figure 44: (A) Bonard and Ganière calculations for EBIC profiles in GaN at 5kV, assuming $1\mu\text{m}$ diffusion length. (B) Similar calculations, for $L=50\text{nm}$.

Figure 45: Bonard and Ganière computations for GaN at 5kV, with $L=50\text{nm}$, $s=10^6\mu\text{m}^{-1}$, and $d_j=50\text{nm}$.

Figure 46: Donolato-Luke data (squares) for $d_b=30\text{nm}$, $s=10^6\mu\text{m}^{-1}$, $L=1\mu\text{m}$, analyzed with the Ong technique (dashed line).

Figure 47: Schematic band diagram of CL processes.

Figure 48: CL emission efficiency as a function of $\eta$.

Figure 49: Schematic CL collection schemes.

Figure 50: Schematic illustration of a monochromator.

Figure 51: Depth distribution of electrons in a GaN LED at typical SEM voltages.

Figure 52: CASINO simulations of diffusive probe broadening at 1keV on GaN.

Figure 53: CASINO simulations of diffusive probe broadening at 10keV on GaN.
Figure 54: CASINO simulations of diffusive probe broadening at 30keV on GaN. 125
Figure 55: Diffused CASINO data from above, with superimposed R_{K\alpha} ranges. 126
Figure 56: TEM schematic. .......................................................... 128
Figure 57: Schematic of TEM diffraction geometry. ................................. 129
Figure 58: H-bar type cross-sectional TEM sample, as prepared by FIB. 134
Figure 59: Illustration of FIB lift-out process. .............................. 135
Figure 60: CL and EL results of Bunker et al. (A) CL peak shift with reverse bias; (B) CL peak shift with SEM \textit{i}_b; (C) EL peak shift with drive voltage. 140
Figure 61: Quantities used for computer model. .................................. 142
Figure 62: Finite mesh scheme for band-edge potentials. ............... 143
Figure 63: Band edges (thick lines) and wavefunctions (dashed lines) for example case. ................................................................. 149
Figure 64: Charge density (dashed line) derived from wavefunctions. 149
Figure 65: Electron-electron derived field \textit{E}_{ee} (dashed line) derived from charge density. .......................................................... 150
Figure 66: Potential \textit{V}_{ee} (dashed line) derived from \textit{E}_{ee} ......................... 150
Figure 67: Illustration of self-consistent iteration to convergence. ........... 152
Figure 68: Illustration of QW density-of-state functions. ...................... 153
Figure 69: Illustration of QW-confinement effect on electron (or hole) energies. 153
Figure 70: Fermi level \textit{E}^F and quasi-Fermi levels \textit{E}^{QF} in a semiconductor. 154
Figure 71: Illustration of number of states computation. ...................... 155
Figure 72: Demonstration of orthogonality rules. .............................. 156
Figure 73: Total QW field as a function of reverse bias. ....................... 158
Figure 74: CL spectrum, model vs. experiment. Experiment conditions: 30keV, 4nA beam, -13.75V reverse bias. Model: \textit{L}_Z=2.5nm, \textit{n}_{2D}=10^8/cm^2, \textit{E}^{POL}=0MV/cm (i.e., flatband) ................................................................. 159
Figure 75: Calculations of \lambda_{PEAK} as functions of \textit{L}_Z and \textit{E}^{POL}. \textit{E}_{gap}=2.28eV, \textit{n}_{2D}=10^8/cm^2. ................................................................. 161
Figure 76: Peak shift with increased injection (screening) density. \textit{E}^{POL}=1MV/cm, \textit{E}_{gap}=2.28eV................................................................. 162
Figure 77: Modeled effect of screening level and $E^{pol}$ on peak wavelength.

$L_z=2.5\,\text{nm},\ E_{gap}=2.28\,\text{eV}$ ................................................................. 162

Figure 78: Experiment vs. model for reverse-bias CL. $L_z=2.5\,\text{nm},\ n_{2D}=10^8/\text{cm}^2,\ E_{gap}=2.28\,\text{eV}$ .................................................................................................. 163

Figure 79: Effect of increasing $n_{2D}$ in reverse bias. $L_z=2.28\,\text{eV},\ E^{pol}=-1.3\,\text{MV/cm}$. 164

Figure 80: Blue LED peak shift with current at 30keV............................................. 165

Figure 81: 30keV, 300pA peaks before and after CL irradiation. ......................... 166

Figure 82: Green LED peak shift at 30keV. (A) Logarithmic and (B) linear scales. 167

Figure 83: (A) CER peak shift vs. $V_a$; (B) electron concentration vs. $Z$ and $V_a$. 170

Figure 84: Blue LED electroluminescence spectra. ................................................. 171

Figure 85: Green LED electroluminescence spectra............................................. 172

Figure 86: Summarized EL peak shift data. ............................................................. 173

Figure 87: Peak shifts vs. calculated $n_{2D}$. $\tau=1\,\text{ns},\ L=250\,\mu\text{m}$.................. 175

Figure 88: (A) Encapsulated LED; (B) Blue LED emitting under bias. ................. 177

Figure 89: (A) FB-2100 focused ion beam; (B) Stages, micromanipulator, and W-injection systems. ................................................................. 179

Figure 90: (A) Sample on header and modified stub; (B) Sample ready for FIB. ... 179

Figure 91: LED (without mechanical polishing) attached to T0 header; SEM image.
 ........................................................................................................................................ 180

Figure 92: Detail of Figure 91; SEM image. .............................................................. 181

Figure 93: Detail of Figure 92, showing FIB-milled window; SEM image. .............. 181

Figure 94: Mechanically polished LED with FIB-milled window; SEM image. ...... 182

Figure 95: (A) The top surface is clean and the layers can be clearly seen; (B) After milling a nearby feature, the top surface was covered in redeposition. SIM images, 40keV. .................................................................................. 183

Figure 96: Effect of FIB re-deposition on LED I-V characteristics. ....................... 184

Figure 97: (A) thermionic and (B) field emission SEMs, and associated electronics.
 ..................................................................................................................................... 185

Figure 98: (a) Fit of x-dependence to Equation (5-36). Solid line is Monte Carlo data; dashed line is fit by Equation (5-36). (b) Fit of z-dependence to Equation
(5-37). Solid line is Monte Carlo data; dashed line is fit by Equation (5-37). (a) and (b): GaN, $E_b=5$ keV, $d_b=10$ nm.  

Figure 99: (a) Monte Carlo derived data for GaN. (b) Data for AlGaAs. Solid symbols are for $\sigma_x$, and open symbols for $\sigma_z$. $E_b$ is beam energy and $d_b$ beam size. The solid line is $\sigma_x$ from Ref. [], and the dashed line $\sigma_z$ from Ref. [].  

Figure 100: X-EBIC data obtained for analysis. Beam conditions are listed in Table IV.  

Figure 101: (A) SE; (B) X-EBIC images showing examined area. Labels (1)-(4) are the areas of subsequent analysis. Both images at 5keV, ~135pA.  

Figure 102: EBIC curves, normalized, taken from area 4 of the above micrographs.  

Figure 103: Exponential fits for area 4; (A) 4keV (B) 7keV.  

Figure 104: $L_s(x)$ calculations for area 4 at different accelerating voltages.  

Figure 105: 4keV $L_s(x)$ data for areas 1-4.  

Figure 106: (A) Donolato-Luke fit for area 4, 4keV; (B) area 4, 7keV. Circles are data, solid lines are theoretical fits to the data. $s=10^{-2}$ $\mu$m$^{-1}$.  

Figure 107: Comparisons of $L$ calculations for different $s$ values. (A) 4keV, area 4; (B) 7keV, area 4.  

Figure 108: (A) Experiment vs. calculations for $L_e=40$nm, $L_h=70$nm, $s=10^{-2}$ $\mu$m$^{-1}$, with varying $d_j$. (B) $L_e=40$nm, $L_h=70$nm, $s=10^{-2}$ $\mu$m$^{-1}$, $d_j=20$nm, with and without normalization. 4keV, area 1.  

Figure 109: Bonard and Ganière model applied to InAs/GaSb superlattice data...  

Figure 110: Bonard-Ganière fits for (A) 4keV, area 4 data; (B) 7keV, area 4 data.  

Figure 111: (A) 4keV, area 4 data and Bonard and Ganière fits; (b) 7keV, area 4 data and Bonard and Ganière fits.  

Figure 112: Ong et al. analysis of 4keV data, area 4, n-side. +$s$ are data excluded by $x>1.5R$ criteria, ◻$s$ are data used for fit, and the solid line a linear fit.  

Figure 113: 6keV, area 4, n-side data. + symbols are data excluded by $x>1.5R$ criteria, ◻$s$ are data used for fit, and the solid line a linear fit.
Figure 114: EBIC model computations and data for sample area 4, at 4keV........ 220
Figure 115: EBIC model computations and data for sample area 4, at 5keV........ 221
Figure 116: EBIC model computations and data for sample area 4, at 6keV........ 222
Figure 117: EBIC model computations and data for sample area 4, at 7keV........ 223
Figure 118: Thermionic SEM data for sample at 5keV........................................... 224
Figure 119: 3keV, 1.0nA EBIC image of blue LED. Dashed area in analyzed in
Figure 120........................................................................................................... 225
Figure 120: Bonard and Ganière analysis of Figure 119........................................ 226
Figure 121: Schematic of CL system used in this work........................................ 231
Figure 122: Photograph of mirror-sample assembly ............................................ 232
Figure 123: Lens-fiber assembly............................................................................ 233
Figure 124: (Left) R1527P PMT; (Right) Housing built for PMT............................. 235
Figure 125: USB2000 spectrometer....................................................................... 236
Figure 126: SpectraPro spectrometer/monochromator......................................... 237
Figure 127: 436.5nm Hg-line at different slit widths............................................. 238
Figure 128: (A) SE image of LED for plan-view analysis; (B) I-V curve from LED.

SEM image, SE mode, beam conditions 10kV, 680pA; blue LED sample...... 240
Figure 129: (A)-(C) Progressively higher magnification PV-EBIC images of LED.. 241
Figure 130: (A) SE; (B) EBIC; (C) Panchromatic CL images of LED. ....................... 242
Figure 131: (A) SE; (B) EBIC; (C) Panchromatic CL images at high magnification.
......................................................................................................................... 243
Figure 132: (A) EBIC; (B) Panchromatic CL. Arrows indicate example areas that
show dark contrast in both modes. SEM images; 10keV, 50pA. In (B),
brightness and contrast have been modified to make the dark areas more clear.
......................................................................................................................... 244
Figure 133: Voltage-resolved spectrum acquired from blue LED....................... 245
Figure 134: Images and spectrum from blue LED............................................... 246
Figure 135: Spectra and spectrally resolved images taken at 20kV, 1.5nA............ 247
Figure 136: (A) SE; (B) Panchromatic CL images of violet wafer in cross-section. 249
Figure 137: (A) SE; (B) Panchromatic CL images of violet wafer cross-section. ... 249
Figure 138: Profiles across panchromatic images of cross-sectional wafer sample. ................................................................. 250

Figure 139: Plan-view and cross-sectional spectra of violet wafer. Beam conditions: 10keV, 1.4nA. ..................................................... 251

Figure 140: (A) 3.39±0.06eV; (B) 3.06±0.05eV; (C) 2.27±0.03eV; (D) SE + CL composite. SEM images; beam conditions: 10keV, 1.4nA. ..................................................... 252

Figure 141: 10keV current-resolved spectra from cross-sectional blue wafer. .......................................................... 253

Figure 142: Re-normalized current-resolved CL data; 10keV, blue wafer cross-section. ............................................................ 254

Figure 143: Peak areas and I_{PEAK} \propto i_b^m dependencies. ................................................................................................. 255

Figure 144: Peak areas and I_{PEAK} \propto i_b^m dependencies for the UV peaks. ................................................................. 256

Figure 145: 5keV current-resolved CL data. ......................................................................................................................... 257

Figure 146: Re-normalized 5keV current-resolved CL data. ........................................................................................................ 257

Figure 147: Peak areas and I_{PEAK} \propto i_b^m dependencies. ................................................................................................. 258

Figure 148: (A) SE and (B) Panchromatic CL images of FIBed cross-section. SEM images; 10keV, ≈200pA. ................................................................. 261

Figure 149: Voltage resolved CL for mechanically polished area. ................................................................................................. 262

Figure 150: Voltage resolved CL data for 40keV FIB-milled area. ................................................................................................. 263

Figure 151: Voltage-resolved CL data for 10keV FIB-polished area. ................................................................................................. 264

Figure 152: 5keV, 5.00nA CL data for the three surface conditions. ................................................................................................. 265

Figure 153: 20keV, 1.25nA CL data for the three surface conditions. ................................................................................................. 266

Figure 154: Images of FIBed structure: SE image; buffer layer (3.54eV); GaN band edge (3.37eV); QW and from impurity-bound-exciton (overlapping emission energies, 3.17eV); Yellow-band (2.25eV). All are ±30meV bandpass. ................................................................. 268

Figure 155: 10pA, 10keV spectra for the three surface conditions. ................................................................................................. 269

Figure 156: 10,000pA, 10keV spectra for the three surface conditions. ................................................................................................. 270

Figure 157: Values of the m-exponent for the three surface conditions and the three major peaks. ................................................................................................. 271
Figure 158: (A) SE; (B) EBIC of 10keV FIBed window; (C) EBIC of 40keV FIBed window. 7keV, 100pA. .................................................................................................................. 273

Figure 159: (A) 10keV FIB; (B) 40keV FIB areas. Images are X-EBIC at 4keV, 138pA. ................................................................................................................................. 278

Figure 160: Bonard and Ganière fits to EBIC data from FIB-milled areas. X-EBIC at 4keV, 138pA. (A) 10keV FIB; (B) 40keV FIB. ........................................................................... 279

Figure 161: Conceptual model that explains the CL and EBIC data on the differently prepared surfaces. NR=non-radiative recombination........................................... 282

Figure 162: Simulated interaction ranges of Ga$^+$ ions of different energies with GaN. All are plotted on the same scale................................................................. 283

Figure 163: (A) SE and (B) EBIC images of LED; in the EBIC image, a particular defected area is indicated by an arrow. SEM images, 10keV, 40pA.................. 285

Figure 164: (A) Tungsten cap deposited over site of interest and rough FIB cuts; (B) Tilting to cut sample almost free; (C) Attach needle; (D) Sample cut free and lifted out. SIM images, 40keV, <100pA................................................................. 286

Figure 165: (A) SE; (B) EBIC images of LED. A defect is marked with arrow in (B). 10keV, 40pA. ............................................................................................................ 287

Figure 166: (A) Bright field; (B) Centered dark field; (C) Zone-axis SADP; (D) Two-beam SADP; (E) Computer simulated SADP. TEM data, Topcon 002B instrument, 200keV ................................................................. 288

Figure 167: Block diagram of the imaging circuit........................................................... 370

Figure 168: Circuit schematic diagram of the imaging circuit........................................ 371
List of Tables

Table I: Physical properties of wurtzite III-nitrides. .......................................................... 4
Table II: Values of polarization constants in III-nitrides. .................................................. 37
Table III: Operational modes of cathodoluminescence. ................................................. 120
Table IV: Experimental conditions for cross-sectional EBIC .................................. 194
Table V: Estimate minimum allowable x-values to avoid depletion region overlap 197
Table VI: L\textsubscript{h} and L\textsubscript{e} values (in nm) estimated using the exponential method. ....... 199
Table VII: L\textsubscript{h} and L\textsubscript{e} values (in nm) estimated using the Donolato-Luke technique. 203
Table VIII: L values (in nm) calculated assuming high or low surface recombination. .......................................................... 204
Table IX: L\textsubscript{h} and L\textsubscript{e} values (in nm) estimated using the Bonard and Ganière technique. .......................................................... 212
Table X: L values (in nm) calculated assuming high or low surface recombination. .......................................................... 214
Table XI: Example moments method data; from Donlato........................................ 215
Table XII: Method of moments calculations for area 4, n-side ...................................... 216
Table XIII: L\textsubscript{h} and L\textsubscript{e} values (in nm) estimated using the Ong et al. technique. Values marked $\emptyset$ were incalculable with this technique. All values had $\alpha \approx 0$ and $s \approx 0$ .......................................................... 218
Table XIV: Minority carrier properties of examined sample........................................ 229
Table XV: Empirical spectrometer bandpasses at 436.6nm ..................................... 238
Table XVI: m-values for emission peaks at E\textsubscript{0}=10keV ............................................ 256
Table XVII: m-values for emission peaks at E\textsubscript{0}=5keV ............................................ 258
Table XVIII: Effect of surface and beam energy on EBIC collection ....................... 274
Table XIX: L and s values measured at 4keV for the three different surface preparation techniques. .......................................................... 280
1. Introduction

Gallium nitride (GaN) is a wide-direct-gap semiconductor first commercially fabricated in the mid-1990’s at Nichia Chemical in Japan by Nakamura. Although GaN had been studied as a luminescent material in the 1960’s, it was not possible at that time to grow device-quality epitaxial layers. GaN crystallizes in either the wurtzite (hexagonal) or zincblende (cubic) crystal structures; the wurtzite form is stable at room temperature and pressure.

GaN is of commercial interest because its direct band gap makes it potentially a very efficient light emitter, and its bandgap of ~3.4eV places its band-edge photoemission at the border between the violet and ultraviolet ranges of the spectrum. The related materials AlN and InN can be alloyed with GaN to tune the bandgap from ~2eV at the indium rich end to ~6eV at the aluminum rich end. In this fashion, light emitting devices from the deep green to deep ultraviolet have been produced from the (In,Ga,Al)N materials system, the so-called III-nitrides.

The III-nitride alloys are being pursued for use as light-emitting-diode (LED) materials for two main reasons:

1. As LEDs, they have the potential to revolutionize the lighting industry; solid-state lighting from LEDs is much more efficient than incandescent lighting (78 vs. 12 lm/W, respectively, according to Edmond et al.), and LEDs have lifetimes well over 10,000 hours. Although fluorescent lights are approximately as efficient as LEDs and a more mature technology, fluorescent lights contain mercury metal vapor and are extremely harmful to the environment.

2. Second, the optical data storage industry is pursuing the development and perfection of GaN-based laser diodes (LDs). As optical data storage density can increase as the square of the decrease in laser wavelength, switching from the current infrared/red lasers used in compact disk (CD) or digital versatile disk (DVD) drives to blue GaN-based technology could conceivably quadruple the data storage density.
Because of its wide bandgap, high thermal conductivity, strong bond strength, and very high electron mobility (potentially 2400 cm$^2$/V-sec, as predicted by Farvacque$^7$), GaN is also under development for high-temperature, high-power, or high-frequency applications, such as high-electron-mobility transistors$^8$ (HEMTs). Commercialization of GaN HEMTs is expected to begin$^9,10,11$ in 2006 or 2007.

Gallium nitride crystals are usually grown by one of three techniques: metal-organic chemical vapor deposition (MOCVD), molecular beam epitaxy (MBE), and hydride vapor phase epitaxy (HVPE)$^5$. True bulk growth of gallium nitride is currently unfeasible, due to the extremely high nitrogen overpressure required to maintain stoichiometry$^5$. As such, bulk wafers of GaN are not commercially available for homoepitaxy, forcing the used of heteroepitaxy. Quasibulk thick GaN layers grown by HVPE on other crystals are sometimes used as homoepitaxy substrates. Typical heteroepitaxy substrates are $\alpha$-Al$_2$O$_3$ (sapphire) or 6H-SiC. Due to lattice and thermal mismatch, GaN epilayers grown on these substrates contain large densities of threading dislocations and other defects.

In a typical light emitting diode (LED) structure, a threading defect density of $10^8$~$10^{10}$/cm$^2$ will be found; despite this density, the GaN-based LEDs still have quite efficient light emission$^4,12$; in a semiconductor such as GaAs, this defect density would completely quench all light-emitting action. GaN is relatively insensitive to dislocations; this issue will be discussed below.

Because the scientific community does not yet well understand the exact relationship between GaN based semiconductor devices and the defects within the crystal, this research was undertaken to try to elucidate the exact nature of the influences of defects upon GaN LED action.
2. Literature Review

2.1 Overview of gallium nitride devices

2.1.1 Gallium nitride properties and growth

InN, GaN, and AlN are direct-bandgap semiconductors that crystallize in the wurtzite (Figure 1(A)) or zincblende (Figure 1(B)) crystal structures.

Figure 1: Wurtzite (A) and zincblende (B) crystal structures$^{13}$. Dark gray is Ga, light gray N.

As the wurtzite polytype is thermodynamically stable at room temperature and pressure, and all commercial devices are all grown in the wurtzite polytype, it will be the focus of this discussion. The major physical properties of the wurtzite-structured III-nitrides are listed in Table I.
Table I: Physical properties of wurtzite III-nitrides\textsuperscript{14,15,16}.

<table>
<thead>
<tr>
<th>Property</th>
<th>AlN</th>
<th>GaN</th>
<th>InN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thermal Conductivity (W/m-K)</td>
<td>320</td>
<td>120</td>
<td>80</td>
</tr>
<tr>
<td>Bond strength (eV)</td>
<td>2.88</td>
<td>2.20</td>
<td>1.93</td>
</tr>
<tr>
<td>Dielectric constant/e(_0)</td>
<td>9.5</td>
<td>8.5</td>
<td>15.3</td>
</tr>
<tr>
<td>Bandgap (eV)</td>
<td>6.2</td>
<td>3.4</td>
<td>1.9</td>
</tr>
<tr>
<td>Lattice constant a (nm)</td>
<td>0.3112</td>
<td>0.3189</td>
<td>0.3548</td>
</tr>
<tr>
<td>Lattice constant c (nm)</td>
<td>0.4982</td>
<td>0.5186</td>
<td>0.576</td>
</tr>
</tbody>
</table>

Industrially, GaN is produced by MOCVD, as this process is controllable, reasonably fast, capable of producing atomically-flat layers, and gives material with good physical properties. In MOCVD, a gallium (or other III-element) precursor organometallic gas is reacted with ammonia (NH\(_3\)) on a heated substrate, which results in the deposition of the III-nitride material from the gas. A typical gallium precursor is trimethylgallium (TMG)\textsuperscript{14}. DenBaars\textsuperscript{14} quotes the reaction below for the growth of GaN in MOCVD:

\[
\text{Ga(CH}_3\text{)}_3(\text{vap.}) + \text{NH}_3(\text{vap.}) \rightarrow \text{GaN(solid)} + 3\text{CH}_4(\text{vap.})
\] (2- 1)

As described by Nakamura et al.\textsuperscript{1}, MOCVD growth of GaN is typically performed in a two-flow type reactor for better growth uniformity.

Growth by MOCVD (or other techniques) must be heteroepitaxial, as bulk GaN substrates are not readily available. Typical growth substrates are sapphire and 6H-SiC, oriented in the [0001] direction. This results in GaN growth in the ±[0001] direction. Unfortunately, none of the substrates available show good lattice matching or thermal-expansion-matching to GaN or its alloys; Figure 2 shows the relationship between the III-nitride a-direction lattice constants and bandgaps. The lattice constant of 6H-SiC is also indicated; sapphire is off the scale at ~0.48nm.
However, the situation with sapphire is not as grim as it appears from the chart, as GaN's lattice grows rotated ~30° around $c$ relative to the sapphire; the effective mismatch is ~15%, which is still larger than the ~3% for 6H-SiC$^5$.

Because of this lattice mismatch, during epitaxial growth, the GaN nucleates as islands that grow across the substrate and impinge upon one another. This results in the formation of slightly-misoriented subgrains within the epitaxial film. At the boundaries of these subgrains, threading dislocations form with their line direction $u$ parallel to the $c$ axis of the crystal. This is schematically illustrated in Figure 3.
In order to reduce this density of dislocations, a thin "buffer layer" of AlGaN is usually grown at low temperature directly on the substrate, before the actual device epilayer is grown, as originally demonstrated by Yoshida et al.\textsuperscript{17,18}. This was found to improve crystal quality by suppressing transfer of dislocations from the interface to the epilayer\textsuperscript{14}, and is now standard practice in the growth of III-nitrides.

Other techniques have also been used to attempt to eliminate or reduce dislocations in GaN. The primary technique is epitaxial lateral overgrowth (ELO)\textsuperscript{19}. In this technique, a buffer layer and a thin layer of GaN will be grown in the standard method. Then, a mask of silicon oxide or silicon nitride will be deposited onto the GaN layer, typically as stripes running along a low-index GaN crystallographic direction such as $<1\overline{1}00>$. The mask will typically have a width 2~3$\mu$m and a pitch of 5~10$\mu$m. Once the mask is in place, GaN growth will be continued; new GaN will grow on the GaN between the mask stripes, and then laterally overgrow the masked areas. These are referred to as "windows" and "wings," respectively. Although the dislocation density in the windows will be $10^8$~$10^{10}$/cm$^2$ as expected from normal
MOCVD epitaxy, the dislocation density in the wing areas can be as low as $10^6$/cm$^2$, which is useful for LDs or other very-dislocation-sensitive structures$^{20}$.

The exact density of the dislocations in the wings depends strongly on the growth mode of the laterally-overgrown segments. This is discussed by Hiramatsu et al.$^{21}$ Hiramatsu et al. deposited $<1\bar{1}00>$ type ELO stripes, and then re-grew the GaN at different temperatures and pressures. Higher temperature and lower pressure were found to favor growth by $\{11\bar{2}0\}$ and $\{0001\}$ type facet; lower temperature and higher pressure, by $\{11\bar{2}2\}$ type facet. This is illustrated in Figure 4.

![Figure 4: Illustration of faceted growth behavior; adapted from Hiramatsu et al.$^{21}$](image)

This is of interest because the dislocation propagation from the windows to the wings is different in the growth modes. According to Hiramatsu et al., in the $\{11\bar{2}2\}$ type growth, threading dislocations can tilt into a basal configuration, propagate laterally across the window, and then tilt back into a threading configuration to reach the surface. In the $\{11\bar{2}0\}$ type growth, the dislocations
simply thread vertically through the window region, and consequently have a much lower density in the wing areas.

Because dislocations are implicated in loss of luminescent efficiency, loss of carrier mobility, and device degradation, many different efforts have been undertaken to reduce their density in devices (LEDs, LDs, HEMTs, etc.). Such techniques include growth on thick pseudo-bulk GaN substrates (homoepitaxy), growth on modified substrates such as grooved sapphire\textsuperscript{22}, or growth on step-free substrates\textsuperscript{23}.

GaN n-type doping is typically achieved by the addition of Si- or Zn-bearing carrier gas to the MOCVD chamber during epitaxy; p-doping is usually achieved by the addition of Mg dopant followed by high-temperature annealing under a nitrogen atmosphere\textsuperscript{14}. Annealing must be performed to break up Mg-H complexes in the GaN; as Mg is a relatively deep dopant (~250meV above the valence band edge), and complexes may still exist despite annealing, full ionization is never achieved in the epilayers. Undoped GaN will be lightly n-type due to nitrogen vacancy native defects.

In summary, GaN and its related alloys have excellent physical properties for light emission and high-power applications. They are grown commercially in the hexagonal wurtzite polytype by MOCVD heteroepitaxy, but can be grown by other techniques. Commercial bulk GaN substrates are still not available, so other techniques must be used to produce low-defect-density material. Doping n- and p-type are usually achieved by Si and Mg.

2.1.2 Gallium nitride devices

The primary commercial application of GaN is in the manufacture of LEDs. These are grown on either SiC or sapphire substrates, in which the n-layer is deeper than the p-layer. Generally, one or more quantum wells (QWs) separate the n- and p-layers.

Quantum wells are manufacturable by MBE, MOCVD, and HVPE. Essentially, a QW is a structure that exists continuously within a semiconductor
crystal, but has a narrower bandgap than the crystal to either side\textsuperscript{24}. For example, a layer of narrow-bandgap InGaN could be grown as a thin layer on a GaN crystal, and then the InGaN layer would be capped by another, still continuous, GaN layer. Typical commercial quantum well widths are 1-5nm. This is illustrated schematically in Figure 5.

![Figure 5: Schematic of QW structure.](image)

If the quantum well is sufficiently thin, the lattice-mismatch resulting from the change in chemistry can be taken up by pseudomorphic strain. Also, as the QW becomes narrow, electrons or holes confined to it will behave in a manner similar to the "particle-in-a-box" from quantum mechanics. That is, an electron or hole will no longer be in an energy continuum as in the bulk crystals cladding the QW, but will be in a discreet, quantized state. At least one state for electrons and one for holes will be available; if the well is deep or the effective masses of the particles high, multiple states will exist.

Quantum wells are used commercially as they confine electrons and holes to an area of particularly chosen bandgap, which causes strong recombination at an
engineered wavelength\textsuperscript{25,26}. QWs also support population inversion, which makes semiconductor lasers more readily manufacturable.

2.2 Dislocations

2.2.1 Overview of dislocations

As discussed above, large densities of threading dislocations form in III-nitride epilayers, especially those grown by heteroepitaxy (which includes all commercial III-nitrides at this time).

Originally, the fact that GaN-based devices could have any luminescent behavior was surprising. GaN films typically contained $>10^9$ dislocations/cm$^2$; LEDs or LDs produced from material such as GaAs required dislocation densities $\sim 10^5$/cm$^2$ or lower for operation.

This led to speculation that dislocations in GaN were electrically inactive, and therefore had no effect on the luminescent behavior. This is best described in the review article of Orton and Foxon\textsuperscript{27}; quoting the article, "it seems that dislocations in GaN do not act as non-radiative centers." This was published in 1998, and as explained in this section and shown in the experimental work presented later in this work (Section 5), their quote is highly inconsistent with the current understanding of dislocations in GaN. Orton and Foxon do, however, make two correct observations: the $\sim 10^{10}$/cm$^2$ dislocation densities are far higher than that observed in conventional semiconductors (they quote $<10^4$/cm$^2$ for LED or LD action in non-nitride conventional III-V and II-VI semiconductors). They also note that dislocations in GaN have mobilities and migration rates perhaps $10^{20}$ times smaller than in comparable GaAs devices. This is likely due to GaN's remarkably high bond strength. As will be explained in the following sections, the true optoelectronic behavior of dislocations in GaN is much more complex, and still not well understood.
2.2.2 Electronic activity of defects

The bandstructure of a crystal arises from the regular arrangement of atoms; even the simplistic Krönig-Penny model predicts discreet energy bands from a simple 1D periodic lattice. Crystallographic defects are defined as areas in a crystal where this symmetry is broken; as such, the bandstructure here will be perturbed in comparison to the bulk.

Dislocations, and defects in general, begin to have significant effects on optoelectronic properties when their defect states lie within the crystal's bandgap. This is illustrated for the case of a generic semiconductor in Figure 6. Here, a defect lies within the bandgap and results in a non-zero density of states \( Z(E) \) within the bandgap. When the defect lies within one of the bands, the effect on the density of states will be much smaller.

Figure 6: Illustration of defect states within a semiconductor. \( E \) represents energy, \( \text{CB} \) the conduction band, and \( \text{VB} \) the valence band.

When states exist in the bandgap, they can act as traps for carriers, which can result in reduced mobility due to ionized impurity scattering, and can also hold
the carriers in a metastable state for a period of time longer than the normal carrier lifetime.

Minority carriers will be particularly affected by these states. Let us assume a defect near the middle of the bandgap (or even near the top of the bandgap) in an n-type semiconductor. Due to the n-doping, the Fermi level will be very near the top of the bandgap. As such, the electron occupancy at the energy of the defect will be high, effectively 1. It can thus be assumed that any existing traps are occupied by the majority carrier (e\textsuperscript{-} in this case). The minority carrier (h\textsuperscript{+} in this case) will be coulombically attracted to the charged defect site; if the defect can capture the minority carrier, an electron-hole pair (EHP) recombination will likely result, causing the loss of the minority carrier, which will no longer be available for device processes such as light emission.

These processes will often be non-radiative and have a much higher recombination rate than the bulk band-edge or free-exciton recombinations; in an LED, this would result in a severe depletion of injected minority carriers without light emission, quenching the LED action, at least in the vicinity of the defects. This minority carrier capture will not be from band-edge to band-edge, so if the process is radiative, it will give off lower-energy photons than expected. In indirect bandgap materials such as Si, it is even possible for radiative transitions to arise at the dislocations.

2.2.3 Electronic activity of dislocations in GaN: Theory

Because of the surprisingly high luminescent efficiency of LEDs containing >10\textsuperscript{10} dislocations/cm\textsuperscript{2} (For example, Lester et al.\textsuperscript{12}), it was initially assumed that dislocations in GaN were electrically inactive. Early calculations using the local-density approximation for intrinsic dislocations (i.e., assuming no point defect or impurity segregation) in GaN predicted that these dislocations would have no bandgap states\textsuperscript{30,31}.

The view of dislocations continued to evolve with time. Continuing with local density theory, Elsner et al.\textsuperscript{32} calculated the behavior of dislocations in which gallium
vacancies (V\textsubscript{Ga}) and oxygen atoms on nitrogen sites (O\textsubscript{N}) formed complexes (V\textsubscript{Ga-}
O\textsubscript{N}) that segregated to threading edge dislocation lines; calculations showed that these dislocations would, indeed, be electrically active. They further hypothesized that these bandgap states could be responsible for so-called yellow-band (YB) emission centered at \(~2.2\text{eV}\) observed in large numbers of experiments on GaN.

Other groups\textsuperscript{33,34} concluded that even intrinsic dislocations should have bandgap states, although the exact nature would depend on doping. Arslan and Browning report a conflicting result, in which intrinsic dislocations were not predicted to have bandgap states\textsuperscript{35}; although dislocations bearing gallium vacancies were predicted to be electrically active. Because Arslan and Browning used multiple-scattering simulations, it may simply be that these simpler calculations are not capable of the level of detail available to the contradicting reports mentioned above that used more rigorous theoretical tools such as ab-initio local density approximations.

Lymperakis et al.\textsuperscript{36} used high-resolution transmission electron microscopy (HRTEM) to determine the arrangement of atoms at dislocation cores in GaN. They observed full-core dislocations. They claim that a typical assumption would be that, with no broken bonds in the full-core structure, the dislocations would be expected to be electrically inactive. However, using the HRTEM atomic arrangements as input, they then used density functional theory to calculate the electronic structures of the dislocations. This led to the conclusion that the enormous strain fields around the dislocations cause metallization of the Ga-Ga bonds and induces deep bandgap states, even in the absence of segregation. The review of Blumenau et al.\textsuperscript{37} summarized density functional theory calculations to conclude that edge dislocations possess states in the top half of the gap, and that V\textsubscript{Ga}-O\textsubscript{N} complexes explain correlations between the YB and dislocations.

In addition to the possibility of carrier trapping and sub-bandgap luminescence, dislocations are often associated with leakage currents. In addition to bandgap states which should result in strong non-radiative recombination, these
screw dislocations were predicted to be conductive paths, and could result in problems such as leakage currents\textsuperscript{38}.

In summary, most calculations agree that dislocations in GaN, intrinsic or otherwise, should have bandgap states and may be associated with the YB emission.

2.2.4 Electronic activity of dislocations in GaN: Experiment

A number of different experimental techniques have been used to try to probe the properties of dislocations in GaN. Electron energy loss spectroscopy (EELS) is a thin-foil scanning TEM (STEM) technique that can directly probe the joint density of states (JDOS) through the thickness of the thin foil. This is accomplished by either performing spectroscopy directly on the valence portion of the JDOS, referred to as valence EELS (VEELS), or by analyzing the shifts and shape changes of the core-loss peaks, which is referred to as energy loss near edge spectroscopy (ELNES)\textsuperscript{39}. VEELS can sometimes be interpreted directly, but evaluation of ELNES requires computational modeling. This will be described in more detail below.

Other techniques to probe the electronic structure of a semiconductor are based on examination of the optical behavior; cathodoluminescence (CL) excites a semiconductor with a kilovolt range electron beam. EHPs will be injected into the sample, and when these EHPs recombine, photons may be emitted. CL spectroscopy measures the wavelength distribution of light emitted by the sample. Photoluminescence (PL) excites EHPs in the sample with photons of energy greater than the bandgap, and thus also injects EHPs. When radiative recombination occurs, PL spectroscopy analyzes that signal. Examination of the electroluminescence (EL) of forward-biased LEDs or LDs can also yield information about the behavior of the defects within the device.

Other techniques are based on the scanned probe microscope; this typically involves scanning a small, conductively-coated atomic-force-microscope (AFM) cantilever and probe across the sample surface and examining how the phase or amplitude of the cantilevers' oscillation changes with position.
Experiments on dislocations in GaN will be discussed in this order: (1) a review of the luminescence techniques (EL, CL, PL) will be presented, followed by (2) EELS experiments on dislocations, (3) a review of scanned-probe techniques applied to dislocations, and then (4) discussion of experiments that do not fall into one of the other three categories.

First, luminescence techniques. Among the earliest results attempting to correlate defects to luminescence is that of Ponce et al.\textsuperscript{40} Here, GaN islands were grown on sapphire under different conditions, in order to yield large (10~50\(\mu\)m) and small (~4\(\mu\)m) GaN islands. These were examined by PL and CL. The large-crystallite sample was observed to have a much lower YB emission during PL at 2K (Figure 7); in CL microscopy (see Section 3.2.5) the YB emission was observed to concentrate around the edges of the islands (Figure 8), which Ponce et al. interpreted to mean an associated with extended defects near the edges of the islands.
Figure 7: Effect of temperature and crystallite size on PL & CL spectra. (a) GaN sample with small crystallites, 2K; (b) Large crystallites, 2K; (c) Small crystallites, 300K.

Reused with permission from Ref. 40. Copyright 1996, American Institute of Physics.
Figure 8: CL images of large crystallites. (left) SEM image; (center) Band-edge CL; (right) Yellow-band CL.

Reused with permission from Ref. 40. Copyright 1996, American Institute of Physics.

Choi et al.\textsuperscript{41} performed similar experiments, by combining PL and CL with TEM-based dislocation-density measurements. They found results that clearly imply that dislocations act as non-radiative defects; specifically, PL intensity of the QW emission dropped with dislocation density (although at a very sub-linear rate), and panchromatic CL imaging showed bright and dark areas; the fraction "dark" increased with dislocation density.

CL results explained in terms of dislocations have been obtained; Tomiya et al.\textsuperscript{20} performed CL of ELO GaN laser diodes, and resolved individual dark features -- both threading and in-plane -- that are interpreted to be dislocations that arrest radiative recombination. At lower spatial resolution, Martin et al.\textsuperscript{42} found the "wing" areas of an ELO-grown sample to have significantly higher CL intensity that the window regions, which was interpreted in terms of the lower dislocation density in the wings resulting in less dislocation-associated non-radiative recombination.

Macht et al.\textsuperscript{43} used a photoelectrochemical KOH etch to remove bulk nominally-undoped n-GaN, leaving behind small nanopillars, each containing a single dislocation. Thus, the bulk non-dislocated GaN was etched away and only GaN very near a dislocation remained. PL examination showed an orders-of-
magnitude reduction in the YB emission, indicating that the YB was not associated strongly with the dislocations. Inoue et al.\textsuperscript{44} grew dislocation-free GaN nanopillars on a (111)-Si substrate and found that the CL band-edge emission was orders of magnitude higher than in a similar, but dislocated, GaN film.

AFM can be used to identify surface pits in GaN, which are generally considered to correspond to dislocations threading the surface. Rosner et al.\textsuperscript{45} performed AFM on an area of an n-type GaN film, and surface pits (and, by inference, dislocations) were mapped. CL of this same area revealed bright and dark regions. They superimposed the surface-pit information on the CL image, and found a near-perfect correspondence between the dark CL areas and the surface pits, which implies that the dislocations cause the dark features in CL. Further, Rosner et al. treated the minority carrier diffusion problem and estimated that the distribution of injected excess minority holes, $\Delta p$, around a recombination-enhancing defect such as the dislocation would be of the form:

$$
\Delta p = g_0 \tau \left[ 1 - \exp \left( \frac{-r}{L_h} \right) \right] \tag{2.2}
$$

Here, $\Delta p$ is the density of minority holes, $g_0$ the beam-induced generation rate of electron-hole pairs, $\tau$ the carrier lifetime, $r$ the distance from the dislocation core, and $L_h$ the minority carrier diffusion distance. They applied this equation as a convolution filter to the etch-pit image, and found that for the case $L_h=250\text{nm}$, the convolved AFM image and the CL image corresponded almost perfectly. This correspondence is re-printed as Figure 9. This is interpreted to mean that the dislocations are non-radiative sites of enhanced recombination rate. However, because the 250nm diffusion length is very near to generation radius of the 10kV electrons used to excite the CL, this $L_h$ value must be interpreted that the generation volume, not the true diffusion length, is controlling the feature size. Thus, $L_h=250\text{nm}$ is a measure of the SEM excitation of the sample rather than measurement of the true minority carrier diffusion length.
Figure 9: Correlation of AFM and SEM-CL data regarding dislocations.

Left column: (A) AFM image of GaN layer; (B) surface pits taken from AFM data; (C) valley contrast from AFM data; (D) SEM-CL image of same area; (E) AFM+SEM/CL data superimposed. Right column: (A) SEM-CL image; (B) AFM surface-pit image treated with 250nm convolution filter. Reused with permission from Ref. 45. Copyright 1997, American Institute of Physics.

Godlewski et al.\textsuperscript{46} examined the in-plane distribution of CL in GaN layers and found the band edge luminescence was strongly arrested at subgrain boundaries,
where dislocation density is highest; however, they found YB emission to be relatively homogenous.

Direct correlation of dislocations to luminescence is very difficult, as it requires a STEM equipped for CL, as only TEM/STEM techniques directly image dislocations. Although AFM or SEM can image features such as surface pits, which are all but certainly associated with dislocations, it takes TEM/STEM to directly and unambiguously observe a dislocation. The Yamamoto research group\textsuperscript{47,48,49} has achieved this feat in the study of GaN. Their GaN work reviewed here is in plan-view, rather than cross-section, but was able to study both threading and basal type dislocations in GaN MQW LED structures. Their most significant results can be summarized as, (1) edge and mixed threading dislocations were associated with the quenching of the GaN band edge CL emission, (2) basal plane screw dislocations did not quench the CL emission, and (3) no dislocations were associated with YB emission. It seems likely that screw-type threading dislocations went unexamined as the $\mathbf{g} \cdot \mathbf{b} = 0$ invisibility criterion in TEM (see Section 3.3) would render screw dislocations in plan-view specimens invisible. Figure 10 is reprinted from Yamamoto et al.\textsuperscript{47} and illustrates the recombination activity of defects.
Figure 10: (S)TEM-CL image of GaN-based LED, showing dislocation activities. (A) CL spectrum of thin plan-view TEM specimen of a MQW GaN-based LED. (B) TEM image showing dislocations. (C) Free-exciton (FX) CL image showing dislocation quenching. (D) Yellow-band (YB) CL image, showing no dislocation contrast.

Reused with permission from Ref. 47. Copyright 2003, American Institute of Physics.
Now, EELS results on GaN dislocations will be reviewed. An electron energy loss spectrum is obtained by placing a thinned sample into a TEM or STEM instrument; the electron beam, typically at 80-400keV, passes through the specimen. In a thin specimen, most of the electrons will not suffer an inelastic scattering event and will thus still carry the incident beam energy; let us assume 200keV for the sake of discussion. The beam energy will not be a δ-function, but due to the energy spread of the electron gun, will have a FWHM of 0.1-5eV, depending on the exact type of gun and monochromation. Although most electrons pass cleanly through the thin foil sample, some number of the beam electrons will suffer inelastic scattering and emerge with energies less than 200keV. If the post-specimen electron beam is dispersed with a magnetic prism onto a multichannel detector, a spectrum of the fraction vs. electron energy can be measured. A beam electron suffers energy-loss by imparting the lost energy to one or more electrons in the sample; thus, the spectrum obtained by observing the emergent energies of the primary beam is directly related to the possible energy-consuming process in the sample. Examples would be core-shell ionization or the generation of an electron-hole pair.

EELS, as mentioned above, probes the JDOS of the solid. Low-loss EELS probes the direct band-to-band optical transitions; core-loss data, via comparison to atomistic models, can also yield information about the local DOS. Figure 11 demonstrates three possible transitions that could be observed in a VEELS spectra: one below the bandgap, one at the bandgap, and one beyond the bandgap. By considering that a transition could originate at any point in the VB and terminate at any point in the CB, or at the defect, it is seen how a broad spectrum of valence energy-loss response can arise. The exact shape of the EELS spectrum will be derived from the VB, defect state, and CB densities of states.
Core-loss EELS is conceptually the same; however, the transition is from a core-shell electron level, which is sharply defined in energy, unlike the electrons in the VB measured by VEELS. In a perfect crystal, the core electron will be excited into the CB, and can take values of energy determined by the CB density-of-states. This results in a sudden peak and slow decay in the EELS spectrum; if a defect is present, it can modify the density of states and thus shape of the EELS peak, leading to changes in the ELNES. Further, a lower-energy state exists (the defect) into which the core-electron could be excited, which will move the onset of the peak to a lower energy; this is called the chemical shift.

Optical characterization techniques, such as ellipsometry, give excellent energy resolution (<<1eV) but poor spatial resolution (>1mm). STEM-EELS gives moderate energy resolution (0.3-1 eV) and Ångstrom-to-nanometer-scale spatial resolution\textsuperscript{50}. Low loss EELS will be resolved to ~10nm, due to the delocalization of the valence and conduction band wavefunctions. Further, as a general rule, it is found that optical and EELS data compare quite well, the only major difference being the poorer energy resolution of the EELS data. Unlike optical techniques, EELS can probe crystallographic defects -- such as dislocations -- at a highly localized level, or obtain optical data from localized defect-free regions.
A number of papers have studied dislocations in GaN via nitrogen K-edge core-loss spectroscopy. The N-edge was chosen because it allows unoccupied conduction band or midgap states to be probed\textsuperscript{51}. A change in the conduction band edge or a large number of midgap states at defects will result in a chemical shift of the N-K edge near the defect site; changes to the conduction band DOS should manifest themselves as changes to the near-edge structure. Both were studied.

Arslan and Browning\textsuperscript{52} used a combination of multiple-scattering models and experimental EELS to study the chemical shift and fine structure of the N-K edge around dislocations in GaN. Their model found that intrinsic dislocations should not show a chemical shift, whereas extrinsic dislocations – modeled in their case as decorated with gallium vacancies – showed a -1.9eV chemical shift.

In contradiction to Arslan and Browning, Fall et al.\textsuperscript{51} showed, using a combination of density-functional theory and experiment, that intrinsic full-core dislocation structures would have midgap states. Lazar et al.\textsuperscript{53} found a large chemical shift of the N-K edge in the buffer layer, between the GaN and substrate, due to the high density of defects in the buffer region, indicating electrically active defects in this region.

In GaN, the Ga 3d peak and plasmon peak are very closely spaced. Sanchez et al.\textsuperscript{54} state that the plasmon peak is at ~20eV, and the Ga 3d peak at ~24eV. They used the intensity and shifts of the 3d peak to measure the distribution of Ga in InGaN quantum wells, and also used it to show segregation of In to threading dislocations; this is speculated to be a failure mechanism in GaN-based optical devices.

By directly examining the low-loss spectrum in the region of the bandgap energy, direct measurements of electronic defects and DOS at the bulk or defects can be made. Bandgap fitting procedures can be applied to defected regions of the crystals. Bangert et al.\textsuperscript{55} measured spectra from the defected GaN near the buffer layer, and found significant signal below the expected cutoff at the 3.4eV bandgap. This shows clearly that low-loss EELS can directly measure midgap states. Changes in the shape of the EELS curve – resulting in changes in optical absorption
– were also observed near dislocations\textsuperscript{55,56,57}. Strong midgap signals near 2.4eV – similar in energy to the yellow band defect often observed in EL, CL, or PL studies of GaN – were also observed near dislocations\textsuperscript{55,56}. Edge and screw dislocations in GaN films were found to show sub-gap absorption\textsuperscript{57}.

Modern GaN devices utilize a buffer layer between the heteroepitaxy substrate and the epilayer; this buffer layer is highly defective. Large differences between the bulk and buffer areas are apparent in VEELS data, as the bandgap becomes completely swamped by midgap defects states in the buffer region, and the midgap states begin at around 1.75eV\textsuperscript{(55)}. Subgrain boundaries also show a similar smearing of the DOS and large numbers of defect levels\textsuperscript{55}.

Absorption in the buffer could be a device-performance issue; modern InGaN LEDs have transparent substrate material and depend on reflectors under the substrate to collect much of the emitted light. Absorption in the buffer layer will reduce the external quantum efficiency.

To summarize the EELS experiments on GaN, sub-bandgap states have been observed using VEELS and nitrogen K-edge ELNES. These have often been interpreted in terms of the YB defect. However, EELS data has not been directly related to optical experiments (i.e., CL) on individual defects.

Now, AFM-based techniques applied to GaN will be reviewed. In contrast to EELS, which probes through the volume of a thin foil specimen, AFM-based techniques directly examine the surface of a specimen. Scanning capacitance microscopy (SCM) was used by Hansen et al.\textsuperscript{58} to study the electrical properties of dislocations threading the surface of MOCVD grown GaN. They observed that the C-V curve shifted towards positive voltages in the vicinity of a dislocation, indicating negative charge around the dislocations. Although a scanned probe technique is not chemically-sensitive and will not determine if the dislocation observed is intrinsic or extrinsic, it does clearly show the electrical activity. The nominally undoped GaN films had a background n-carrier concentration $\sim5\times10^{16}/\text{cm}^3$, which indicates n-behavior. Carrier traps in an n-type material would be expected to show a negative

25
charge. Schaadt et al.\textsuperscript{59} also used SCM to study AlGaN/GaN transistor structures and arrived at essentially the same conclusions as Hansen et al.\textsuperscript{58}.

Hsu et al.\textsuperscript{60} used a technique they call a scanning current-voltage microscope (SIVM), in which an AFM-type tip acts as a Schottky contact to the sample; they examined GaN/AlGaN HEMT material. They were able to show that screw dislocations acted as reverse-bias leakage paths, whereas mixed or edge dislocations did not. Dislocations that act as conductive paths will necessarily have bandgap states, indicating once again the nature of dislocations in GaN. The same group used MBE-GaN on thick GaN-template substrates, in conjunction with TEM and SIVM, to further show that the leakage current varied greatly depending on growth stoichiometry\textsuperscript{61}. Under Ga-rich growth, excess Ga accumulated at the surface termination points of the threading screw dislocations, and the Ga-rich material showed \(~1000\times\) greater leakage current than Ga-poor material. Thus, it was concluded that excess Ga drastically changes the dislocations' core structure, leading to spectacularly changed electrical properties.

Simpkins et al.\textsuperscript{62} used similar techniques to show enhanced conductivity at dislocations, and degraded conductivity around dislocations, in p-GaN. They attributed this behavior to segregation of Mg to dislocations and depletion of the dopant from the area around the dislocations. This is an indication of the relative unimportance of the exact behavior of intrinsic dislocations, as any practical GaN material will contain both accidental (i.e., O-impurity, N-vacancy) and intentional (i.e., dopant) point defects, which will dominate the dislocations' behavior. Li et al.\textsuperscript{63} found that reverse bias leakage current varied exponentially with voltage, and the pre-exponential factor scaled with the square of the screw-component dislocation density. This is a slightly different result from that of other reports, which indicated that the pure-screw nature of a dislocation related to reverse bias leakage, as opposed to simply having a screw component (i.e., screw or mixed). Krtschil et al.\textsuperscript{64} used scanning surface-potential microscopy of differently-doped GaN layers to conclude that the dopant's identity, rather than doping level or consequent Fermi level, determined the electrical properties of the dislocations, which is a further
argument towards the non-intrinsic, point-defect-decorated nature of dislocations in GaN.

Various other techniques have also been used to study the effects of dislocations in GaN. Look and Sizelove\textsuperscript{65} developed a model based on charged dislocation scattering to predict the in-plane electron mobility vs. temperature and free carrier density vs. temperature, with the dislocation density as an input parameter. The model accurately predicted results for MOCVD GaN on sapphire; as the assumption of the model was the electrical activity of dislocations dominated both carrier density and charged line scattering of carriers, this strongly implies the trapping nature of dislocations in the materials measured.

Wang et al.\textsuperscript{66} studied diffusion of surface metallization into AlGaN/GaN HEMT structures. TiN was found to form from the surface, diffuse down the dislocation, and to the AlGaN/GaN heterostructure, which formed an ohmic contact with the 2-dimensional electron gas at the heterointerface. This is further evidence of the highly non-intrinsic nature of dislocations in practical GaN materials.

Electron-beam induced current (EBIC, which will be described in detail in Section 3.2) is an in-SEM technique that can also shed light on the recombination activity of dislocations. Albrecht et al.\textsuperscript{67} examined a Schottky-type n-AlGaN device grown by MBE with combined EBIC, CL and TEM. They found that increasing beam current during the EBIC experiment caused dark spots to appear on the sample, and further increases in beam current lead to increased contrast in the EBIC image. The appearance of the dark spots was reversible, by reducing the beam current, and therefore not a beam-damage artifact. These dark spots were seen to be dark in panchromatic CL imaging and had the same density as screw dislocations observed in TEM. The contrast of the dark-spot-defects in EBIC increased as the sample temperature was reduced. The dark spots also coincided with pits observed by SEM imaging. The sudden appearance of contrast with increasing injection, and the increase of contrast with decreasing temperature, was interpreted to mean that the dislocations became active as the density of carriers increased. As the EBIC contrast will be proportional to inverse of carrier lifetime in the defected area, the
lowered temperature or increased injection density were thus resulting in shorter carrier lifetimes around the dislocation. They thus deduced that the dislocations acted as shallow acceptor-like states with an acceptor energy of 20meV.

Electron holography (EH) is a technique based on TEM. A field emission gun TEM has high coherence of the electron beam; this can be utilized to perform interferometric techniques. In off-axis TEM EH, the sample is arranged to occupy approximately one-half the field of view of the TEM screen. An electron-biprism is inserted at the selected area aperture position; this biprism will typically be a <1μm diameter quartz fiber, coated with a conductive layer, and held at a positive voltage ~10-100 volts. Part of the TEM beam will pass through the sample, and part will pass through the vacuum next to the sample. The beam that passed through the sample will have suffered a phase shift relative to the reference beam. The biprism causes the two beams to recombine on the viewing screen, and the shift in phases will result in an interference pattern. This pattern can be computationally treated to obtain the value of the internal potential for the sample, if the thickness is known. If the thickness is unknown but constant, or varying in a known fashion, the changes in internal potential can be found. EH allows the measurement of the mean inner potential (MIP) of a material; the MIP of semiconductors is typically ~15 volts, and (most importantly) the MIP is sensitive to the bending of the bands in the semiconductor. If a defect such as a dislocation bears an electrical charge, there will necessarily be band-bending in the locality of the defect in order to keep the Fermi level constant. EH can measure this band bending. Indeed, a number of experiments performed on plan-view TEM samples of GaN films found significant negative charge on threading dislocations. The general result from these reports was that threading dislocations in n-type Si-doped GaN layers hold of charge of approximately 2 electrons/|c|, where c is the perpendicular unit-cell lattice parameter (0.52nm).
2.2.5 Dislocations: summary of literature

As discussed in the above sections, there is a large amount of conflicting literature relating to the behavior of dislocations, particularly threading dislocations, in gallium nitride and its related alloys. As these III-nitride materials are currently produced by heteroepitaxy and have significant densities of threading dislocations, a detailed understanding of the effect of this dislocations on the optoelectronic properties of devices fabricated from III-nitride devices is necessary.

The literature on the topic is rather contradictory; dislocations have been asserted to be non-electronically-active\textsuperscript{27,30}, yet various experiments (such as the STEM-CL of Yamamato et al.\textsuperscript{47}, and the many other reports discussed above) clearly show that dislocations do have optoelectronic activity.

The ability to study dislocations in GaN-based devices, and quantify both their density and behavior, is necessary to provide feedback to improve crystal growth procedures.

One of the main unanswered questions is the relation of dislocations to the yellow-band luminescence observed in CL, PL, and EL examination of GaN. Although various reports (i.e., \textit{ab-initio} simulation by Elsner et al.\textsuperscript{32}, or EELS experiments by Brockt and Lakner\textsuperscript{74,75}) associated \textasciitilde2.2eV defect levels with the presence of dislocations, other observations (i.e., STEM-CL by Yamamato et al.\textsuperscript{47}) indicate that the yellow-band is homogenous and not related to particular extended defects, such as dislocations.

In short, dislocations have clearly been shown to be non-radiative recombination sites and trapping sites, but experimental evidence of their association with the YB emission is ambiguous.

2.3 Other III-nitride defects

2.3.1 The yellow-band defect

Cathodoluminescence or photoluminescence examination of GaN typically shows a broad peak centered around 550nm (2.25eV); this is referred to as the
The atomistic explanation of the YB emission is still a matter of some debate; a consensus seems to be building toward gallium-vacancy related states\textsuperscript{76}. Reschikov and Morkoc\textsuperscript{76} reviewed the luminescence effects of point defects in III-nitrides and they conclude that the yellow luminescence results from a transition from either the conduction band or shallow donor to a deep acceptor, and quote the acceptor's energy at 0.8~0.9eV. A schematic illustration of this situation is shown as Figure 12.

![Figure 12: Possible model of the YB emission related states.](image)

Firstly, they state that the YB is centered at 2.20-2.25eV, nearly Gaussian with FWHM of 350-450meV, and structureless even at very low measurement temperatures. In 1976, Pankove and Hutchby\textsuperscript{77} reported the results of implantation of 35 different ions into GaN; almost all of them resulted in a broad PL peak around 2.15eV, which indicates that implantation damage, rather than the identity of the implanted species, was responsible for the PL peak. Positron annihilation studies in conjunction with TEM\textsuperscript{78} indicate that gallium vacancies are present in n-GaN in large concentration, and are present in the grain interiors, as opposed to grain boundaries or on dislocations. The integrated intensity of the YB peak in PL experiments has been found to have a nearly perfect correlation with $V_{Ga}$ concentration\textsuperscript{79}. Further,
the YB emission drops significantly or disappears with p-doping\textsuperscript{77}, and $V_{Ga}$ are not expected to exist in large concentrations in p-GaN. Although other candidates exist, such as carbon impurities, the preponderance of evidence points at $V_{Ga}$ or $V_{Ga}$-containing complexes being the source of the YB.

2.3.2 Stacking faults

Because of the relatively small difference in free energy between the wurtzite and zincblende polytypes of GaN, stacking faults can form easily. In (0001)-grown GaN films, stacking faults will mostly exist in the buffer layer. In non-(0001)-grown films, however, such as those grown in (1120) to avoid polarization effects, stacking faults often appear throughout the films. EELS examination\textsuperscript{55,80} of buffer layers show significant sub-bandgap signal; this was observed for both hexagonal\textsuperscript{55} and cubic\textsuperscript{80} GaN. This sub-bandgap peak is attributed to the large density of stacking faults and inclusions or opposite-polytype material in the buffer layer.

Unlike dislocations, in which EELS shows sub-bandgap electronic states but CL shows no emission, stacking faults have been shown to clearly give defect-associated luminescence. Ponce’s group\textsuperscript{81,82} grew non-polar (1120)-type GaN layers, which contained large quantities of stacking faults. They examined ion-milled samples in SEM-CL at low voltage (5kV) and high spatial resolution, and then observed the exact same areas in TEM.

Figure 13 shows example data from Ref. 81.
In the CL spectrum, distinct emission peaks at 3.474, 3.413, and 3.288-3.340eV can be seen. The 3.474eV peak is attributed to the donor-bound exciton D⁰X, and is the bulk, non-defect emission. In the right part of the figure, the CL maps at the different sub-D⁰X peaks are seen, and compared to the TEM diffraction-contrast image which shows stacking faults directly. It can be seen that the D⁰X emission is suppressed by the stacking faults, whereas the different 3.288 to 3.418eV emission are coming from the faults. By using diffraction contrast to determine the specific types of stacking faults present, they were able to correlate the crystallography of the faults to the particular emission lines.

In short, stacking faults are both clearly sites for defect recombination, and emit sub-bandgap luminescence.

2.4 Polarization and segregation effects

2.4.1 Polarization: theory

The wurtzite phases of the III-nitrides are non-centrosymmetric; as such, the III-nitrides exhibit significant spontaneous (sp) and piezoelectric (pz) polarizations.
Here, we will discuss these polarizations, following the treatment of Bernardini and Fiorentini.83,84,85,86,87,88,89

First, the macroscopic polarization $P$ of a solid is defined, and will have units of $\text{C/m}^2$. $P$ will be the sum of the spontaneous polarization in the equilibrium structure $P^{sp}$, and the strain-induced piezoelectric polarization $P^{pz}$. This piezoelectric polarization is defined from strain $\varepsilon$ and the piezoelectric constants tensor $e_{ij}$:

$$P_{ij}^{pz} = \sum_j e_{ij} \varepsilon_j$$

(2-3)

In commercial GaN-based devices, crystal growth proceeds along the $c$ direction of the unit cell; due to lattice mismatch at features such as the QWs, there will be pseudomorphic strain in the structures. This results in the strain $\tilde{\varepsilon}$ in the QWs or cladding layers.

Due to this $c$ growth, the polarization component of the most interest will be $P_{3}^{pz}$, along the $c$ axis. Wurtzite has three non-vanishing, independent piezoelectric tensor components, $e_{33}$, $e_{31}$, and $e_{15}$. From equilibrium lattice parameters $c_0$ and $a_0$, and strained lattice parameters $c$ and $a$, the values of strain can be found as:

$$\varepsilon_1 = \varepsilon_2 = (a - a_0)/a_0$$
$$\varepsilon_3 = (c - c_0)/c_0$$

(2-4)

From these quantities, the polarization $P_{3}^{pz}$ is found:

$$P_{3}^{pz} = e_{33}\varepsilon_3 + e_{31}(\varepsilon_1 + \varepsilon_2)$$

(2-5)

The $e_{15}$ component enters through shear strain and is thus neglected in this pseudomorphic case.

Polarizations in materials are relevant as they lead to electrostatic fields; polarization and field are related by the equation:

$$\tilde{E} = \frac{\tilde{D} - \tilde{P}}{\varepsilon_1 \varepsilon_0}$$

(2-6)
Where $\varepsilon_r$ and $\varepsilon_0$ are the relative and absolute dielectric constants, and $\mathbf{D}$ the displacement field. If $\varepsilon$ is the total dielectric constant for the material, and $\mathbf{D}=0$, then $\mathbf{E}=\mathbf{P}/\varepsilon$. Displacement $\mathbf{D}$ arises from the free charge distribution; if $n(\mathbf{r})$ is the electron distribution and $p(\mathbf{r})$ the hole distribution, then:

$$\nabla \cdot \mathbf{D} = q[p(\mathbf{r}) - n(\mathbf{r})]$$

(2-7)

Where $q$ is the elementary electronic charge. (2-7) is the Poisson equation in vector form.

The PZ polarization comes from the deformation of the low-symmetry lattice, resulting in a moment within the unit cell. The spontaneous polarization occurs because the wurtzite materials have a non-zero polarization in even their unstrained equilibrium geometry. This will not lead to a measurable electrostatic field in a large sample. In a large sample at non-zero temperature, thermally-generated electron-hole pairs will be continuously formed. They will drift under the action of $\mathbf{P}^{sp}$ and its consequential $\mathbf{E}^{sp}$; this will result in the formation of a displacement $\mathbf{D}$. When $\mathbf{D}=-\mathbf{P}^{sp}$, the spontaneous polarization is fully screened, and no macroscopic electrostatic field exists. Because $\mathbf{D}$, $\mathbf{E}$ and $\mathbf{P}$ must all vanish outside a sample, the spontaneous polarization is cancelled by surface states in such a way that:

$$\int_{z=h}^{z=0} \{q[p(\mathbf{r}) - n(\mathbf{r})] - \nabla \cdot \mathbf{P}^{sp}\}dz = 0$$

(2-8)

In which $h$ is the width of the surface region. Bernardini and Fiorentini suggest that $h$ will be ~100s of Ångstroms; thus, the free carriers can form a two-dimensional electron (hole) gas (2DEG, 2DHG) at the surface to screen the polarization. Ridley suggests that absorbed ions plus the migrating electrons will cancel the bulk electric fields. In a nanostructure, however, differences in spontaneous polarization across an interface result in non-cancelled polarization.

In the III-nitrides, the values of the piezoelectric or spontaneous polarizations tensors are anomalously high, approximately one order of magnitude greater than those encountered in II-VI or other III-V semiconductors. As such, the effects of
these fields in GaN-based devices will be profound. Continuing to follow the
treatment of Fiorentini et al., it is possible to examine the magnitudes of these
fields within a semiconductor nanostructure. First, it is assumed that the
semiconductor consists of a finite superlattice of alternating cladding layers C and
active layers A (i.e., quantum wells). The piezoelectricity-induced electric fields $E^{pz}$
in the active or cladding layers can be found from:

$$
\begin{align*}
\bar{E}_A^{pz} &= 4\pi L_C (\bar{P}_C^{pz} - \bar{P}_A^{pz})/(L_C\varepsilon_A + L_A\varepsilon_C) \\
\bar{E}_C^{pz} &= 4\pi L_A (\bar{P}_A^{pz} - \bar{P}_C^{pz})/(L_C\varepsilon_A + L_A\varepsilon_C)
\end{align*}
$$

In which case $\varepsilon_{A,C}$ is the dielectric constant and $L_{C,A}$ the thicknesses. Thus, in the
general case, if $P_A \neq P_C$, an electric field will be present. Fiorentini et al. further
mention three special cases:

1a. Active layer lattice-matched to the substrate: $P_A = 0$

1b. Cladding layer lattice-matched to the substrate: $P_C = 0$

2. $L_A = L_C$: then $E_A = -E_C$

3. $L_A \ll L_C$: then $E_C \approx 0$ and $E_A^{pz} = 4\pi P_A^{pz}/\varepsilon_A$

It is also suggested that, to the first order, the piezoelectric field in alloys can
be interpolated using a Vegard-like law; for example, for an Al$_x$Ga$_{1-x}$N layer, with PZ
tensors $e^{GaN}$ and $e^{AlN}$, and strain $\tilde{\varepsilon}$, the PZ polarization would be:

$$
\bar{P}^{pz} = \left[ x e^{AlN} + (1-x)e^{GaN} \right] \tilde{\varepsilon}
$$

The spontaneous polarizations also treated in a similar manner. In Figure 14,
the spontaneous polarizations for the III-nitrides are plotted as a function of lattice
constant; the data is after Ref. 84.
Fiorentini et al.\textsuperscript{84} also provide expression for the spontaneous-polarization-derived electric fields $E^{sp}$:

$$
\vec{E}^{sp}_A = 4\pi L_C (\vec{P}^{sp}_C - \vec{P}^{sp}_A) (L_C \varepsilon_A + L_A \varepsilon_C)
$$

$$
\vec{E}^{sp}_C = 4\pi L_A (\vec{P}^{sp}_A - \vec{P}^{sp}_C) (L_C \varepsilon_A + L_A \varepsilon_C)
$$

Another Vegard-like rule is provided for the calculation of spontaneous polarization of alloys Al$_x$In$_y$Ga$_{1-x-y}$N:

$$
\vec{P}^{sp} = x\vec{P}^{sp}_{AlN} + y\vec{P}^{sp}_{InN} + (1 - x - y)\vec{P}^{sp}_{GaN}
$$

As a general rule, the MQW structures will have both heterointerfaces (causing $E^{sp}$) and strain (causing $E^{pz}$), and the electric field in cladding or active layers will be:

$$
\vec{E}_{A,C} = \vec{E}^{sp}_{A,C} + \vec{E}^{pz}_{A,C}
$$

Estimates of values for the spontaneous polarizations and PZ tensor components are given by Ref. 83 and are reproduced in Table II below.
Recalling the III-nitride lattice parameters listed in Table I, and in conjunction with the data in Table II, two important cases become apparent:

1. An AlGaN active layer:
   a. The strain will be small, as $\Delta a$ from AlN to GaN is ~2%; because the strain is small, $E^{pz}$ will be small
   b. The difference in $P^{sp}$ is large (~95% difference), so $E^{sp}$ will be large

2. An InGaN active layer:
   a. The strain will be large, as $\Delta a$ from InN to GaN is ~10%; because the strain is large, $E^{pz}$ will be large
   b. The difference in $P^{sp}$ is small (~10% difference), so $E^{sp}$ will be small compared to $E^{pz}$

In these two cases, one polarization component will be dominant, but they are both of the same order and it would be improper to neglect either one in calculations\textsuperscript{86}.

The above discussion outlines the cause and magnitudes of polarization fields in III-nitrides. The question now becomes, "What effects do polarization fields in III-nitride nanostructures have?"

The primary effect is that of the quantum confined stark effect, QCSE, as described by Miller et al.\textsuperscript{91}. Consider an InGaN SQW structure, with or without an internal electric field across it, Figure 15.
The first, and most apparent, consequence is that the electron-hole recombination energy is lower in the QCSE case than in the flatband case. In other words, the QCSE causes a redshift of the photons resulting from recombination in the quantum well.

A second result of the QCSE is a reduction in quantum-mechanical oscillator strength; that is, the recombination time for a given EHP injected into the QW increases. This is apparent when the wavefunctions $\psi$, $\phi$ of the carriers are considered, as illustrated in Figure 16.
In this case, the wavefunctions are $\psi_e(z)$ for the electrons and $\phi_{h^+}(z)$ for the holes. The sample's XY extent is assumed large in relation to the Z-thickness $L_z$ of the quantum well (~$10^{-4}$ m compared to ~$10^{-9}$ m in practical GaN devices), so that in-plane effects are neglected, turning the problem into a one-dimensional case. Oscillator strength $a$ is represented as:

$$a = \oint_{z=-\infty}^{z=+\infty} \psi_e^* \phi_{h^+} \, dz$$  \hspace{1cm} (2-14)

As is apparent from Figure 16, the centroids of the wavefunctions in the QCSE case do not overlap, and as such the overlap integral of the wavefunctions will be lowered. As a consequence, the oscillator strength will also be lowered.

In practice, this results in slower optical processes in the quantum wells, such as lasing. By slowing the radiative (desirable) recombination paths, but not necessarily the non-radiative (undesirable) paths, the QCSE may negatively impact optical quantum efficiency by allowing the non-radiative paths to deplete the EHPs from the quantum wells.

It is possible for free carriers to counteract, to at least a degree, the effect of the polarization fields. Electrons and holes will pool to opposite sides of the
quantum well. In the case of Figure 16, the polarization field applied to the quantum well is negative (i.e., it points in the -z direction). The attraction between the electrons and holes, however, results in a positive field (pointing in +z) and as a result the two fields will act to counterbalance each other, at least to some degree.

The level of injection needed to screen the polarization field depends upon a number of factors, and will be explored more in Section 5.1. For a quick example, Fiorentini et al.\(^8\) calculated that an injection level of \(2 \times 10^{13}\) EHP/cm\(^2\) would not quite return to flatband a 5-nm GaN QW between Al\(_{0.2}\)Ga\(_{0.8}\)N cladding layers.

This screening effect will die off as the carriers that cause it recombine; doping, however, may result in the transfer of carriers from the cladding layers into the QW, and thus result in some degree of screening that will not decay due to recombination back toward equilibrium\(^8\).

An additional complication is that of the potential change due to the electric field across a QW. Consider Figure 17, a simple case of a QW that is \(L_z\) wide and has a total field of \(E\).

![Figure 17: Magnitude of potential drop across a QCSE quantum well.](image)

This shows a potential drop \(q \cdot V\), which is to a first approximation \(L_z \cdot E\). However, this potential drop cannot exceed, or even equal, that of the bandgap.
Consider an n-doped type GaN layer; when \( q \cdot V \approx E_{\text{gap}} \) (where \( E_{\text{gap}} \approx 3.4 \text{eV} \) for GaN), the Fermi level will approach the band edge and a large quantity of carriers will be introduced into the QW, which will cause screening, and thus limit the potential change to \( <E_{\text{gap}} \). The formation of a 2DEG or a 2DHG is possible when these self-doping effects are high enough. This will become most pronounced in QWs with high fields, QWs that are wide, or in thick MQW superlattices, as the potential change across the many QWs will be additive. These effects are discussed in detail by Bernardini and Fiorentini\(^86\). Ridley\(^90\) also examines this issue, although applied to the case of a HEMT, rather than a QW-LED structure; Ridley provides some rules for approximation and explores the possibility of the formation of a 2DHG. Numerical calculations of Kong et al.\(^92\) also show how polarization effects lead to self doping and huge carrier densities \((\sim 10^{20}/\text{cm}^3)\) at an AlGaN/GaN interface.

All of these QCSE derived phenomena will be made more pronounced when a quantum well is larger; as has been shown by various calculations\(^86\), doubling the width (for example, from 2 to 4 nm) would result in several orders of magnitude reduction is oscillator strength, as well as a significant (\(\sim 100\text{s of meV}\)) redshift of emission. The consequences of the QCSE are expected to be highly super-linear with respect to \(L_Z\).

### 2.4.2 Segregation: theory

Solid-phase immiscibility was first predicted for the InN-GaN system by Ho and Stringfellow\(^93\) in 1996. They explained that the large difference in atomic spacing (i.e., lattice parameter) between GaN and InN would lead to a solid-solid miscibility gap. Their predicted phase diagram is given as Figure 18.
The most important feature of this graph is that, indeed, a solid-solid miscibility gap exists. The dashed lines indicate a spinodal. For example, at 800°C, a typical InGaN QW growth temperature, the solubility of InN in GaN is ~6%; this indicates that growth of blue or green devices, which require larger additions of InN, will likely not result in homogenous alloys. Thus, it is predicted that InGaN grown in commercial MOCVD or MBE processes will be metastable and can suffer phase-separation into InN-rich and GaN-rich regions.

Later results\textsuperscript{94,95} indicated that the biaxial stress present in InGaN QWs may have a significant effect on the exact shape of the spinodal diagram; this is illustrated schematically in Figure 19.
Although the exact degree of shift will be a function of the details of the growth conditions, QW thickness, and strain state, it does seem clear that stress can have a significant affect on the overall thermodynamics of the situation.

What, exactly, will result from the prediction of immiscibility is a more complicated question. First, it would be anticipated that InN-rich areas in the QWs would have lower bandgaps than the surrounding GaN-rich areas; depending on the size and dimensionality of the InN-rich regions, quantum disks or quantum dots (QDs) could potentially form. Certainly, if the structures have nanometer-scale dimensionalities, quantum confinement effects will result, even if the morphology is not truly dot-like.

If nm-scale InN-rich areas were to form, it would be expected that free carriers or excitons within the quantum well should thermalize into those narrow-bandgap regions. Recombination from these quantum-confined regions would therefore be redshifted relative to the non-segregated case. Figure 20 shows an illustration of such a hypothetical bandstructure.
This effect will resemble the polarization-induced QCSE. As discussed above, in a QW affected by an internal electric field and subject to the QCSE, injecting more carriers will result in stronger screening of the internal field by the carriers and thus a lessening of the QCSE, resulting in a blueshift of the emission relative to the low-injection limit.

In a QW that has segregation-induced areas of smaller bandgap, an increase in injection density will result in a filling of the lowest energy states in the segregated areas' density of states first, and a progressive filling of their DOS. This will eventually lead to complete saturation of this DOS, and spilling of the injected carriers into the QW. As injection increases yet more, the QW DOS will fill even further. In effect, the quasi-Fermi-levels for the electrons and holes will move farther apart, resulting in a blueshift of the emission from the segregated QW with increasing injection.
Thus, both the QCSE resulting from polarization and state-filling due to the presence of segregation conceptually result in a blue-shift of emission with increasing injection, such as from EL drive current of CL or PL excitation density.

The III-nitride literature has long debated whether InN-segregation or polarization fields dominate the optical processes in QWs; compelling arguments can be made for either case. However, the correct answer appears to be "it depends;" the experimental evidence pointing to this conclusion shall be reviewed in the next section.

2.4.3 Polarization vs. segregation: experiment

The literature regarding the polarization-vs.-segregation issue contains many different reports, often contradictory to each other. In the following sections, we shall discuss three broad classes of experiments addressed at these issues:

1. Experiments intended to prove the existence of InN-segregation and QDs inside QWs, or to measure them

2. Experiments intended to prove the existence of polarization fields in QWs, or to measure them

3. Experiments that measure the effects of segregation or polarization, and attempt to explain these effects in terms of one or both phenomena

First, a review of experiments intended to prove the existence of InN-segregation and the so-called quantum dots. These studies are performed by attempting to image the crystallographic structure. These structural imaging techniques are all some variation of TEM; the most common variant is HRTEM. Other variations include Z-contrast STEM, convergent beam electron diffraction (CBED), energy-dispersive X-ray spectroscopy in STEM (STEM-EDS), STEM-EELS mapping, and energy-filtering TEM (EFTEM), which is a variation on EELS.

The studies of localized luminescence used either CL or near-field scanning optical microscopy (NSOM) techniques such micro-photoluminescence (µPL) or EL-imaging.
Kisielowski et al.\textsuperscript{96} were the first to apply HRTEM with the intent of mapping In-distribution in InGaN QWs. They examined a 5nm thick In\textsubscript{0.43}Ga\textsubscript{0.57}N QW grown at 823°C; in comparison to the results of Ho and Stringfellow\textsuperscript{93}, this is well into the unstable, spinodal regime of the miscibility gap, so significant In-redistribution would be expected. The QW was grown on GaN and capped by a thin (30nm) Al\textsubscript{0.1}Ga\textsubscript{0.9}N layer. Samples were prepared by ion-milling and KOH etching, and were examined in an 800keV HRTEM. They contend that the ion-milling + KOH etching preparation procedure provides an atomically flat sample surface with ~1nm thick amorphous layer. Because amorphous regions or thickness variations can bias quantitative HRTEM results, this is an important consideration. They observed fluctuations of scattering potential 1~2nm in size inside the InGaN QW, whereas these fluctuations are not present in either the GaN or AlGaN layers. They conclude that there exist local variations in In-content within the QW.

Ruterana et al.\textsuperscript{97,98,99} also used HRTEM to study In-segregation in devices. Their early work\textsuperscript{97} showed little difference between the quality of underlying GaN layers from the In\textsubscript{x}Ga\textsubscript{1-x}N layer if x was small, ~0.1 to 0.2. As x increased, the quality of the QW layer decreased. Later work used careful strain-analysis from the numerical evaluation of HRTEM images\textsuperscript{98}. They determined, again, that In-segregation was present, in agreement with other works. They also determined that the average In-concentration in the clusters was ~25%; later works indicated clusters approaching 100% In.

A series of papers by Neubauer et al.\textsuperscript{100,101,102,103} examined the effects of growth parameters on In-segregation. The first conclusion reached is that, for constant precursor flow parameters during growth, increases in substrate temperature resulted in reduced In-incorporation into the film\textsuperscript{100}. Fluctuations in local In-content were also found in all examined samples. They also suggest that, because the PL emission peak is of lower energy than expected from the overall In-composition of the active layers, this redshift may indicate the QCSE is simultaneously present with the segregation.
They further performed PL studies on the same samples, and concluded that polarization effects, not segregation effects, dominated the optical properties of the QWs. However, the QWs grown in this study varied from 5.2 to 17.6 nm in thickness, which are rather larger than QWs used in commercially interesting devices. As the QCSE becomes much more pronounced in thick wells, it comes as no surprise that these particular devices showed QCSE dominance. By growing a thick (~100 nm) InGaN layer, they were able to show a significant effect of strain on cluster In-content; clusters near the GaN/InGaN interface, where strain was highest, had significantly increased peak In contents compared to those in the relaxed part of the layer. As a general result, they found that growth technique or conditions did not have a large effect on the overall In-segregation behavior. Lastly, and most interestingly, they noted both the short-range In-segregation already mentioned, and lower-intensity, longer-range fluctuations ~100 s of nm in extent. Later, luminescence experiment results will be discussed that show length scales consistent with this result. Other papers have reported essentially the same results as those reviewed above.

Stirman et al. used in-TEM CBED to examine local lattice parameter, which was then used to determine local In-concentration. However, they performed these experiments on 10 nm thick InGaN layers, which are rather thicker than the QWs used in practical devices. Indeed, the pseudomorphic strain would likely be relaxed in a 10 nm thick well, and the growth modes of the segregated regions would likely be different. In principle, however, if this technique was implemented in a TEM with sufficient resolution to probe areas of segregation in a thin QW, and the samples were sufficiently thin, CBED could be a valid means of measuring local In-content. However, thin samples will have very low dynamical contrast from CBED, and it might not be possible to get contrast within the disks from a sample that is too thin.

Z-contrast (ZC) STEM, also known as atomic number contrast or high-angle annular dark field (HAADF) imaging, images with electrons scattered to high angle, as is roughly sensitive to the square of the atomic number. It is thus sensitive to changes in composition, such as In-segregation. Barnard et al. and Watanabe et
al.\textsuperscript{110} used ZC-STEM to image InGaN quantum wells. Barnard et al.\textsuperscript{109} drew no conclusions regarding In-segregation from their ZC-STEM data, but Watanabe et al.\textsuperscript{110} did observe segregation in QWs.

Fay et al.\textsuperscript{111} used EFTEM and STEM-EDS to seek In-segregation; they saw In-fluctuations in the EFTEM image, but felt point-defect artifacts may be to blame. EDS had insufficient signal-to-noise to draw any conclusions regarding In-segregation.

Takeguchi et al.\textsuperscript{112} examined 10nm-thick InGaN wells using Z-contrast, HRTEM, and electron holography. They found certain regions showed significant segregation of In to the lower part of the wells, resulting in a two-step well structure in which the deeper area was richer in In. Given the thicknesses of these wells compared to the 2~3nm common in commercial devices, it is not clear if their results are generally applicable.

The topic of artifacts must now be discussed, as it throws doubt on all of the above results. Although thermodynamics does suggest, perhaps even require, the formation of In-rich areas, the effect of the high-voltage, high-brightness electron beam necessary for lattice-resolution imaging on the crystal is more than sufficient to cause agglomeration of point defects, which can be confused for growth-induced In-segregation. Various examples will now be discussed.

Among the earliest reports is that of Fay et al.\textsuperscript{111}, who mentioned in passing that point-defect clustering must be removed from EFTEM images in order to obtain true elemental maps. Even before then, in the original paper discussing experimental evidence for segregation, Chichibu et al.\textsuperscript{113} correlated conventional TEM data to optical data. Their TEM micrographs showed a mottled dark-spot contrast, which they attributed to QDs, and they explained their optical data in light of these QD observations. Chichibu et al. admitted that the dark spot contrast "could have been formed during the ion-milling process," but that "the structure can still be explained as replica of In-rich region \textit{[sic]} since InN is softer than GaN\textsuperscript{113}." However, as shown in Figure 20, reprinted from Ref. 113, the dark spot contrast is present in both the InGaN QWs and the nominally 100% GaN cladding layers, which throws
doubt on the Chichibu et al. explanation of simple replication. Ponce et al.\textsuperscript{114} observed that argon-ion-milling, the most common specimen preparation technique for HRTEM, can cause In-droplet formation on the surface of the specimen, which could be confused for segregation.

![Figure 21: Micrograph purported to show In-segregation.](image)

Figure 21: Micrograph purported to show In-segregation.

Reused with permission from Ref.113. Copyright 1996, American Institute of Physics.

However, more important results are those in which segregation is observed to form under the irradiation of the electron beam, in the range of one to several minutes elapsed time. Smeeton et al.\textsuperscript{115} found, if they captured a HRTEM image within a few seconds of exposing the sample area to the beam, subsequent strain analysis showed little or no segregation in the QW; however, if the sample was irradiated for a few minutes, very large In-segregations were measured. By correlating their XRD and TEM data, they concluded that there was little In-content-fluctuation in the samples they examined, in contrast to most other published InGaN HRTEM results. Smeeton et al.\textsuperscript{116} continued this work with carefully prepared HRTEM samples by wedge polishing on diamond lapping films followed by 3.2keV Ar\textsuperscript{+}-ion-milling. The low Ar\textsuperscript{+} ion energy and incidence angle should result in relatively little preparation-induced specimen damage, although any ion-based
technique will result in some amount of lattice damage. Images recorded within 20 seconds of sample exposure to the TEM beam showed effectively no segregation, whereas after 210 seconds at 200kV, areas of ~60% "InN" were observed. Because this damage was worse at the bottom of the QWs, they suggested the possibility that the strain-induced PZ fields may play a role in the mechanism. Again, they conclude that no gross clustering of In is present in the un-exposed specimens. O'Neill et al.\textsuperscript{117} found effectively the same result in what they call "moderate" electron doses of ~100A/cm\textsuperscript{2}; they suggest that the problem will be far worse at analytical TEM electron doses.

Li et al.\textsuperscript{118} quantified these effects, and also found qualitatively different behavior. They observed, in contrast to the results of Smeeton et al.\textsuperscript{115,116}, that In-inhomogeneity was indeed present within the first 20 seconds of electron beam exposure. They then examined different areas of the same quantum well as a function of time; this data is reprinted as Figure 22.

![Graph](https://via.placeholder.com/150)

Figure 22: Two different areas of an InGaN QW, showing segregation vs. time. Reused with permission from Ref.118. Copyright 2005, American Institute of Physics.
As can be seen in Figure 22, In-concentration in the two different regions is effectively constant for the first ~50sec; after that, region A increases to nearly 90% In, whereas region B remains roughly constant. Li et al.\textsuperscript{118} admit that the possibility of damage within the first 20sec cannot be completely excluded, but that the near-constant behavior up to 50sec makes this unlikely. They thus conclude that In-clustering is present in their as-grown material, and that experiments in which the HRTEM data is collected in <~60 seconds of beam exposure should be valid.

The most comprehensive study of this topic is that of Bartel et al.\textsuperscript{119} They claim that their sample preparation procedures eliminate surface roughness variations larger than atomic steps, and thus the thickness artifacts are minimized. At 150kV, they performed a time-resolved experiment and found that instabilities in the microscope and environment can mimic the results noted above\textsuperscript{116,117,118}. Analysis of the statistical fluctuations of signal within a single image lead them to note that the standard deviations of measured indium contents within the quantum wells were significantly larger than those observed in the GaN matrix; this indicates that the statistical In-composition deviates from a simple Poisson distribution expected from miscible alloys (i.e., AlGaAs), and indicates that the In-concentration is not homogenous. However, neither is it directly measurable from the data, due to averaging through the foil thickness. This data was compared to 800kV data and found the same trends. They then concluded that reliable In-fluctuation measurements can be obtained from 150kV data, but that microscope or environmental instabilities, as well as sample-preparation surface roughness, can introduce noise as large as the signal itself.

Continuing their experiments to 800kV, several advantages of the higher-energy analysis were noted. Firstly, the regions of near-constant contrast in thickness-defocus space are larger than at lower kV. Second, they found that the InGaN samples themselves were stable under the beam for up to 2 minutes, likely due to the smaller inelastic cross-section of high-kV electrons. Thicker crystals can be imaged, which reduces strain relaxation and other thin-foil effects. They found
highly repeatable data for reasonable exposure times (~2min) with the 800kV instrument.

In short, the work of Bartel et al.\textsuperscript{119} concluded that valid measurements of In-clustering can be made with HRTEM if thickness and defocus variation are controlled for, but that this is quite difficult in the 150-300kV range, where most HRTEMs operate. They imply that the changes in the QWs observed by the other authors may be due to microscope instabilities, but as the other authors\textsuperscript{115,116,117,118} report monotonic increases in In-clustering with time, it seems that any microscope instabilities are superimposed upon true beam-induced clustering.

In summary, TEM techniques, predominately HRTEM, are used to attempt to prove the existence of In-segregation and QD formation within InGaN quantum well devices. Although thermodynamics strongly suggests cluster formation, and various TEM studies have purported to show their existence, damage caused by the energetic electron beam, and thickness-defocus instabilities in the sample or microscope itself, means that most of these studies must be looked at suspiciously.

That concludes techniques to directly image QDs. Now, direct measurement of internal electrical fields will be discussed. The technique used to directly measure internal electric fields is off-axis electron holography in the TEM. This was done on InGaN QWs by a number of different groups\textsuperscript{120,121,122,123,124,125,126,127}.

The earliest results\textsuperscript{121,122} made estimates based upon the PZ field in a 1.5nm QW of In\(_{0.52}\)Ga\(_{0.48}\)N, in which they calculated \(E_{PZ} \approx -4\)MV/cm, based on stress and PZ arguments, and measured fields of the same value via EH. It appears these early results ignored the effects of spontaneous polarization. Cai et al.\textsuperscript{125} found a field of \(\approx -2.2 \pm 0.6\)MV/cm in a 2.7nm wide In\(_{0.18}\)Ga\(_{0.82}\)N QW, and also began to observe substructure in the EH data, specifically "bumps" in the inner potential curve at the top and bottom interfaces of the QW. They attributed the anomalies at the interfaces to interface dipole charges, polarization sheet charges, and free carriers\textsuperscript{126}.

An important result, combining EH with CL data, is that of Stevens et al.\textsuperscript{127} Here, In\(_{0.13}\)Ga\(_{0.87}\)N layers of thicknesses from 2 to 10nm were grown, and the peak emission wavelength measured by CL. The fields within the QWs were then
measured with EH. It was found that the peak CL wavelength decreased rapidly as the QW width increased from 2 to ~6nm; the trend fell in between a flatband (0MV/cm) square-well computation and that for a well suffering the QCSE with a 2.2MV/cm internal field. Beyond ~6nm, however, the CL emission peak changed very little. This was matched by the EH data, which showed a constant internal field of ~1.7MV/cm up to 6nm width; beyond ~6nm, the electric field decreased abruptly to ~0MV/cm at 10nm thickness, which indicates relaxation of the strain. TEM imaging of the thick (i.e., 10nm) QWs showed spotty contrast near the bottom interfaces, which could be explained as misfit dislocations, which is consistent with strain relaxation. They further note that the critical thickness for relaxation in In$_{0.1}$Ga$_{0.9}$N would be ~10nm$^{(108)}$, and they thus contended that the collapse of the EH-measured internal fields and CL-measured QCSE for the >6nm QWs was due to relaxation of the stresses in the QWs, and thus a reduction of the PZ field toward zero. This is shown in Figure 23.
To summarize electron holography, internal electric fields have been measured in InGaN QWs, and other related III-nitride structures. However, agreement between theory and experiment is not perfect, although trends are consistent. Due to issues such as sample preparation, data reduction, and imaging artifacts arising from dynamical diffraction, EH is only progressing slowly as a GaN-related experimental technique.

Figure 23: Effect of quantum well width on potential vs. position, internal field, and emission energy.

Reused with permission from Ref. 127. Copyright 2004, American Institute of Physics.
The last part of this section will review experimental results, mostly optical, that can be interpreted in terms of either internal fields and the QCSE, or in terms of In-segregation and QD formation.

Many different authors\textsuperscript{113,128,129,130,131,132,133,134} reported results in which an increase in excitation density -- whether through EL, PL, or CL -- resulted in a blueshift of the peak position. This can be interpreted in terms of carriers in the quantum well being separated by the internal field and screening the field in return, bringing the QW closer to flat-band conditions. However, because of band-tail state filling, blue-shifts with excitation density can also be interpreted in terms of the In-segregation model.

Since GaN devices tend to be grown in the (0001) direction with the n-side below the p-side, applying a reverse bias to the device will tilt the QWs toward flat-band, and possibly past flat-band, if the reverse bias is large enough. This result, again attributed to internal fields and the QCSE, was found by various authors\textsuperscript{132,133,135,136,137,138}.

These results are similar in concept to the earliest direct measurement of deliberately induced internal electrical fields in semiconductor heterostructure (Caridi et al.\textsuperscript{139}, in 1990), which involved the measurements of \(~10^5\) V/cm fields in zincblende grown III-V semiconductors grown in the piezoelectric (111) direction.

It is also predicted from the QCSE that, with all other factors being equal, a thicker QW will have a lower-energy (redshifted) emission in comparison to a thinner QW. This has, indeed, been observed\textsuperscript{128,131,134,140}. An important corollary to this result is that, as the QWs become very wide, the redshift will level off at some particular emission energy, rather than redshifting monotonically with width (for example, Ref. 128). This could potentially result from one of two reasons. First, as the width of the well increases at constant internal electric field, the voltage drop across the well will approach that of the GaN bandgap, which will result in band-bending limitation to the voltage drop permissible across either a SQW or summed across MQWs. Secondly, as the width of the QWs increase, the stress in the QW can relax due to the formation of misfit dislocations in QWs that are above the critical
thickness of relaxation, thus resulting in a lessened internal PZ field due to stress reduction. Wang et al.\textsuperscript{141} grew In\textsubscript{0.11}Ga\textsubscript{0.89}N QWs of various thicknesses (1.4-3.9nm) and examined them by excitation-density resolved PL. They concluded that, perhaps surprisingly, the QD-like-effect of peak-shoulder-formation was more pronounced in wider wells; blueshifts were observed as excitation increased, but this could, of course, be interpreted in terms of either the QCSE or In-segregation. Quantum efficiency was found to improve with increased injection density, which is harder to explain in light of the segregation hypothesis, but reasonable considering that the QCSE can lengthen radiative recombination times and allow more non-band-edge recombination. Increased injection screens the QCSE and should reduce the radiative lifetime in comparison to the non-radiative.

Chichibu et al.\textsuperscript{134} noted that, for constant $x_{\text{In}}$, as the width of the quantum well increased, PL intensity decreased. This could be interpreted in terms of the QCSE either directly quenching luminescence, or the increased recombination time allows more non-radiative recombination to take place.

Further, as doping of the barrier layers increases, blueshifting of the emission peak is observed for otherwise identical structures, under given excitation parameters\textsuperscript{134}. This is because doping-derived carriers from the barrier layers can migrate into the quantum wells; the presence of these carriers will result in screening of the electrical field, in a matter similar to that of excitation-density experiments discussed above. This doping-effect is difficult to explain by the In-segregation hypothesis.

Imaging studies give evidence for In-segregation, but do not as of yet have the resolution to confirm or refute the nm-scale In-rich regions purported to exist from TEM studies. Chichibu et al.\textsuperscript{134} discussed a CL spectrum taken from a 10x10\textmu m area; the width of the QW peak is much wider than that of any spot-spectrum; further, the spot-spectra all have different peak values. They infer this to mean that the spot-spectra are from areas of different In-concentration. In a separate paper, Chichibu et al.\textsuperscript{142} present CL micrographs (reprinted as Figure 24) taken on an In\textsubscript{0.03}Ga\textsubscript{0.97}N SQW sample. (The sample was grown as cubic-GaN to
eliminate polarization fields while, according to Chichibu et al.\textsuperscript{142}, maintaining the In-segregation effect.)

![3-nm-thick In\textsubscript{0.03}Ga\textsubscript{0.97}N SQW](image)

Figure 24: CL emission showing different energies with position.
Reused with permission from Ref. 142. Copyright 2000, American Institute of Physics.

The two micrographs shown are for different emission wavelengths, and show non-coincident bright and dark areas, which is interpreted by the authors to be evidence of localizations of In-segregation. Experimental conditions of 3kV incident beam energy were noted; the authors contend\textsuperscript{143} a spatial resolution \(\approx 60\) nm, which may be slightly optimistic. Additional CL data was reported by Ponce et al.\textsuperscript{114}, in which QW spectra were found to consist of a high-energy main peak and a lower-energy shoulder peak. Images taken at these two energies gave opposite contrast, indicating that the contributions to the overall peak shape were, indeed, from different areas. They reported experimental conditions of 5-10kV accelerating voltage, which is consistent with the \(\approx 0.25\) to \(0.50\) \(\mu\text{m}\) resolution observed in the image data. \(\mu\text{PL}\) data in the near-field scanning optical microscope (NSOM) gave similar results\textsuperscript{144,145}, in which the shift of the PL peak was mapped at the \(\approx 100\) s of nm scale, and again attributed to In-segregation.
CL peak shifting was shown to correlate with directly-measured In-concentrations in a 180nm-thick InGaN layer. Martin et al.\textsuperscript{146} used an electron-probe microanalyzer (EMPA) equipped with a CL collection system and wavelength-dispersive X-ray spectrometers (WDS) for compositional measurement. They found that areas richer in In had lower CL intensity and lower-energy CL peak wavelength. However, Martin et al.’s results are for a thick InGaN layer, not QWs, and will have spatial resolution approaching 1\(\mu\)m for the compositional measurements; thus, the In-segregation measured therein are not quantum-level. Gerthsen et al.\textsuperscript{103} (above) noted In-fluctuations \(\sim\)100nm using TEM; the CL and \(\mu\)PL results discussed above seem more consistent with Gerthsen et al.’s result than the "quantum dot" hypothesis.

All of those optical-technique results have spatial resolutions from \(\sim\)100 to \(\sim\)1000nm. Thus, they may be imaging broad, low-frequency spatial variations in In content, but no published results have directly and un-ambiguously imaged emission resulting from nm-scale In-fluctuations, as purported from TEM studies. The closest results to this are those involving the use of a light-blocking shadow-mask over the InGaN layer, in order to increase artificially the resolution of the probing technique. In the above-reviewed publications, even the "narrow" emission lines obtained from spot-CL measurements were \(\sim\)30meV compared to \(\sim\)90meV for the wide-area scan\textsuperscript{143}.

By using lithography to place a metal shadow-mask on an In\textsubscript{0.15}Ga\textsubscript{0.85}N QW sample, Schomig et al.\textsuperscript{147} claimed to be able to achieve \(\mu\)PL spatial resolution below the diffraction-limited spot size. As the aperture size decreased from 25\(\mu\)m to 175nm, the PL spectrum changed from a broad, wide Gaussian-like emission peak to a spectrum made up of sharp, meV-range narrow individual lines, with the number of lines falling as the aperture size was reduced. As excitation increases, QCSE arguments would suggest a shift of both the ensemble peak and individual lines to higher energy as the internal electric field was screened. In the work of Schomig et al., the ensemble peak position moved to higher energy, but the energies of the individual peaks did not change. Individual lines had FWHM as small as 0.8meV.
These results led Schomig et al.\textsuperscript{147} to conclude that individual In-rich localization centers were responsible for the discreet lines in the $\mu$PL spectra. Seguin et al.\textsuperscript{148} used shadow masking + CL to achieve substantially the same results.

Time resolved experiments, typically time-resolved photoluminescence (TRPL), involve the pulsed excitation of the sample and the observation of the decay of the luminescence after the termination of the excitation pulse. Many reports\textsuperscript{105,134,144,145,149,150,151} used this technique. The primary use of TRPL is to measure the decay time $\tau$ of the photoluminescence. Changes in $\tau$ result from either changes in recombination mechanisms or spatial separation of the electron and hole wavefunctions due to QCSE.

Chichibu et al.\textsuperscript{134,150} noted that, for constant $x_{\text{In}}$, an increase in QW thickness resulted in longer recombination decay times. This was interpreted in terms of the QCSE; as the well became wider, the EHPs were progressively more confined to opposite sides of the well, resulting in a lower oscillator strength.

In a LED with a 2.5nm SQW of approximate composition In$_{0.25}$Ga$_{0.75}$N, Chichibu et al.\textsuperscript{105} reported that the PL decays (log(intensity) vs. time) at the individual emission energies are described by a stretched-exponential function, not a pure exponential function, and they attribute this to quantum confinement, rather than the QCSE. This is similar to the results of Ichimiya et al.\textsuperscript{149}, who observed the same phenomena in an In$_{0.07}$Ga$_{0.93}$N thick (200nm) layer; they attributed the change in decay time with energy to the relaxation of higher-energy carriers to localized states.

Bell et al.\textsuperscript{152} performed time-resolved cathodoluminescence (TRCL) on In$_{0.13}$Ga$_{0.87}$N SQW devices, with QW widths of either 6 or 8nm. In this experiment, the quantum wells were excited by a 130ns-long pulse of a 5kV, 400pA SEM beam. The QW cathodoluminescence was measured during the pulse and after the end of the pulse. Thus, the QWs' responses to excitation and decay from excitation were observed. Bell et al.'s\textsuperscript{152} results are shown in Figure 25.
In the 6nm quantum well, from the moment excitation begins until the cut-off of excitation, a small redshift of the QWs’ CL is observed. When the beam is cut off, a further redshift is observed as the CL decays toward zero. The redshift during excitation is interpreted as the beam-induced carriers thermalizing into the quantum-confined In-rich areas, which dominates over any QCSE-screening-induced blueshift. The redshift after excitation cutoff is interpreted to be due to the loss of QCSE screening as the carriers causing the screening recombine and are lost.

In the 8nm quantum well, a blueshift is observed from the moment of excitation until excitation cut-off. This is interpreted to be a result of the QCSE being screened by the continuously increasing number of injected carriers. Similarly to the 6nm-case, when the beam is cut off, the QW emission redshifts, again attributed to the loss of free carriers to recombination and a loss of screening.

Thus, Bell et al.\textsuperscript{152} concluded that In-segregation effects are more pronounced in the narrower QW, whereas the QCSE is more pronounced in the wide QW. Unfortunately, this experiment was performed on wells that are rather
wider than those used commercially, and only on violet-range devices ($X_{\text{In}} \approx 13\%$). Although interesting, the generally applicability of these results is somewhat ambiguous.

Nakamura et al.$^{153,154,155,156,157}$ performed a series of experiments that shed further light on the QCSE-vs.-QD issue. By growing GaN, InGaN, or AlGaN layers in non-polar directions, the QCSE is removed entirely, and the relative magnitudes of the QCSE-vs.-QD effects can be found. They grew a-plane ($11\bar{2}0$) oriented films, in contrast to the standard (0001)-type films. They found that the non-polar films had no appreciable blueshift with excitation, and had much higher quantum efficiencies than polar films of the same In-content. This indicates that blueshifting with excitation is primarily due to QCSE, and that the elongated recombination time associated with wavefunction separation due to the QCSE allows more carriers in the QWs to recombine non-radiatively at defects.

Further, by comparing dilute InGaN to dilute AlGaN alloys, it was found that even slight additions of In resulted in significantly improved quantum efficiency. This implies that In-segregation, even in dilute alloys, is enough to trap carriers or excitons away from non-radiative sites and reduce the rate of carrier loss to defects. So-called "semipolar" ($10\bar{1}3$) growth gave results in-between the other two cases, but was easier to manufacture than the ($11\bar{2}0$) type films.

2.4.4 Polarization vs. segregation: summary

To summarize this section, both large internal electric fields and the QCSE, and areas of InN-GaN demixing, are expected to exist in InGaN devices. Large numbers of studies have been published that attempted to study the relative optoelectronic impact of the QCSE vs. the InN-GaN miscibility gap. In short, reports such as those of Bell et al.$^{152}$ using time-resolved CL, Schomig et al.$^{147}$ using shadow-masked $\mu$PL, or Nakamura's group using non-polar growth, indicate that phase-separation must exist and result in quantum confinement of carriers, but perhaps not true quantum "dot" formation. However, due to problems associated
with HRTEM imaging, these areas of quantum confinement are difficult to observe unambiguously. Internal electrical fields and the QCSE, however, have been directly observed, and are clearly also responsible for some of the optoelectronic effects observed in III-nitride devices. The relative balance of these two effects, however, is not currently predictable \textit{a priori}, although thicker wells will, to a first approximation, tend toward QCSE domination and thinner wells to QD domination.

2.5 Summary of literature
From the above, these major points should be reiterated:

1. Dislocations in GaN are recombination sites, but their optoelectronic character is still subject to debate.
2. The yellow band appears to be a point-defect-related phenomena, but some papers have tried to associate it with dislocations.
3. Polarization fields and InN-GaN immiscibility are competing hypotheses to explain the optical character of InGaN QWs; however, it appears the two phenomena exist in complement.
3. Experimental Techniques

3.1 Scanning electron microscopy

Scanning electron microscopy uses a focused, raster-scanned beam of low-to mid-energy (0.1~30 keV) electrons to probe the surface and shallowest few microns of a specimen. Figure 26 illustrates a schematic SEM.

The condenser and objective lenses are magnetic coil lenses. Through the action of a magnetic field on the moving charged electrons, they provide focusing. Not shown are the objective aperture, which adjusts the size and convergence angle of the final spot, and the scan coils, which provide the XY rastering action of the beam.
At the top of a SEM is an electron source; this will either be a thermionic or field emission source\(^{158}\). The source will be characterized by its brightness. Brightness \(\beta\) is defined as current in the beam, \(i_B\), divided by the product of beam area, \(\pi d_b^2/4\), and solid angle, \(\alpha^2\). \(\alpha\) is the beam convergence angle; \(\pi \alpha^2\) is the solid angle, in steradians; \(d_b\) is the diameter of the beam spot, assumed circular.

\[
\beta = \frac{4i_B}{\left(\pi d_b \alpha^2\right)^2}
\quad (3-1)
\]

For gun brightness, this is measured at the first crossover, not at the sample.

Thermionic sources are tungsten wire filaments or LaB\(_6\) single crystals. A tungsten filament will be the simplest and least expensive type of electron source, and can provide very large beam currents. LaB\(_6\) sources are brighter, allowing slightly higher-resolution work than tungsten. However, neither thermionic source can provide resolution as high as field emission sources.

Field emission sources also come in one of two main varieties: cold-field-emission or Schottky emission. In field emission, an oriented, sharply pointed single crystal of tungsten is subjected to a large electric field, which can reach 10MV/cm\(^{158}\). This causes electrons to tunnel out of the metal and into the vacuum beyond. In cold-field-emission, the gun is maintained at room temperature. Schottky emission requires heating the source to \(~1800K\), which gives thermal assistance to the field-emission effect\(^{159}\). Because of the quantum-mechanical nature of field emission, the surface cleanliness and vacuum integrity of the sources are important to beam stability. Cold field emission sources will have poor (\(\sim 5\%)\) emission stability, and Schottky sources slightly better (\(\sim 2\%)\)\(^{158}\). As such, field emitters require ultra-high vacuum in the gun chamber, and cold field emitters must be flashed (heated) often to drive off adsorbed gasses. Field emitters have far smaller source-sizes than thermionic guns, and \(~3\) orders of magnitude higher brightness, and thus require less demagnification in the optics. However, despite the higher brightness, they have lower significantly absolute beam currents.
Field emission sources’ much higher brightness and smaller source-size result in higher current-densities on the sample, and thus higher resolution. A Schottky instrument will have much higher potential beam current than a cold field emitter, but slightly larger spot size and energy spread. In short, field emission sources result in a lower-current, less-stable, more expensive, but higher resolution, very bright instrument in comparison to thermionic sources.

After the electron beam is generated, it is accelerated to an energy of ~0.1 to 30keV. The beam energy will be a choice of the operator, selected in terms of beam-specimen interactions and their effects on the experiment being performed. Choice in this respect will be discussed in the section on beam-injection techniques (Section 3.2) below.

The accelerated beam must then be de-magnified onto the sample. The condenser and objective lenses perform this. Situated just above the objective lens is the objective aperture, which reduces the size and convergence angle of the final beam. The objective lens is focused by changing its operating current, which moves the crossover point of the beam up or down the optical axis. When the crossover is precisely on the surface of the sample, the beam is considered focused and the image produced by the microscope will be most crisp.

Increasing the excitation of the condenser lenses and/or use of a smaller objective aperture will result in a smaller spot-size on the sample, but also a smaller beam current on the sample. This leads to the most fundamental trade-off in SEM operation: lower signal to noise ratios will be the unavoidable consequence of higher resolution operation, if all other factors (such as voltage and working distance) are held constant.

Increasing the operating voltage will result in both higher currents and smaller spot sizes, if all other factors are held equal. However, higher voltages sample larger volumes of the specimen, as will be discussed in the beam-injection techniques section below. Using a shorter working distance can reduce spot size, but decreases the available depth-of-focus. The most important constraint on
working distance is the presence of collectors for signals such as CL or EBIC, which take up space in the chamber and prevent short working distances from being used.

Various detectors are used to produce SEM images. The most common is the Everhart-Thornley detector (ETD), which collects large numbers of secondary electrons (SE) from throughout the sample chamber via a weak electrostatic field, and also detects backscattered electrons (BSE) that enter the detector in direct line-of-sight trajectories. Also used are backscattered electron detectors, which are sensitive only the BSEs.

ETD images, usually (if slightly inaccurately) referred to as SE images, are primarily sensitive to topography, as SEs are defined as those having energy <50eV, and therefore come from the area immediately below the surface of the sample. BSE images are mostly sensitive to average atomic number, as heavier matrices backscatter more beam electrons than lighter matrices.

Other imaging modes include specimen current (SC), X-ray mapping, CL, and EBIC. In SC, a high-bandwidth picoammeter is placed between the sample and electrical ground. Changes in topography or average atomic number will changes the yield of SE and BSE. The product of SC \( (i_{SC}) \), SE current \( (i_{SE}) \) and BSE current \( (i_{BSE}) \) must be equal to the beam current \( (i_B) \) in order to maintain charge neutrality. In practice, imaging with the SC signal is much like imaging with an inverted BSE signal. The charge neutrality equation is:

\[
  i_{SC} = i_B - (i_{SE} + i_{BSE})
\]

(3-2)

Variations in \( i_{SE} \) or \( i_{BSE} \) result in \( i_{SC} \) contrast.

When the energetic electron beam hits the sample, X-rays, both characteristic and continuum, are produced. Mapping the intensity of a characteristic X-ray line vs. position allows the presence of the atom associated with the characteristic X-ray to be inferred. CL and EBIC will be discussed in their own sections below.

In short, SEM is a near-surface technique suitable for bulk samples. It can image sample chemistry through BSE imaging and X-ray mapping, and can image sample topography through SE or SC imaging. Lower voltages give better surface
sensitivity, but at the cost of spot size and beam current. High resolution is achieved at the sacrifice of signal to noise ratio.

SEM is valuable in the characterization of semiconductors, such as the III-nitrides, because of the electron beam's ability to inject electron-hole pairs into the sample; these beam-injection techniques will now be discussed.

3.2 Beam-injection techniques

3.2.1 Beam-injection overview

3.2.1.1 Generation of electron-hole pairs

The CL and EBIC techniques predicate themselves upon the injection of excess minority carriers by the incident electron beam. As a beam electron with energy in the range of hundreds of eV to hundreds of keV enters and travels through a solid material, many different inelastic processes cause energy loss of the electron, eventually bringing it to rest\textsuperscript{158,160}. Phenomena such as core-level ionization, secondary electron emission, or phonon generation are well known in SEM literature, and occur in all solid materials under electron bombardment. In semiconductors and insulators, an additional effect is possible: electrons can be promoted across the bandgap, from the valence to conduction bands, leaving behind a hole in the valence band. This effect results in the generation of electron-hole pairs (EHPs) in the material. Except for the rare cases of an intrinsic or perfectly compensated material, either p- or n-type doping will be dominant before the introduction of beam-induced electron hole pairs. In n-type material, of course, holes are the minority carrier, and vice-versa. The situation of beam-injection of electron-hole pairs for the example case of an n-doped semiconductor is illustrated in Figure 27.
Figure 27: Generation of EHP by electron beam. $e^-$ denotes electrons, $h^+$ denotes holes.

The spatial rate of beam-induced EHP generation is denoted $g(r)$, where $r = (x, y, z)$. Units for $g(r)$ are EHPs/cm$^3$-sec. There are various methods to approach the evaluation of $g(r)$. An early attempt at quantification by Donolato$^{161}$ involved the modification of the depth-dose curve approach of Everhart and Hoff$^{162}$.

Under the case of cylindrical symmetry, Donolato found that $g$ could be expressed as:

$$g(r, z; R) = \frac{g_0}{R} \frac{\Lambda(z/R)}{2\pi\sigma^2(z,R)} \exp\left[-\frac{r^2}{2\sigma^2(z,R)}\right]$$

Where $g$ is a function of radial distance $r$ and depth $z$ in the sample, with parameter $R$. $R$ is the electron range, and thus is related to beam energy and material properties. This expression gives EHPs generated per second per unit volume. The $g_0$ term is the total generation rate of EHPs/second. The $\sigma$ term accounts for the finite SEM beam spot size (assumed Gaussian) and the lateral spread of carriers;
σ^2(z,R)=0.36d^2+0.11z^3/R , where d is the SEM spot size. The term \( \Lambda \) Donolato\textsuperscript{161} took from Everhart and Hoff\textsuperscript{162}, and is defined as:

\[
\Lambda(z/R) = \begin{cases} 
0.6 + 6.21(z/R) - 12.4(z/R)^2 + 5.69(z/R)^3 & \text{for } 0 \leq z/R \leq 1.1 \\
0 & \text{otherwise}
\end{cases}
\]  
(3-4)

This is an empirical depth-dose term. In the original paper, R was taken as the Everhart-Hoff range. Another convenient form for R is the K-O range. The R\textsubscript{K-O} equation is usually written as\textsuperscript{158,163}:

\[
R_{K-O} = \frac{0.0276 \cdot A \cdot E_0^{1.67}}{Z^{0.89} \cdot \rho}
\]  
(3-5)

In which case A is the average atomic weight in the sample, in g/mol; \( E_0 \) is the incident electron beam energy, in keV; \( Z \) is the average atomic number, and \( \rho \) the density, in g/cm\(^3\). This yields a value \( R_{K-O} \) in \( \mu \)m. In this work, the term R without subscripts will imply that it is the \( R_{K-O} \) range.

The \( g_0 \) term can be calculated various ways; the most commonly used approximation, from Newbury et al.\textsuperscript{164}, is:

\[
g_0 = \frac{E_0(1-n)j_B}{E_{eh} \cdot q}
\]  
(3-6)

Here, \( g_0 \) is the mean rate of carrier generation, EHPs/sec. \( E_0 \) is the electron beam energy (eV, not keV), \( n \) the backscatter coefficient, \( j_B \) the beam current, \( q \) the elementary charge, and \( E_{eh} \) the average energy to create an electron-hole pair. The \( E_{eh} \) term is ill-defined; as a general rule, it is three times the bandgap of the material\textsuperscript{165}, or \( \approx 10 \) eV for the case of GaN.

This expression (3-6) will have a number of approximation-related errors associated with it. Firstly, the \( (1-n) \) term, meant to account for the loss of beam electrons to backscattering, will not account for EHPs generated by an electron that
backscatters after creating some EHPs, but rather, only for electrons whose first interaction with the solid is to backscatter. A proposal for an improved value of the (1-n) term is discussed in a few paragraphs.

The $E_{eh}$ term is a simple mean-energy, which replaces detailed treatment of the whole interaction process. As the ensemble of interactions – core level ionization, plasmon generation, phonon generation, EHP generation, etc. – would be computationally intractable, this assumption, although rough, is reasonable. It is important to realize that this $E_{eh} = 3E_{GAP}$ approximation was derived for SEM-regime energies; application of this equation to the STEM case (80+keV) or very-low-voltage case is suspect, but no better approximations exist for those regimes.

An example calculation using Donolato's technique is given below; In this case, $E_{BEAM}=5$keV, $n=0.27$, $I_b = 20$ pA, $E_{eh} = 10$ eV. $R$ is taken as 200nm (the $R_{K-O}$ value for 5keV electrons on GaN) and $d$ as 50nm. The material (GaN) enters into the treatment solely through $R$ and $E_{eh}$. The total generation rate $g_0$ in this case is $4.5 \times 10^{10}$ EHPs/sec. Figure 28 shows the generation contours, $\log_{10}(g)$, where $g$ has units of EHP/cm$^3$-sec for this case.
Figure 28: Generation contours, derived from Donolato’s treatment, for GaN at 5kV. Calculated using Mathcad.

However, this Donolato\textsuperscript{161} calculation is a rough approximation; although acceptable at the time of publication in 1981, modern computational techniques allow for much more detailed and accurate electron-beam/solid interaction profiles to be calculated. This will be illustrated below.

The best approach to the injection of carriers is to use Monte Carlo modeling of the beam interaction\textsuperscript{166,167,168}. Shown below as Figure 29 is a simulation from the Monte Carlo software CASINO\textsuperscript{169}, which illustrates the energy deposition of a 5kV electron beam in GaN, summed in the direction normal to the page.
Figure 29: CASINO data for a 5kV beam incident on GaN, where \( R_{K-O} = 200 \text{nm} \). Note that the vast majority of the energy is deposited with nanometers of the point of beam incidence, and the 5% contour is actually at \( z \approx \frac{1}{2} R_{K-O} \).

Again, this method shows that most of the energy is deposited very near the point at which the beam enters the sample. However, the qualitative shape and quantitative results are quite different in comparison to the Donolato calculation above (Figure 28).

The energy dissipation contour data of Figure 29 can also be converted to carrier generation data; this was done in a multi-step method. First, the total energy dissipation (measured in arbitrary units) was found by summing the entire dataset used to construct Figure 29. This sum was then normalized to equal 1, and then the
value at each data point scaled by the same factor. For a given \( g_0 \) value (as used for the Donolato-type calculation above; viz, \( 4.5 \times 10^{10} \) EHP/sec), the total \( g_0 \) can be multiplied by the weighted energy dissipation fraction at each data point to get the generation rate at that point. By then dividing by the spatial extent (volume) of a datapoint in cm\(^3\), the carrier generation rate, in EHPs/cm\(^3\)-sec, can be found. This is shown for the case of 5kV electrons with 20pA beam current incident on GaN in Figure 30.

![Figure 30: Generation rate, plotted as Log\(_{10}(g\ [EHPs/cm^3\cdot sec])\), for a 5kV, 20pA SEM beam incident on GaN, using the CASINO data from Figure 3.](image)

This is considered the most accurate way to estimate the number of EHPs being generated, and their spatial distribution, and will be used for the remainder of
this work for quantitative EHP analyses. Multiplication of Figure 30 by the minority carrier lifetime ($\tau \sim 10^{-9} \text{sec}$) will allow rough approximation of the actual injection level, in EHPs/cm$^3$.

In order to improve the (1-n) factor, simulations of electron backscattering were performed. The (1-n) case treats an electron as taking its full incident energy out of the sample when it backscatters; this is obviously not correct. The CASINO software can produce data of fraction of BSE vs. backscattered energy, which allows the factor of correction to be improved.

Consider a function of fraction of backscattered electrons (as a fraction of total backscatter yield) vs. energy, $H(E)$. Let $H(E)$ be normalized such that:

$$
\int_{E=0}^{E=E_0} H(E) dE = 1
$$

(3-7)

Where $E_0$ is the incident energy of beam electrons.

Then it is possible to find the average energy $\langle E \rangle$ carried away by a backscatter electron:

$$
\langle E \rangle = \int_{E=0}^{E=E_0} E \cdot H(E) dE
$$

(3-8)

A typical $H(E)$ curve, derived from CASINO simulations for GaN at 10keV, $10^5$ trajectories, is shown as Figure 31.
Figure 31: BSE fraction vs. backscatter energy

From this curve, the value $<E>$ is found to be 6.47keV. By analyzing voltages 2, 4, 6... 30 keV for GaN, $10^5$ trajectories each, and performing the $<E>$ calculation for each, a curve of $<E>$ vs. $E_0$ and a linear fit to the data (constrained to an $<E>=0$ intercept at $E_0=0$), is given as Figure 32.
The $g_0$ equation (3-6) can therefore be modified to read:

$$g_0 = \frac{E_0(1-\langle n \rangle)_B}{E_{ch} \cdot q} \quad \text{(3-9)}$$

In which the value $\langle n \rangle$ is defined as:

$$\langle n \rangle = n \cdot \frac{\langle E \rangle}{E_0} \quad \text{(3-10)}$$

The ratio $\langle E \rangle/E_0$ is found to be very nearly 0.65, from the linear fit shown in Figure 32. Values of $n$ calculated at the same time give ≈0.27. Thus, $\langle n \rangle ≈ 0.176$ for GaN in the 1-30keV regime, and the (1-$\langle n \rangle$) factor shall be taken as 0.82 for the remainder of this work.

This modified equation (3-9) should provide an improved value for generation rate; the difference will, however, only be ~6%.
3.2.1.2 The low-injection condition

Once carriers are generated, they are subject to diffusion and drift. Usual experimental conditions require that “low-injection conditions” be used; this means that the number of injected minority carriers is less than equilibrium number of majority carriers.\(^{170,171}\)

Low injection is important to, first, minimize the perturbation of the majority carrier density; this allows diffusion of majority carriers to be neglected in analyses, which makes the mathematics tractable. That is, the minority carriers diffuse under their concentration gradient, with diffusion coefficient \(D_e\) or \(D_h\) for electrons and holes, respectively. The (comparatively) small number of excess majority carriers will drift to follow the diffusive motion of these minority carriers, in order to maintain charge neutrality. Thus, the motion of the minority carriers and majority carriers is simply characterized by \(D_{\text{minority}}\). If high injection conditions obtain, the diffusion problem becomes much more complex, and is characterized by a weighted diffusion coefficient\(^{170}\) \(D_{\text{effective}}=2D_eD_h/(D_e+D_h)\). Second, excessive injection of minority carriers may saturate trapping and recombination sites, leading to changes in the optoelectronic behavior of the specimen.

Following the technique of Bunker\(^{172}\), it is possible to roughly estimate where the threshold from low to high injection is, based upon beam energy, beam current, and materials parameters. Equation (3-9) for \(g_0\) above is used to calculate the number of EHPs/sec injected. By multiplying this number by the estimated minority carrier lifetime, the number of steady-state EHPs is estimated. If diffusion and drift are neglected, dividing this value by the interaction volume gives the number of injected EHPs per unit volume; this can be compared to the doping level. For GaN, the K-O range varies approximately as:

\[
R_{K-O}^{\text{GaN}}(E) \approx 0.0136E^{1.67}
\]

(3- 11)

For \(R\) in microns if \(E\) is in keV. This assumes \(<\mathcal{Z}>=19\), \(<\mathcal{A}>=41.5\) g/mol and \(\rho=6.15\) g/cm\(^3\).
Continuing to follow Bunker, this electron range can be taken as the diameter (not radius) of an interaction sphere; the volume of this sphere is taken as the interaction volume. Thus, with an assumption of $10^{-9}$ second minority carrier lifetime and 27% backscatter fraction, a chart of injection level as a function of beam energy and current can be drawn, as in Figure 33.

![Figure 33: Approximate injection level of EHPs in GaN as a function of beam current and voltage.](image)

This technique is of course approximate, as it spatially averages over the interaction volume, and ignores carrier redistribution due to drift or diffusion. An important non-intuitive result of this model is that lower beam voltages give higher injection levels for a constant beam current. This is because the number of injected carriers increases linearly with beam voltage, but the interaction volume increases
as about the 5th power of beam voltage. This is because $R_{K-O}$ scales as $E^{1.67}$, and interaction volume scales as $R_{K-O}^3$, or equivalently, as $E^{5.01}$. Thus, changes in the beam energy at constant current will cause the density of injected carriers to vary at $\sim 4^{th}$ power.

3.2.1.3 Diffusion and drift

Once carriers are injected, they then undergo diffusion and drift. Diffusion is the motion of carriers under the action of a concentration gradient, and drift the motion of carriers (which are of course charged) in an electric field.

Diffusion is treated with the continuity equation\textsuperscript{161} for minority carriers. This is illustrated for the case holes of concentration $p(r)$, with minority carrier lifetime $\tau$, diffusion coefficient $D$, and generation rate $g(r)$:

$$D \nabla^2 p(r) - \frac{1}{\tau} p(r) = -g(r) \quad (3-12)$$

This equation can be interpreted as “Diffusion - recombination = -(generation).” The $D \nabla^2 p(r)$ term represents the diffusion, $p(r)/\tau$ recombination, and $g(r)$ generation. The actual motion of the carriers will be at a velocity defined by the average kinetic energy of the carriers\textsuperscript{170}, obtained from the thermal state of the lattice. Quantitatively, $1/2m v^2=3/2kT$. Their random scattering off the lattice will result in a diffusive motion characterized by the coefficient $D$.

Boundary conditions to this differential equation depend on the geometry of the specimen. A plan-view sample containing a p-n junction is shown in Figure 34.
In this case, the depletion region is drawn with vanishing thickness, for simplicity. The boundary condition at the top surface, following Donolato\textsuperscript{161}, will be:

$$\left( \frac{\partial p}{\partial z} \right)_{z=0} = sp$$

(3-13)

Here, s is defined as $v_S/D$. D again is the diffusion coefficient; $v_S$ is the surface recombination velocity of the free surface for holes with units of cm/sec. This surface recombination velocity accounts for the recombination and trapping activity of the surface, which, due to broken symmetry, impurities, and other band-structure effects, can be markedly different from that of the bulk.

At the depletion region, the boundary condition given by Donolato is:

$$p)_{z=h} = 0$$

(3-14)

This accounts for the fact that, at the depletion region, the internal electric field will collect all the minority carriers, leading to vanishing concentration $p(r)$ at $z=h$.

If, for another example, a Schottky diode in plan-view was being examined, the diffusion equation (3-12) would be unchanged; the boundary conditions,
however, would be different. In this case, a condition of $p=0$ would obtain at $z=0$, because of the depletion region on the top of the sample. The semiconductor would be treated as semi-infinite in $z$, and under the assumption of injection only at the top surface, a condition of $p=0$ at $z \rightarrow \infty$.

Various approximations and analytical models exist, as discussed above, for the estimation of carrier concentrations as a function of injection and surface activity; however, accurate values will require numerical simulations.

The most important parameter for the characterization of minority carrier behavior in semiconductors is the minority carrier diffusion length $L$; $L$ is related to diffusion coefficient $D$ and minority carrier lifetime $\tau$ as:

$$L = \sqrt{D \tau}$$  \hspace{1cm} (3-15)

The value of $L$ will be related to the nature of the semiconductor (direct or indirect), the concentration of both majority and minority carriers, and most importantly, defect density. For example, consider an n-type semiconductor containing a mid-gap defect state, as shown in Figure 35.
Because the Fermi level sits above the defect level, the occupation probability of the defect for an electron state is very high; effectively 1. Thus, the defect site will quickly trap an electron. When a hole ($h^+$) diffuses or drifts past this defect site, there is a possibility that it will be promoted to the defect site and will recombine with the electron, removing the minority hole from the semiconductor. This recombination phenomena shortens the diffusion length of the holes, leading to a shorter lifetime $\tau$ and concomitantly shorter diffusion length $L$.

Obviously, if the hole-trapping cross section of the defect increases or the density of defects increases, the minority carrier diffusion length will decrease. Techniques to measure $L$ and relate it to microstructure are of great interest in semiconductor technology. This will be discussed quantitatively in the Section 3.2.3.

Typical defects in hexagonal GaN that reduce carrier diffusion length in this fashion are dislocations, stacking faults, inversion boundaries, and cubic-material inclusions, which were discussed in Chapter 2.
Drift is the motion of the carriers under an electric field. For an electric field of strength \( E \), the drift velocity \( v \) of the carriers will be \( v = \mu E \), for carrier mobility \( \mu \). \( \mu \) will be different for both electrons and holes; typically, \( \mu_h \) is lower than \( \mu_e \), as holes have a larger effective mass \( m^* \) in most semiconductors. \( \mu \) will be strongly influenced by the crystal quality; it is well known that the presence of point defects such as ionized dopant atoms (much less extended defects like dislocations or stacking faults) will cause many orders of magnitude of difference in mobility. Diode action (electroluminescence) and the EBIC experimental technique are both controlled by the drift of carriers in devices.

3.2.1.4. Recombination statistics

Once carriers are injected into the sample by the electron beam, they begin to recombine with a characteristic lifetime \( \tau \). Various different processes result in recombination, such as direct-band-edge recombination across the gap, or defect recombination at a localized site. These processes will be controlled by Shockley-Read-Hall\(^{170,173,174} \) statistics. For some generic process \( j \), the lifetime \( \tau_j \) (sec) and recombination rate \( R_j \) (sec\(^{-1} \)) will be defined as:

\[
\tau_j = \frac{1}{R_j} = \frac{1}{\sigma_j N_j v_{th}}
\]

Where \( \sigma_j \) is the capture cross-section of the site \( j \) (cm\(^2 \)), \( N_j \) is the density of sites \( j \) (cm\(^{-3} \)), and \( v_{th} \) the thermal velocity of carriers (cm/sec). For radiative recombination processes 1,2,3,...,m, the total lifetime rate \( \tau^{RAD} \) will be:

\[
\frac{1}{\tau^{RAD}} = R^{RAD} = \frac{1}{\tau_1} + \frac{1}{\tau_2} + \frac{1}{\tau_3} + \cdots + \frac{1}{\tau_m}
\]

Similarly, for non-radiative recombination types 1,2,3,...,n, the total non-radiative lifetime \( \tau^{NR} \) will be:

\[
\frac{1}{\tau^{NR}} = R^{NR} = \frac{1}{\tau_1} + \frac{1}{\tau_2} + \frac{1}{\tau_3} + \cdots + \frac{1}{\tau_n}
\]
The total recombination lifetime $\tau$ will be:

$$\frac{1}{\tau} = R = \frac{1}{\tau_{RAD}} + \frac{1}{\tau_{NR}} = R_{RAD} + R_{NR} \quad (3-19)$$

In this case, $nR=n/\tau=dn/dt$, where $n$ is the number of minority carriers and $t$ time. These recombination lifetimes and rates control the contrast in CL and EBIC images. CL microscopy will image the ratio of radiative vs. non-radiative recombination rate with position, with areas of high $\tau_{NR}/\tau_{RAD}$ giving strong CL signal levels. EBIC microscopy will image areas of high vs. low overall recombination rate.

3.2.2 Introduction to electron beam induced current

Electron beam induced current (EBIC) is a technique that allows the measurement of electrical properties of semiconductors in an SEM or STEM instrument.

When the electron beam is incident on the sample, EHPs are generated as discussed above. In a semiconductor containing either a p-n junction or a Schottky junction, an internal electric field will be present within the depletion region of the semiconductor. This is illustrated in Figure 36. At the p-n junction, the electrons and holes see a significant concentration gradient. This leads to diffusion across the junction and recombination with the carriers on the opposite side. This loss of free carriers leads to the exposure of bare dopant ions, leading to a charge density $\rho$ in the depletion region.
Figure 36: Illustration of the depletion region.

The charge density $\rho$ will have units of C/cm³; $E$ will have units of V/cm. This charge density leads to an associated electric field, via the Poisson equation:

$$\frac{\partial^2 V}{\partial X^2} = \frac{\partial E}{\partial X} = -\frac{\rho(X)}{\varepsilon \varepsilon_0}$$  \hspace{1cm} (3-20)

For electric field $E$, electrostatic potential $V$ (volts), distance $X$ (cm), semiconductor relative permittivity $\varepsilon$ and permittivity of free space $\varepsilon_0$.

By integrating $\rho$ with respect to $X$, the electric field is obtained. Schematic plots of both $\rho$ and $E$ are shown in Figure 36, above.

Any beam-injected carriers that enter the depletion region will encounter this electric field, and drift will separate the EHP. The electron will be swept to the n-
side, and the hole to the p-side. This separation will result in an excess electron on the n-side, which can flow out of the ohmic contact into an external circuit, and similarly for the separated hole. Thus, the separating action of the internal electrical field on the EHPs can result in the flow of a current $I_{\text{EBIC}}$ through an external circuit.

This current will depend on a number of factors, first and foremost by how many electron-beam-induced carriers are generated near the electric field of the depletion region; “near” is defined as approximately one diffusion length $L$. Secondly, the number of carriers that are trapped at defects and recombine, and are thus prevented from being separated by the electric field and contributing to the current. This is directly related to the local value of overall recombination rate $R$.

An EBIC experiment could be performed in the manner of Figure 37.

---

**Figure 37: Schematic EBIC experiment.**
Figure 38 illustrates the four basic types of EBIC experiments. EBIC can be performed in plan-view or cross-sectional geometries, on either p-n junction or Schottky junction type devices. The case of Figure 37 is a cross-section p-n junction type experiment.

![Figure 38: Schematic of the four main EBIC modes.](image)

$I_b$ denotes the electron beam current; $I_{cc}$ is the EBIC current.

Figure 38: Schematic of the four main EBIC modes. $I_b$ denotes the electron beam current; $I_{cc}$ is the EBIC current.

Reused with permission from Ref. 170. Copyright 1982, American Institute of Physics.

Plan-view EBIC is mostly useful for the evaluation of defects. As discussed above, defects will increase the recombination rate, and reduce the recombination lifetime, of minority carriers. As a result, if EHPs are generated near a defect, fewer of them, in comparison to injection away from defects, will survive long enough to diffuse to the depletion region and contribute to the EBIC signal. Thus, areas near a recombination-enhancing defect will appear dark in an EBIC micrograph (lower $I_{EBIC}$).
and areas away from defects will appear bright (higher $I_{\text{EBIC}}$), thus resulting in defect contrast.

Donolato presented a model for the contrast of a generic defect$^{176}$ or a surface-perpendicular dislocation$^{177}$ in plan-view EBIC. Consider a point defect located at position $x=0$ and depth $z=\alpha$ in a semi-infinite semiconductor; let it further be assumed that the semiconductor has infinite surface recombination velocity. Donolato treated the question, "what current of minority carriers is collected at the surface?" which, combined with $v_s=\infty$, is tantamount to the case of a Schottky junction. The defect is assumed to have a recombination lifetime $\tau'$ which is much shorter than $\tau$ in the bulk; this means carriers injected near the defect will have a lower probability of collection at the surface.

Donolato developed equations for the width, $w$, and contrast, $c$, of the EBIC linescan image across the defect. Although the approximations necessitated by the computational techniques available at the time of publication make the exact quantitative results unimportant, two major qualitative conclusions were drawn and are considered to be generally applicable$^{176,177}$:

1. The contrast and width of the defect image were found to depend strongly upon the range of the electrons in the solid; that is, to depend strongly on the beam energy, reaching peak contrast and minimal width when $R\approx\alpha$.

2. The contrast and width of the defect image were almost totally unaffected by the diffusion length of the minority carriers in the sample; indeed, for calculations in which $L\rightarrow\infty$, contrast and image width were scarcely affected.

Although point (1) is intuitively obvious, point (2) is surprising. Donolato further continued this theme$^{178}$ by observing that point (2) was found to be true in experiment, and solving carrier diffusion equations to further show that defect image width was independent of diffusion length. This case of a point-defect was expanded to that of a dislocation, either parallel to the surface$^{161,179}$ or at an angle to the surface$^{180}$. Again, the major conclusions were that the quality of the image depended on the depth of the defect and energy or spot-size of the beam, not the
diffusion length. However, as will be shown later, point (2) in fact only holds for medium to high beam energies; i.e., ~10keV or more.

In a similar vein, Pasemann performed mathematical evaluations of combined CL and EBIC to determine the depth of a dislocation\textsuperscript{181}, evaluation of dislocations perpendicular to the surface of a Schottky or p-n diode\textsuperscript{182}, and the CL or EBIC contrast of a buried dislocation parallel to the surface of a Schottky device\textsuperscript{183}. However, these techniques are not applicable in practice to GaN. The first requires that the electron range be much less than the depth of the p-n junction\textsuperscript{182}, which would require $E_0 \sim 1$keV in GaN-type LEDs, due to the shallowness of the junction; the second requires that the radius of the dislocation-affected area be much less than the minority carrier diffusion length, which will not obtain in short-L materials such as GaN\textsuperscript{183}. Kittler and Seifert discussed the use of temperature- and current-resolved EBIC to evaluate defect properties\textsuperscript{184}.

3.2.3 Mathematical models for cross-sectional EBIC
3.2.3.1 Introduction

Cross-sectional EBIC (X-EBIC), unlike plan-view EBIC (PV-EBIC), is primarily used to quantify the minority carriers' diffusion lengths $L_e$ or $L_h$, and the recombination velocity $v_s$ of the surface created by cross-sectional sample preparation. The minority carrier diffusion length $L$ is a dominant parameter that controls many properties in semiconductor devices; although $v_s$ is not a device parameter per se, it significantly impacts the extraction of $L$ data. Accurate measurement of $L$ and $v_s$ is a vital characterization step for semiconductor devices. SEM-EBIC experiments of this type can be performed on cross-sectional samples, where the SEM beam is parallel to the p-n or Schottky junction, or on plan-view samples containing a Schottky contact. Figure 37 demonstrates the cross-sectional p-n junction SEM-EBIC geometry, which will be our focus here.

With samples prepared for cross-sectional SEM examination, such that the electron beam is parallel to the p-n junction, the EBIC mode of SEM shows a strong
peak at the p-n junction and an exponential-like falloff to either side. Many different publications have proposed different mathematical models to analyze this cross-sectional SEM-EBIC data for the extraction of $v_s$ and $L$ values. This section will first review these models in order to compare and contrast them.

Some commercially important materials, such as GaN, have very short diffusion lengths in comparison to the materials originally examined by EBIC, such as silicon. As such, this section will also examine theoretically how the different published models fare when confronted with diffusion lengths much shorter than those originally intended. In Section 5.2, experimental data will be used to compare the different models in practice. As diffusion lengths and the electron range in the solid (related to SEM beam energy) decrease, the spot size of the SEM begins to have an important effect. We will thus also examine theoretically in this section, and experimentally in Section 5.2.3, how the SEM's spot size affects the behavior of the published models.

3.2.3.2 The exponential model

Of the various equations that have been proposed to relate the profile of the measured EBIC current $I(x)$ to the minority carrier diffusion length $L$, the most commonly used is the simple exponential model, first discussed by van Roosbroeck:

$$I(x) = Ae^{-x/L}$$  \hspace{1cm} (3-21)

Here, $x$ is defined as in Figure 37 and $A$ is an arbitrary fitting constant. Unfortunately, only under very restrictive experimental conditions will the exponential equation (3-21) be valid. In particular, $v_s$ must be zero, and $L$ and $x$ must both be much larger than the electron range $R$, because the electron generation is assumed to be a point source, which can only be approximated when the experimental values of $x$ are much larger than the electron range $R$ and diffusion length $L$. 
3.2.3.3 Donolato's technique

A less restrictive relation is that of Donolato\(^{186}\), Equation (3-22):

\[
I(x) = \frac{2}{\pi} \sum_{k=0}^{k=\infty} \frac{k}{\mu^2} \left\{ \exp\left( -\frac{k^2 \sigma^2}{2} \right) 
- 0.57 \exp\left( \frac{\sigma^2}{2L^2} - \mu z \right) \frac{s}{s + \mu} 
\times \text{erfc}\left( \frac{\sigma}{\sqrt{2}} \left( \frac{\mu - z}{\sigma^2} \right) \right) \right\} \sin(kx)dk
\]

\(3-22\)

In which \(\sigma = R/\sqrt{15}\), \(z = 0.3R\), and \(\mu = (k^2 + L^2)^{-1/2}\); \(k\) is an arbitrary variable of integration. The normalized surface recombination parameter is \(s = \nu_s/D\), for diffusion coefficient \(D\). This equation is based upon a spherical Gaussian generation function, and accounts for surface recombination with the factor \(s\). This technique has been used to analyze data for GaN light emitting diodes\(^{187}\). This equation is best used by first obtaining an \(R\) value, typically from the Kanaya-Okayama range\(^{158,163}\) \(R_{K-O}\) calculation. Then, the experimental data should be compared to the calculated trace for different values of \(L\) and \(s\), until the best fitting combination of \(L\) and \(s\) is found. Using this method, Gonzalez et al.\(^{187}\) found values of \(L\) in the range of 55 to 80 nm for GaN, with negligible surface recombination (i.e., \(s \approx 0\)). This is in contrast to typical values for semiconductors such as silicon, which may have diffusion lengths of tens or hundreds of microns (for example, see Donolato\(^{188}\) or Cavallini et al.\(^{189}\)). Gonzalez et al. also applied the exponential equation to the same data, and calculated values of 75-114 nm for \(L\), which are somewhat larger. Because of the restrictions on the exponential equation, the values calculated via the Donolato method likely are much more accurate. Specifically, the electron range \(R\) was much larger than the diffusion length \(L\), so the point-source assumption of the exponential equation fails. It is examined later, in Section 3.2.4, if even the Donolato equation is valid in this regime. According to Luke et al.\(^{190}\), Donolato-derived EBIC data will be valid if \(R < 4L\), which was the case for Gonzalez et al.\(^{187}\). We will also examine this
Although much better than the exponential relation, the Donolato equation is not without drawbacks. Most significantly, it cannot be applied where the generation volume overlaps the depletion region; this is typically taken as $x \geq R/2$, but may extend to larger values of $x$ if the depletion region is wide in comparison to $R$; see Figure 37.

This forbidden $x$ regime is where the measured $I(x)$ curve is strongest, and therefore least affected by experimental noise. This low-noise data must be neglected to use the Donolato equation. This effect is exacerbated by short diffusion lengths, as the falloff of the EBIC curve occurs much closer to the depletion region.

In GaN, the diffusion length is so short that even at very low SEM beam voltages the electron range will be several times larger than the diffusion length. At 5kV, the $R_{K-O}$ value is $\approx 200$nm, or about three to four times the $L$ values measured by Gonzalez et al. With typical picoameters or current amplifiers used for EBIC measurements, only two to three orders of magnitude of dynamic range of data are usually obtainable in practice. Indeed, Gonzalez et al. began fitting their data at an $x$ value that gave $I \approx 0.5$, which means the data with the best signal-to-noise was neglected. If, in contrast, $L >> R$, then the neglected region $x < R/2$ will only be a tiny fraction of the curve, and a strong signal will still be available for analysis.

In principle, this problem could be eliminated by measuring the EBIC data at even lower SEM voltages, which results in far smaller $R$ values. However, below about 5kV the penetration depth of the electrons is so low that peak of the depth-dose curve may lie within the surface damaged layer, and the injected electron-hole pairs do not penetrate into the bulk semiconductor. Further, the induced current (which is proportional to the accelerating voltage, Equation (3-9)) becomes immeasurably low. The small SEM spot-size needed to measure accurately the nm-range profiles exacerbates this problem, as spot size and beam current are directly related. Thus, reducing the spot size reduces the beam current significantly. Further, spot size increases with decreasing beam energy, which implies even
smaller beam currents must be used to obtain small spot sizes at low beam energies. Thus, reducing the accelerating voltage to improve resolution will result in a serious reduction in the signal-to-noise ratio.

3.2.3.4 Luke's technique

Luke et al.\textsuperscript{190,191,192,193} used the Donolato model to calculate theoretical profiles of $I(x)$ and $dI(x)/dx$, which were referred to as EBIC and DEBIC, respectively. Central to their calculations was the quantity $L_s$, the effective diffusion length. In their work:

$$L_s(x) = -\frac{I(x)}{dI(x)/dx} = \frac{EBIC}{DEBIC}$$

These $L_s$ profiles were calculated for various values of $L$ and $s$, at different electron ranges (accelerating voltages). The asymptotic value of the experimental $L_s$ was used as the starting point for a successive approximation method to obtain $L$. In the presence of surface recombination, the theoretical $L_s$ curves were found to have minima that occur near $x \approx L$. The ratio of the asymptotic $L_s$ value to this minimum $L_s(x)$ value was used to find $s$, by comparison to the pre-calculated curves, as an increase in $s$ will result in a sharper minima.

These techniques are based on the Donolato equation and will suffer the same drawbacks, particularly with respect to the $x \geq R/2$ restriction. Further, when confronted with experimental data, this technique is reputed to fare poorly\textsuperscript{194}, as noisy, discrete datapoints do not give $L_s(x)$ curves of the expected shape, especially when only the tail of the curve can be used; this is explored experimentally later (Section 5.2).

Luke also modified this technique to account for the finite spot size of the SEM probe. If the SEM Gaussian probe is described as having an $e^{-1/2}$ half-width of $\sigma_w$, the new Donolato-Luke equation is.
The full-width half-maximum beam spot size $d_b$ will be $2\sqrt{2 \ln 2}$ times this $\sigma_w$ value, or $d_b \approx 2.355 \sigma_w$. In Equation (3-24), Luke's notation (as published in Ref. 193) has been translated to be consistent with Donolato's original notation. Under conditions of small $L$ or small $R$, the beamwidth becomes a non-negligible factor, but the literature often neglects it. Although (3-24) is a refined version of (3-22), it still suffers from the fact that (3-24) can still only be applied to regions where $R$ does not overlap the depletion region.

Luke et al.\textsuperscript{190} concluded that these techniques could be applied when the electron range was as much as four times the diffusion length; that is, $R \leq 4L$. Although only derived for the case of these techniques used on cross-sectional samples, this rule has been used as a validity criterion for the examination of plan-view Schottky-contact samples analyzed with a different mathematical technique\textsuperscript{195,196,197}. It is not clear that this is valid. Luke\textsuperscript{193} concluded that these techniques could also be used when the beamwidth was equal to or less than the diffusion length.

The effect of $R$ and $d_b$ are expected to have significant effects because of the $x \geq R/2$ restriction. Firstly, for short-$L$-materials, the $x \geq R/2$ criterion must be modified to $x \geq (d_{DR} + R/2)$, where $d_{DR}$ is the extension of the depletion region into the p- or n-side being examined. Although not an important correction when $d_{DR}$ is much less than $R$, when $d_{DR}$ is comparable, such as for small $R$ or low doping (large $d_{DR}$), it will push the unusable $x$ regime to larger values. In the analysis yielding $R \leq 4L$, Luke et al.\textsuperscript{190} found that as the $R/L$ ratio increased, the $L_x/L$ curves broadened, leading to
less-pronounced minima. This was because the minima themselves were in the neglected $x < R/2$ region. These calculations assumed an $L$ value on the order of 1000nm in Si or GaAs; in a material such as GaN, where $L$ might be $\approx 50$nm, and the extension of the depletion region on the order of 100nm, the entire non-asymptotic part of the curve may disappear into the forbidden $x$ region, resulting in complete inapplicability of the Luke EBIC-DEBIC technique; this is not surprising, as most published and established EBIC techniques assume $L$ values significantly longer than those that would be encountered in GaN.

3.2.3.5 Bonard and Ganière's technique

Bonard and Ganière derived a new expression for the EBIC function $I(x)$, which accounted for generation within and across the depletion region. This complication is neglected by all other literature models.

The Bonard and Ganière expression, $\eta(x)$, for the EBIC intensity is given in Equation (3-25):

$$
\eta(x) = \frac{1}{\pi} \int_{k=0}^{k=\infty} \left[ s_h \Theta_h (k, x) + s_e \Theta_e \left( k, -(x + d_j) \right) \right] \text{d}k + 0.5 \left[ \text{erf} \left( \frac{x + d_j}{\sigma_x} \right) - \text{erf} \left( \frac{x}{\sigma_x} \right) \right]
$$

(3-25)

Where $s_{h,e}$ is the $s$ value for holes or electrons, $d_j$ is the full width of the depletion region, and:
\[ \Theta_i(k, x) = \frac{1}{k^2 + s_i^2} \text{erfc} \left( \frac{\sigma_x}{2} \left( \mu(k, L_i) - \frac{2x}{\sigma_x^2} \right) \right) \times \exp \left[ \frac{\sigma_x \mu(k, L_i)}{2} \right]^2 \times \exp \left[ -\frac{\mu(k, L_i)x}{1 + k^2 \sigma_z^2} \right] \times \left[ 1 + 3s_i \sigma_z - k^2 \left( s_i \sigma_z^3 + 3\sigma_z^2 \right) \right] \]  

(3-26)

In which case \( i \) denotes either electrons or holes. The variables \( \sigma_x \) and \( \sigma_z \) are related to the injected carrier generation function of the beam within the sample. Bonard and Ganière used the following EHP generation function in the derivation of (3-25) and (3-26):

\[ h(x, z) = \frac{1}{2\sqrt{\pi} \sigma_x \sigma_z^3} \exp \left( -\frac{x^2}{\sigma_x^2} \right) z^2 \exp \left( -\frac{z}{\sigma_z} \right) \]  

(3-27)

In this case, an \((x,y,z)\) generation function \( g(x,y,z) \) has been summed along the \( y \)-axis to provide the two-dimensional projection \( h(x,z) \). The original paper by Bonard and Ganière studied Al\(_{0.4}\)Ga\(_{0.6}\)As semiconductors. For this material, they quoted values of:

\[ \sigma_x = \sqrt{0.05 + 0.0001E_0^{2.78}} \]
\[ \sigma_z = \sqrt{0.02 + 0.003E_0^{1.5}} \]  

(3-28)

Where \( E_0 \) is the energy of the SEM beam. The original paper\(^{194} \) did not unequivocally define the units of \( \sigma_{x,z} \) or \( E_0 \), but they are presumably \( \mu \text{m} \) and keV, respectively. Computation of \( \sigma_x \) and \( \sigma_z \) values for GaN, and comparison of values computed for Al\(_{0.4}\)Ga\(_{0.6}\)As to those of Equation (3-28), are described in Section 5.2.2. The Bonard and Ganière technique appears to be the most advanced and rigorous EBIC treatment available in the literature, and will be the focus of the discussion in Section 5.2.

This technique, however, does not provide a "normalized" answer; the \( \eta(x) \)
The equation above (3-25) does not necessarily have a peak value of 1.0, as is expected from normalized EBIC data. $\eta_{\text{max}}$ is typically $\approx 1$ under the conditions modeled here, but can vary to values slightly more or less than 1.0. As such, the actual Bonard and Ganière equation used in the calculation of theoretical curves in Section 3.2.4.2 is $I(x) = \eta(x)/\eta_{\text{max}}$. The differences between the normalized and raw Bonard and Ganière equations will be examined in comparison to experimental data in Section 5.2.3.2.

3.2.3.6 Method of moments

The method of moments\cite{188,189} involves solving an pair of simultaneous equations for the unknowns $s$ and $L$. First, two EBIC curves must be acquired at different accelerating voltages; these two curves are treated as functions of $Z$, where $Z=0.41R$, for electron range $R$. For two curves 1 and 2, the moments $m_1$ and $m_2$ are found from the EBIC traces $I_1(x)$ or $I_2(x)$ as:

$$m(Z) = \int_{x=0}^{x=x_{\text{max}}} [I(x, Z) x] dx$$  \hspace{1cm} (3-29)

Once the $m_1$ and $m_2$ values are obtained, $s$ and $L$ can be extracted from:

$$m_1 = L^2 \left[ 1 - \left( \frac{sL}{(1+sL)} \right) \exp\left(-\frac{Z_1}{L}\right) \right]$$  \hspace{1cm} (3-30)

$$m_2 = L^2 \left[ 1 - \left( \frac{sL}{(1+sL)} \right) \exp\left(-\frac{Z_2}{L}\right) \right]$$

These are two non-linear equations in the unknowns $s$ and $L$, and solvable numerically from known $m_1$, $m_2$, $Z_1$ and $Z_2$. The fundamental equation from which this technique is derived assumes a point generation of EHPs at a depth $Z$ below the surface. Because of the point-source approximation, this technique would only be expected to be applicable when $x$ and $L$ are significantly larger than $R$. In fact, Donolato\cite{188} specifically suggests that this technique will only be valid in the range $R/L<0.5$. Thus, it would be expected that this method would not be applicable to the
short-diffusion length regime. This will be tested in Section 5.2.3.4.

3.2.3.7 Other techniques

Guermazi et al.\textsuperscript{198} propose an EBIC dependency of the form:

\[
I(x) \propto \exp(-x^2 / L^2)
\]  

(3- 31)

This equation does not explicitly account for either generation volume or surface recombination. Gonzalez et al.\textsuperscript{187} also applied Equation (3-31) to their GaN data; values of L significantly larger than those obtained from either the Donolato or van Roosbroeck equations were obtained; viz., 99-250 nm. There is no reason to presume that this technique will be of general applicability, and it will not be discussed further.

Ong et al.\textsuperscript{199,200,201,202} propose an equation of the form:

\[
I = k x^\alpha \exp\left(-\frac{x}{L}\right)
\]  

(3- 32)

\[
\ln\left[I(x)/x^\alpha\right] = \ln(k)-\frac{x}{L}
\]  

(3- 33)

In this case, \( \alpha \) is a linearization constant. By plotting \( \ln[I(x)/x^\alpha] \) as a function of x for \( \alpha=0 \), they anticipate a plot that is concave upwards. By adjusting the value of \( \alpha \) between 0 and \( \approx-0.6 \), the concavity of the plot will change. Once the concavity is minimized, i.e., the \( R^2 \) value of a linear fit of the data is maximized, the value of \( \alpha \) is considered to have been found. According to Ong et al.\textsuperscript{199,200,201}, L will be the negative of the slope of this best-fit line; the value \( S=sL=v_sL/D \) can then be found from the relation\textsuperscript{202}:

\[
S = -17.074 + 4.67 \sqrt{2\ln\left(\frac{499.74}{\alpha + 0.63}\right)}
\]  

(3- 34)

This equation is stipulated to only be applicable to the regime \( x > 2L \); as a result, this technique requires fitting a curve to the tail of the EBIC data, where signal-to-noise will be worst. This is published as a material-independent technique.
They also stipulated that for electron range R, the equation must be fit where \( x > 1.5R \). Further\(^{202}\), best accuracy will be obtained when the fit is made in the \( x \)-range \( 2L < x < 11L \). This technique will be examined in Section 5.2.

Ong and Wu\(^{203,204}\) improved upon this technique with a new technique, based upon the Donolato\(^{186}\) technique (Section 3.2.3.3). Although intended to be applied to the analysis of thin layers in which the distance from the junction to the contact is shorter than the diffusion length, it will not be useful when the R value is also comparable to the layer thickness, due to the overlap of the interaction volume with the depletion region or the metal contact. As the GaN LEDs examined by X-EBIC in this work have very thin p-sides, this technique\(^{203,204}\) was not applied. Luke\(^{205}\) expanded on previous work in thin layers\(^{206}\) to find an equation for use in the case where the thickness of the semiconductor is comparable to that of \( L \), and accounts for the surface and contact recombination velocities. Unfortunately, this technique\(^{205}\) is based on a derivation for a point-source of EHPs\(^{206}\) and therefore unlikely to be applicable to situations where \( R \sim L \). These two techniques will therefore not be examined further.

3.2.4 Theoretical comparison of cross-sectional EBIC models

Now, EBIC curves will be computed for long- and short-L cases, using the above-discussed techniques, with the example case of 5keV electrons incident on GaN. 5keV is chosen as a small-interaction-volume case (\( R=200\text{nm} \)), but which has been shown in practice\(^ {187}\) to provide acceptable signal-to-noise in GaN-based LEDs. These theoretical computations will allow us to compare the general behavior of the different EBIC models, and form hypotheses regarding their behavior in practice when confronted with short-L-materials such as GaN LEDs.

3.2.4.1 The Donolato-Luke methods

For the following discussions, the examples all assume an \( R \) electron range of 200nm for Donolato or Luke calculations, which is equivalent to 5keV electrons incident on GaN. These curves were calculated using Mathcad 12; the code used is
listed in Appendix 1.

The Donolato-Luke Equation (3-24) gives a series of different \( I(x) \) curves as \( s \) varies from \( 10^{-2} \ \mu m^{-1} \) to \( 10^6 \ \mu m^{-1} \), as shown in Figure 39, calculated for values of \( R=200\text{nm} \) and \( L=1\mu m \). As seen in Figure 39, when \( s \) decreases below about \( 10^{-1} \ \mu m^{-1} \), the curvature of the \( I(x) \) curve changes minimally, indicating that this regime approximates \( s=0 \). Shortly beyond about \( s=10^2 \ \mu m^{-1} \), the curves again saturate, indicating that this is the approximately \( s=\infty \) regime. Thus, values of \( 10^{-2} \) and \( 10^6 \ \mu m^{-1} \) will be used to approximate the \( s=0 \) and \( s=\infty \) cases.

In this \( R=200\text{nm} \), \( L=1\mu m \) case, EHPs are injected to a depth much shallower than the diffusion length, and the surface can strongly effect their diffusion. Figure 39 is the classical EBIC result, as expected from the discussions of Opdorp\textsuperscript{207} or Donolato\textsuperscript{186}.

![Figure 39: Calculated EBIC curves for \( R=200\text{nm} \), \( L=1\mu m \). The \( s \) values are labeled on the curves. Calculated using the Donolato-Luke equation, with \( d_b=0\text{nm} \).](image)

If the minority carrier diffusion length \( L \) decreases from \( 1\mu m \) to 50nm, the
qualitative behavior of the curves changes significantly. Figure 40 shows curves calculated at R=200nm for values of $s=10^{-2} \, \mu m^{-1}$ or $10^6 \, \mu m^{-1}$, with L values of L=1\mu m or 50nm. As seen above, the L=1\mu m curves showed a significant change from $s=10^{-2}$ to $10^6 \, \mu m^{-1}$, as expected. However, for L=50nm, the change from the low to high s case is far less pronounced. This indicates that it will be difficult to extract the surface recombination from experimental data when the electron range is significantly larger than the diffusion length. This behavior is because in the L=50nm case, unlike the L=1\mu m case, the EHPs are being injected deeply enough that most recombine before reaching the surface.

![Figure 40: Comparison of curves for long and short diffusion length, under constant R=200nm and d_b=0nm. Dashed curves are for L=50nm; solid lines are for L=1\mu m. Calculated using the Donolato-Luke equation.](image)

Luke\textsuperscript{190,191,192,193} discussed the analysis of L and s data by plotting theoretical curves of $L_s(x)/L$, where $L_s$ is the effective diffusion length (Section 3.2.3.4 above).
For low-\( s \) calculations, the curve of \( L_s(x)/L \) will be approximately flat; as \( s \) increases, a minima in \( L_s(x)/L \) appears in the \( x \approx L \) region. In the normally assumed case for \( L >> R \), the minima becomes much sharper as \( s \) increases. In the new case for short \( L \), however, the peak necessary for the Luke-style analysis disappears. Figure 41 shows the analysis for varying \( R \) values and \( s = 10^6 \mu m^{-1} \), at constant \( L = 50nm \). As \( R \) increases, the Luke-type peak disappears. In addition to the softening of the peak minima, the fact that data for \( x < R/2 \) must be neglected means that the entire pre-asymptotic region of the \( L_s/L \) curve is lost for \( R >> L \). This problem will be exacerbated if the finite size of the depletion region must be accounted for, pushing the forbidden \( x \) values to the right.

![Figure 41: Luke-technique effective diffusion length analysis. Here, \( L = 50nm \), \( s = 10^6 \mu m^{-1} \), \( d_b = 0nm \). The labels on the curves are the electron ranges used to calculate the curves, and approximate beam energies corresponding to those ranges.](image-url)
The beam energy \( E_0 \) values were calculated from the K-O range \( R \) as \( E_0 = \left[ \frac{R}{13.7} \right]^{0.598} \) with \( R \) in nm and \( E_0 \) in keV.

Figure 42 shows curves for different \( s \) values, at constant \( d_b=0 \text{nm} \), \( L=50 \text{nm} \) or \( 1 \mu\text{m} \), and \( R=200 \text{nm} \). Even by assuming zero depletion region width, the broadening of the \( s \)-related minima and the loss of the \( x<R/2 \) regime makes the \( L=50 \text{nm} \) curves’ differences very subtle, compared to that expected in the large-\( L \) case. In the presence of experimental noise, it is doubtful the \( s=0 \) or \( s=\infty \) regimes could be differentiated using the Luke technique.

![Figure 42: Effect of \( s \) and \( L \) on effective diffusion length curves \( L_s(x)/L \). Solid lines are for \( L=1 \mu\text{m} \), and dashed lines \( R=50 \text{nm} \). \( R=200 \text{nm} \) and \( d_b=0 \text{nm} \) for all curves.](image)

SEM spot size also has a significant effect; finite spot size exacerbates the effects of short diffusion length. Figure 43 shows plots of \( s=10^6 \mu\text{m}^{-1} \) data where \( R=200 \text{nm} \) and \( L=50 \text{nm} \). As spot size increases from \( d_b=0 \text{nm} \) to \( d_b=100 \text{nm} \), the larger spot size begins to smear the EBIC curves, adding a level of complexity to any attempt at curve fitting to extract \( s \) and \( L \) values. Figure 43 indicates that
significant broadening of the curves occurs when $d_b \sim L$, indicating that $d_b$ should be kept less than or approximately equal to the diffusion length to use this technique. For long diffusion lengths, the beamwidth effect is much less pronounced. A plot like Figure 43 for $L=1\mu m$ would show superimposed curves, with no measurable change in the $d_b=0$ to 100nm range.

![Figure 43: Effect of increasing spot size $d_b$ on Donolato-Luke curve generation. Conditions $R=200nm$, $L=50nm$, $s=10^6 \mu m^{-1}$.](image)

3.2.4.2 The Bonard & Ganière method

The following analyses, using the technique of Bonard and Ganière\textsuperscript{194}, Equation (3-25), are calculated using $\sigma_x$ and $\sigma_z$ values derived in Section 5.2.2 for GaN at 5keV. Values of depletion width $d_j=50nm$ were assumed. The rightmost
edge of the depletion width is taken as $x=0$; the leftmost edge as $x=-d_j$. For the purposes of illustration, electrons and holes were considered to have the same $L$ and $s$ values. These $L$ and $s$ equalities will obviously not be generally true, but are useful for the sake of demonstration. Mathcad 12 and Maple 9.5 codes used are given in Appendix 2. Mathcad is generally the preferred software, but failed to converge properly when $\sigma_x$ was large and $L$ small; thus, Maple was used for large $\sigma_x$ calculations under small $L$ conditions. Both software packages gave identical solutions under all other situations.

For long diffusion lengths, $L=1\mu m$, the calculated EBIC curves are as expected from classical EBIC theory; Figure 44(A) shows $L=1\mu m$ curves calculated for $d_b=0$nm at $s=10^{-2}$, $10^1$ and $10^6$ $\mu m^{-1}$, and is exactly the behavior expected. Conversely, Figure 44(B) shows curves for $L=50$nm.
Figure 44: (A) Bonard and Ganière calculations for EBIC profiles in GaN at 5kV, assuming 1μm diffusion length. (B) Similar calculations, for L=50nm.
The differences resulting from variations in $s$ are far less pronounced. Also noticeable is the gradual curvature in the area near $x=0$, which is the depletion region. This broadening effect is due to the short minority carrier diffusion length in comparison to the depletion region width. The combination of the depletion-region-associated broadening and deep injection in comparison to $L$ (minimizing the surface effects) results in the small variations with $s$.

Spot size has a similar effect; for $L=1\ \mu\text{m}$, there is no noticeable difference in the curves for $d_b$ up to several hundred nanometers. For $L=50\text{nm}$, the $d_b=300\text{nm}$ curve falls below the $d_b=0\text{nm}$ curve for small $x/L$, but then broadens to fall above the $d_b=0\text{nm}$ curve beyond $|x|/L\sim 3$. This is seen in Figure 45, for the case of $s=10^6\ \mu\text{m}^{-1}$. It is speculated that this is due to the broad injection at $d_b=300\text{nm}$ injecting more carriers far from the active region, in comparison to $d_b=0\text{nm}$ for small $x$, and conversely, the injection of more carriers near the active region at larger $x$.

![Figure 45: Bonard and Ganière computations for GaN at 5kV, with $L=50\text{nm}$, $s=10^6\ \mu\text{m}^{-1}$, and $d_j=50\text{nm}$.

\begin{center}
\includegraphics[width=0.5\textwidth]{figure45.png}
\end{center}
3.2.4.3 The Ong et al. method

The Ong technique\textsuperscript{199,200,201} (Equations (3-32) and (3-33)) can be tested theoretically by applying it to curves generated via the Donolato-Luke or the Bonard and Ganière techniques. Figure 46 shows Donolato-Luke data for $R=200\text{nm}$, $d_b=30\text{nm}$, $L=1\mu\text{m}$ and $s=10^6\ \mu\text{m}^{-1}$, which has been analyzed by the Ong technique; as expected, the best-fit $\alpha$ value is found to be -0.6, if $s=10^6\ \mu\text{m}^{-1}$ is taken as $s=\infty$. Using Bonard and Ganière-technique generated data, instead of Donolato-Luke as the starting point, provides an identical analysis.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure46.png}
\caption{Donolato-Luke data (squares) for $d_b=30\text{nm}$, $s=10^6\ \mu\text{m}^{-1}$, $L=1\mu\text{m}$, analyzed with the Ong technique (dashed line).}
\end{figure}

This leads to the conclusion that the Donolato-Luke, Bonard and Ganière, and Ong et al. techniques, are all equivalent when $L>>R$.

However, when short $L$ curves are used as input data, the Ong technique fails. For example, for a theoretical curve calculated with the Donolato-Luke equation under conditions of $R=200\text{nm}$, $d_b=30\text{nm}$, $L=50\text{nm}$ and $s=10^6\ \mu\text{m}^{-1}$, the Ong et al. analysis yields an $\alpha$ value of +0.3, which is forbidden by the Ong et al. analysis.
Using Bonard and Ganière calculated data yields a similar analysis, but with an even worse $\alpha$ value of +0.8. With Donolato-Luke data of $R=200\text{nm}$, $L=50\text{nm}$ and $s=10^{-2}\text{ \mu m}^{-1}$, the Ong et al. technique would predict an $\alpha$ value of 0.0, but in fact the best fit curve is found for $\alpha=-0.4$ to -0.6. In this case, however, the change in the $R^2$ value from $\alpha=0.0$ to $\alpha=-0.6$ is from 0.9995 to 0.9999. This further indicates that in this case s-sensitivity is very poor. However, from the analysis of the Donolato-Luke data via the Ong et al. technique, L values for the short-L-case were found within $\pm10\%$, despite the drawbacks regarding $s$.

Thus, the Ong technique is hypothesized to be adequate for estimating L in short-diffusion-length materials, but not useful for the extraction of $s$.

### 3.2.4.4 Theoretical analysis of methods: summary

The intent of the above analyses was to predict the applicability of the different mathematical models for X-EBIC to experimental data from samples with short diffusion lengths.

In light of the above analysis, the hypothesis is formed that the Bonard and Ganière technique will be the most useful for minority carrier diffusion length and surface recombination velocity analyses of experimental data derived from GaN-based LED samples. This hypothesis is experimentally examined in Section 5.2.

Further, as first noted by Opdorp and confirmed by the theoretical analyses above, it is hypothesized that very low beam energies $E_0$ will be necessary to extract the surface recombination parameter $s$. This will, in particular, lead to a significant loss in signal-to-noise ratio in the experimental data.

### 3.2.5 Cathodoluminescence

We now move away from EBIC and to the other main in-SEM beam injection technique, Cathodoluminescence (CL). CL is an optical technique in which a sample is induced to emit light by electron bombardment, either in SEM or STEM. As discussed above, the energetic electron beam excites EHPs within the
semiconductor crystal. These EHPs recombine with a time constant $\tau$; contributing to $\tau$ are the factors for radiative recombination $\tau^{\text{RAD}}$ and non-radiative recombination $\tau^{\text{NR}}$, where $1/\tau=1/\tau^{\text{RAD}}+1/\tau^{\text{NR}}$. By definition, radiative recombinations produce a photon, and non-radiative recombinations do not. As radiative and non-radiative recombinations compete for the total number of EHPs, the intensity of the CL signal will be related to $\tau^{\text{NR}}/\tau^{\text{RAD}}$. That is, for a strong CL signal, $\tau^{\text{RAD}}$ should be much shorter than $\tau^{\text{NR}}$, indicating more radiative events occur than non-radiative.

A number of different processes are possible in CL; these are outlined schematically in Figure 47, as discussed by Refs. 159, 208, and 209.

![Figure 47: Schematic band diagram of CL processes.](image)

$\bullet =$Electron  $\bigcirc =$Hole

The processes 1-9 are now described:
1. **EHP generation.** This is the excitation of an electron from the valence band into an empty state in the conduction band. The initial and final states are not necessarily at the band edges.

2. **Thermalization.** This is not a recombination process, but rather a prerequisite to recombination. Here, the hot hole rises to the VB edge and the hot electron drops to the CB edge. This process has a time scale \( \sim \) picoseconds.

3. **Band-to-band.** Here, an electron drops directly from the conduction band to the valence band. The energy \( E_{\text{CL}} \) of the emitted photon is \( E_{\text{CL}} \approx E_{\text{gap}} \).

4. **Exciton recombination.** Excitons are particle-like associations of an electron and hole; for exciton binding energy \( E_{\text{ex}} \), \( E_{\text{CL}} \approx E_{\text{gap}} - E_{\text{ex}}/n^2 \), where \( n \) is the principal quantum number of the exciton. Excitons are spectroscopically denoted \( X \). Free exciton recombination will often be denoted \( F_X \), and bound exciton recombination \( B_X \).

5. **Donor-bound exciton.** An exciton binds to a donor atom. A free hole then recombines with the donor-bound exciton. \( E_{\text{CL}} \approx E_{\text{gap}} - E_D^0 \). This process is denoted \( D_0^0X \).

6. **Acceptor-bound exciton.** An exciton binds to an acceptor atom. A free electron then recombines with the donor-bound exciton. \( E_{\text{CL}} \approx E_{\text{gap}} - E_A^0 \). This process is denoted \( A_0^0X \).

7. **Donor-acceptor pair.** Here, an electron and a hole on a donor site and acceptor site recombine. The strength and energy of this transition will depend strongly on the separation \( r \) between the atoms and material dielectric constant \( \varepsilon \). \( E_{\text{CL}} \approx E_{\text{gap}} - E_D - E_A - q^2/(4\pi\varepsilon r) \). This process is typically denoted \( D\text{AP} \).

8. **Radiative defects.** This could denote the 1s-like level of a vacancy, the 4f electrons in a rare-earth impurity, or other orbital-like states.

9. **Non-radiative defects.** Here, the electron and hole recombine, but the energy and momentum carried by the EHP is taken up by phonons and no photon is produced.
Excitons are important to GaN-based materials, as GaN has a particularly high exciton binding energy and excitons can, under some circumstances, be present at room temperature in GaN.

Two basic models exist for excitons: Frenkel and Wannier-Mott\textsuperscript{209}. Frenkel excitons are strongly bound EHPs and highly localized; Wannier-Mott excitons are loosely bound and have sizes on the order of or greater than that of the unit cell\textsuperscript{210,211,212}. In low-temperature experiments, excitonic lines (process (4) above) can sometimes be resolved at energies slightly lower than that of the main band-edge peak (3) in CL or PL experiments. The exciton binding energy $E_{ex}$ noted above can be calculated from a hydrogenic-like equation:

$$E_{ex} = \frac{\mu q^4}{2\hbar^2\varepsilon^2}$$

(3-35)

Where $q$ is the electronic charge, $\hbar$ is Dirac’s constant, $\varepsilon$ is the dielectric constant of the material, and $\mu$ is the reduced effective mass of the exciton. For electron and hole effective masses $m_e^*$ and $m_h^*$, $\mu$ can be found:

$$\mu = \frac{m_e^* m_h^*}{(m_e^* + m_h^*)}$$

(3-36)

Exciton energies in GaN are ~0.26 eV\textsuperscript{213}; this rather large binding energy indicates they could, under some circumstances, exist even at room temperature.

Once EHPs are generated, and once the recombination processes listed above generate photons, it is necessary for the photons to leave the sample before they can be collected and analyzed. There are two factors that prevent generated photons from escaping the sample: absorption and refraction.

The absorption loss $f_A$ is due to the exponential decay of light intensity passing through a medium; the well-known Beer-Lambert law $I(z) = I_0 \exp[-\alpha z]$ describes the attenuation of light intensity, where the absorption coefficient $\alpha$ is a...
materials parameter that is also a strong function of wavelength. Yacobi and Holt\textsuperscript{209} give an approximate calculation for $f_A$, based on a point source model:

$$f_A = \frac{L_{CL}(\text{observed})}{L_{CL}(\text{generated})} = \int_{z=0}^{z=\infty} L_{CL}(z)\exp(-\alpha z)dz = \frac{1}{1 + \alpha L} \quad (3-37)$$

Here, $L_{CL}$ is the generated CL intensity at some given wavelength, and $\alpha$ the absorption coefficient at that wavelength. $L$ is the minority carrier diffusion length. Accurate modeling of the internal absorption will require numerical treatment of the generation, diffusion, recombination, and absorption.

Yacobi and Holt\textsuperscript{209} also discuss losses due to refraction. Because the samples of interest to CL work typically have high indices of refraction $\eta$, the refraction loss $f_R$ will be a major contribution to loss of CL signal in an experiment. To derive $f_R$, first recall Snell's law for refraction:

$$\eta_1 \sin(\theta_1) = \eta_2 \sin(\theta_2) \quad (3-38)$$

For incident and refracted rays of angles $\theta_1$ and $\theta_2$ in materials of refractive indices $\eta_1$ and $\eta_2$. Further recall that when the rays are normal to the interface ($\theta_1=\theta_2=0$) and $\eta_2=\eta_{\text{vacuum}}=1$, the reflected intensity $R$ is given by:

$$R = \left(\frac{n-1}{n+1}\right)^2 \quad (3-39)$$

This is valid to the first order if $0<\theta_1<\theta_C$, where $\theta_C=\arcsin(1/\eta)$. Total internal reflection is the most important limiting factor, as with the large mismatch in $\eta$ between the sample and vacuum, most of the light generated will be trapped within the sample and not contribute to the measurable CL signal. Yacobi and Holt derive the expression for $f_R$ as:
\[ f_R = (1 - R) \frac{1}{4\pi} \int_{\theta=0}^{\theta_c} 2\pi \sin(\theta) d\theta \approx \left[ 1 - \left( \frac{\eta - 1}{\eta + 1} \right)^2 \right] \left[ \frac{\eta - \sqrt{\eta^2 - 1}}{2\eta} \right] \]  
(3-40)

Where all symbols are defined as above.

The important conclusions to derive from this treatment are the critical angle and \( f_R \) value of GaN, which are 21.1° and 2.6%, if GaN is assumed to have an index of refraction \( \eta = 2.78 \), which Lin et al.\textsuperscript{214} found at the band-edge energy of 3.4eV. Longer wavelengths give lower \( \eta \) values for GaN, and should give slightly higher \( f_R \) efficiencies. The \( f_R \) curve from \( \eta = 1.0 \) to 4.0 is given in Figure 48.

![Figure 48: CL emission efficiency as a function of \( \eta \).](image)

An important implication of Figure 48 is for the case \( \eta = 1 \), which would indicate generation of photons in free space; this gives the cone semiangle \( \theta_c = 90° \) and \( f_R = 50\% \), indicating the implicit assumption in the above derivation that only
photons leaving in a direction away from the substrate are counted. As samples are rarely mounted on reflective substrates during CL experiments, and there will be significant absorption losses through the backside of the sample, these assumptions are reasonable.

Considering the above discussions, it should be clear that CL emission is not an efficient process. First, there are many competing radiative and non-radiative processes that will reduce the intensity of a given emission. Second, the light may be subject to absorption through the thickness of the sample. Third, only 1~10% of the generated, non-absorbed photons will actually leave the sample surface.

Thus, the conclusion can be reached that in order to perform CL experiments, the number of photons generated and collected should be maximized.

Lowering the temperature of the specimen will increase the ratio $\tau_{NR}/\tau_{RAD}$; this will result in stronger CL emission. As an example, Onuma et al.\textsuperscript{156} used variable-temperature TRPL to show that, although $\tau_{RAD}$ stayed approximately constant for 300K to liquid helium temperatures, the $\tau_{NR}$ could drop by as much as three orders of magnitude. This type of phenomena will mostly be due to freeze-out of carriers into the defect states.

Of course, liquid nitrogen or liquid helium cryostats are not always available. Another possibility is to improve the collection efficiency of the CL hardware, in order to maximize the available signal from a given number of emitted photons.

Various forms of CL collection apparatus are possible; some examples are shown in Figure 49.
(A) The sample sits at the focus of an elliptical mirror, which collimates the light directly into a detector. (B) Sample sits at one focus of an elliptical mirror, and a fiber optic which feeds to an ex-vacuo detector sits at the other focus. (C) A parabolic mirror and lens focus light onto a fiber. (D) A fiber or lens + fiber system sit directly above a tilted sample, and collect the light.

Of the hypothetical systems discussed in Figure 49, (A) will be the most efficient, but also the most expensive and complicated. Because of the relatively low cost and good robustness, option (C) was used in this work.

Figure 49 indicates "detector" as a generic quantity. For CL, a number of different detector types are possible. The simplest detector would be a photodiode,
which produces a change in voltage when struck by light. However, such a detector will not be particularly efficient, as there will be little gain in the signal.

A better choice would be a photomultiplier tube (PMT)\textsuperscript{215}. PMTs are advantageous\textsuperscript{216,217} because of their low-noise internal gain of $\sim 10^6$ to $10^7$, and high bandwidth $\sim 1$GHz. Photodiodes (PDs) have much higher quantum efficiency than PMTs ($\sim 85\%$ compared to $<40\%$), but they have no internal gain, so an external amplifier must be used. The noise associated with this amplifier will be larger than the PD signal unless the photodiode is operated in the shot-noise-limited regime; thus, PDs are most useful in high-light-level applications\textsuperscript{217}. Avalanche photodiodes (APDs) have internal gain $\sim 50$, and thus still require amplification, although not to the degree of a photodiode. Thus, APDs are suited to light levels slightly lower than that of PDs. PMTs are thus best suited for very-low-light levels, down to tens of photons per second, or for high bandwidth applications. However, in the deep infrared, PMT quantum efficiency is $\sim 0\%$, so cooled Ge or InGaAs photodiodes and lock-in amplification must be used\textsuperscript{218}.

When performing CL imaging with a photodiode or a PMT, the mode of CL is called panchromatic. If $I(x,y,\lambda)$ is the emission intensity of the sample at some wavelength $\lambda$ at some point $(x,y)$, $A(\lambda)$ is the transmission loss of the CL collection optics as a function of $\lambda$, and $B(\lambda)$ is the detector's quantum efficiency as a function of $\lambda$, the panchromatic CL intensity $P(x,y)$ will be:

$$P(x,y) = \int_{\lambda=\lambda_{\text{min}}}^{\lambda=\lambda_{\text{max}}} [1 - A(\lambda)]B(\lambda)I(x,y,\lambda) d\lambda$$ (3-41)

As such, the panchromatic signal will be relatively strong, as all detectable emission wavelengths will contribute, weighted by the system's wavelength response. Panchromatic imaging is limited, however, as it will not differentiate between emissions; for example, the yellow-band, quantum well, or band-edge emissions in
GaN all contribute to the signal. Thus, panchromatic imaging is most useful for examining a sample in order to map non-radiative defects\textsuperscript{219}.

In order to image with particular wavelengths, or in order to take spectra, the incoming light must be dispersed by wavelength. This is accomplished using a monochromator, as illustrated schematically in Figure 50.

Figure 50: Schematic illustration of a monochromator.

Here, light enters through the entrance slit, which ensures that light only travels down the optical axis of the device. Mirrors then focus the light onto a diffraction grating. Only light of a particular wavelength bandpass will be able to escape through the exit slit; the rest is lost inside the monochromator. Thus, by slewing the grating, a particular wavelength can be chosen. By opening or closing the entrance slit, more or less light is allowed into the instrument, but it falls off-axis, and makes the grating-slewed selection broaden. Opening or closing the exit slit will control the bandpass of light that can exit the monochromator to the detector, which will typically be a photomultiplier tube.

It is possible to remove the exit slit; then, a dispersion plane is formed, which will contain the various wavelengths of light from the incident beam, resolved into a
rainbow-like spectrum. Placing a CCD camera at the dispersion plane will allow the entire spectrum that falls on the CCD to be collected at once.

The advantage to CCDs is that they can integrate an entire spectrum very quickly, and require no grating scanning. The advantage to PMTs over CCDs is that they have lower noise and higher sensitivity, especially in the green-to-UV range. CCDs will be superior for spectroscopy when signal-to-noise ratio is high or the sample is beam-sensitive and the spectrum must be acquired quickly. PMTs are superior for very-low-light-level spectroscopy.

A monochromator collects data as a function of wavelength \( \lambda \) for a particular bandpass \( \Delta \lambda \). Wavelength can be converted to energy by:

\[
E(\text{eV}) = \frac{1239\,(\text{nm}\,-\,\text{eV})}{\lambda\,(\text{nm})}
\] (3-42)

This implies that, in order to convert the intensity of a spectrum from wavelength domain to energy domain, the intensity should be scaled by a factor \( \lambda^2 \), due to the much smaller \( \Delta E \) for a given \( \Delta \lambda \) as \( \lambda \) becomes large; i.e., \( \Delta E \) from \( \lambda=300 \) to \( \lambda=301\,\text{nm} \) is 14meV; \( \Delta E \) for \( \lambda=500 \) to \( \lambda=501\,\text{nm} \) is 5.0meV.

By selecting a particular wavelength bandpass with the monochromator, and using the PMT’s output as an imaging signal, a spectrally resolved image can be obtained, as opposed to a panchromatic image. This will have poorer signal-to-noise compared to panchromatic imaging, but is useful for spatially mapping particular emission peaks.

Table III briefly outlines the different modes of CL.
Table III: Operational modes of cathodoluminescence.

<table>
<thead>
<tr>
<th>Mode</th>
<th>Detector</th>
<th>Monochromator</th>
<th>Beam</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spectroscopic imaging</td>
<td>PMT</td>
<td>Fixed</td>
<td>Scanned</td>
</tr>
<tr>
<td>Panchromatic imaging</td>
<td>PMT</td>
<td>None</td>
<td>Scanned</td>
</tr>
<tr>
<td>Serial spectroscopy</td>
<td>PMT</td>
<td>Scanned</td>
<td>Fixed or scanned</td>
</tr>
<tr>
<td>Parallel spectroscopy</td>
<td>CCD</td>
<td>Fixed</td>
<td>Fixed or scanned</td>
</tr>
</tbody>
</table>

As mentioned, signal-to-noise is always the overriding concern during CL experiments, due to the loss of carriers to non-radiative recombination, the loss of photons to internal absorption ($f_A$), and the difficulty with which photons escape the sample ($f_R$). One option that is not always available for increasing the usable CL signal is that of increasing the EHP generation rate $g_0$, Equation (3-9). As the specimen parameters $E_{eh}$ and $<n>$ are fixed, the two operator-controllable parameters are $E_0$ and $i_b$. Typically, $E_0$ will be fixed by the type of experiment being performed. In a GaN LED in plan view, for example, variations in $E_0$ will change the penetration depth and thus the layer that is being sampled. Figure 51 shows simulated data for a GaN LED structure.
Figure 51: Depth distribution of electrons in a GaN LED at typical SEM voltages. Left: CASINO\textsuperscript{169} simulation of EHP generation of 10keV electrons incident on GaN. Right: CASINO curves for 10, 20 and 30keV electrons incident on 2\(\mu\)m of GaN on a SiC substrate. 5x10\textsuperscript{5} simulated electrons per curve; curves normalized to have integrated area =1.

Figure 51 indicates that, to sample the quantum wells, the beam energy must be maintained in the 10~20keV range; to sample the p-side only, the beam energy should be well under 10keV.

Beam current is also not a free parameter. As mentioned in Section 3.1, increasing beam current leads to increases in SEM spot size, and will degrade the spatial resolution of images. This is particularly true at low keV or long working distances, which are often used in CL. Thermionic SEMs will have larger spot sizes than field-emission SEMs, especially toward the high-\(i_b\) range of operation. Further, an SEM will have an absolute limit to the beam current it can put on the sample; this may be \(\sim 0.1\text{nA}\) to \(\sim 1\text{\mu A}\), depending on the specific microscope design. Large beam currents can also contaminate or damage the sample, through hydrocarbon surface contamination, charging, heating, knock-on, or filling traps.

However, voltage-resolved and current-resolved studies are valuable ways to probe the details of the sample’s radiative transitions.
Because \( g_0 \) is proportional to \( E_0 \) and \( i_b \), the (nominal) number of EHPs generated can be kept constant by keeping the product \( E_0i_b \) constant. By varying the voltage, and adjusting \( i_b \) to keep the power constant, depth-resolved studies can be performed. Changes in spectra or images with \( E_0 \) at constant beam power indicate changes occurring due to depth. These changes can be due to a combination of changes in the sample with depth and changes in \( f_A \) with depth.

Current resolved studies are also very valuable. By keeping \( E_0 \) constant, the depth of excitation does not change. Therefore, changes in the spectra with changes in \( i_b \) will be due to saturation of recombination sites with increasing EHP generation rate.

It is valuable to plot the integrated intensity of a CL (or PL) peak, \( I_{\text{PEAK}} \), as a function of excitation density (\( i_b \) for CL). The effects of excitation intensity were originally studied in terms of photoluminescence; in the PL case, beam current \( i_b \) is replaced by laser power \( L \). Taguchi et al.\(^{220}\) and Schmidt et al.\(^{221}\) studied CdTe emissions as functions of laser excitation intensity \( L \); numerically, \( L \) was taken as the EHP generation rate\(^{220}\) or laser power\(^{221}\).

Taguchi et al.\(^{220}\) derived an \( I_{\text{peak}} \propto L^m \) law for excitons bound to neutral acceptors, where \( m=1.5 \); this was in good agreement with their experimental data. Schmidt et al.\(^{221}\) performed a more detailed analysis of the problem of \( I_{\text{peak}} \) vs. \( L^m \). They found three different regimes of behavior:

1. **Only a small fraction of carriers form into excitons; most recombination is of free electrons and holes, or via donor or acceptor sites.** In this case, \( m=2 \) for free- and bound exciton recombinations, and for the free-to-bound transitions \( eA^0 \) or \( hD^0 \), \( m=1 \).

2. **Most EHPs form excitons, and most recombination is via excitons.** In this case, \( m=1 \) for excitonic lines and \( m=0.5 \) for free-to-bound transitions.

3. **Neither free-to-bound nor excitonic recombination dominates.** In this case, \( \log(I_{\text{peak}}) \) vs. \( \log(L) \) plots will be curved, and no single \( m \) law is found for a given line.
In the case where $N_D^0$ and $N_A^0$ were not constant with $L$, Schmidt et al. did not find general rules regarding $I_{\text{peak}}$ vs. $L^m$ behavior. They did, however, generalize their theoretically derived rules and compare them to experiment. This led them to conclude that $1 < m < 2$ for free- and bound-exciton emission and $m < 1$ for free-to-bound and donor-acceptor pair emissions.

To summarize cathodoluminescence, it is similar to EBIC in that it is a localized beam-injection technique, and it is similar to photoluminescence in that the sample is excited into radiative recombination and the signal of interest is that of the light emitted. CL microscopy is useful for mapping emissions associated with particular defects or the bulk, and CL spectroscopy is used for studying the emission behavior as a function of depth, power, temperature or time.

### 3.2.6 CL and EBIC spatial resolution issues

In 1978, Donolato\textsuperscript{176} published a study in which simulations were used to estimate the effect of minority carrier diffusion length on CL or EBIC spatial resolution. Donolato concluded that $L$ had negligible effect on the width or contrast of an EBIC image; rather, defect depth and SEM beam energy (and consequent electron range) was found to dominate. This is because the carrier concentration falls off at a faster-than-exponential rate; specifically, for minority carrier density $\Delta n$:

$$\Delta n \propto \frac{1}{r} \exp\left[-\frac{r}{L}\right]$$

(3-43)

This was well accepted for some time; in 2002, Norman\textsuperscript{222} performed 2keV cross-sectional SEM-CL of GaAs/AlGaAs superlattices and demonstrated a resolution of 20nm. Norman also commented that various published papers (which Norman did not cite specifically in Ref. 222) had assumed that spatial resolution would be related to $L$, despite the calculations of Donolato\textsuperscript{176}.

However, more recent work by Phillips et al.\textsuperscript{159,223} has indicated that a transition region exists between the Donolato-type $L$-independent behavior and
diffusion-limited behavior. The arguments of Phillips et al. followed two parallel tracks to arrive at that conclusion. First, they performed plan-view CL on GaN epilayers, either undoped or Si-doped, under conditions of 1keV and ~nanometers spot-size. This gave clear dislocation images in CL, and by taking linescan data across the dislocations, fitting exponential curves to the linescan allowed them to estimate diffusion lengths of ~50nm or ~90nm in the Si-doped and nominally undoped materials.

Second, calculations using CASINO\textsuperscript{169} software modified to broaden the energy-deposition profiles by Equation (3-43), showed significant broadening of low-keV interactions volumes by diffusion, but almost no broadening of high-keV interaction volumes. Similar calculations have been performed as part of this work, and are presented as Figure 52, Figure 53, and Figure 54.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{Figure52.png}
\caption{CASINO simulations of diffusive probe broadening at 1keV on GaN.}
\end{figure}
Figure 53: CASINO simulations of diffusive probe broadening at 10keV on GaN.

Figure 54: CASINO simulations of diffusive probe broadening at 30keV on GaN.

These figures were computed for $10^4$ trajectories per simulation, for 20x20x20 element meshes and 1nm spot sizes. (This coarse meshing, compared to more typical 100x100x100 meshes, was necessary to assure convergence.) The target was assumed to be GaN. For Figure 52, a 1keV probe (R=14nm) was then diffused assuming L=50nm and $v_s \rightarrow 0$ cm/sec. Figure 53 and Figure 54 were computed for
10keV ($R=640\text{nm}$) and 30keV ($R=4000\text{nm}$) probes, and then diffused using $L=1000\text{nm}$ and $v_s \rightarrow 0 \text{ cm/sec}$.

As can be seen, the 1keV probe is broadened by a factor of ~5 by the diffusion process; the 10keV probe gains significantly larger tails, but the most significant core-region is broadened by only a few percent. The 30keV probe is negligibly changed by the $L=1000\text{nm}$ diffusion.

Figure 55 shows the same data as above, for the diffused $L>0\text{nm}$ cases, but with the $R$ ranges superimposed as circles tangent to the beam incidence point. The $R$ ranges are marked as red circles. It can clearly be seen that the diffused interaction volume is well approximated by $R$ for the high-energy cases, but the interaction volume for the low-energy case is clearly dependant almost entirely on diffusion.

![Figure 55: Diffused CASINO data from above, with superimposed $R_{K-O}$ ranges.](image)

In short, the $L$-independence predicted by Donolato $^{176}$ and defended by Norman $^{222}$ is clearly an important but special case, not the general case. Although the super-exponential decay of the diffusion profile means that even very long diffusion lengths will have negligible effects on probes of ~10-20keV or more, it is clear that low-energy probes, ~1keV, are purely diffusive in nature, as proven experimentally by Phillips et al.$^{223}$ A transition region of 1 to 10keV also seems to exist in which case the CL and EBIC spatial resolution moves from primary-electron-range dominated to minority-carrier-diffusion dominated.
3.3 Transmission electron microscopy

The transmission electron microscope (TEM) is the oldest EM technique; it is also the preferred technique for the direct crystallographic analysis of defects. A TEM will have an electron source, like those found in an SEM. After the electron gun, through, will be a much larger accelerating voltage. TEMs operate in the 80 to 1000+ keV range\textsuperscript{224,225}; 200keV is most common.

After the gun and high voltage will be the illumination stage; the purpose here is to place either a parallel beam or a highly convergent spot onto the specimen. The specimen is placed inside the objective lens, typically with a side-entry holder. After the objective lens is the projection system, and finally the detectors. Figure 56 illustrates schematically a TEM system.
The illumination system, consisting of the electron source, condenser lenses, and condenser aperture, provide the parallel or convergent illumination for the objective lens.

The sample is thin (1~500nm) to allow electrons to pass cleanly through it, and is inserted between the halves of the objective lens. The electrons of the beam enter the sample, and are either transmitted directly through, diffracted to a Bragg angle $\theta_B$, inelastically scattered to a random angle, or absorbed by the specimen$^{226}$.

Below the objective lens is the projection stage, which consists of an intermediate lens, a diffraction lens, and projector lenses. When the TEM is in imaging mode, the projection stage will magnify the image plane of the objective.

Figure 56: TEM schematic.
lens -- and, thus, an image of the sample -- onto the viewing screen or other
detector. When the instrument is in diffraction mode, the projection stage magnifies
the back focal plane of the objective lens -- and, thus, the diffraction pattern from the
sample -- onto the viewing screen.

The diffraction geometry of a parallel electron beam passing through a thin
sample is illustrated in Figure 57; the treatment follows Hammond\textsuperscript{227} and
Edington\textsuperscript{228}.

![Figure 57: Schematic of TEM diffraction geometry.](image)

In Figure 57, \(R\) is the physical separation of the diffraction pattern spots in the
detector, \(L\) the effective camera length of the TEM diffraction system, \(0\) and \(g\) the
diffraction vectors, and \(\theta_B\) the Bragg angle of diffraction. By the small angle
approximation, \(\theta_B=R/L\). From this relation and the Bragg law \(n\lambda=2d_{hkl}\sin(\theta_B)\), the
spacing \(d_{hkl}\) of the lattice planes that the diffraction vector \(g\) arises from can be
found. For the purposes of TEM, \(n\), the diffraction order, is always taken as 1. \(\lambda\) is
the relativistic wavelength of the electrons\textsuperscript{225}, and \(\lambda_{200keV}=0.00273\)nm. This leads to:

\[
d_{hkl} = \frac{\lambda L}{R}
\]  

(3-44)
By measuring the distance $R$ between $\mathbf{0}$ and $\mathbf{g}$, and measuring the TEM's $L$ value with a specimen of know crystal spacing, the spacing of the plane (hkl) can be calculated.

The major strength of TEM that it directly images crystal defects$^{229}$. In particular, dislocations in a crystal appear as dark lines in TEM images. By tilting the TEM specimen to a two-beam condition, where only the $\mathbf{0}$ and a single $\mathbf{g}$ diffraction vector are strongly excited, it dislocation contrast becomes pronounced. More importantly, dislocations which satisfy the condition $\mathbf{g} \cdot \mathbf{b} = 0$, for dislocation burger's vector $\mathbf{b}$, will be invisible, allowing the dislocation type to be determined. Additionally, densities of dislocations can also be measured.

To summarize TEM, it is a thin-foil technique employing high-energy electrons. By depending upon diffraction, TEM allows crystallographic defects in crystals to be quantitatively examined.

3.4 Focused ion beam

The focused ion beam (FIB) is a technique similar in principle to SEM. However, instead of an electron beam, it is a beam of ions that is projected onto a spot on the sample's surface. Although many different focused-ion-beam types exist$^{230,231,232,233}$, by far the most common for FIB is $\text{Ga}^+$. These gallium ions are produced by a liquid metal ion source (LMIS). In a LMIS, a blunt field-emitter tip is coated with a metal that has a high surface tension and low vapor pressure; this metal is heated above its melting point and placed near a kV-range potential. This will draw the liquid metal out into a cone with an apex $\sim 5\text{nm}$, and the associated high electric field will cause atoms in the liquid to field evaporate and atoms in the vapor to field-ionize$^{231,232}$. The source, as for a SEM, is characterized by brightness $\beta$:

$$\beta = \frac{I_s}{A_s \Omega_S}$$  \hspace{1cm} (3- 45)
Where \( I_s, A_s \) and \( \Omega_s \) are the current, area, and solid angle of the source. The typical brightness for a LMIS is \( \sim 10^6 \) A/cm\(^2\)-sr. The effective source size is on the order of nanometers to tens of nanometers, and the energy spread will typically be \( \sim 5-10\text{eV} \).

These ions will then be accelerated to a typical energy of 0.5~40keV. Below the LMIS and accelerating potential in the column will be a series of lenses; FIB uses electrostatic lenses, rather than the electromagnetic lenses seen in SEM. This is because of the large mass of ions compared to electrons, and their slow speed. The fields needed for magnetic lenses would be difficult to generate; thus, cylindrically symmetric electrostatic lenses are used. These lenses suffer particularly from chromatic aberration, and because of the large (5~10eV) energy spread of a LMIS, chromatic aberration is often the limiting factor for spot size.

Spot size and beam current are controlled by mechanical apertures that block more or less of the beam, as dictated by the needs of the operator. Typical beam currents range from \( \sim \)picoamps to tens of nanoamps; spot sizes from \( \sim \)5nm to 500nm are typical under these conditions. Octupole lenses are used to generate the scan field.

The reason FIB is of interest lies in the interaction of the beam with the sample. Although an electron beam, particularly at TEM energies, can cause knock-on type point-defect damage in a specimen, or may interact with gasses in the vacuum to etch or deposit onto the sample, the FIB beam, in contrast, is inherently destructive. This is due to the huge difference in mass between an ion and an electron.

The first application of a FIB beam (let us consider a Ga\(^+\) beam for the remainder of the discussion) is because it will produce a large number of secondary electrons; as a strong and surface-sensitive signal, SEs are useful for forming images. If the FIB beam is raster scanned and a SE detector is fitted to the sample chamber, the FIB can act as a scanning ion microscope (SIM), analogously to an SEM. Because the ion beam penetrates only shallowly into the sample, and is subject to ion-channeling effects, it provides strong crystallographic contrast on
polycrystalline specimens; in addition, the SE yield dependencies provide excellent topographic contrast. Thus, the SIM is useful as a high-resolution, high-contrast imaging technique in materials science\textsuperscript{233}.

FIB ions hitting the sample cause lattice damage by displacing atoms from their equilibrium positions, and from the implantation of gallium atoms into the sample. At an dose level $\sim 10^{14}$ ions/cm\textsuperscript{2}, As\textsuperscript{+} was found to cause enough surface damage as to render amorphous silicon surfaces\textsuperscript{230}; doses of Ga\textsuperscript{+} on the same order should amorphize other common semiconductors. Kato\textsuperscript{235} reviewed the literature on amorphous layer formation and performed FIB-TEM experiments, and found that 30keV Ga\textsuperscript{+} ions incident on Si produced an amorphous layer $\sim$20-30nm deep. This value again should be typical for Ga\textsuperscript{+} beams incident on other semiconductors.

The most important beam-specimen interaction in FIB, however, is that of sputtering. The keV-range Ga\textsuperscript{+} beam can physically remove atoms from the sample by sputtering; control of this process is used for micromachining, which is the primary application of FIB\textsuperscript{230,231,232,233}.

Through computer control of the scan of the Ga\textsuperscript{+} beam, boxes or other patterns can be cut into a sample. For example, to cut a line in a specimen, the FIB beam would scan back-and-forth along the desired line location. Every time the beam passes over a point, incident Ga\textsuperscript{+} atoms will sputter away a small number of surface atoms, eventually leaving behind a milled line. The rate at which atoms are removed from the surface is dependant upon the target material, the energy of the incident ions, and the current density of the incident ions.

In addition to sputter-milling, FIB can chemically etch specimens if a reactive gas (H\textsubscript{2}O, Cl\textsubscript{2}, I\textsubscript{2}, XeF\textsubscript{2}, etc.) is allowed to adsorb onto the sample surface\textsuperscript{232,234,235}. The ion beam cracks the gas, forming highly reactive fragments which bind to and volatilize the surface atoms, leading to enhanced milling rates.

By introducing other gasses, deposition can be performed. W(CO)\textsubscript{6} and C\textsubscript{7}H\textsubscript{17}Pt are the most commonly used\textsuperscript{232}, which deposit W or Pt metal, contaminated with C, O, and Ga. Although not as conductive as pure metal, these deposits are electrically conductive. More importantly, they are high-atomic-weight, and resist
By a combination of deposition and sputtering, the most important application of FIB is realized – the preparation of TEM specimens. As discussed in Section 3.3, TEM specimens must be very thin, ~100nm. FIB is an ideal tool for this application.

The most simple means of preparing a TEM specimen is that of the "H-bar" cross section. First, the semiconductor wafer or device to be prepared is cut to <3mm in length and polished to ≤50μm, and affixed to a standard 3mm TEM half-grid sample holder. A W or Pt strap, ~20x5μm, is deposited over the area of interest to protect it. The sample is then thinned from the top and bottom mechanically-polished-surface by a FIB beam at normal incidence. Progressively smaller beam currents (spot sizes) are used, until the sample is ~100 to 300nm thick for TEM examination. This is illustrated in Figure 58. As the electron-transparent membrane is thus milled at glancing incidence, the FIB damage will be relatively low. Low-keV final milling is used to reduce the damage further.
The other major type of TEM sample preparation by FIB is "lift-out". In lift-out, a bulk specimen is inserted into the FIB chamber, a W or Pt strap is deposited across an area of interest, and a small (~10x10x2 μm) membrane is cut from the sample; this segment will be left attached by only a small bridge of material. The sample is tilted ~60° to cut the bottom of the membrane free from the substrate. A micromanipulated tungsten needle is then inserted into the chamber and attached to the top of the membrane by metal deposition. The last bridge of material is then cut, freeing the segment, now attached only to the micromanipulator. This membrane is then moved to a 3mm TEM half grid, and attached to the grid by metal deposition. Finally, the ~2 μm thick specimen is thinned to electron transparency with progressively finer ion beams. A tutorial for the lift-out technique is presented by Rahman et al. Lift-out is illustrated schematically in Figure 59.
Figure 59: Illustration of FIB lift-out process.

(A) Trenches are cut, almost freeing the membrane. (B) The needle is attached, the membrane is cut free, and lifted out. (C) The membrane is attached to the TEM grid and the needle cut free. (D) The membrane is thinned to electron transparency.

H-bar sample preparation is simpler and has a higher success rate, but is more difficult to perform with site-specificity, as the feature of interest must but near the center of the mechanically polished section. Lift-out is more complicated and requires a FIB equipped with a micromanipulator, but is trivial to perform with site-
specificity. Even the "low" success rate of lift-out can far exceed 90% with a well-established protocol.

FIB can also be used to make so-called "SEM cross-sections," by milling a trench into a bulk sample and polishing one of the faces, or by polishing a cleaved or mechanically prepared face. The examination of such FIB-milled cross-sections with CL and EBIC will be discussed later, in section 5.4.

To summarize FIB, a source of Ga⁺ ions produces a keV-range beam, which can then be scanned under computer control across a sample surface to image, remove material, or sputter material, all down to the nanometer scale. FIB is particularly valuable for the preparation of TEM and SEM specimens of specific geometry or of a specific site.

3.5 Analytical techniques summary

Scanning electron microscopy is a relatively simple, but powerful, technique for the examination of materials' near-surface structure and properties. The examination of a semiconductor by SEM invariably results in the injection of electron-holes pairs into the sample; this can be utilized to perform the beam injection techniques of electron beam induced current and cathodoluminescence. EBIC allows the examination of electronic properties of the sample, including defect recombination and minority carrier transport properties. Although there exists a significant mathematical literature for the analysis of minority-carrier diffusion length and surface recombination by X-EBIC, a systematic evaluation of the techniques has not previously been published. Sections 3.2.3 and 3.2.4 above attempt to fill in this gap by critically evaluating these models; Section 5.2 below subjects these models to experimental tests. CL is a powerful technique for the examination of optical properties of semiconductors and defects, and can yield information about recombination pathways and their identities. Transmission electron microscopy is a powerful crystallographic probe. The focused ion beam is a microscope in its own right, but most useful as a sample-preparation tool for the other discussed techniques.
4. Motivation and goals

4.1 Motivation

Chapters 1 and 2 outlined how the III-nitride semiconductors are important commercial materials with major applications in the solid-state lighting and high-frequency amplifier industries. However, due to lattice-mismatched heteroepitaxial growth, these devices contain significant densities of defects. It is, indeed, amazing that these devices are in any way operational when their dislocation densities are compared to those of conventional III-V semiconductors. However, the natures of the defects within GaN and related materials are not well understood.

Chapter 3 outlined a number of analytical methods -- particularly CL and EBIC -- that can be used to probe optoelectronic defects in semiconductor thin-films and devices. CL is useful for probing the optical properties of defects or other recombination pathways; EBIC is a powerful probe of local electronic properties and recombination rates.

This research was undertaken to apply the techniques of CL and EBIC, supported by other techniques such as TEM and FIB, to probe the localized optoelectronic character of the bulk and of defects in GaN-based quantum-well light emitting diodes grown on SiC substrates.

The goals of this work can be broken up into a number of sub-sections, listed below.

4.2 Modeling of the polarization fields in GaN

Polarization fields in GaN-based QWs drastically affect the light-emission properties of the device; Bunker et al.\textsuperscript{133,172} performed a series of CL experiments meant to elucidate the presence and magnitude of polarization fields in GaN-based LEDs. Section 5.1 of this work outlines a quantum-mechanical computer model intended to aid the interpretation of the SEM-CL experiments of Bunker et al. Also presented are experiments performed using electroluminescence injection density, and a mathematical model developed to help interpret the EL results.
4.3 Development of experimental apparatus and techniques

The remainder of this work involves CL and EBIC experimentation. However, CL and EBIC are not trivial techniques to employ, and experimental apparatuses were designed and built for these experiments. Although not a goal per se of this work, the development, implementation, and continuing improvements of these apparatuses are a necessary prerequisite to the later experiments, and consumed the majority of the man-hours represented by this work. The experimental apparatuses are described in Sections 5.2.1 and 5.3.1. In addition, procedures for preparation of specimens from fully-processed commercial LED packages are described.

4.4 Cross-sectional EBIC

Among the most important optoelectronic properties of a semiconductor device are the minority carrier diffusion lengths and surface recombination velocities. As such, the critical review and theoretical analysis of X-EBIC mathematical models outlined above in sections 3.2.3 and 3.2.4 was undertaken. The theoretical analysis leads to the hypothesis that the model of Bonard and Ganière\textsuperscript{194} will fare best when confronted with experimental data. It was also hypothesized that, despite the drawbacks associated with loss of signal-to-noise ratio, low-keV EBIC will be mandatory for quantitative analysis. Section 5.2.3 describes the acquisition of experimental cross-section EBIC data from GaN-based LEDs, and the analysis of the data via the different theoretical models. In addition to the L and s data acquired, conclusions are drawn regarding the validity and value of the different theoretical techniques.

4.5 CL and EBIC analysis of defects

The two dominant defects in current-generation GaN-based LEDs are threading dislocations and the yellow-band defect. In order to study these two defects, CL in both plan-view and cross-sectional configurations, as well as plan-
view EBIC, were performed. The experimental procedures and results are described in Section 5.3.

4.6 Analysis of FIB-induced damage

Ultra-high-resolution CL and EBIC will require examination of the samples in a STEM equipped with CL and EBIC detectors. Preparation of samples from LED devices or wafers will require FIB-based sample preparation, as described in Section 3.4. However, the exact effects of the FIB beam damage on the CL and EBIC response of GaN is not known. Studies of FIB-prepared cross-sections, in comparison to mechanically polished surfaces, are described in Section 5.4.

4.7 Combined CL/EBIC and TEM

By combining SEM-CL/EBIC examination of the optoelectronic behavior of defects with lift-out FIB preparation of TEM samples, the crystallographic nature of defects can be determined by TEM and compared to the optoelectronic nature elucidated in CL and EBIC. In Section 5.5, a procedure for using TEM to examine defects identified and characterized by CL and EBIC is described.
5. Experiments, results, and discussion

5.1 Modeling of polarization fields

5.1.1 Experimental results

The purpose of modeling the polarization fields in GaN-based LEDs is to allow comparison of shifts expected from polarization effects to shifts observed in experiment. Bunker et al.\textsuperscript{133,172} performed SEM-CL spectroscopy on a green-range GaN-based LED. Bunker's three main results were:

1. A blueshift of 52meV (11nm) in CL emission peak (at constant SEM-beam conditions) as reverse bias increased up to ~14V, and a redshift as reverse bias increased beyond ~14V.
2. A blueshift of 52meV (11nm) in CL emission peak as the SEM beam current was increased at constant accelerating voltage at zero reverse bias.
3. A blueshift of 40meV (10nm) in EL emission peak as EL drive current increase from 2.1 to 5.0V.

The results of Bunker et al. are reprinted as Figure 60.

Figure 60: CL and EL results of Bunker et al.\textsuperscript{133} (A) CL peak shift with reverse bias; (B) CL peak shift with SEM \textit{i}_b; (C) EL peak shift with drive voltage.

Reused with permission from Ref. 133. Copyright 2005, American Institute of Physics.

Bunker et al. concluded that these results were consistent with the action of a piezoelectricity-induced internal electric field and the concomitant QCSE in the LED's quantum wells. They performed a rough calculation, assuming that the 14V
reverse bias returned the QWs to flatband, and that the entire 14V-bias dropped across the depletion region. By dividing the 14V bias by the expected depletion region width, a field $E_{QW} \approx -1.0 \pm 0.2$MV/cm was calculated. Identical experiments were then performed with a blue LED, and no blueshifts were observed within the spectrometer resolution of ~2nm.

The modeling presented below was undertaken in order to attempt to infer more accurate values of $E_{QW}$.

5.1.2 Computer model

First, it was necessary to devise a simplified model for the physical structure of the QW LED. They were assumed to be SQW devices with a QW width $L_Z$, and to consist of three layers, GaN/InGaN/AlGaN, where the quantum well was InGaN and the other layers cladding. Bandgaps of GaN and AlGaN were taken as 3.40 and 4.00eV; $E_{\text{gap}}$ of the InGaN was calculated for each individual case examined. The band structure used for the calculations is drawn in Figure 61.
Figure 61: Quantities used for computer model.

The offsets of the conduction to valence bands A₁ and A₂ were calculated after the InGaN bandgap $E_{\text{gap}}$ was known; the ratio of $A₁/A₂$ was taken as 0.36.

The system was taken to be 20nm, running from -10 to +10nm Z-coordinates, with the QW centered at $Z=0$. When an electric field $E$ was applied to the system, it was taken as a slope on the QW, which pivoted about the $Z=0$ point. Thus, the band edges in the GaN or AlGaN would be moved up or down from the values listed above (but stay horizontal). Electron and hole effective masses were assumed constant as $m^{*}_e=0.2m_e$ and $m^{*}_h=1.2m_e$, for electron rest-mass $m_e$. Thus, the entire computation was performed within the effective-mass approximation.

The QW electron and hole states can be found by solving the stationary Schrödinger equation:

$$\nabla^2 \psi = \frac{2m^{*}}{\hbar^2} \left[ V(\vec{r}) - E \right] \psi$$  \hspace{1cm} (5-1)
By assuming that the QW width is much smaller than its in-plane extent, the Schrödinger equation reduces to its 1D case:

\[
\frac{\partial^2 \psi}{\partial z^2} = \frac{2m^*}{\hbar} [V(z) - E] \psi
\]

(5-2)

Here, \( V \) is the band-edge potential, \( \psi \) the wavefunction, and \( E \) the energy of the state. The potential of the band-edges can be meshed as a vector of discrete values \( V = V_i \) at points \( Z_i \) in a manner illustrated in Figure 62.

![Figure 62: Finite mesh scheme for band-edge potentials.](image)

If the assumption is made that only states confined to the QW are of interest, then at \( Z = Z_{\text{max}} = 10\text{nm} \) and \( Z = Z_{\text{min}} = -10\text{nm} \) the values of \( V(Z) \) can be taken as infinite; this will result in a wavefunction value \( \psi = 0 \) at those points. Because the values of the boundary points of the problem are known, it is possible to solve the Schrödinger equation as a finite-difference two-point boundary value problem (2PBVP). Kincaid and Cheney\textsuperscript{241,242} give a technique for solving this 2PBVP.

Following the treatment of Kincaid and Cheney, let the second-order differential equation \( x^{(2)} \) be defined by:
\[
\begin{align*}
x'' &= f(t, x, x') \\
x(A) &= \alpha; \quad x(B) = \beta
\end{align*}
\] (5-3)

The interval \([A, B]\) is discretized into \(n\) points \(t_0, t_1, t_2\ldots t_{n+1}\), where \(t_0 = A\) and \(t_{n+1} = B\). The spacing \(h = (B-A)/(n+1)\). If the approximated value of \(x(t_i)\) is taken to be \(y_i\), then the discrete system of equations is:

\[
\begin{bmatrix}
y_{i-1} - 2y_i + y_{i+1} \\
y_0 = \alpha \\
y_{n+1} = \beta
\end{bmatrix}
= f\begin{bmatrix}t_i, y_i, \left(2(y_{i+1} - y_{i-1})h^{-2}\right)\end{bmatrix} \quad (1 \leq i \leq n) \tag{5-4}
\]

Let the linear case be assumed, where:

\[
f(t, x, x') = u(t) + v(t)x + w(t)x'
\] (5-5)

Then, the abbreviations \(a, b, c\) and \(d\) can be introduced:

\[
\begin{align*}
a_i &= -1 - \frac{1}{2}hw_{i+1} \\
b_i &= -h^2u_i \\
c_i &= -1 + \frac{1}{2}hw_i \\
d_i &= 2 + h^2v_i
\end{align*}
\] (5-6)

Which results in a system of equations:

\[
\begin{bmatrix}
d_1 & c_1 \\
a_1 & d_2 & c_2 \\
a_2 & d_3 & c_3 & \ddots \\
\vdots & \ddots & \ddots & \ddots \\
a_{n-2} & d_{n-1} & c_{n-1} \\
a_{n-1} & d_n & c_n & \ddots \\
a_{n-1} & d_n & c_n & \ddots & \ddots & \ddots
\end{bmatrix}
\begin{bmatrix}
y_1 \\
y_2 \\
y_3 \\
\vdots \\
y_{n-1} \\
y_n
\end{bmatrix}
= \begin{bmatrix}
b_1 - a_0\alpha \\
b_2 \\
b_3 \\
\vdots \\
b_{n-1} \\
b_n - c_n\beta
\end{bmatrix}
\] (5-7)

In the case of the Schrödinger equation, the substitutions are made that \(x = \psi, \ x' = \partial\psi/\partial z, \ x'' = \partial^2\psi/\partial z^2\), \(t = z\), \(A = Z_{\text{min}}\), \(B = Z_{\text{max}}\), \(\alpha = \beta = 0, \ h = a, \ u(t) = w(t) = 0, \ v(t) = 2m [V(z) - E]/\hbar, \ a_i = c_i = -1, \ b_i = 0, \) and \(d_i = 2 + m^*a^2[V_i - E]/\hbar^2\). The term \(a\) is introduced as the spacing of the mesh points in the quantum mechanical problem.
This allows the linear system to be re-written as:

\[
\begin{bmatrix}
2 + \frac{2m^* a^2}{\hbar^2} (v_1 - E) & -1 &  &  \\
-1 & 2 + \frac{2m^* a^2}{\hbar^2} (v_2 - E) & -1 & \\
-1 & -1 & 2 + \frac{2m^* a^2}{\hbar^2} (v_3 - E) & \\
&  &  & \\
&  &  & \\
\end{bmatrix}
\begin{bmatrix}
\psi_1 \\
\psi_2 \\
\psi_3 \\
\vdots
\end{bmatrix} = \begin{bmatrix}
0 \\
0 \\
0 \\
\vdots
\end{bmatrix} \tag{5-8}
\]

Which can equivalently be re-written in its final form as:

\[
\begin{bmatrix}
V_1 + \frac{\hbar^2}{ma^2} - E & -\frac{\hbar^2}{2ma^2} &  &  \\
-\frac{\hbar^2}{2ma^2} & V_1 + \frac{\hbar^2}{ma^2} - E & -\frac{\hbar^2}{2ma^2} & \\
-\frac{\hbar^2}{2ma^2} & -\frac{\hbar^2}{2ma^2} & V_1 + \frac{\hbar^2}{ma^2} - E & \\
&  &  & \\
&  &  & \\
\end{bmatrix}
\begin{bmatrix}
\psi_1 \\
\psi_2 \\
\psi_3 \\
\vdots
\end{bmatrix} = \begin{bmatrix}
0 \\
0 \\
0 \\
\vdots
\end{bmatrix} \tag{5-9}
\]

This solution (5-9) is also given (without derivation) by Ferry\textsuperscript{243}, and is abbreviated as:

\[
[S] \bar{\psi} = \bar{0} \tag{5-10}
\]

The term \(E\) appears only on the diagonal matrix elements of \([S]\); thus, it follows that if \([I]\) is the identity matrix, \(E\) can be removed from \([S]\), forming \([S']\):

\[
[S'] \bar{\psi} = \left[\bar{S}^* \right] - E[I] \bar{\psi} = \bar{0} \tag{5-11}
\]

It then follows that:

\[
[S^*] \bar{\psi} - E[I] \bar{\psi} = \bar{0} \tag{5-12}
\]

By grouping the identity matrix and the wavefunction, (5-13) is found:

\[
[S^*] \bar{\psi} - E[I] \bar{\psi} = [S^*] \bar{\psi} - E \bar{\psi} = \bar{0} \tag{5-13}
\]
This equation is useful because for some matrix \([A]\), with eigenvalue \(\lambda\) and associated eigenvector \(x\), the standard eigenvalue problem\(^{244}\) is defined to be:

\[
[A] \vec{x} = \lambda \vec{x}
\]  \hspace{1cm} (5-14)

It is clear that Equation (5-13) can then be cast as an eigenvalue problem:

\[
[S^*] \vec{\psi} = E \vec{\psi}
\]  \hspace{1cm} (5-15)

In this case, the energy \(E\) of some state \(i\) is an eigenvalue of the \([S^*]\) matrix, and the associated eigenvector \(\psi\) is the wavefunction of state \(i\).

This derivation is valuable as \(E\) and \(\psi\) can be determined using standard linear algebra eigenvector and eigensystem numerical methods.

In order to test this technique, the standard particle-in-a-box\(^{245}\) and quantum harmonic oscillator\(^{246}\) problems were solved. These problems were chosen because the solutions for \(E\) and \(\psi\) are know analytically, and the known solution can be compared to the numerical solution. This model gave less than 1% error for the first 6 energy levels of each case, when using \(n=100\). The model was then applied to a finite potential well of depth 0.2ev and width \(L_z=4\text{nm}\), in which case \(m^*\) was taken to be 0.1, 1.0 or 10\(m_e\). In this case, the energies of the bound states, and leakage of the wavefunctions into the barriers, followed the expected trends in \(m^*\).

The eigen-routines used are those of Pozo\(^{247}\). The downloaded routines, and the entirety of the model, were written in the C++ programming language.

For the actual GaN-polarization solutions, \(n=300\) was used, meaning 300x300 element matrices \([S^*]\) were solved. This means that 300 eigenvalues and 300 eigenvectors were returned by the linear algebra routines. However, because of the \(V=\infty\) boundary conditions at \(Z_{\text{min}}\) and \(Z_{\text{max}}\), only those wavefunctions whose energy is lower than the lowest edge of the quantum well are meaningful. In the calculations described below, wavefunctions in which \(E\) is above the GaN or AlGaN CB edge (for electrons) or below the GaN or AlGaN VB edge (for holes) are neglected from further consideration. Because numerical algebra routines are not quantum mechanics-specific, it is necessary to multiply the eigenvectors by a constant factor in order to turn them into wavefunctions. The factor is obviously the
integrated area under the eigenvector-squared graph; this is to ensure that the wavefunctions calculated numerically satisfy the normalization condition of quantum mechanics:

\[
\int \int \int \psi^* \psi \, dV = \int_{Z=Z_{\min}}^{Z=Z_{\max}} \psi^* \psi \, dZ = 1
\]  

(5-16)

The term \( \psi^* \) is the complex conjugate of \( \psi \); the triple integral is the general case, and the single integral is a simplification based upon the assumption that the system being studied here remains confined between \( Z_{\min} \) and \( Z_{\max} \) and can be treated one-dimensionally.

As discussed in Section 2.4.1, the injection of EHPs into a QW with a polarization-induced field \( E_{\text{POL}} \) will result in screening of the field by the spatially separated carriers. The separation of the centroids of the electron wavefunction \( \psi \) and the hole wavefunction \( \phi \) to opposite sides of the QW will result in a non-zero charge density function \( \rho(Z) \). Through the Poisson equation, this electron-electron interaction will result in a carrier-induced field \( E_{ee} \) and a carrier-induced potential \( V_{ee} \):

\[
\frac{d^2 V_{ee}}{dZ^2} = \frac{dE_{ee}}{dZ} = -\frac{\rho(Z)}{\varepsilon_{GaN} \varepsilon_0}
\]

(5-17)

It is simple to calculate \( \rho(Z) \):

\[
\rho(Z) = q n_{2D} \left( -|\psi(Z)|^2 + |\phi(Z)|^2 \right)
\]

(5-18)

Here, \( n_{2D} \) is the sheet density of injected carriers (EHPs/cm\(^2\)), \( \psi \) and \( \phi \) are normalized electron and hole wavefunctions, and \( q \) the elementary charge. \( |\psi|^2 \) and \( |\phi|^2 \) are the probability densities of the electrons and holes.

For a system with known \( \psi \) and \( \phi \) -- as calculated above -- and a user-input \( n_{2D} \), this charge density is found. From this charge density, and the assumption that the total electric field at the boundary of the system \( E(Z_{\min})=0 \), the electric field \( E_{ee} \) due to carriers can be calculated by simple integration:
\[ E_{ee}(Z) = -\frac{1}{\varepsilon_{GaN} \varepsilon_0} \int_{Z' = Z_{\min}}^{Z' = Z} \rho(Z')dZ' \]  

Once the function \( E_{ee}(Z) \) is known, the electron-electron potential \( V_{ee} \) can be found. With the assumption \( V_{ee}(Z_{\min}) = 0 \), this yields:

\[ V_{ee}(Z) = \int_{Z' = Z_{\min}}^{Z' = Z} E_{ee}(Z')dZ' \]  

This Poisson calculation was performed for the simulated GaN-based LED structures after the above-described Schrödinger calculations were complete. \( \varepsilon_{GaN} \) was taken to be 8.9. Under the assumptions of \( E_{ee}(Z_{\min}) = V_{ee}(Z_{\min}) = 0 \), and because the number of electrons is equal to the number of holes (=\( n_{2D} \)) resulting in a \( \rho(Z) \) curve whose integral from \( Z_{\min} \) to \( Z_{\max} \) is zero, it follows that \( E_{ee}(Z_{\max}) = 0 \) but that in general \( V_{ee}(Z_{\max}) \neq 0 \).

For these screening calculations, the approximation was made that all \( n_{2D} \) carriers resided in the lowest-subband, \( \psi_1 \) or \( \phi_1 \). Because of the strong decay of the Fermi function, this approximation is reasonable.

Figure 63 through Figure 66 demonstrate a sample calculation for the bands, wavefunctions, charge density, \( E_{ee} \) field and \( V_{ee} \) potential for a QW with \( L_z = 2.5 \text{nm}, \) \( E_{POL} = -1.0 \text{MV/cm}, \) \( n_{2D} = 10^{13}/\text{cm}^2, \) and \( E_{gap} = 2.3 \text{eV}. \) Note that the wavefunctions pool on opposite sides of the QW, the charge density follows the shape of the wavefunctions, \( E_{ee} \) peaks in between the wavefunctions in the center of the well, and \( V_{ee} \) increases monotonically, with a non-zero value (~25mV) at \( Z_{\max}. \)
Figure 63: Band edges (thick lines) and wavefunctions (dashed lines) for example case.

Figure 64: Charge density (dashed line) derived from wavefunctions
Figure 65: Electron-electron derived field $E_{ee}$ (dashed line) derived from charge density.

Figure 66: Potential $V_{ee}$ (dashed line) derived from $E_{ee}$

Once $V_{ee}$ was found, it was used to perform self-consistent iterations of the Schrödinger and Poisson equations. The Schrödinger equation was modified such that:

$$V(Z) = V_{band}(Z) + V_{ee}(Z)$$  \hspace{1cm} (5-21)
In this case, $V(Z)$ is the screened band-edge and $V_{\text{band}}$ the initial band-edge, whose shape is based upon the polarization field $E^{\text{POL}}$. In practice, the above-mentioned matrix calculations were performed using the modified version of $V(Z)$, Equation (5-21), but with $V_{ee}=0$ for the first iteration, before the first wavefunction set was calculated for input into the $V_{ee}$ equation.

Thus, the Schrödinger solver obtained the wavefunctions $\psi$ and $\phi$; this led to a solution of the Poisson equation yielding a value for $V_{ee}$, which was fed back into the Schrödinger solver for a second iteration, and continued until convergence. Thus, the Schrödinger and Poisson equations were iterated self-consistently for convergence. This resulted in calculation of screened wavefunctions and band-edges $V(Z)$.

An example, for a 5nm-QW with a 0.5MV/cm field and $n_{2D}=10^{14}$EHP/cm$^2$ is shown in Figure 67. As is apparent, convergence was achieved for this very high injection level after only 11 iterations. For simplicity, only the conduction band is shown; the valence band mirrors it.
Figure 67: Illustration of self-consistent iteration to convergence.

Once convergence is achieved, the bandstructure, the subband state energies (eigenvalues), and subband wavefunctions (eigenvectors), are known for the case of some injection level \( n_{2D} \), some QW width \( L_z \) and QW material of bandgap \( E_{\text{gap}} \). Then, in order to properly model the CL emission, it is necessary to model the density of states and the optical transitions between the EHPs.

The DOS function \( Z(E) \) for a QW is a stair-step function, each step being of magnitude \( m^*/\pi \hbar^2 \) being associated with a single confined state\(^{25,248} \). Figure 68 is a schematic drawing of this \( Z(E) \) function for electrons; \( E_{\text{CB}} \) is the conduction band edge, and \( E_1, E_2, \ldots \) are the subband energies \( E \) calculated by the above Schrödinger solution. This shape arises from the fact that the electrons are quantum-confined in the \( Z \)-direction of the QW, but are free to move in the \( XY \)-plane of the QW. This is illustrated in Figure 69, adapted from Davies\(^{25} \). The QW confinement will be quantitatively discussed below.
After density of states $Z(E)$ is known, the next step is to calculate the number of states. This requires the Fermi function for the system. A GaN-LED under electron beam bombardment for cathodoluminescence will contain excess electrons.
and holes; it is not at equilibrium. Thus, quasi-Fermi (QF) levels are a more appropriate form of calculation. QF levels are illustrated schematically in Figure 70:

![Figure 70: Fermi level $E^F$ and quasi-Fermi levels $E^{QF}$ in a semiconductor.](image)

Let the Fermi function for either electrons or holes be defined as $F(E)$, and the density of states as $Z(E)$. Then, for the case of the electrons as an example,

$$N^{TOT} = \int_{E=E_{cs}}^{E=\infty} \frac{dN_{e^-}}{dE} dE = \int_{E=E_{cs}}^{E=\infty} Z_{e^-}(E)F_{e^-}(E) dE \approx \int_{E=E_1}^{E=E_{top}} Z_{e^-}(E)F_{e^-}(E) dE \quad (5-22)$$

Here, $N^{TOT}$ is the total number of states, $E_1$ the energy of the lowest state, and $E_{TOP}$ is the top of the QW, i.e., the energy where the confined states give way to the continuum. The Fermi function is:

$$F(E) = \frac{1}{1 + \exp \left[ \frac{E_{F,QF} - E}{kT} \right]} \quad (5-23)$$

Again using the conduction band as an example, this mathematical relation is illustrated in Figure 71; $N^{TOT}$ is the area under the $dN(E)/dE$ curve:
The question remains, "How to determine the quasi-Fermi level?" Once the QF level is known, the total number of states can be calculated. However, as discussed above, injected-sheet-carrier density $n_{2D}$ (See Section 2.4.1 and Ref. 83) is an input parameter of the problem. Thus, $N^{TOT}=n_{2D}$ was known, and it became possible to find $E^{QF}$ for both the electrons and holes from known $Z(E)$ and $N^{TOT}$.

A starting $E^{QF}$ was guessed, and a five-point Gauss integration routine solved the approximate version of Equation (5-17). This yielded a value $N^{(1)}$; treating $N^{TOT}-N^{(1)}$ as a non-linear equation in $E^{QF}$; this was solved by the Newton-Raphson method$^{241,242}$. Thus, from the user-input $n_{2D}$ and the known shape of $Z(E)$, $E^{QF}_{e}$ and $E^{QF}_{h+}$ were found.

Because luminescence arises from the transition of an electron from a CB to VB state, the next step of the calculation of the luminescence is to determine the oscillator transition probability between CB and VB states. For some electron subband state $\psi_i$ and some hole subband state $\phi_j$, the oscillator strength $a_{ij}$ is:

$$a_{ij} = \left( \int_{Z=Z_{\text{min}}}^{Z=Z_{\text{max}}} \psi_i^* \phi_j dZ \right)^2$$

(5-24)

In a symmetrical quantum well, states with different quantum numbers $i$ and $j$ would have a zero oscillator strength, and those with $i=j$ would have $a=1$. However, because the QWs in this problem are not symmetrical, it is necessary to perform the
integration to determine $a_{ij}$ for each $i$, $j$ combination. This so-called "orthogonality rule" is illustrated schematically in Figure 72.

![Figure 72: Demonstration of orthogonality rules.](image)

For both the density/number of states calculations mentioned above and the $a_{ij}$ calculations, only those states $i$ or $j$ confined to the quantum well were considered; for deep triangular wells resulting from large $E^{POL}$, this number could be very large. It was determined that cutting off at a maximum of $i,j=8$ had no appreciable effect on the final result (due to the fast drop off of the Fermi function) and this simplification of $i,j \leq 8$ was used.

Next it was assumed (as implicit in the effective mass approximations used above) that the subbands were quantized in growth-direction momentum $K_Z$ and parabolic in the in-plane momentum $K_{XY}$. By taking the energies $E_i$ and $E_j$ of the subbands $i$ and $j$, an effective band-gap separating the $i,j$ subbands is $E_{gap}'=E_i - E_j$. If we assume conservation of momentum and that electron-hole transitions must be vertical in energy-momentum space, the energy of a transition occurring at momentum $K_{XY}$ is found as:

$$E_{photon} = E'_{g} + \frac{\hbar^2 K^2_{XY}}{2m_e^*} + \frac{\hbar^2 K^2_{XY}}{2m_h^*}$$  \hspace{1cm} (5-25)
From all of the above, it is now possible to calculate the relative intensity of an emission of some given energy E. In short, by summing from $K_{XY}=0$ to the $K_{XY}$ value that gives either the electron or hole state in the energy continuum, the intensity arising at E from two subbands $i,j$ will be:

$$I = \left[ \frac{\text{#States in } i \text{ at } K_{XY}}{\text{#States in } j \text{ at } K_{XY}} \right] \left[ \text{transition strength} \right]$$

$$I = [Z_i (E_{CB}) F_{e-}(E_{CB}) \Delta E_{CB} Z_j (E_{VB}) F_{h+}(E_{VB}) \Delta E_{VB} \left\langle |i,j| \right\rangle^2]$$  
(5-26)

The calculation is performed for bins of $\Delta K_{XY}$, which results in energy steps of $\Delta E_{CB,VB}$. By finding this I(E) data for every combination $i,j$ and summing these contributions, the overall I(E) curve is found. This I(E) curve is the predicted cathodoluminescence spectrum, based upon the factors $L_Z$, $E_{\text{gap}}$, $E^{\text{POL}}$, and $n_{2D}$.

Although the calculation could be well-approximated by restricting it to the case of $e_1 \rightarrow h_1$ case (i.e., where $i=1$ and $j=1$), all confined states up to $i,j=8$ were used.

By guessing different values of $E_{\text{gap}}$ for a known $L_Z$, etc., and comparing the peak CL emission predicted to the measured flatband voltage, it is possible to find an estimate of the $E_{\text{gap}}$ value. This allowed the bandgap to be inferred from an assumed QW width and known flatband emission peak.

In calculations dealing with reverse bias, $E^{\text{POL}}$ was replaced with $E^W$, the total field in the well. $E^W$ was calculated from $E^{\text{POL}}$, the depletion region width $W$, and applied reverse bias $V_a$:

$$E^W = E^{\text{POL}} - \frac{V_a}{W}$$  
(5-27)

The standard formula for depletion region width is:

$$W(V_a) = \sqrt{\frac{2 \varepsilon_0 \varepsilon_{\text{GaN}} (N_a^- + N_d^+) (V_0 - V_a)}{q N_a^- N_d^+}}$$  
(5-28)

Here, $N_a^-$ is the number of ionized acceptors, and $N_d^+$ the number of ionized donors. $V_0$ is the built-in potential. $V_0$ is found from:

$$V_0 = \frac{kT}{q} \ln \left( \frac{N_d^+ N_a^-}{n_i^2} \right)$$  
(5-29)
The value \( n_i^2 \) is the intrinsic carrier concentration. Total donor and acceptor levels in the n- and p-sides were taken to be \( 10^{18} \) Si/cm\(^3\) and \( 10^{20} \) Mg/cm\(^3\), respectively. This resulted in calculated ionized acceptor level estimates of \( N_d^+ = 9 \times 10^{17} \) Si/cm\(^3\) and \( N_a^- = 1 \times 10^{19} \) Mg/cm\(^3\) at room temperature. \( V_0^{GaN} \) was calculated to be 3.26V, and \( V_0^{AlGaN} \) of 3.87V. The overall \( V_0 \) was approximated as the average, 3.66V. From these dopant and \( V_0 \) values, \( W \) could be found as a function of reverse bias, \( W(V_a) \), allowing the field on the well \( E^W \) to be calculated for some given \( E^{POL} \) and \( V_a \).

Figure 73 shows example calculations for \( E^W \) under the assumptions above, for \( E^{POL} = +1, 0, \) or \(-1\) MV/cm. Note that, as calculated by Bunker et al. above, a reverse bias of \( \approx -15 \)V gives flatband condition for an initial -1MV/cm well.

![Figure 73: Total QW field as a function of reverse bias.](image)

From this, it is possible to calculate curves of emission peak as a function of reverse bias, under assumptions of the flatband emission peak (which allows calculation of \( E_{gap} \)) and input values of \( L_z, E^{POL}, \) and \( n_{2D} \).
That discussion concludes the outline of the model developed; the computer
code is re-printed in Appendix 3. The above section’s salient points can be
summarized briefly:

1. From input parameters of L<sub>Z</sub>, E<sub>gap</sub>, E<sup>POL</sup>, and n<sub>2D</sub>, predicted emission peaks
can be calculated.

2. From input parameters of E<sup>POL</sup>, n<sub>2D</sub>, and the experimental λ<sub>PEAK</sub>, E<sub>gap</sub> can be
found, allowing subsequent evaluation of the other dependencies.

5.1.3 Results of model & comparison of model to experimental data

The first output of the model to consider is that of the CL spectrum shape
predicted. An example is shown in Figure 74.

![Figure 74: CL spectrum, model vs. experiment. Experiment conditions: 30keV, 4nA
beam, -13.75V reverse bias. Model: L<sub>Z</sub>=2.5nm, n<sub>2D</sub>=10<sup>8</sup>/cm<sup>2</sup>, E<sup>POL</sup>=0MV/cm (i.e.,
flatband)](image)

The short-wavelength (high-energy) side of the peak shows good agreement
between the model and experiment; the long-wavelength (low-energy) side shows
poorer agreement. Indeed, when the ~2nm resolution of the spectrometer used to
acquire the data in Figure 74 is considered, the error in the short-wavelength side is only \( \sim 3 \text{nm} \) (15meV), which is surprisingly good considering the number of assumptions and approximations in the model. As is obvious from Figure 74, the error on the other side is more severe. The error is interpreted to derive mostly from the neglecting of quantum-confinement to In-rich areas, which will be of non-uniform size and composition, and therefore give a distribution of recombination energies below that of the lowest-quantized \( e_1 \rightarrow h_1 \) transition. Further, if excitons are present, their binding energy would result in slight narrowing of the effective gap, and this was also ignored.

It is a prediction of the QCSE that a wider well, with all other factors equal, will suffer a far more significant redshift than a thinner well. The QWs of the studied devices are expected to be 2.5nm wide; by assuming that the \( \lambda_{\text{PEAK}} \) at maximum blueshift corresponds to the flatband condition, and using this to determine \( E_{\text{gap}} \), then determining what \( E^{\text{POL}} \) gives the zero-reverse-bias \( \lambda_{\text{PEAK}} \) gives a direct estimate of the built-in \( E^{\text{POL}} \) in the device. Both the modeled effect of \( L_Z \) on \( \lambda_{\text{PEAK}} \) as a function of \( E^{\text{POL}} \), and an estimation of \( E^{\text{POL}} \) for the device studied by Bunker et al., are shown in Figure 75.
Figure 75: Calculations of $\lambda_{\text{PEAK}}$ as functions of $L_Z$ and $E_{\text{POL}}$. $E_{\text{gap}}=2.28\text{eV}, n_{2D}=10^8/\text{cm}^2$.

Thus, from this computer model, the device studied by Bunker et al. is inferred to have a polarization field of $\approx -1.33\text{MV/cm}$, compared to Bunker's estimate of $-1.0\pm 0.2\text{MV/cm}$. Here, the negative sign indicates the field points toward the substrate, as discussed in section 2.4.1.

As mentioned previously, the spectrometer used to measure the flatband emission peak has a resolution of $\approx 2\text{nm}$. If this is considered to be the dominant source of error, this leads to an estimated error of $\pm 0.15\text{MV/cm}$. However, given the approximations inherent to the model, $\pm 0.15\text{MV/cm}$ will likely be a lower limit to the actual error level.

Screening of the QCSE is seen to be well-modeled by this method. Figure 76 shows the peaks calculated for a $L_Z=2.5\text{nm}$ QW with $E_{\text{gap}}=2.28\text{eV}$ and $E_{\text{POL}}=1\text{MV/cm}$; Figure 77 shows the shift in CL peak with screening level $n_{2D}$ for a 2.5nm QW with different levels of $E_{\text{POL}}$. 
Figure 76: Peak shift with increased injection (screening) density. $E^\text{POL}=1\text{MV/cm}$, $E_\text{gap}=2.28\text{eV}$

Figure 77: Modeled effect of screening level and $E^\text{POL}$ on peak wavelength. $L_2=2.5\text{nm}$, $E_\text{gap}=2.28\text{eV}$

In particular, Figure 76 matches qualitatively to the data of Bunker et al. in Figure 60(C). Although an accurate conversion of $n_{2D}$ units to beam current is not possible,
the two graphs show the same trend. This indicates that the effects observed by Bunker et al.\textsuperscript{133} are indeed consistent with the QCSE calculations observed here. Also of interest, both charts show that the blueshift found by increasing injection from $10^8$ to $5 \times 10^{13}/\text{cm}^2$ is about the same as that found by increasing injection from $5 \times 10^{13}$ to $10^{14}$; i.e., almost all of the screening is in the last order of magnitude.

Lastly, we will compare $\lambda_{\text{PEAK}}$ as a function of $V_a$ obtained by Bunker et al. to reverse-bias calculations using the present model. In Figure 78, data from Bunker et al.\textsuperscript{133} is presented as open circles; the error bars indicate the resolution of the spectrometer. The solid lines indicate predicted $\lambda_{\text{PEAK}}$ curves for $n_{2D}=10^8/\text{cm}^2$, $E_{\text{gap}}=2.28\text{eV}$ and $L_Z=2.5\text{nm}$ for different $E_{\text{POL}}$ values. The dashed curve indicates the predicted $\lambda_{\text{PEAK}}$ values under the case $E_{\text{POL}}=-1.3\text{MV/cm}$, which was the value inferred from the above discussion.

![Figure 78: Experiment vs. model for reverse-bias CL. $L_Z=2.5\text{nm}, n_{2D}=10^8/\text{cm}^2$, $E_{\text{gap}}=2.28\text{eV}$.](image)

Figure 78: Experiment vs. model for reverse-bias CL. $L_Z=2.5\text{nm}, n_{2D}=10^8/\text{cm}^2$, $E_{\text{gap}}=2.28\text{eV}$
The experimental $\lambda_{\text{PEAK}}$ vs. $V_a$ data shows a significant "lag:" there is no peak shift until the reverse bias reaches a level of about -7V, whereas the modeled data predicts peak shift for any application of reverse bias. Although none of the calculated reverse bias curves fit the experimental data well, the best-fitting curve is the -1.3MV/cm curve; however, as the -1.3MV/cm value was derived from the point of maximum blueshift, this agreement is a direct consequence of the model.

Surprisingly, the computer model predicts relatively little difference in the reverse-bias curves as a function of injection density $n_{2D}$, Figure 79.

![Figure 79: Effect of increasing $n_{2D}$ in reverse bias. $L_z=2.28\text{eV}, E_{\text{POL}}=-1.3\text{MV/cm}$](image)

5.1.4 Higher-resolution CL injection studies

The results presented by Bunker et al.$^{133,172}$ were obtained with ±2nm spectrometer resolution; the improved cathodoluminescence system described in detail below (Section 5.3) was used to obtain higher-resolution current-resolved CL peak shift data from blue and green LED devices. This improved CL system both allowed acquisition of data at lower SEM beam currents, and allowed variation of the spectrometer resolution by adjusting the slit widths, which allowed continuous variation of the resolution vs. collection efficiency.
Peak shift data for the blue LED, taken with a 0.86nm wavelength bandpass (±0.006eV at 460nm) for beam currents of 100 to 30,000pA at 30keV accelerating voltage, is presented as Figure 80.

![Figure 80: Blue LED peak shift with current at 30keV.](image)

The most obvious feature of Figure 80 is a redshift of the peak energy with respect to beam current, which is the opposite effect expected from either the QCSE or In-segregation hypotheses. However, this counter-intuitive result can be explained in terms of Figure 81.
Figure 81: 30keV, 300pA peaks before and after CL irradiation.

Figure 81 shows CL data acquired from the blue LED sample at 30keV and 300pA, but before and after the 1,000, 3,000, 10,000, and 30,000pA acquisitions. It is clear from the graph that the peak redshift in Figure 80 is due to electron beam damage of the sample, rather than an actual QCSE or segregation related effect.

The data for the green LED, as reported by Bunker et al., shows a very clear blueshift with injection current. This data is presented in Figure 82.
It be seen from this data that the green emission peak is approaching a saturating value; further, the data of Bunker et al. covered only ~0.5 orders of magnitude of beam current, whereas this data covers ~3 orders of magnitude.
From the computer model, it would be expected that lower beam currents would produce little blueshift, and larger beam currents would produce progressively larger blueshifts (Figure 77). The data of Figure 82 does not show this behavior; the interpretation is that the low-beam current cases in Figure 82 inject $n_{2D}$ levels larger than the low-$n_{2D}$ cases calculated in the model. As SEM beam currents are reduced, it would be expected that the peak value would be reduced to a minimum value associated with the zero-screening case. However, due to the signal/noise levels of the spectroscopic data, acquisition of curves at $i_b<100pA$ was not possible with the given slit widths. Further, it would be expected that modeling sufficiently large sheet densities would return the QWs to flatband, and no further peak shift would be expected, which would result in a saturation like that found in Figure 82(B).

5.1.5 Discussion of computer model

The computer model replicates all of the major consequences expected from the QCSE: (1) redshift with increasing field, (2) redshift with increasing QW width, and (3) blueshift with increasing injection level.

As for increased injection density, the orders of injection evaluated by the model ($10^8$ to $10^{14}/cm^2$) indicated an expected blueshift of $\sim$5nm ($\sim$25meV) in the simulated structure; this is on the close order of the value observed in Bunker et al.'s current-resolved-CL experiment, of $\sim$50meV. Given the approximations and simplifications in the computer model, this is reasonable agreement.

The major deviation between the model and experiment is that of reverse-bias; as seen in Figure 78, the fit between the modeled $\lambda_{\text{PEAK}}(V_a)$ curve and that from experiment is rather poor. The experimental data undergoes a negligible change with $V_a$ up to a point, but beyond that point the experimental $\lambda_{\text{PEAK}}(V_a)$ data has much higher curvature than the modeled data.

This phenomena was also observed by Kaplar et al. They studied 2.0 or 3.0nm thick In$_{0.07}$Ga$_{0.93}$N QWs grown by MOCVD on sapphire. They performed contact-electro-reflectance (CER) under differing levels of reverse bias, and found results like those found by Bunker et al., despite the different optical evaluation.
technique. In particular, Kaplar et al.'s CER peaks were found to redshift past flatband under large reverse bias, much like the CL peaks of Bunker et al. However, the devices studied by Kaplar et al. showed no measurable blueshift up to the flatband voltage (-17.6V, in their case). This is shown in Figure 83(A). Their interpretation was that due to the doping of the p- and n-GaN barriers, in the $10^{18}$ to over $10^{19}$/cm$^3$ range, the reverse bias was unable to cause band-bending until the doping-induced carriers were fully depleted from the well by the reverse bias. The depletion of their calculated number of carriers is shown in Figure 83(B).
Figure 83: (A) CER peak shift vs. $V_a$; (B) electron concentration vs. $Z$ and $V_a$.

Reused with permission from Ref. 138. Copyright 2004, American Institute of Physics.

A further complication of the physical system that was neglected from the present model is that of doping; this would be most apparent at low levels of $n_{2D}$, as doping-induced carriers in the QWs will screen the $E^{POL}$ to some degree. However, as can be seen from Figure 77, even with doping neglected, most of the blueshift due to screening appears in the last 1-1.5 orders of magnitude of injection, indicating that this simplification should have little practical effect.
5.1.6 Electroluminescence: analytical model and experiment

As shown by Bunker et al. and presented in Figure 60(c), the green LED showed a peak blueshift with increasing EL drive current, but a blue LED did not.

Similar experiments were repeated in this work, using blue or green GaN-based LEDs. The monochromator-spectrometer described in Section 5.3.1 was used to acquire high-resolution electroluminescence spectra; blue or green LEDs like those used by Bunker et al. were placed in an integrating sphere and driven at $1\mu\text{A}$ to $5\text{mA}$ currents, and spectra were obtained. Figure 84 and Figure 85 show spectra for the blue and green LEDs, respectively.

![Figure 84: Blue LED electroluminescence spectra.](image-url)
Figure 85: Green LED electroluminescence spectra.

The slight change in peak-shape for the 1μA and 10μA curves in Figure 85 is due to widening of the monochromator slits to increase the signal/noise level. The normalized intensities in the graph were corrected for the change in absolute signal level resulting from the change in slit width.

This data is summarized in Figure 86.
The green LED suffers a blueshift of \( \sim 10 \text{nm} (\sim 43 \text{meV}) \) and the blue LED a blueshift of \( \sim 2 \text{nm} (\sim 10 \text{meV}) \).

The injection levels these drive currents result in was estimated by deriving and solving an ordinary differential equation. If the drive current is \( I \), and the LED is assumed to be square in cross-section with edge lengths \( L \), then the injected current density \( J \) in the QW will be:

\[
J = \frac{I}{L^2}
\]

The units of \( J \) will therefore be amps/cm\(^2\). A more useful quantity is \( \hat{J} \), the injected number of EHPs/sec-cm\(^2\). It follows that:

\[
\hat{J} = \frac{I}{q \cdot L^2} = \frac{J}{q}
\]
Let it be assumed that the change in the injected carrier density $n_{2D}$ is equal to the density of injected carriers $\dot{J}$ minus the rate of carrier decay, which will be $n_{2D}/\tau$ for lifetime $\tau$. Thus, the number of carriers in the QW will be the solution of the ordinary differential equation:

$$\frac{dn_{2D}}{dt} = \dot{J} - \frac{n_{2D}}{\tau} \quad (5-32)$$

d$n_{2D}/dt$ will have dimension of EHP/sec-cm$^2$. This differential equation was solved by standard methods and yields the solution $n_{2D}$:

$$n_{2D} = \tau C e^{-t/\tau} + \dot{J} \tau \quad (5-33)$$

Here, the solution has units of EHP/cm$^2$; $C$ is a constant of integration. For the steady-state case, $t \rightarrow \infty$, so $n_{2D} = \dot{J} \tau$, and $C$ need not be found. If the transient case is being studied, $C$ will be:

$$C = \frac{n_{2D}^{(0)} - \dot{J} \tau}{\tau} \quad (5-34)$$

$n_{2D}^{(0)}$ is the density of carriers in the well before injection begins, i.e., the doping-induced carrier density. For the case of GaN, where the expected value for $\tau$ is $<<1\mu$sec, the steady-state will obtain in well under one second. As the spectra acquired in the EL experiment were taken over several minutes, it is clear that the transient case can be neglected. However, the assumption implicit in this entire treatment is that injected carriers dominate over $n_{2D}^{(0)}$.

Although the actual values of the lifetime $\tau$ for these samples are not known, and will probably be a weak function of $n_{2D}$ and thus require numerical solution of (5-32) for an exact solution, it seems reasonable to assume a value of $\tau \sim 1$ns, based on the TRPL data of Chichibu et al.$^{105}$ and Martinez et al.$^{249}$ As the devices studied were of known size, $L \approx 250\mu$m, it is possible to convert the EL drive current graph Figure 86 above into an equivalent graph based upon injected $n_{2D}$; Figure 87:
Figure 87: Peak shifts vs. calculated $n_{2D}$. $\tau = 1$ns, $L = 250\mu$m.

Because of the roughness of the $\tau \sim 1$ns value, these curves could easily be expected to shift one or two orders of magnitude left or right. However, the emission peaks begin shifting in the green device at $n_{2D}$ values roughly on the order of the $n_{2D}$ values calculated for screening in the CL numerical model above, indicating general agreement between the different experimental and computational approaches.

5.1.7 Summary of polarization fields study

Between the experimental results of Bunker et al. $^{133,172}$ and the experimental and computational results presented here, it becomes clear that the polarization hypothesis explains the peak-shifting behavior of InGaN-based LEDs quite well. The hypothesis of In-segregation and quantum confinement at local areas of narrow bandgap could also be used to explain the peak blueshifting with EL or CL excitation density; however, the reverse bias behavior observed by Bunker et al.$^{133,172}$, Kaplar
et al.	extsuperscript{138}, and the present results are outside of the phenomenological bounds of the In-segregation hypothesis.

Thus, by combining CL, EL, and computer modeling, this work has given strong support to the hypothesis that polarization-induced internal electric fields and the concomitant QCSE have a strong influence on the optical properties of In-rich InGaN QWs. However, it is not possible to conclude from this work that In-segregation has no influence on the optical properties. Given the large low-energy tail of the QW peak in the CL experimental data, and the discussion in Section 2.4, it seems likely that In-segregation effects are simultaneously present with the studied polarization effects.

5.2 Cross-sectional EBIC

5.2.1 Sample preparation and experimental techniques

As discussed in Sections 3.2.2 to 3.2.4, cross-sectional EBIC allows the evaluation of the minority carrier diffusion length $L$ and the surface recombination velocity $v_s$ of a Schottky or p-n type device. In practice, $s$, where $s = v_s/D$, is more commonly determined than $v_s$.

As outlined in Figure 37 (on page 86), X-EBIC requires an SEM sample with an exposed cross-sectional face, electrical connections to the p- and n-sides of the device through ohmic contacts, and some means of measuring the current $I_{\text{EBIC}}$.

Sample preparation to expose the cross-sectional face requires, first, obtaining bare, unpackaged LEDs. Such devices were obtained in this study by taking commercial, fully-packaged LEDs, as shown in Figure 88, and removing the encapsulating plastic and connecting legs.
Figure 88: (A) Encapsulated LED; (B) Blue LED emitting under bias.

The procedure used for this extraction involved a number of steps:

1. Cut off the legs near where they meet the plastic, leaving <5mm.
2. Grind the plastic on 400-grit sandpaper, by grabbing the remaining leg stubs with needle-nose pliers, to remove as much of the encapsulating plastic as possible without grinding into the LED or the gold wire attaching it to the second leg. About a 2x2x2mm cube of material, centered on the LED, was left when this stage was completed.
3. For mechanical cross-sections, the following steps were performed:
   a. The 2mm-cube would be waxed onto a sample-holder stub for an Allied Techprep polishing wheel.
   b. On the polishing wheel, 30μm-diamond lapping pads were used to grind through the remaining metal of the leg, the reflector cup, and encapsulating plastic, until about 10μm of plastic remained before the edge of the LED itself. It was imperative that the LED top-surface be perpendicular to the polishing wheel at this point.
   c. 6μm diamond lapping pads were used to grind ~25-50μm into the LED; this stage was stopped when the polishing line had advanced about half-way from the edge of the LED to the gold wire's bond-pad.
   d. 1 and 0.5μm pads were used to remove about 15 and 3μm of material. The 0.5μm pad should leave a mirror-finish on the sample. Lastly, the sample would be polished on a 0.05μm pad for about 20 minutes.
4. Samples were then placed in a folded piece of filter paper, making a basket, and the paper placed into a sealable polyethylene sample cup. The cup was then filled with enough Dynasolve 210 or CH₂Cl₂ solvents to cover the sample by about 1cm. The seal on the cup was then double-sealed with electrical tape to minimize solvent evaporation. All steps involving these solvents (Steps 4-5) were performed under a fume hood.

5. The sample was soaked in Dynasolve 210 or CH₂Cl₂ for 24-72 hours. After this time, the basket of filter paper was removed and placed in a funnel sitting in the throat of a chemical waste container. The sample was rinsed with acetone, ethanol and methanol to remove all traces of Dynasolve 210 or CH₂Cl₂.

6. The soaking in Dynasolve 210 or CH₂Cl₂ softens the encapsulating plastic; at this point, the sample was moved to a stereoscope and the plastic peeled away with fine tweezers and a sharp hobby knife, leaving behind the bare LED. The gold bond wire often separated from the sample at this point, despite care to preserve it.

7. The bare die was cleaned in Dynasolve 210 or CH₂Cl₂, then acetone, ethanol and methanol, than attached to a T0-type header using silver epoxy (Circuit Works CW2400) by using a fine pair of tweezers and a stereoscope. The epoxy was cured at 50°C for 12-36 hours.

8. If the bond wire was lost, a new bond wire would be re-attached at this point, bridging from the bond-pad on the sample to the isolated post of the T0 header, thus completing the circuit. If the original wire was preserved, it would be attached to the post with silver epoxy, followed by another curing step.

9. The sample was then complete, and ready for electrical testing (I-V curve measurement) and EBIC analysis.

If a FIB cross-section was desired, the sample prepared as above to step #9 would be a suitable starting point. The sample would be placed into the Hitachi FB-
FB-2100 FIB (Figure 89) in plan-view, and a suitable area of interest found by imaging with the 10keV (less-damaging) beam. Sample mounting required fabrication of a modified stub for the FIB’s SEM-type stage; this stub along with an LED on a T0 header is shown in Figure 90. Because the FB-2100 is a single-beam instrument, all imaging must be performed with the (destructive) ion beam. Thus, all imaging was performed at the lowest magnification possible and for the shortest possible time.

Figure 89: (A) FB-2100 focused ion beam; (B) Stages, micromanipulator, and W-injection systems.

Figure 90: (A) Sample on header and modified stub; (B) Sample ready for FIB.
Once an area of interest on the edge of the GaN layer was found, a protective tungsten cap, \(\sim 15 \times 5 \times 1 \mu m\), was deposited with the 40keV, 80\(\mu m\)-aperture, \(\sim 300pA\) beam. This cap always was placed on, or partially on, the GaN's metallization. A window was then milled into the cross-section at 40keV, using progressively lower and lower beam currents, terminating at the 40keV, 20\(\mu m\) aperture beam (<10pA), into the metallized GaN. After completing the 40keV polishing, the sample could be tilted \(\sim 5^\circ\) off plan-view and the EBIC surface cleaned with a 10keV polishing beam. The cleaning parameters used were 10keV, 80\(\mu m\) aperture, and 120 seconds of milling at 0.5\(\mu sec\) pixel dwell time. The beam was defocused slightly before the final 10keV cleaning step.

This also allowed samples to be prepared with no mechanical polishing (Steps 3a-3d above). To do so, a large trench was cut into the SiC before the FIB polishing of the GaN face; this allowed the SEM beam access to the cross section.

Micrographs of finished LEDs, ready for X-EBIC examination, are presented as Figure 91-Figure 94.

![Figure 91: LED (without mechanical polishing) attached to T0 header; SEM image.](image)
Figure 92: Detail of Figure 91; SEM image.

Figure 93: Detail of Figure 92, showing FIB-milled window; SEM image.
FIB milling of the sample could result in a serious artifact. As atoms sputter off the surface under the action of the ion beam, they scatter and can deposit on any surface within their line-of-sight. This is called redeposition, and can be cleaned by scanning the deposited surface with a low-current beam for a few seconds or minutes. Figure 95 shows the same surface, before and after redeposition.
Figure 95: (A) The top surface is clean and the layers can be clearly seen; (B) After milling a nearby feature, the top surface was covered in redeposition. SIM images, 40keV.

The effect of redeposition on these LEDs is to cause a shunt-resistance path for the current around the diode itself; I-V curves for an LED before FIB, after FIB, and after using FIB to progressively clean away the redeposition, is given in Figure 96.
Thus, FIB uncontrolled for redeposition turned the diode into a kΩ-range resistor; cleaning of the redeposition returned the sample to a diode condition, but not to its pre-FIB state.

X-EBIC was performed using either the Hitachi S-3200N thermionic-emission SEM or the JEOL JSM6400F cold-field-emission microscope; these are shown in Figure 97.
Small electrical plugs were used to hold the T0 headers and make contact to the headers' legs. These plugs were attached to BNC electrical vacuum-feed-throughs to pass the EBIC signal out of the SEM chambers. Keithley 4801 low-noise BNC signal cables were used to carry the EBIC current to a Keithley 428 current amplifier. The Keithley 428 can apply a gain ranging from $10^3$ to $10^{11}$ volts/amp to the input current signal, and thus produces an output voltage signal. Most EBIC currents measured were between 100pA and 10nA, so the Keithley 428 was operated in the $10^{10}$ to $10^7$ V/A gain ranges. Noise pickup in the LED and in-chamber cabling were the limiting factors and constrained the useful gain to $\leq 10^{10}$ V/A; typical noise levels were ~10-100pA in the EBIC experiments. The noise was periodic, and consisted of 60Hz noise and higher-frequency noise on top of the 60Hz signal. As the SEMs synch their fast-scan raster to the electrical mains, this noise spectrum resulted in vertical lines that the EBIC signal was superimposed upon. As such, high-$E_0$ and high-$i_b$ operation of the SEM during EBIC experiments was preferred, in order to swamp the noise with the EBIC signal. Low-$E_0$ and low-$i_b$ operation resulted in periodic noise in the EBIC signals.

The EBIC experiments performed in this work were at $E_0$ values above the unity yield point; thus, the specimen current from the sample will be negative. As pointed out by Opdorp\textsuperscript{207}, it is thus important to make sure that the n-side of the LED
is grounded, as opposed to the p-side. If the n-side is grounded and the p-side connected to the amplifier, the specimen current flows to ground before it can reach the amplifier and contribute to the signal. If the p-side is grounded, the specimen current can’t cross the depletion region to enter the p-side and reach ground; it must pass through the amplifier to reach ground and will contribute to the apparent EBIC signal. Experimentally, it was found that n-grounding did, indeed, completely suppress specimen current observed by the EBIC amplifier. All data presented was taken in this fashion.

The voltage output from the Keithley 428 was conditioned by a home-made "imaging box," which used a series of operational amplifiers to apply gain and level adjustments to the EBIC signal; the final op-amp in the circuit provided an optional inversion of the signal. A detail description and a circuit diagram of the imaging box are given in Appendix 4. After conditioning in the imaging box, the signal was passed to the auxiliary input port of the instrument, and used to form an EBIC image. For EBIC images which were to be quantitatively evaluated, the grayscale of the image was carefully adjusted using the gain and level controls of the imaging box to ensure there were no over- or under-saturated pixels, which causes loss of dynamic range. EBIC data was converted to quantitative data with ImageJ 1.33u software\textsuperscript{250}; the data derived from ImageJ was normalized using a spreadsheet. Normalization was performed by taking the average value of the signal far away from the junction as $i_{\text{min}}$, and the single highest point as $i_{\text{max}}$. The normalized intensity $I(x)$ at any point $x$ was then $I(x) = (i(x) - i_{\text{min}})/(i_{\text{max}} - i_{\text{min}})$. Quantitative curve fitting was performed using worksheets written in either Mathcad 12 or Maple 9.5 software, presented in Appendices 5 and 6. Beam currents were measured using a Faraday cup and either a Keithley 614 electrometer or a GW Electronics Model 31 specimen current meter. Spot sizes were estimated from 20%-80% profiles measured across either (100) silicon cleaved edges or droplets of gold evaporated on carbon.
5.2.2 Determination of input parameters for Equation (3-25)

As discussed in Section 3.2.4.4, the computer modeling of the different EBIC experimental techniques leads to the hypothesis that the model of Bonard and Ganière\textsuperscript{194} will quantify experimental EBIC curves most accurately in the case of L<<R, which is expected in GaN LEDs.

However, Equation (3-25) has two terms, $\sigma_x$ and $\sigma_z$, which are related to the depth- and lateral-dose curves for the material at particular beam conditions. The original paper's dependencies were given in Equation (3-28); however, as the original paper studied Al$_{0.4}$Ga$_{0.6}$As, these dependencies will be irrelevant to the GaN work presented here. A technique was developed for the computation of $\sigma_x$ and $\sigma_z$ for arbitrary material, beam energy, and beam spot size.

In this work, Monte Carlo SEM interaction simulation software CASINO\textsuperscript{169} was used to find the distributions of EHP generation. CASINO was chosen because it outputs the distribution of energy deposition of the electron beam in the solid as a function of (x,y,z). It also explicitly includes spot size $d_b$ in its simulation. By summing the distribution at each (x,z) value along the y direction, a distribution of energy deposition as a function of (x,z) is found. By assuming that the generation of EHPs is directly proportional to the energy lost\textsuperscript{165,170}, the normalized distribution of energy loss is taken to be equal to the normalized distribution of EHP generation.

All CASINO calculations were performed for $10^6$ electron trajectories per simulation, with energy distribution calculated for a 100x100x100 x,y,z mesh. The energy loss function utilized was that of Joy and Luo\textsuperscript{251}. Incident electron beam diameters $d_b$ were 1, 10, 100, and 1000 nm. Beam energies $E_b$ were 5, 10, 20 and 30 keV. The materials chosen for analysis were Al$_{0.4}$Ga$_{0.6}$As, of density 4.8 g/cm$^3$, and GaN of density 6.2 g/cm$^3$. Al$_{0.4}$Ga$_{0.6}$As was chosen to be consistent with the original paper\textsuperscript{194}, in order to compare the current results to those of Bonard and Ganière. GaN was chosen as a second example to provide the $\sigma_x$ and $\sigma_z$ values necessary for the examination of the GaN LEDs in this work. In principle, the values for any semiconductor could be calculated in a similar fashion.
The Monte Carlo data was fit to the theoretical function, Equation (3-27), by first performing a separation of variables on Equation (3-27):

\[
\begin{align*}
h(x, z) &= k(x) j(z) \\
k(x) &= A \exp \left( -\frac{x^2}{\sigma_x^2} \right) \\
j(z) &= B z^2 \exp \left( -\frac{z}{\sigma_z} \right)
\end{align*}
\] (5-35)

Here, A and B are constants. It is apparent that the Monte Carlo data can be fit separately in x and z. If the data at any x position is summed along z, this will yield a function k(x). Similarly, if z-data is summed along x, a function j(z) is found.

The function k(x) can be linearized as:

\[
\ln[k(x)] = \ln(A) - \frac{x^2}{\sigma_x^2}
\] (5-36)

By plotting the Monte Carlo data k(x) as ln[ k(x) ] against x^2, the value of \( \sigma_x \) can be found as the square root of the negative reciprocal of the slope. Similarly, the function j(z) can be linearized as:

\[
\ln[j(z)] - 2\ln(z) = \ln(B) - \frac{z}{\sigma_z}
\] (5-37)

Thus, plotting the Monte Carlo data j(z) as ln[j(z)]-2ln(z) against z, \( \sigma_z \) is found as the negative reciprocal of the slope.

Figure 98 illustrates this procedure; in Figure 98(A), data for GaN with \( E_b=5 \) keV and \( d=10 \) nm is analyzed in terms of k(x), and Figure 98(B) illustrates the analysis for \( \sigma_z \). In this case, \( \sigma_x=60 \) nm and \( \sigma_z=15 \) nm. Values of \( \sigma_x \) and \( \sigma_z \) for GaN and Al_{0.4}Ga_{0.6}As are presented in Figure 99(A) and (B), respectively. The lines in Figure 2(b) correspond to the \( \sigma \) values quoted in the original reference\(^{194}\).
Figure 98: (a) Fit of x-dependence to Equation (5-36). Solid line is Monte Carlo data; dashed line is fit by Equation (5-36). (b) Fit of z-dependence to Equation (5-37). Solid line is Monte Carlo data; dashed line is fit by Equation (5-37). (a) and (b): GaN, $E_b=5$ keV, $d_b=10$ nm.
Figure 99: (a) Monte Carlo derived data for GaN. (b) Data for AlGaAs. Solid symbols are for $\sigma_x$, and open symbols for $\sigma_z$. $E_b$ is beam energy and $d_b$ beam size. The solid line is $\sigma_x$ from Ref. [194], and the dashed line $\sigma_z$ from Ref. [194].

It can be seen that the present Al$_{0.4}$Ga$_{0.6}$As data matches the Bonard and Ganière data well for the 20 and 30 keV beam energies (<18% difference at 20 keV), but matches the Bonard and Ganière data poorly at 5 or 10 keV (55% difference at 5 keV; if the 1 $\mu$m spot-size case is disregarded, 100% difference). In particular, their
low-keV $\sigma_x$ data matches the Monte Carlo data best for large SEM spot sizes, indicating a possible bias in the experimental data from which their values were derived. The interpretation is that the Bonard and Ganière dependencies are unsuitable at lower beam energies, but better at 20 or 30 keV. This is likely due to the sharply peaked depth-dose function in real materials, and the importance of spot size at low beam energies. The energy-loss function$^{251}$ used is expected to be valid to beam energies as low as $\sim 1$ keV$^{(158,252)}$, so the deviations between this model (at $\geq 5$ keV) and the experiment of Bonard and Ganière$^{194}$ are likely not due to systematic errors in the Monte Carlo simulation. It also is concluded that, by fitting the Monte Carlo simulation data to Equation (3-27), better values of $\sigma_x$ and $\sigma_z$ are found. The spot size $d_b$ dependency, in particular, is an important variable to control for. High resolution SEM-EBIC requires low voltage operation of the instrument$^{222}$; at low voltage, the generation volume is sufficiently small that the SEM’s spot size may become the dominant factor affecting EBIC resolution, especially if a small spot size field-emission SEM is not used.

In the work of Li et al.$^{253}$, InAs/GaSb superlattices were examined via SEM-EBIC. A derivative of the Bonard and Ganière technique was used. Li et al. performed Monte Carlo simulations to obtain InAs/GaSb values of $\sigma_x$ and $\sigma_z$ at 7 keV of 180 nm and 15 nm, respectively, albeit for a slightly modified version of Equation (1), $h(x,z)$. Although these values are thus not directly comparable to the current work, they are on the order of the values obtained independently in this work, which implies that both techniques will be accurate to within some tens of percent, or better. Unfortunately, Li et al. did not describe the method they used to obtain their values. The present work fills this gap by providing a procedure by which any researcher can obtain the $\sigma$ values needed to apply the Bonard and Ganière EBIC technique to their work.

As discussed in Section 3.2.4.4, the Bonard and Ganière$^{194}$ technique is probably the best method for the extraction of $L$ and $s$ from SEM-EBIC data of p-n junction samples. However, this method requires that variables relating to the lateral
and depth range of the EHP generation in the sample be known. The above discussion illustrates how Monte Carlo simulations can be used to quickly and accurately extract these input values for use in the analysis of SEM-EBIC data.

5.2.3 Cross-sectional EBIC: results

Cross-sectional EBIC data taken from a diamond-polished blue LED, acquired at 4, 5, 6, and 7keV in the field emission instrument, is shown in Figure 100. All four images are from the exact same area of the LED, and the bright-and-dark areas in the different-keV EBIC images are consistent. For all EBIC images and graphs presented in this section, the left side is the n-side of the diode, and the right, the p-side.

Figure 100: X-EBIC data obtained for analysis. Beam conditions are listed in Table IV.
From Figure 100, the EBIC response is seen to be inhomogeneous; specifically, the p-side shows a sharp cut-off but the n-side shows bright-and-dark areas of varying extension. Kuroda et al. observed this same effect, and attributed it to diffusion of Mg (the p-side dopant) down dislocations into the n-side, causing partial compensation of the n-doping and thus changing the minority carrier lifetime in the n-side. Although Kuroda et al.'s conclusion is certainly possible, it is not immediately clear that it is necessarily correct. However, this bright-and-dark area effect was observed in all samples, and is not believed to be an artifact. It is possible that future work involving FIB-lift-out and TEM could address the topic.

Figure 101(A) shows a 5keV SE micrograph of the area used to acquire the EBIC images in Figure 100; in Figure 101(B), the 5keV data is presented again, but with four 100nm-tall boxes drawn on the micrograph, and labeled 1-4. These areas 1-4 are the subject of the subsequent quantitative analysis. Careful comparison of EBIC and SE images indicates that the p-n junction is ≈250nm below the top contact. Figure 102 shows EBIC data, extracted from area 4, for the different accelerating voltages.

![Figure 101](image)

Figure 101: (A) SE; (B) X-EBIC images showing examined area. Labels (1)-(4) are the areas of subsequent analysis. Both images at 5keV, ~135pA.
Figure 102: EBIC curves, normalized, taken from area 4 of the above micrographs.

As can be seen, the EBIC curves at a given area broaden gradually with increasing accelerating voltage. Data for the different micrograph conditions is presented in Table IV.

Table IV: Experimental conditions for cross-sectional EBIC

<table>
<thead>
<tr>
<th>E₀ (keV)</th>
<th>d₀ (nm)</th>
<th>i₀ (pA)</th>
<th>R (nm)</th>
<th>σₓ (nm)</th>
<th>σᶻ (nm)</th>
<th>Δn (cm⁻³)</th>
<th>i_PEAK (pA)</th>
<th>i_PEAK/i₀ (exp.)</th>
<th>i_PEAK/i₀ (pred.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>14</td>
<td>140</td>
<td>138</td>
<td>41.6</td>
<td>12.3</td>
<td>2.1 x 10¹⁷</td>
<td>500 ± 100</td>
<td>4 ± 1</td>
<td>328</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>135</td>
<td>200</td>
<td>62.1</td>
<td>14.6</td>
<td>8.3 x 10¹⁶</td>
<td>1000 ± 200</td>
<td>7 ± 2</td>
<td>410</td>
</tr>
<tr>
<td>6</td>
<td>21</td>
<td>190</td>
<td>271</td>
<td>78.9</td>
<td>22.1</td>
<td>5.6 x 10¹⁶</td>
<td>3000 ± 1000</td>
<td>15 ± 5</td>
<td>492</td>
</tr>
<tr>
<td>7</td>
<td>16</td>
<td>100</td>
<td>351</td>
<td>104</td>
<td>29.9</td>
<td>1.6 x 10¹⁶</td>
<td>2500 ± 500</td>
<td>25 ± 5</td>
<td>574</td>
</tr>
</tbody>
</table>
Here, $E_0$, $d_b$, $i_b$, and $R$ have their normal meanings. $\sigma_x$ and $\sigma_z$ are the Bonard and Ganière range parameters, computed using the techniques discussed in the previous section. Because $d_b$ values $\sim$nanometers are extremely difficult to quantify, the value of $d_b$ for all subsequent quantitative analyses, and the calculation of $\sigma_x$ and $\sigma_z$ values presented in Table IV, $d_b \approx 30$nm was used. Given the size of the electron interaction volumes in GaN at the $E_0$ values used, and the Gaussian nature of the probe, differences in $d_b$ on the order of 10-30nm will be entirely negligible.

The electrical current $i_{PEAK}$ is the current associated with the peak of the X-EBIC signal. This was measured by monitoring the voltage output of the current amplifier with an oscilloscope and then applying the V/A gain factor of the current amplifier to the peak voltage measured on the oscilloscope screen. The ratio $i_{PEAK}/i_b$ (exp.) is the experimental ratio of beam current to peak EBIC current. The ratio $i_{PEAK}/i_b$ (pred.) is the predicted $i_{PEAK}/i_b$ ratio, based on the assumption that the total number of electron-hole pairs generated, $g_0$, calculated by Equation (3-9), is collected by the junction. It can clearly be seen that the experimental value is much smaller than the predicted value. This will be discussed in Section 5.4.2.

The value $\Delta n$ represents the estimated electron-hole pair injection level; values $<10^{18}$ cm$^{-3}$ were taken to be low-injection conditions. As such, it is seen that the current experimental conditions are estimated to be near but not past the high-injection limit. Although low-injection is an assumption of the quantitative EBIC analysis models, it was necessary to use high beam currents to obtain X-EBIC data with sufficient signal-to-noise for quantitative evaluation, despite the risk of approaching high injection. As shown by the comparison of experimental to predicted $i_{PEAK}/i_b$ values, relatively few EHPs are actually injected into the bulk semiconductor in comparison to the predicted number. The ratio of the predicted $i_{PEAK}/i_b$ to experimental $i_{PEAK}/i_b$ values drops precipitously with $E_0$; the ratio is 23:1 at $E_0=7$keV and 82:1 at $E_0=4$keV. This implies that the loss of carriers is due to the surface damaged layer. It is thus safe to assume that high injection does not obtain in any of the experimental conditions above, as the predicted value is $\sim 0.1$ times the low-injection limit, and the majority of the carriers are lost to the damaged layer and
never enter the diffusion problem, thus pushing the actual number of carriers even lower. This effect will become more and more pronounced as $E_0$ decreases, as is seen by the variation of $i_{PEAK}/i_b$ (exp.) with $E_0$.

Secondly, as pointed out by Berz and Kuiken\textsuperscript{171}, when high-injection is localized only within a small segment the generation region, it will not affect the EBIC curves' behavior. Because of the rapid $(1/r)e^{-r/L}$ type falloff of minority carrier density from the point of injection, a slight high-injection case should decay into the low-injection case after only a very short distance. Thus, even if high injection did locally obtain, which is not considered to be the case here, it could be ignored.

Thus, although the calculated $\Delta n$ values give the appearance of borderline high injection, the actual situation is taken to be well within the low-injection case.

These values of $E_0$ (4, 5, 6, and 7 keV) were chosen because 4keV was the lowest energy that produced an EBIC signal greater than the instrumental noise level, and 7keV was considered to be trending to the large-R case; specifically, $R$, $\sigma_x$ and $\sigma_z$ at 7keV are more than twice that at 4keV, and approximately four times our \textit{a priori} estimates of diffusion length ($\leq 100$nm).

Doping levels $N_d$ and $N_a$ for this material were taken to be $10^{18}$ cm$^{-3}$ and $10^{19}$ cm$^{-3}$, respectively. From these doping levels, standard p-n junction mathematics allows the width of the depletion region $d_j$, and its extent $d_n$ and $d_p$ into the n- and p-sides, to be calculated. Here, $d_j \approx 60$nm, $d_n \approx 54$nm, and $d_p \approx 6$nm. As discussed in Section 3.2.3, several of the X-EBIC mathematical techniques require that the generation volume not be allowed to overlap the depletion region. Thus, the minimum values allowed for the different $R_{k-o}$ conditions in the p- and n-sides are presented in Table V.
Table V: Estimate minimum allowable x-values to avoid depletion region overlap

| $E_0$ (keV) | $|x_{\text{min}}|$ (nm) | n-side | p-side |
|------------|--------------------------|--------|--------|
| 4          | 123                      | 74     |
| 5          | 154                      | 105    |
| 6          | 190                      | 141    |
| 7          | 230                      | 181    |

With those preliminaries, it is now possible to continue on to the analysis of the X-EBIC data with the different mathematical models discussed in Section 3.2.3.

5.2.3.1 Exponential method

This method involves the fitting of a simple exponential to the EBIC data; examples for 4 and 7 keV, area 4, are shown in Figure 103.
Figure 103: Exponential fits for area 4; (A) 4keV (B) 7keV.
The extracted L values for the four areas at the different $E_0$ conditions are listed in Table VI:

Table VI: $L_h$ and $L_e$ values (in nm) estimated using the exponential method.

<table>
<thead>
<tr>
<th>$E_0$</th>
<th>Area 1</th>
<th>Area 2</th>
<th>Area 3</th>
<th>Area 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>keV</td>
<td>n-side</td>
<td>p-side</td>
<td>n-side</td>
<td>p-side</td>
</tr>
<tr>
<td>4</td>
<td>68</td>
<td>43</td>
<td>69</td>
<td>39</td>
</tr>
<tr>
<td>5</td>
<td>103</td>
<td>32</td>
<td>100</td>
<td>35</td>
</tr>
<tr>
<td>6</td>
<td>100</td>
<td>96</td>
<td>91</td>
<td>93</td>
</tr>
<tr>
<td>7</td>
<td>81</td>
<td>79</td>
<td>74</td>
<td>72</td>
</tr>
</tbody>
</table>

The $L$ values are found to be on the order expected from previous GaN work; specifically, $\sim$50-150nm. The $L_e$ values in the p-side of the diode increase as $E_0$ increases, whereas the $L_h$ values in the n-side of the diode are more constant with $E_0$. This is likely because of the presence of the metal contact at $x \approx +250$nm, which will interfere with the diffusion and generation processes and bias the results on the p-side.

Although the exponential method is the fastest and most straightforward technique to use, and provided $L$-values on the anticipated order, it will be necessary to compare exponential-technique results to those of the other techniques to comment on its accuracy.

5.2.3.2 Donolato-Luke methods

First, analysis by the Luke method of finding $L_s(x)$, Equation (3-23), were attempted. As an example, the $L_s(x)$ curves for the different accelerating voltages at area 4 are shown in Figure 104; in Figure 105 are the $L_s(x)$ curves for areas 1-4 in the 4keV case.
Figure 104: $L_s(x)$ calculations for area 4 at different accelerating voltages.

Figure 105: 4keV $L_s(x)$ data for areas 1-4.
It can be seen from Figure 104 and Figure 105 that the smooth $L_s(x)$ curves predicted by Luke et al.\textsuperscript{190,191,192,193} do not obtain in practice, at least not with the signal-to-noise levels presented by low-$E_0$, low-$i_b$ operation necessary for high resolution X-EBIC of GaN. The $L_s(x)$ curves are, in fact, noise-dominated at $x$-values even slightly away from the $x=0$ center. Despite the noise, it seems that the n-side gives $L_s$ asymptotic values of $\approx 50$-100nm; however, this value is too ill-defined to make further refinement of the approximation, as discussed by Luke et al., physically meaningful. The p-side does not give an asymptotic value of $L_s$, but rather decays to zero. This is likely due to the thinness of the p-side and the presence of the ohmic contact at $x \approx +250$nm.

As such, it can be concluded that in this type of LED device, the Luke et al. technique is not a useful means of determining minority carrier properties. If lock-in type measurement techniques were used to reduce the experimental noise values or allow lower-keV operation, it might be possible to perform a Luke-style analysis on the n-side to estimate $L$, but as shown in Section 3.2.4.1, the sensitivity of the peak-minima-finding method of Luke et al. is lost when $L < \approx R$, as is clearly the case here.

However, the Donolato\textsuperscript{186} technique can be used in this case; below, curves are calculated using the Donolato equation as modified for SEM beam diameter by Luke, Equation (3-24).

Figure 106 shows example data for the n- and p- sides of area 4, calculated using the 4 and 7keV data. In each graph, the solid line is the theoretical fit, and the open symbols the experimental data. Table VII shows the extracted $L$ values in the n-side ($L_n$) and p-side ($L_p$) from the four areas for the four accelerating voltages, for the n- and p-sides. All values are calculated assuming $s=10^{-2}\ \mu m^{-1}$, which is tantamount to the $v_s=0$ case. The reason for this assumption will be discussed below.
Figure 106: (A) Donolato-Luke fit for area 4, 4keV; (B) area 4, 7keV. Circles are data, solid lines are theoretical fits to the data. $s=10^{-2} \mu m^{-1}$. 

202
Table VII: $L_h$ and $L_e$ values (in nm) estimated using the Donolato-Luke technique.

<table>
<thead>
<tr>
<th>E0 (keV)</th>
<th>Area 1 n-side</th>
<th>Area 1 p-side</th>
<th>Area 2 n-side</th>
<th>Area 2 p-side</th>
<th>Area 3 n-side</th>
<th>Area 3 p-side</th>
<th>Area 4 n-side</th>
<th>Area 4 p-side</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>150</td>
<td>50</td>
<td>120</td>
<td>50</td>
<td>127</td>
<td>50</td>
<td>85</td>
<td>50</td>
</tr>
<tr>
<td>5</td>
<td>130</td>
<td>35</td>
<td>105</td>
<td>33</td>
<td>125</td>
<td>35</td>
<td>90</td>
<td>40</td>
</tr>
<tr>
<td>6</td>
<td>130</td>
<td>73</td>
<td>105</td>
<td>70</td>
<td>130</td>
<td>75</td>
<td>100</td>
<td>70</td>
</tr>
<tr>
<td>7</td>
<td>83</td>
<td>92</td>
<td>70</td>
<td>56</td>
<td>80</td>
<td>65</td>
<td>80</td>
<td>65</td>
</tr>
</tbody>
</table>

Firstly, it can be seen from the data that the absolute values of the calculated curves do not necessarily match those of the experimental data, even if the slopes match. Similarly, adding a term to the equation to force matching of the absolute value at some point may cause the slopes to no longer match. Although it is possible to multiply the calculated curve by some factor in order to force the absolute values of the normalized data and normalized curve to match, as has been performed in other works\textsuperscript{172}, it is not immediately clear that this is a valid technique. Indeed, doing so will change the apparent minority carrier diffusion length (slope) of the curve. Instead, the technique applied to this work has been to attempt to maximize the fit to the experimental data, but without the absolute-value renormalization, in order to provide a curve fit that more closely matches Donolato’s and Luke’s original papers.

It is speculated that the reason for this dis-registry is the fact that a technique intended for large minority carrier diffusion lengths ($>>1000$nm) is being applied to a material with very short minority carrier diffusion lengths ($\sim$100nm). The details of the physical model likely do not match the real physical situation well enough to model the curve-fit in fine detail. In particular, the original Donolato derivation\textsuperscript{186} treats the interaction volume as a 3-dimensional Gaussian distribution. As shown in...
the Monte Carlo calculations of Section 3.2.1.1, the actual minority carrier generation is non-Gaussian and sharply peaked near the beam-incidence point.

Further, the poor fitting on the p-side can only be exacerbated by the fact that the p-side is very thin; the ohmic contact at $x\approx+250\text{nm}$ will modify the minority carrier collection in an ill-defined manner, and the movement of the beam off of the sample will also modify the EBIC behavior.

In Figure 106, it can further be seen that when the stipulation is followed that the curve fit not be calculated where the interaction volume overlaps the depletion region, the result is fitting the curves to the noisy tail-ends of the data, as predicted in the theoretical discussion of Section 3.2.3.3.

The above calculations and curves all used $s=10^{-2}\mu\text{m}^{-1}$, which is tantamount to $v_s=0$. As an example, Table VIII Presents values calculated from area 4 data using $s=10^6\mu\text{m}^{-1}$ (tantamount to $v_s=\infty$), presented alongside the $s=10^{-2}\mu\text{m}^{-1}$ values shown above.

Table VIII: L values (in nm) calculated assuming high or low surface recombination.

<table>
<thead>
<tr>
<th>$E_0$ (keV)</th>
<th>$s=10^{-2}/\mu\text{m}$ n-side</th>
<th>$s=10^{-2}/\mu\text{m}$ p-side</th>
<th>$s=10^6/\mu\text{m}$ n-side</th>
<th>$s=10^6/\mu\text{m}$ p-side</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>85</td>
<td>50</td>
<td>250</td>
<td>100</td>
</tr>
<tr>
<td>5</td>
<td>90</td>
<td>40</td>
<td>150</td>
<td>50</td>
</tr>
<tr>
<td>6</td>
<td>100</td>
<td>70</td>
<td>160</td>
<td>90</td>
</tr>
<tr>
<td>7</td>
<td>80</td>
<td>65</td>
<td>100</td>
<td>78</td>
</tr>
</tbody>
</table>

This same data, for the 4 and 7keV cases, is shown graphically in Figure 107.
Figure 107: Comparisons of L calculations for different s values. (A) 4keV, area 4; (B) 7keV, area 4.
Clearly, the $s=10^6 \, \mu m^{-1}$ case fits poorly to the 4keV data, but the fit to the 7keV data for $s=10^{-2}$ or $10^6 \, \mu m^{-1}$ can be made simply by changing the L value. This is most likely an indication that, first, the actual s-value of these materials is very small, and secondly, that the 4keV has a sufficiently small interaction volume to make analysis of the s-value valid, unlike the larger-R cases, as at 7keV.

Thus, it seems the Donolato-Luke techniques can be applied to short diffusion length materials such as the GaN-based LEDs studied here; however, complications such as the thin p-layer and surface recombination greatly affect the accuracy of this method. The mismatching of the absolute EBIC values indicates the modeling is far from perfect, however. Clearly, an improvement is needed.

5.2.3.3 The Bonard and Ganière method

Equation (3-25) was presented in Section 3.2.3.5 and evaluated theoretically in Section 3.2.4.2. Based on the theoretical analyses of the different methods, it was hypothesized in Section 3.2.4.4 that the Bonard and Ganière technique would be the best-suited to the analysis of short-L EBIC data, but that the effects of s would still be very difficult to measure.

This technique has been applied to the data acquired in this study; however, two issues were noted with the application of the model to the experimental data. These are illustrated in Figure 108 for the 4keV data, from sample area 1.
Figure 108: (A) Experiment vs. calculations for $L_e=40\text{nm}$, $L_h=70\text{nm}$, $s=10^{-2} \mu\text{m}^{-1}$, with varying $d_j$.  (B) $L_e=40\text{nm}$, $L_h=70\text{nm}$, $s=10^{-2} \mu\text{m}^{-1}$, $d_j=20\text{nm}$, with and without normalization.  4keV, area 1.
First, in Figure 108(A), calculation of the theoretical curve using the depletion region width \(d_j=60\text{nm}\) (as estimated from doping levels) provides a theoretical curve with a flat area in the center, which is not replicated in the experimental data. Two possibilities exist here.

First, the actual depletion region may be narrower than is estimated from the doping levels used in the calculation; as the input values of \(~10^{19}/\text{cm}^3\) \(N_a\) and \(~10^{18}/\text{cm}^3\) \(N_d\) are estimated from nominal manufacturer's doping levels of \(~10^{20}\) and \(~10^{18}/\text{cm}^3\), respectively, and then corrected for the fact that Mg is a very deep donor and therefore only \(~10\%\) of the Mg acceptors will be ionized at room temperature. If the nominal doping levels are wrong, or more Mg acceptors are activated than expected, a narrower \(d_j\) than calculated would obtain in practice.

Second, the original Bonard and Ganière model makes the assumption that all carriers generated in the depletion region are separated by drift before recombination, and therefore contribute to \(I_{\text{EBIC}}\). Complications such as surface recombination, trapping at defects, or interfacial scattering and trapping are ignored. As the QWs will be located within or very near the depletion region in these samples, and cathodoluminescence shows clear recombination from the QWs, it is clear that these effects have a complicating effect on these samples and the full-collection assumption in the model is not strictly correct. The complicating effect of self-absorption of cathodoluminescence is also a possible contributor.

As \(E_0\) (and, thus \(\sigma_x\) and \(\sigma_z\)) increase, this \(d_j\) problem has a less pronounced effect, due to the broadening of the interaction volume and smaller proportion of carriers being generated within the depletion region at any given x-point. It is also possible that near-surface effects are modifying the \(d_j\) value at these very low \(E_0\) (and thus very small \(\sigma_z\)) values.

Li et al.\(^{253}\) used the Bonard and Ganière technique to analyze InAs/GaSb superlattice structures; they observed the flat-portion curve expected and fit it very well with the model. This is re-printed as Figure 109.
This indicates that the model is suitable for the calculation of the depletion region effect in p-n junction devices. Therefore, a tentative conclusion is that some effect in the GaN nanostructure results in the deviation of the model from theory at larger \( d_j \) values.

In Figure 108(B), the effect of the normalized values of \( I(x) = \eta(x)/\eta_{\text{max}} \), as opposed to the raw calculated value \( \eta(x) \), are demonstrated. Unlike in the Donolato-Luke technique discussed above, this effect is very minor. In addition, because of the fact that EBIC experimental data is always normalized to 1.0, the physical processes modeled in the equation that contribute to a non-1.0 peak should therefore be normalized to 1.0 in order to treat the modeled data consistently with the experimental data.

From those discussions, two important points are made: first, \( d_j \) of 20nm is used in subsequent calculations, even though 60nm is expected in these materials. Second, the normalized \( I(x) \) function is used, as opposed to the raw \( \eta(x) \) function, for the curve-fitting of the experimental data.
Figure 110 shows example curve fits for the 4 and 7keV data on sample area 4; this is the same experimental data shown above in Figure 106 and Figure 107. Table IX shows L values obtained for the different experimental conditions using the Bonard and Ganière fitting procedure, all assuming $s=10^{-2} \, \mu \text{m}^{-1}$ and $d_l=20\text{nm}$. 
Figure 110: Bonard-Ganière fits for (A) 4keV, area 4 data; (B) 7keV, area 4 data.
Table IX: $L_h$ and $L_e$ values (in nm) estimated using the Bonard and Ganière technique.

<table>
<thead>
<tr>
<th>$E_0$ (keV)</th>
<th>Area 1 n-side</th>
<th>Area 1 p-side</th>
<th>Area 2 n-side</th>
<th>Area 2 p-side</th>
<th>Area 3 n-side</th>
<th>Area 3 p-side</th>
<th>Area 4 n-side</th>
<th>Area 4 p-side</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>125</td>
<td>50</td>
<td>80</td>
<td>35</td>
<td>100</td>
<td>45</td>
<td>70</td>
<td>40</td>
</tr>
<tr>
<td>5</td>
<td>35</td>
<td>15</td>
<td>55</td>
<td>10</td>
<td>50</td>
<td>10</td>
<td>60</td>
<td>15</td>
</tr>
<tr>
<td>6</td>
<td>40</td>
<td>40</td>
<td>40</td>
<td>45</td>
<td>40</td>
<td>45</td>
<td>60</td>
<td>45</td>
</tr>
<tr>
<td>7</td>
<td>60</td>
<td>20</td>
<td>70</td>
<td>20</td>
<td>70</td>
<td>20</td>
<td>75</td>
<td>20</td>
</tr>
</tbody>
</table>

The above values were calculated assuming zero surface recombination. Above, in the Donolato-Luke calculations, at 4keV, the calculated curves for $s=10^{-2}$ or $10^6 \mu m^{-1}$ showed measurably different behavior, but the behavior is effectively indistinguishable at 7keV. The same trend is seen in the Bonard and Ganière analyses here. Figure 111 shows the 4 and 7keV data for area 4.
Figure 111: (A) 4keV, area 4 data and Bonard and Ganière fits; (b) 7keV, area 4 data and Bonard and Ganière fits.
In Figure 111, d_j=20nm for all calculated curves. The values of L used to find best-fits for the s=10^{-2} or s=10^{6} \mu \text{m}^{-1} cases are given in Table X.

Table X: L values (in nm) calculated assuming high or low surface recombination.

<table>
<thead>
<tr>
<th>E_0</th>
<th>s=10^{-2} \mu \text{m}^{-1}</th>
<th>s=10^{6} \mu \text{m}^{-1}</th>
</tr>
</thead>
<tbody>
<tr>
<td>keV</td>
<td>n-side</td>
<td>p-side</td>
</tr>
<tr>
<td>4</td>
<td>70</td>
<td>40</td>
</tr>
<tr>
<td>5</td>
<td>60</td>
<td>15</td>
</tr>
<tr>
<td>6</td>
<td>60</td>
<td>45</td>
</tr>
<tr>
<td>7</td>
<td>75</td>
<td>20</td>
</tr>
</tbody>
</table>

The value "$>>10,000\text{nm}$" for the n-side at 4keV indicates that no L_h value was able to achieve a good fit at $10^{6} \mu \text{m}^{-1}$ at 4keV. For the 5, 6 and 7keV data, increases of L_h were necessary to achieve an improved fit, but no change was needed to achieve a fit for L_e.

The data in Figure 111 and Table X indicate that, at 5keV or greater, the electron range is sufficiently large in comparison to L to make determination of s unfeasible. However, the very small interaction volume at 4keV means that, with the 50~150nm diffusion lengths being observed for L_h, it is feasible to quantify L_h and s in this case. Again, it is most likely that the L_e values are underestimates of the true values, due to the presence of the ohmic contact at x~+250nm.

Thus, it can be seen that the Bonard and Ganière technique is capable of reasonable quantification of the L and s values even in the GaN LED-type materials examined here. In particular, the 4keV data in conjunction with simulation shows that despite short diffusion lengths, surface recombination can be analyzed. If improved sample preparation or instrumentation allowed operation in the 1-3keV range, it is conceivable that the "classical" X-EBIC behavior (Figure 39) could be regained in these highly non-classical materials. It has also been shown that,
indeed, \( n_s \approx 0 \) in this material, such as measured by other groups using the Donolato technique\(^{187}\) or inferred by groups using near-surface CL\(^{255}\).

5.2.3.4 Method of moments

As discussed in Section 3.2.3.6, the method of moments requires integration to find the first moment of the EBIC curve. Two curves taken at different accelerating voltages \( E_0 \) then have electron ranges \( R_1 \) and \( R_2 \), depth-factors \( Z_1 \) and \( Z_2 \) (\( Z_i = 0.41R_i \)) and moments \( m_1 \) and \( m_2 \).

Equation (3-30) is then solved to yield \( L \) and \( s \) for the sample; (3-30) is repeated here as (5-38).

\[
\begin{align*}
m_1 &= L^2 \left\{ 1 - \left[ \frac{sL}{(1+sL)} \right] \exp \left[ - \frac{Z_1}{L} \right] \right\} \\
m_2 &= L^2 \left\{ 1 - \left[ \frac{sL}{(1+sL)} \right] \exp \left[ - \frac{Z_2}{L} \right] \right\} 
\end{align*}
\]

(5-38) must be solved numerically, as it has no closed-form solution. In the original paper\(^{188}\) examined a silicon device at 20 and 30keV, and found the \( m \), \( R \), \( Z \), \( L \), and \( sL \) values listed in Table XI.

Table XI: Example moments method data; from Donolato\(^{188}\).

<table>
<thead>
<tr>
<th>( E_0 )</th>
<th>( R )</th>
<th>( Z )</th>
<th>( m )</th>
<th>( L )</th>
<th>( sL )</th>
<th>( s )</th>
</tr>
</thead>
<tbody>
<tr>
<td>keV</td>
<td>( \mu \text{m} )</td>
<td>( \mu \text{m} )</td>
<td>( \mu \text{m}^2 )</td>
<td>( \mu \text{m} )</td>
<td>---</td>
<td>( \mu \text{m}^{-1} )</td>
</tr>
<tr>
<td>20</td>
<td>3.2</td>
<td>1.3</td>
<td>335</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>6.6</td>
<td>2.7</td>
<td>465</td>
<td>98</td>
<td>46</td>
<td>0.47</td>
</tr>
</tbody>
</table>

A numerical solver was written using Maple 9.5 software, and is presented in Appendix 7; it was tested by using the data from Donolato as an example. With the input data of Donolato's \( m \) and \( Z \) values, it yielded \( L = 97 \mu \text{m} \) and \( sL = 43 \), indicating agreement with Donolato's results.
However, this solver was not found to operate properly when analyzing the data obtained experimentally in this work. This is not unexpected, as Donolato suggests this method will only be valid when R/L<0.5, and in our case, R/L>>0.5. For example, for the n-side of area 4 with E₀=4 or 7keV, the extracted data is presented in Table XII.

Table XII: Method of moments calculations for area 4, n-side

<table>
<thead>
<tr>
<th>E₀ (keV)</th>
<th>R (nm)</th>
<th>Z (nm²)</th>
<th>m (nm)</th>
<th>L (nm)</th>
<th>sL</th>
<th>s (μm⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>138</td>
<td>57</td>
<td>3321</td>
<td>-44</td>
<td>-0.17</td>
<td>0.0039</td>
</tr>
<tr>
<td>7</td>
<td>351</td>
<td>144</td>
<td>12154</td>
<td>-44</td>
<td>-0.17</td>
<td>0.0039</td>
</tr>
</tbody>
</table>

Obviously, values of L<0 are non-physical. In the silicon material example given by Donolato¹⁸⁸ (Table XI), the factor Z/L very nearly zero, so the term \( \exp(-Z/L) \) gives a value very nearly 1. However, in the case of the GaN LED in the present work, if we assume L values ~50nm, Z/L ~ 1 to 3. Thus, \( \exp(-Z/L) \) approaches zero and suppresses the sL/(1+sL) terms in Equation (5-38), which appears to cause the convergence issues and again shows how large R values in comparison to L reduce the surface-sensitivity of an EBIC experiment.

It thus concluded that this method of X-EBIC quantification is wholly unsuitable for GaN and similar short-L materials. Beam energies ~2keV (R=43nm) would be necessary to use the method of moments. As it does not rigorously model the carrier generation in the manner of the Bonard and Ganière technique, even then it would not be an accurate technique.

5.2.3.5 Ong et al. method

The method of Ong et al. is discussed in Section 3.2.3.7; an example Ong et al. analysis of the 4keV data of area 4 is shown in Figure 112.
As stipulated by Kurniawan and Ong\textsuperscript{202}, the fit is begun at \(x=1.5R=-208\text{nm}\). Although they suggest extending the fit to \((x/L)_{\text{END}}\approx 11\) in order to maximize accuracy, in this case the fit was extended until the EBIC data reached the noise floor of the data. The data in Figure 112 was fit to Equation (3-33) with a slope of 0.0147\text{nm}^{-1} and \(\alpha=0.0\). This indicates \(L\approx 68\text{nm}\) in this analysis, and \(s=0\). This implies that \((x/L)_{\text{END}}\approx 7\).

However, the original paper describing the Ong et al. technique\textsuperscript{199} suggests that the curve generated by Equation (3-32) will be concave-upward in the presence of \(v_s\), and making \(\alpha\) more negative will straighten this curve. Figure 112 shows a slight concave-downward shape, indicating behavior outside of the range of behavior expected by the model. The Ong et al. model could not be applied to the p-side of any of the data in this experiment, as the \(x>1.5R\) requirement put the area of the fit
wholly into the experimental noise level. Neither could the Ong et al. model be
applied to the 7keV data, again because the $x>1.5R$ requirement pushed the fit into
the experimental noise.

All Ong et al. fits showed a best-fit correlation at $\alpha=0.0$, if the range of $\alpha$ was
restricted to $-0.63<\alpha<0.0$. In actuality, better fits would have been achieved with
$\alpha>0.0$, which is forbidden by the Ong et al. theory (Equation (3-34)). L values
derived from these fits are presented in Table XIII.

Table XIII: $L_h$ and $L_e$ values (in nm) estimated using the Ong et al. technique.
Values marked $\varnothing$ were incalculable with this technique. All values had $\alpha\approx0$ and $s\approx0$.

<table>
<thead>
<tr>
<th>$E_0$ (keV)</th>
<th>Area 1</th>
<th>Area 2</th>
<th>Area 3</th>
<th>Area 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n-side</td>
<td>p-side</td>
<td>n-side</td>
<td>p-side</td>
</tr>
<tr>
<td>4</td>
<td>65</td>
<td>$\varnothing$</td>
<td>52</td>
<td>$\varnothing$</td>
</tr>
<tr>
<td>5</td>
<td>57</td>
<td>$\varnothing$</td>
<td>62</td>
<td>$\varnothing$</td>
</tr>
<tr>
<td>6</td>
<td>50</td>
<td>$\varnothing$</td>
<td>52</td>
<td>$\varnothing$</td>
</tr>
<tr>
<td>7</td>
<td>$\varnothing$</td>
<td>$\varnothing$</td>
<td>$\varnothing$</td>
<td>$\varnothing$</td>
</tr>
</tbody>
</table>

It is clear that the Ong et al. technique was not able to calculate many of the L
values in the table. Further, for the 5 and 6keV data, such as shown for 6keV, n-
side, area 4 in Figure 113, the model must be fit to the noisiest tail end of the data.
The datapoints fit in Figure 113 are clearly the noisiest, and the slope of the fit area does not match the slope of the rest of the curve. It is thus seen that the 5 and 6keV data fits for L are very suspicious and should not be taken as truly quantitative data.

The most recent Ong et al. paper suggests that this technique will be most accurate when $Z/L < 0.1$. Taking $Z \approx 0.41R$, in the present case $Z \approx 57$ to 144nm, but $L < 100$nm, indicating $Z/L \approx 0.5$ to 2. Thus, it comes as no surprise that the Ong et al. technique is highly inappropriate to the experimental situation of GaN LEDs. Although, as shown in the theoretical calculations of Section 3.2.4, the Ong et al. technique should work very well for large-L materials, this work has clearly shown with both theoretical and experimental tests that it is inapplicable to small-L materials.
5.2.3.6 Comparison of techniques

EBIC is the best technique for the determination of $L$ and $s$ in LED-type devices; thus, no independent technique is available to provide "true" values of $L$ and $s$ in these materials for comparison to the data extracted above. However, the different techniques can be compared to each other. Figure 114 through Figure 117 shows comparisons of the different fits to the n-side of area 4 at the different $E_0$ values.

![Figure 114: EBIC model computations and data for sample area 4, at 4keV.](image)
Figure 115: EBIC model computations and data for sample area 4, at 5keV.
Figure 116: EBIC model computations and data for sample area 4, at 6keV.
The Ong et al. fit could not be applied to any of the p-sides, nor could it be applied to the n-side of the 7keV sample (Figure 117).

Although there are variations in the slope of the log(I(x)) vs. x curves as x becomes larger in magnitude, the area of greatest practical interest in the area within ~1L of the peak, as this is where the light-emitting active region is located. It is clear from Figure 114 through Figure 117 that only the Bonard and Ganière and exponential methods actually fit to this region, and only the Bonard and Ganière method provides a good fit in this region. Although there are issues associated with the normalization of the Bonard and Ganière type data, and the effect of the depletion region, it is clear that the Bonard and Ganière technique is the most useful for evaluation of minority carrier diffusion lengths in GaN LEDs, and by operation of the SEM at or below 4keV, it becomes possible to quantitatively evaluate not just diffusion length L, but the surface recombination parameter s as well, using the Bonard and Ganière method.
5.2.3.7 X-EBIC with thermionic SEM

Data was taken on the same sample, but using the Hitachi S-3200N thermionic SEM. An accelerating voltage of 5keV and a beam current of 1.0nA was used; the spot size was measured to be 100-150nm. The data and a Bonard and Ganière fit for \( \text{L}_h = 65\text{nm}, \text{L}_e = 70\text{nm}, d_p = 20\text{nm}, \text{ and } s = 10^{-2} \text{ } \mu\text{m}^{-1} \), are shown in Figure 118.

![Figure 118: Thermionic SEM data for sample at 5keV.](image)

The \( \text{L}_h \) value obtained is similar to the values obtained with the field-emission instrument and discussed above. The \( \text{L}_e \) value is somewhat larger than obtained in the above analyses; this appears to be due to the more gradual fall-off of the p-side data in comparison to the field-emission data. The most likely explanation for this is that the thermionic SEM has much larger beam-tails and these, in conjunction with
the much higher beam current, resulted in a very gradual drop-off of the EBIC signal with large positive values of $x$.

The high beam currents associated with the thermionic instrument make operation at very low accelerating voltages possible; Figure 119 shows an EBIC micrograph taken at 3keV, 1.0nA (compare to 4-7keV and ~0.1nA for the field emission data above).

Figure 119: 3keV, 1.0nA EBIC image of blue LED. Dashed area in analyzed in Figure 120.

Figure 119 was taken with $d_b\approx270$nm. The area marked by the dashed box was analyzed by the Bonard and Ganière model to estimate the diffusion lengths as $L_e\approx L_h\approx125$nm, with $d_j=20$nm and $s=10^{-2} \mu m^{-1}$. $\sigma_x$ and $\sigma_z$ values for $E_0=3$keV and $d_b=270$nm were calculated and used for the fit. This is shown in Figure 120.
Examining the actual EBIC current by monitoring the current amplifier output with an oscilloscope indicates that the peak EBIC signal was also \( \sim 1\text{nA} \), indicating that the ratio of \( i_{\text{peak}}/i_B \) was \( \approx 1 \); for 3keV, and under the assumption of full collection, the anticipated \( i_{\text{peak}}/i_B \) ratio would be \( \approx 250 \). This indicates that only \( \sim 0.4\% \) of the anticipated number of injected carriers actually contributing to EBIC. This is certainly due in part to the action of the surface damaged layer, but the broadened interaction volume resulting from the large spot size will also mean that even if the center of the spot is on the depletion region, many of the injected carriers will be injected so far from the depletion region (distance \( \gg L \)) that they will not be collected before recombining.

This rough calculation, in conjunction with the values of \( L \) that are large in comparison to those measured with the field emission instrument, and the noise level apparent in the EBIC image, all lead to the conclusion that simply using a low
$E_0$ value and a higher beam current are not desirable means to achieve improved EBIC data. First, the low beam energy and concomitant large spot size will result in a loss of resolution in the EBIC image. Although it was possible to model the $\sigma_x$ and $\sigma_z$ interaction parameters as a function of spot size and achieve a good fit, when the depletion region extent and minority carrier diffusion lengths are much smaller than the spot size, both accuracy and absolute signal level should be improved by the use of a small-spot-size field-emission SEM. In principle, a latest-generation Schottky instrument could be used, which could provide nA-range beam currents in a small probe at low accelerating voltage. However, in this case, sufficiently large beam currents at very small interaction volumes would run the risk of high-injection.

5.2.4 Cross-sectional EBIC: conclusions

It has been shown that, cross-sectional EBIC is a powerful technique for the examination of the minority carrier properties in GaN-based devices. In particular, minority carrier diffusion length and surface recombination velocity can be estimated by the comparison of SEM-acquired X-EBIC curves to theoretical models. It has also been shown both theoretically and experimentally that none of the models available in the literature can be applied naively to GaN-based LED samples, due to the very short L-values and thin p-side extent. However, current results lead to the conclusions that the model of Bonard and Ganière is the most accurate for the purpose of extracting $L$ and $s$ from the area near the active region of a GaN LED. It has also been shown how Monte Carlo modeling can be used to determine input parameters for the Bonard and Ganière model.

The most important conclusion is that the Bonard and Ganière model was able to extract both the minority carrier diffusion length and the surface recombination parameter when the X-EBIC data was taken at 4keV beam energy. Because of the short minority carrier diffusion lengths in the samples examined, very small interaction volumes are required, and 4keV was empirically shown to be small enough to make the EBIC curves responsive to changes in $s$ in these materials.
Luke et al.\textsuperscript{190} suggested, based on computational arguments, that if \( R \leq 4L \), then the Donolato-Luke technique could be validly applied. In these experiments, \( L \) was found to be on the order of 50-100nm. In the 4keV case, \( R \approx 138\)nm, and at 5keV, \( R \approx 200\)nm. It was at 4~5keV that the Donolato-Luke and Bonard and Ganière techniques began to transition from the "large \( R \)" to "small \( R \)" case, as indicated by the transition from surface-insensitivity to being surface-sensitive. Thus, it has been shown empirically that, at least in GaN-based LEDs, the \( R \leq 4L \) rule is approximately accurate, although perhaps slightly optimistic; \( R \leq 2L \) is suggested as a more conservative rule. However, the minimum \( E_0 \) (and concomitant \( R \)) value that will provide usable signal-to-noise ratios should always be chosen, in order to maximize the surface sensitivity of the data and minimize broadening of the interaction volume. This minimum \( E_0 \) will be a function of surface quality, noise pickup due to defects and shunt resistance in the device itself, and the noise level inherent to the instrumentation, particularly the in-chamber EBIC cabling. As such, the minimum \( E_0 \) value will have to be determined empirically for every given experimental run.

Further, the very short values of the p-side minority carrier diffusion lengths \( L_e \) found in this work, regardless of the quantification technique, indicates that the top-side contact \( \sim 250\)nm from the p-n junction is probably with the diffusion of the carriers in the p-side and biasing the EBIC data. This is especially clear when the poor fit of the Donolato-Luke technique curves on the p-side is compared to the much better fit on the n-sides (see Figure 106(A)). Thus, it should be concluded that measurements of \( L \) in such a confined region will at best be estimates of the value. Because the Bonard and Ganière technique can be applied to the area nearest the junction, unlike the other techniques, it should again be least-inaccurate for evaluation of diffusion lengths in this situation.

In materials with diffusion lengths \( L \sim 50 \) to 100nm, such as is the case here, thermionic SEM can be used, but the small spot size of a field emission SEM will greatly improve quantification.
Lastly, if we take the 4keV Bonard and Ganière analyses are being the most accurate, and average across all four analysis areas, the minority carrier properties of this sample can be summarized as.

Table XIV: Minority carrier properties of examined sample

<table>
<thead>
<tr>
<th>L_h</th>
<th>L_e</th>
<th>s</th>
</tr>
</thead>
<tbody>
<tr>
<td>92±15 nm</td>
<td>42±6 nm</td>
<td>&lt;10^{-2} μm^{-1}</td>
</tr>
</tbody>
</table>

5.3 Cathodoluminescence and plan-view EBIC

5.3.1 Sample preparation and experimental techniques

Cathodoluminescence spectroscopy and microscopy are valuable techniques for the evaluation of the bulk and defect optical properties of semiconductors. Plan-view EBIC is included in this section, rather than with cross-sectional EBIC above, as PV-EBIC, like CL, is a defect-imaging and -evaluation technique.

Samples for CL and PV-EBIC could be prepared in the exact same manner as X-EBIC samples above. Alternately, the wire-bonding step could be neglected, as samples for CL alone do not require a live electrical circuit (although EBIC will, of course, be impossible in a non-electrically-functional sample). Wafer material could also be cleaved to ~5x5mm sections and examined via CL. As the wafer material used in this study was not metallized, EBIC could not be performed on it.

For cross-sectional CL, wafer or LEDs could be prepared. LEDs were prepared as for X-EBIC, above. Wafer was polished by grinding on an Allied Techprep polishing wheel with 30, 6, 1, 0.5 and 0.05μm diamond lapping films. The 30μm step was used to remove ~100-200μm or more of material. The 6-0.5μm steps were used to remove an amount of material equal to 3x the grit size of the pad before; i.e., the 1μm pad was used to remove at least 18μm of material (3x6μm=18μm). This ensures removal of all polishing damage imparted by the previous step. 0.05μm films were used to polish the sample for ~10-20 minutes, but
did not remove measurable amounts of material. The 1, 0.5 and 0.05\(\mu\)m steps were always performed with the motion of the grinding platen parallel to the QW direction.

For PV-EBIC, samples were mounted and examined in a manner identical to that of X-EBIC; the only difference, of course, being the orientation of the sample under the SEM beam. PV-EBIC was performed in either the JEOL JSM6400F or Hitachi S-3200N instruments. PV-EBIC in the Hitachi could be performed in a dedicated EBIC apparatus or within the CL apparatus.

The CL apparatus was designed, built, and optimized as a major part of this work. The CL apparatus was implemented solely in the Hitachi SEM; it was chosen not to perform CL in the JEOL FE-SEM for two reasons:

1. **Vacuum level.** The CL apparatus was built from off-the-shelf parts; the \(\sim10^{-5}\) torr pressure in the Hitachi’s sample chamber was not impacted by these components, but the \(\sim10^{-7}\) torr pressure in the JEOL sample chamber would have been incompatible with many of the parts.

2. **Beam stability.** As mentioned above, a cold-field-emission SEM will have an unstable beam current due to the field-emission process taking place at the source. Due to the long acquisition times of CL spectra, and to a lesser extent images, this instability would manifest as uncorrectable artifacts in CL data acquired. (Indeed, high-resolution CL is usually performed using a Schottky-type instrument\(^{223}\).)

Additionally, the large beam current of the Hitachi S-3200N allowed injection-density CL studies to be performed over more orders of magnitude than would have been possible in the FE-SEM.

The CL system is illustrated schematically in Figure 121; the individual sub-components will be described in detail below.
As is apparent from Figure 121, the CL system consists of three gross parts: the mirror-sample assembly, the lens-fiber assembly, and the detectors. The mirror-sample assembly is shown in the photograph of Figure 122.
The XYZ micromanipulator is a Newport corporation MS-125-XYZ micropositioner, which was custom-ordered to have $10^{-6}$ torr vacuum compatibility. The mirror is taken from a small flashlight, and was ground on fine-grit sandpaper to remove the bottom ~40% of the mirror. A hole, ~2mm in diameter, was drilled above the focus of the mirror. The backside of the mirror was painted with silver paste (SPI product #5063) to prevent charging. The baseplate of the stage, sample holder, and the plate attaching the mirror to the micromanipulator were custom machined from aluminum.

Figure 123 shows a photograph of the lens-fiber assembly.
The black tube in Figure 123 is a series of Edmund Industrial Optics C-mount style lens tubes, consisting of a 25mm lens holder (Edmund part #56-353), an extension tube (54-629), a fine-thread focusing tube (03-625), and a 9mm thick-lens mount (54-622). The extension tube had a hole drilled in it to vent the area between Lens 1 and Lens 2 to vacuum. The 25mm lens, Lens 1, is a plano-convex 25mm diameter 75mm focal length synthetic fused silica lens with antireflective coating (NT46-268); the 9mm lens, Lens 2, is a plano-convex 9mm diameter, -13.5mm focal length lens of the same material and coating (46-311).

The third lens is a Thor Labs F260SMA-A fiber-optic collimator, special ordered to be vacuum-compatible to $10^{-6}$ torr. It has a numerical aperture NA=0.16 and a quoted wavelength range of 400-600nm. A custom aluminum adaptor plate connected the threading on the snout of the F260SMA-A to the C-mount 9mm lens holder. A hole was drilled in the plate to vent the area between Lens 2 and the collimator.

The fiber optic vacuum feed-through is an Avantes FC-VFT-UV1000, which contained a 1000μm UV-visible fiber optic, and had SMA-905 connectors on both ends. The NA of this fiber was matched to that of the collimator mentioned above.

A number of different detectors were available; all used SMA-905 inputs to couple to the vacuum feed-through.
Two different PMTs were used for panchromatic imaging; their input windows were held directly in front of the vacuum feed-through's fiber-optic output. The first PMT was a Hamamatsu H6780-20 photomultiplier module; this PMT supplied a current output, and had a gain range of 0-10⁶. However, as this was an infrared-sensitive PMT, it showed significant dark counts, and was replaced by an improved PMT, a Hamamatsu R1527P, which had UV-visible sensitivity and very low dark current. The gain range for the R1527P was 0-10⁷; however, the operating voltage of the PMT that gave maximum signal/noise was 980 volts, corresponding to ~5x10⁶ gain.

The R1527P was a bare photomultiplier tube, and as such a custom housing was built for it (Figure 124). This consisted of an aluminum block fitted with an SMA-905 adaptor (Thor Labs SM1SMA), a magnetic shield to protect the electron trajectories in the PMT from interference (Hamamatsu E989), a high-voltage socket with voltage divider (Hamamatsu C8991), and an aluminum Faraday cage surrounding the high-voltage socket. With the Faraday cage properly grounded, the dark current of the assembly was ~100pA at 980V. The high voltage socket required a 15-volt input supply, and a 0.000 to 1.250 volt control voltage, which were provided by a 15 volt battery and potentiometer circuit housed in a home-made PMT power supply.
The aluminum block and Faraday cage were in contact and thus grounded to each other; the Faraday cage, in turn, was grounded to the chassis ground of a Keithley 428 current amplifier. The R1527P was used in current-output mode, and connected to a Keithley 428 current amplifier. Imaging was performed in the same manner as an EBIC signal, described in Section 5.2.1.

An Ocean Optics USB2000 mini-spectrometer (the same spectrometer described by Bunker et al.\textsuperscript{133}) was available for spectroscopy; however, its UV sensitivity was poor and its noise level very high, and was only used when very large SEM beam currents excited significant CL signal. The USB2000 is shown in Figure 125, and was controlled by a laptop computer.
Used from high signal levels down to near-single-photon levels was an Acton SpectraPro 2150i monochromator/spectrometer. This was a 150mm nominal-focal-length Czerny-Turner grating monochromator, equipped with a 1200 lines/mm 500nm-blaze grating. Slit widths were continuously variable from <10 to 3000μm, which gave quoted spectral resolutions of <0.04nm to ~12nm. The detector was another R1527P PMT, in an Acton PD471 housing. In spectroscopy the signal was processed by an Acton SpectraHub, coupled to the SpectraPro and a laptop computer.

The SpectraPro 2150i was also used for spectrally-resolved imaging. In addition to operating as a scanning spectrograph, it could also operate in monochromator mode. This was accomplished by slewing the grating to illuminate the PMT with a particular central wavelength. The PMT output was then input into a Keithley 428 and imaged as an EBIC or panchromatic signal, described above. The SpectraPro is shown in Figure 126.
Figure 126: SpectraPro spectrometer/monochromator.

The 436.6nm Hg-line from fluorescent lighting was used to test the calibration and resolution of the spectrometer; spectra of this line, recorded at different slit-widths (and thus resolutions), are given in Figure 127.
The FHWM of these curves were determined via Fityk curve-fitting software\textsuperscript{256}, and are given in Table XV.

Table XV: Empirical spectrometer bandpasses at 436.6nm

<table>
<thead>
<tr>
<th>Slit width ((\mu\text{m}))</th>
<th>Peak FWHM (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.074</td>
</tr>
<tr>
<td>50</td>
<td>0.37</td>
</tr>
<tr>
<td>500</td>
<td>2.7</td>
</tr>
<tr>
<td>1500</td>
<td>7.3</td>
</tr>
</tbody>
</table>

At the 10\(\mu\text{m}\) slit width, the peak began showing low- and high-energy shoulders, indicating that the resolution is limited by the structure of the peak at that resolution.

Long acquisition times are necessary for many CL images or spectra; as such, carbon contamination and beam-induced damage are important.
considerations. It was found that oxygen-plasma-cleaning the chamber with an Evactron plasma cleaning unit for approximately 20 minutes before performing CL experiments would eliminate all carbon contamination for several hours. Although LED samples suffered beam-induced damage that changed the intensity and peak position of the CL emission, wafer material did not. This may be due to surface treatments performed in the final LED processing steps introducing states into the near-surface bandgap that were filled by beam-injected carriers, which then persisted with long lifetimes. However, the exact nature of this effect was not explored. Rather, wafer material was used for long-dwell-time CL experiments in order to simply eliminate the effect.

It is possible to estimate the efficiency of the CL signal collection. From Chichibu et al.\textsuperscript{257}, we can estimate the internal quantum efficiency of the QW emission at $\sim$0.10. As the QW emission is of much lower energy than the bandgap of the GaN cladding layers, the absorption coefficient $\alpha$ can be assumed to be $\sim$0. Due to the very short ($\sim$250nm) photon travel lengths, scattering will likely also be negligible. From these considerations and the quantitative discussion in Section 3.2.5, the external quantum efficiency at the QW wavelengths can be estimated as $\sim$0.003. This, however, seems somewhat low, as the electroluminescent external quantum efficiencies of LEDs are often estimated at $\sim$0.10. It is also possible to estimate the collection efficiency using an assumption of LED external quantum efficiency. A sample prepared as for EBIC was placed in the CL setup. Due to handling and removal of the reflective mounting, an assumption of 1% electroluminescent EQE was made. The LED was driven at 1mA of EL forward current, which implies $6.2\times10^{13}$ photons/sec electroluminescence. From the PMT gain setting and manufacturer's specifications, it was possible to estimate that the panchromatic PMT was measuring $\sim1.5\times10^{12}$ photons/sec. This indicates a CL system collection efficiency (in panchromatic imaging) of 2%. Due to the design of the monochromator, spectroscopy or spectroscopic imaging will likely show $\sim$1/10 to 1/2 the efficiency of panchromatic imaging.

239
5.3.2 CL and Plan-view EBIC: results and discussion

5.3.2.1 General features observed in LEDs

Figure 128(A) shows a typical SE image of a plan-view CL/EBIC LED sample; Figure 128(B) shows the I-V curve, indicating good diode characteristics from the sample preparation.

Figure 128: (A) SE image of LED for plan-view analysis; (B) I-V curve from LED. SEM image, SE mode, beam conditions 10kV, 680pA; blue LED sample.

In Figure 129(A)-(C), low- and high-magnification EBIC images of the sample are shown:
In Figure 129(a), the first feature to note is the bright edge of the image; this is the edge of the GaN mesa, which is non-metallized. The darker interior of the LED structure has a thin, optically transparent metallization layer on it, which results in the slight loss of signal in that area. Secondly, large bright- and dark-areas can be seen; in Figure 129(b), more detail of these dark areas is visible. They are \( \sim 5-10\mu\text{m} \) in size, and roughly circular. The EBIC scale has been stretched by the imaging electronics; the actual contrast from bright to dark areas is \( \sim 33\% \), as the EBIC current in the bright areas was \( \sim 6\text{nA} \) and in the dark areas \( \sim 4\text{nA} \). It would be expected that a 10keV beam would induce \( \sim 1000 \) EHPs per incident electron (Equation (3-9)); thus, for 100% EBIC collection efficiency, a 10keV beam current of \( 270\text{pA} \) would lead one to expect \( \sim 220\text{nA} \). The actual EBIC current is 2-3\% of that number. Two major factors will contribute to this reduction; firstly, it is apparent from the large number of dark areas in the micrographs that this sample contains a high density of electronically active defects, which cause recombination and loss of minority carriers and thus loss of EBIC current.

Secondly, the active region of this type of device is \( \sim 250\text{nm} \) deep; the peak of the energy deposition curve of 10keV electrons incident on GaN is at \( \sim 100\text{nm} \). If a minority carrier diffusion length \( L \) of 50nm is assumed in the top p-side, then the peak of the EHP generation is approximately \( 2L \) distant from the p-n junction. For a falloff proportional to \( e^{-nL} \), only a few percent of the injected carriers -- even in the
absence of fast-recombination trapping sites such as defects -- will reach the p-n junction and contribute to the EBIC current. This order-of-magnitude calculation agrees with the measured EBIC current level.

Also of note in Figure 129(B) and Figure 129(C) is that, in addition to the large-dark-spot contrast is a population of small dark spots. A point-count analysis of these spots indicates a density ~10^8 to 10^9/cm^2. Given the size of these spots, <<1μm, and their density, it seems clear that they must be interpreted as threading dislocations in the GaN layer that, as recombination sites, cause local depletion of minority carriers and an associated dark EBIC contrast.

Panchromatic CL is valuable complement to PV-EBIC. PV-EBIC images changes in the overall recombination time; panchromatic CL images overall changes in the non-radiative to radiative rate ratio. As such, both image defected areas, due to enhanced non-radiative recombination at defects. Figure 130 is a SE-EBIC-CL triplet of images of the same device shown above in Figure 128 and Figure 129.

![Figure 130: (A) SE; (B) EBIC; (C) Panchromatic CL images of LED. SEM images; 10keV, 50pA; blue LED sample.](image)

Similar to Figure 130 is Figure 131, which shows the three modes of image but at significantly higher magnification. Dislocations and large-dark-spot defects are seen in EBIC; signal/noise in CL is lower, but effectively the same contrast is seen.
Important to note in Figure 130 and Figure 131 is that the SE images show no detail, other than the edge of the metallization. That is to say, there is no surface topography associated with the CL and EBIC contrast. Thus, it seems clear that the defects associated with the changes in the CL and EBIC contrasts are within the samples, and not manifesting themselves at the surface, at least at the resolution achieved in the SE image.

These types of images lead to the question, "can correlations be drawn between the CL and EBIC contrast?" Clearly seen in Figure 131(B) are two large dark spot on the top of the image; in Figure 130(C), the left EBIC dark spot appears dark in CL as well; the right EBIC dark spot appears slightly dark in CL.

Figure 132, which re-prints the data from Figure 131(B) & (C), shows two particular areas that correlate well; the areas marked → are a pair; as are the areas marked ⊡. Each pair of areas shows dark contrast in both CL and EBIC modes.
Figure 132: (A) EBIC; (B) Panchromatic CL. Arrows indicate example areas that show dark contrast in both modes. SEM images; 10keV, 50pA. In (B), brightness and contrast have been modified to make the dark areas more clear.

Thus, it can clearly be seen that EBIC and panchromatic CL give valuable, complementary information regarding crystal defects that cannot be obtained by SE imaging.

5.3.2.2 Spectral data from LEDs and wafers

At this point, EBIC and panchromatic CL need to give way to CL spectroscopy and spectrally resolved CL imaging. Figure 133 shows a spectrum obtained from a blue LED.
As discussed in Section 3.2.5, voltage-resolved, constant-power spectra use CL as a depth-probing technique. Figure 133 uses 10, 20 and 30keV electrons at constant beam power, this injecting a nominally constant number of EHPs (Equation 3.9) and thus exciting, nominally, the same number of photons. As is clear from the figure, the different beam energies result in very different spectral responses from the sample. In particular, the QW and GaN emission is higher at 20keV than at either 10 or 30, and the YB emission increases with $E_0$. The standard interpretation is in terms of the LED layer structure; the 10keV beam is mostly exciting the thin top p-layer. As reviewed previously, the YB is typically attributed to Ga-vacancies, and these are most common in the n-doped GaN layers. Further, the 10keV beam deposits most of its energy (and EHPs) in the p-side; only those that diffuse to the QW layer, a small fraction, given the short diffusion length, will contribute to the QW peak.

At 20keV, however, the peak of the injection is very near the QW depth, leading to a strong QW signal. Because more carriers are injected into the n-side, the YB emission is increased. Similarly, with the 30keV beam, the peak injection is
well below the QWs, leading to a loss of QW intensity, but the peak is in the YB-producing n-layer, giving the strong YB peak. This is illustrated in Figure 134.

Figure 134: Images and spectrum from blue LED.
(A) SE image; (B) Panchromatic CL; (C) 3.35±0.05eV CL; (D) 2.72±0.02eV; (E) 2.66±0.02eV; (F) 2.60±0.02eV; (G) 2.25±0.03eV; (H) Spectrum. SEM beam conditions: 10kV, 1.0nA.

As seen in Figure 134(H), the spectrum is dominated by the QW peak. The panchromatic image, Figure 134(B), shows essentially the same contrast as the images taken using the QW peak (E) or the tails of the QW peak (D, F). The YB
image (G) is featureless; the GaN band-edge image shows little contrast in comparison to the panchromatic or QW images.

Increasing $E_0$ will result in a degradation of spatial resolution, but will excite the GaN and YB emissions more strongly. Figure 135 shows micrographs of the same sample taken at 20keV.

![Micrographs of sample](image.png)

Figure 135: Spectra and spectrally resolved images taken at 20kV, 1.5nA.

(A) 3.35±0.03eV; (B) 2.75±0.02eV; (C) 2.72±0.02eV; (D) 2.66±0.02eV; (E) 2.60±0.02eV; (F) 2.25±0.03eV; (G) Spectrum; (H) Detail of (G).

It should be noted that the ± values quoted are the resolution values in nm of the spectrometer at the given slit width, then converted from wavelength to energy. The ±0.03eV in Figure 135(A) is at the same slit width, and same ±2.5nm (5nm
bandpass), as the micrographs of (B)-(E). The ±0.03eV resolution at 2.25eV (F) is actually with much wider spectrometer slits, and Δλ=±6nm (12nm bandpass). This varying bandpass factor with energy is due to the fact that conversion of wavelength to energy is non-linear; specifically, E=[1239 (eV-nm)]/λ.

The first result to notice from Figure 135 is that in (A), the GaN emission, the contrast effectively mirrors that of the QW peak emission Figure 135(D). This result has two possible explanations: first, that the defects arresting the QW emission and causing contrast at the QW peak energy also affect the GaN peak; second, if the contrast in the QW peak is due to In-content fluctuations, it is possible that absorption is increased by the fluctuations, resulting in less CL signal leaving the sample when it is generated deeper in the film than the QWs, as would be the case for a large fraction of the CL at 20keV. Given the QWs' narrowness, absorption will be very weak, and the first hypothesis is taken as our interpretation.

Now, the case of cross-sectional CL analysis will be examined. As mentioned above, this requires the polishing of wafer or LED samples to a fine diamond grit (0.05μm) to provide a clean, smooth face for observation. Cleavage sample preparation was not used due to the rough surface topography produced by the cleavage of GaN.

Figure 136 shows a SE image and a panchromatic CL image taken of a wafer of violet LED material.
Figure 136: (A) SE; (B) Panchromatic CL images of violet wafer in cross-section. SEM images; beam conditions 20keV, 1.4nA.

The main feature to notice in Figure 136 is the strong CL emission from the entire thickness of the GaN layer, in addition to the thin bright region near the top of the layer. Figure 137 shows the same area at higher magnification.

Figure 137: (A) SE; (B) Panchromatic CL images of violet wafer cross-section. SEM images; beam conditions 10keV, 230pA.

Figure 137, the bright emission near the top of the wafer is seen much more clearly and sharply.
As discussed in Section 3.2.6, the spatial resolution of beam injection techniques (CL and EBIC) are expected to be dominated by the electron range in the solid in the ~10keV+ range, as is the case here. Analysis of Figure 136(B) and Figure 137(B) yield the following intensity profiles.

Figure 138: Profiles across panchromatic images of cross-sectional wafer sample.

Analysis of these curves indicate approximate FWHM values of the sharp peak of ~0.50μm at 10keV, and ~1.1μm at 20keV; R_{K-O} ranges in GaN at these voltages are ~0.64 and ~2.0μm, respectively, indicating resolution finer than the R_{K-O} range would imply.

Spectra from this sample, taken in plan-view and in cross-section, are presented in Figure 139. Their intensities are normalized separately, as changes in the optical alignment with the shift of sample position will result in changes in collection efficiency, and make absolute changes in intensity misleading.
Figure 139: Plan-view and cross-sectional spectra of violet wafer. Beam conditions: 10keV, 1.4nA.

Most importantly in Figure 139, it can be seen that the PV-spectra shows its ultraviolet peak at ~3.28eV, whereas the cross-section shows a peak at 3.4eV with a 3.28eV shoulder. The 3.4eV emission will be the GaN band-edge, or else a very shallow exciton, perhaps bound to the very shallow Si-donors present in the n-side. The 3.28eV emission is most likely emission from a Mg acceptor-bound exciton; Mg-doping is known to give 3.2-3.3eV range emission\textsuperscript{76}, and in plan-view at 10keV as is the case for Figure 139, almost all of the electron beam energy will be deposited in the heavily-Mg-doped p (top) side.

From the spectra presented in Figure 139, it is possible to select emission bands for spectrally resolved imaging; by imaging with the 3.39, 3.06 and 2.27eV peaks, Figure 140 is generated.
Figure 140: (A) 3.39 ± 0.06 eV; (B) 3.06 ± 0.05 eV; (C) 2.27 ± 0.03 eV; (D) SE + CL composite. SEM images; beam conditions: 10 keV, 1.4 nA.

In Figure 140(D), the SE image and the three spectral bands are presented together in a single image to show the registration in the film thickness between the emissions.

Figure 140(A) shows the GaN emission; like the panchromatic image, it is inhomogeneous in the length of the film, indicating the possible existence of defects such as dislocation cells in the plane of the film. Figure 140(B), the 3.06 eV image is a clear indication of the location of the QWs. Figure 140(C), the YB image, again shows a homogenous emission from the YB, which indicates that the YB does not
seem to be associated with particular locations in the crystal. From Figure 140(D), it appears that the GaN and YB emissions do not reach quite to the top of the film.

5.3.2.3 Quantitative analysis of emission mechanisms

As discussed in Section 3.2.5, current-resolved CL or PL will allow the analysis of the mechanisms associated with particular emissions. In particular, the area of an emission peak $I_{\text{PEAK}}$ will increase proportionally to the beam current $i_b$ at constant $E_0$ as $I_{\text{PEAK}} \propto i_b^m$. Values of $m$ near or greater than 1 indicate excitonic processes, whereas values below 1 indicate free-to-bound processes\textsuperscript{220,221}.

A diamond-polished blue LED wafer sample was examined by current-resolved CL in cross-section; spectra taken at currents of 10 to 10,000pA at 10keV are given in Figure 141.

![Figure 141: 10keV current-resolved spectra from cross-sectional blue wafer.](image)

Data was taken without optical re-alignments, so the relative magnitudes of the CL intensities are directly comparable between curves. Curve intensities at particular beam currents were very repeatable, indicating that neither beam damage nor
optical misalignments occurred with acquisition time, even from large (10,000pA) beam currents, nor with stage shifts to access the Faraday cup for beam-current measurements. In Figure 142, a sub-set of the data from Figure 141 is presented, where the curves were all normalized to the GaN emission peak intensity.

![Figure 142: Re-normalized current-resolved CL data; 10keV, blue wafer cross-section.](image)

In Figure 142, the scaling factors for the curves are listed. As is apparent from Figure 141 and Figure 142, increasing the beam current leads to significant increases in the GaN emission intensity, but much smaller increases in the YB or QW emission intensities. At 10pA, the YB is significantly stronger than the GaN peak; at 10,000pA, the YB is negligible compared to the GaN emission.

Deconvolution of the spectra into Gaussian components yields the data presented in Figure 143 for the YB, QW, and GaN peaks.
Peak deconvolution analysis reveals the presence of two ultraviolet shoulder
peaks just below the GaN emission, at ~3.24eV and ~3.11eV. Reschikov and
Morkoc\textsuperscript{76} and Dingle and Ilegems\textsuperscript{258} attribute these peaks to donor-acceptor-pair
(DAP) emission, although they also suggest that at room temperature a free-to-
bound e-A type transition is more likely. Peak areas vs. current are given in Figure
144 for these two peaks, labeled UV1 and UV2.
Figure 144: Peak areas and $I_{\text{peak}} \propto i_b^m$ dependencies for the UV peaks.

From Figure 143 & Figure 144, the following m-exponent values for the different peaks can be found.

Table XVI: m-values for emission peaks at $E_0=10\text{keV}$

<table>
<thead>
<tr>
<th>Peak</th>
<th>Energy (eV)</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>YB</td>
<td>2.300±0.055</td>
<td>0.32</td>
</tr>
<tr>
<td>QW</td>
<td>2.757±0.049</td>
<td>0.83</td>
</tr>
<tr>
<td>UV1</td>
<td>3.114±0.058</td>
<td>1.5</td>
</tr>
<tr>
<td>UV2</td>
<td>3.236±0.017</td>
<td>1.5</td>
</tr>
<tr>
<td>GaN</td>
<td>3.380±0.008</td>
<td>0.97</td>
</tr>
</tbody>
</table>

Similar experiments were performed at 5keV, but for a smaller range of $i_b$ values. This was necessitated by the longer acquisition times associated with the weaker 5keV signal levels. The data is shown in Figure 145, Figure 146, Figure 147, and Table XVII.
Figure 145: 5keV current-resolved CL data.

Figure 146: Re-normalized 5keV current-resolved CL data.
Figure 147: Peak areas and $I_{\text{PEAK}} \propto i_b^m$ dependencies.

Table XVII: m-values for emission peaks at $E_0=5$keV

<table>
<thead>
<tr>
<th>Peak</th>
<th>Energy (eV)</th>
<th>m</th>
</tr>
</thead>
<tbody>
<tr>
<td>YB</td>
<td>2.243±0.045</td>
<td>0.25</td>
</tr>
<tr>
<td>QW</td>
<td>2.731±0.007</td>
<td>0.81</td>
</tr>
<tr>
<td>GaN</td>
<td>3.382±0.010</td>
<td>1.02</td>
</tr>
</tbody>
</table>

From the value of $m \approx 1$, using the rules of Taguchi et al.\textsuperscript{220} and Schmidt et al.\textsuperscript{221} (Section 3.2.5), the GaN peak appears to be of excitonic character. Similarly, the $m \approx 1.5$ values for the UV1 and UV2 peaks indicate an exciton-type or free-carrier transition. The low value of $m$ for the YB, $m \approx 0.3$, shows a free-to-bound type behavior.

These results are consistent with theory, in particular the YB theory of a deep acceptor-like state being responsible for the emission. The YB intensity increases much more slowly than the others; for example, as seen in Figure 146, at 5kV and 100pA, the ratio of the YB to GaN peak areas is ~26; at 10,000pA, the ratio is ~0.7.
This further indicates that increasing the excitation density causes the YB emission sites to saturate, and the EHPs then recombine using other radiative mechanisms, such as the QW or YB.

An initially surprising result is the relatively low value of the QW m-value. However, as the rules for the assignment of emission processes based upon m-values were derived for 3D bulk semiconductors, not 2D QWs, it is not surprising that an m-value outside of the expected range would be acquired.

It can be concluded from the low m-value for the YB that it is indeed a defect-mediated process, and that the appearance of the YB at low CL, PL, or EL excitation densities and its disappearance at higher excitations are all consistent results.

5.3.3 CL and EBIC: Conclusions

Cathodoluminescence and electron beam induced current in plan-view are powerful probes of defect behavior for semiconductor devices such as LEDs. Combining CL and PV-EBIC allows both optical and electronic behavior of defects to be analyzed, and has been used in this work to demonstrate the high density of threading defects, inferred to be dislocations, that cause enhanced minority carrier recombination and reduce light emission and EBIC current.

CL spectroscopy, on cross section or plan-view samples, allows quantitative evaluation of light-emission mechanisms in semiconductors. Here, voltage-resolved CL spectroscopy has been used to show that YB defect has inhomogeneous depth-distribution; spectroscopic imaging has shown a near-homogenous in-plane YB emission but that QW and GaN emissions are strongly inhomogeneous in their in-plane distribution, indicating that defects locally arrest these emissions. Current-resolved CL has shown that the YB is a free-to-bound transition, which is consistent with the theory\textsuperscript{76} that it arises from a conduction band to deep acceptor transition. The GaN and DAP peaks were shown to be excitonic or free-carrier type in character.
5.4 CL and EBIC of samples cross-sectioned by FIB

As discussed in the sample preparation section above (Section 5.2.1), mechanical grinding and FIB milling can be used separately or in conjunction to prepare cross-sectional samples for CL and EBIC analysis. Additionally, as shown in Section 5.2, the presence of a surface-damaged-layer limits the usable $E_0$ values from which EBIC can be obtained. Further, as discussed in detail by Bunker\textsuperscript{172}, very-high-resolution CL and EBIC will require the use of a STEM instrument, instead of SEM. Preparation of thin samples for CL could be performed using argon-ion-milling, but preparation of electrically live device samples for STEM will require FIB milling (as shown by Bunker et al.\textsuperscript{259}); preparation of CL samples from devices, rather than wafer material, will also require FIB milling.

Unfortunately, the effect of FIB and its associated damage mechanisms on the CL and EBIC response of GaN has not been studied. Indeed, a literature search on the topic returned no publications.

Here, a quantitative study of the behavior of CL and EBIC on mechanically prepared vs. FIB-prepared cross-sections will be discussed.

5.4.1 CL of FIB polished areas

Cross-sectional mechanical polishing followed by FIB-milling of GaN-based LED wafers was performed, on a number of different samples with different QW emission energies. These samples were studied with both voltage- and current-resolved CL. All the different samples showed the same behavior; one particular example (a violet emission wafer) will be discussed here. Figure 148(A) shows a micrograph of a cross-sectionally polished wafer of the violet LED material, into which a window has been milled with FIB.

The FIB window has two sub-areas in it, one which was polished with a low-current 40keV beam, and the other a low-current 10keV beam. In principle, the area of 10keV polish is expected to have a thinner surface damaged layer and less surface damage. \textit{A priori}, it is not known how the FIB polished-vs.-mechanically polished damaged layer thicknesses will compare, although in light of the EBIC
results above which indicate stronger EBIC signals from FIBed areas, the FIB polished surface would be expected to give better CL response.

In actuality, the mechanically polished surfaces were found to give a much stronger CL signal than the FIBed areas. Figure 148(B) shows a panchromatic CL image of the area in (A), and it is clear from the micrograph that the FIBed areas have significantly arrested CL emission compared to the mechanically polished.

Figure 148: (A) SE and (B) Panchromatic CL images of FIBed cross-section. SEM images; 10keV, $\approx 200\mu\text{A}$.

Additionally, it can be seen that the 10keV FIB window has somewhat brighter CL emission than the 40keV FIB window. In order to examine the surface and near-subsurface of a sample with CL, it is necessary to depth-profile by performing voltage resolved CL. Data for the mechanically prepared area indicated by the "Polished" label and arrow in Figure 148(A) is given in Figure 149.
In this dataset, the 5keV spectrum was acquired with 5,000pA beam current; the beam power was kept constant as the voltage changed, which resulted in lower beam currents with increasing beam energy, resulting in a current of 1,250pA at 20keV.

Data for the 40keV-FIB and 10keV-FIB polished areas are presented in Figure 150 and Figure 151.
Figure 150: Voltage resolved CL data for 40keV FIB-milled area.
Figure 151: Voltage-resolved CL data for 10keV FIB-polished area.

Data comparing the three surface conditions are 5keV and 20keV are shown in Figure 152 and Figure 153.
Figure 152: 5keV, 5.00nA CL data for the three surface conditions.
Figure 153: 20keV, 1.25nA CL data for the three surface conditions.

It can be seen from the above graphs that:

1. The mechanically polished areas have stronger signal than the FIB polished areas, particularly for the band-edge (3.4eV) and QW (3.1eV) emissions, but less so in the YB emission (2.3eV).
2. The 10keV FIB-polished area has significantly stronger emissions than the 40keV FIB-milled area, again particularly in the band-edge and QW emissions.
3. For all three surface conditions, deeper injection of the EHPs (higher keV CL) results in a significantly stronger signal, indicating that the surface suppresses CL emission.

Figure 154 shows images of this structure. At 3.54eV is the AlGaN buffer emission. The GaN band-edge or shallow exciton emission is at 3.37eV. At 3.17eV is both the QW emission (as indicated by the high-intensity vertical structure in the
image) and an overlap with a near-UV line, perhaps a DAP emission. At 2.25eV is the yellow band emission.
Figure 154: Images of FIBed structure: SE image; buffer layer (3.54eV); GaN band edge (3.37eV); QW and from impurity-bound-exciton (overlapping emission energies, 3.17eV); Yellow-band (2.25eV). All are $\pm \approx 30$meV bandpass.
It is clear from Figure 154 that all of the emission peaks are suppressed by the FIB milling. In order to examine the physics of the emission mechanisms, and if they changed with FIB milling, current-resolved CL was performed. 10, 100, 1,000, and 10,000pA spectra were acquired at 10keV for each of the three areas. 10 and 10,000pA spectra are shown in Figure 155 and Figure 156.

Figure 155: 10pA, 10keV spectra for the three surface conditions.
Figure 156: 10,000pA, 10keV spectra for the three surface conditions.

It can clearly be seen that, regardless of excitation, the YB emissions are roughly the same, but that the QW and band-edge emissions are significantly suppressed by FIB milling. Further, the 10keV FIB milling is much less detrimental than the 40keV FIB milling.

The $m$-exponent values were extracted for the three peaks in each of the three areas.
Although the m-value for the YB decreases slightly with the aggressiveness of the sample preparation (mechanical to 10keV FIB to 40keV FIB), its actual change is minor (~20%). The differences in the QW and band-edge m-values are also small and appear to be noise-limited. As such, it seems that, at least to a first approximation, the emission processes are not changed by the FIB milling.

From the above CL experiments, it seems that two hypotheses can be proposed to explain the loss of CL intensity with FIB milling.

1. FIB milling increases the thickness of the surface damaged layer, which reduces the number of EHPs that enter the bulk and can then radiatively recombine.
2. FIB milling introduces non-radiative recombination sites into the structure that quench the CL intensity and reduce the number of minority carriers available for radiative recombination. Of course, a combination of (1) and (2) are also possible. The surface damaged layer, as discussed by Wittry and Kyser\textsuperscript{260} and Abrams and Holloway\textsuperscript{261}, is a thin region of disrupted lattice directly on the surface of a semiconductor. EHPs that are injected into this area will immediately suffer non-radiative recombination and will neither diffuse nor contribute to a CL or EBIC signal.

It does not appear possible to distinguish between hypotheses (1) or (2) simply using cathodoluminescence; as such, X-EBIC was used to try to distinguish between surface damaged layer and non-radiative-recombination effects.

5.4.2 X-EBIC of FIB polished areas

Figure 158 shows a blue LED that had been cross-sectioned mechanically, and then had small 10keV and 40keV windows FIBed into it, to provide areas of FIB-prepared surface to compare to the mechanically prepared surface. This is the same sample used in the example cross-sectional analyses presented in Section 5.2; the data presented in Section 5.2 is used as the mechanically-cross-sectioned control experiment to compare to the FIBed cross-section experiments in this section. It was shown above as Figure 94.

SE and EBIC images of the FIBed area are shown in Figure 158.
X-EBIC of this samples shows significant differences between the mechanically and FIB polished surfaces. At 4keV, the expected ratio of the peak EBIC current to the beam current will be 328. For the mechanically polished area, the actual ratio was found to be ~4 (Table IV). For the 40keV FIBed area, this ratio was ~7, and for the 10keV FIBed area, ~29. The conditions are listed in Table XVIII.
<table>
<thead>
<tr>
<th>E₀ (keV)</th>
<th>Surface</th>
<th>i₀ (pA)</th>
<th>iₚ (pA)</th>
<th>iₚ/i₀ (exp.)</th>
<th>iₚ/i₀ (pred.)</th>
<th>Collection</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Mechanical</td>
<td>96</td>
<td>0</td>
<td>0</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>3</td>
<td>40keV FIB</td>
<td>135</td>
<td>0</td>
<td>0</td>
<td>246</td>
<td>0%</td>
</tr>
<tr>
<td>3</td>
<td>10keV FIB</td>
<td>135</td>
<td>1000</td>
<td>7</td>
<td>3%</td>
<td>3%</td>
</tr>
<tr>
<td>4</td>
<td>Mechanical</td>
<td>140</td>
<td>500</td>
<td>4</td>
<td>1%</td>
<td>1%</td>
</tr>
<tr>
<td>4</td>
<td>40keV FIB</td>
<td>138</td>
<td>2000</td>
<td>14</td>
<td>328</td>
<td>4%</td>
</tr>
<tr>
<td>4</td>
<td>10keV FIB</td>
<td>138</td>
<td>4000</td>
<td>29</td>
<td>328</td>
<td>9%</td>
</tr>
<tr>
<td>5</td>
<td>Mechanical</td>
<td>135</td>
<td>1000</td>
<td>7</td>
<td>2%</td>
<td>2%</td>
</tr>
<tr>
<td>5</td>
<td>40keV FIB</td>
<td>111</td>
<td>3000</td>
<td>27</td>
<td>410</td>
<td>7%</td>
</tr>
<tr>
<td>5</td>
<td>10keV FIB</td>
<td>111</td>
<td>6000</td>
<td>54</td>
<td>492</td>
<td>13%</td>
</tr>
<tr>
<td>6</td>
<td>Mechanical</td>
<td>190</td>
<td>3000</td>
<td>15</td>
<td>3%</td>
<td>3%</td>
</tr>
<tr>
<td>6</td>
<td>40keV FIB</td>
<td>88</td>
<td>5000</td>
<td>57</td>
<td>492</td>
<td>12%</td>
</tr>
<tr>
<td>6</td>
<td>10keV FIB</td>
<td>88</td>
<td>10000</td>
<td>114</td>
<td>492</td>
<td>23%</td>
</tr>
<tr>
<td>7</td>
<td>Mechanical</td>
<td>100</td>
<td>2500</td>
<td>25</td>
<td>4%</td>
<td>4%</td>
</tr>
<tr>
<td>7</td>
<td>40keV FIB</td>
<td>101</td>
<td>7000</td>
<td>69</td>
<td>574</td>
<td>12%</td>
</tr>
<tr>
<td>7</td>
<td>10keV FIB</td>
<td>101</td>
<td>10000</td>
<td>99</td>
<td>574</td>
<td>17%</td>
</tr>
</tbody>
</table>

Here, E₀ and i₀ have the normal meanings. The value iₚ is the average of the peak EBIC reading from the oscilloscope monitoring the EBIC amplifier's output. iₚ/i₀ (exp.) is again the experimental value of iₚ/i₀; iₚ/i₀ (pred.) is the predicted value, calculated as the number of EHPs that should be produced for each incident beam electron; this is E₀(1-<n>)/E₁₀H, or equivalently, Equation (3-9) multiplied by q/i₀. Collection is the value iₚ/i₀ (exp.) divided by iₚ/i₀ (pred.), which is what fraction of the nominal number of generated EHPs that are actually collected by the junction.

The error in the i₀ values in Table XVIII is estimated at less than ±2%, as the beam currents were measured by focusing the SEM beam into the aperture of a
Faraday cup and measuring the absorbed current from the stage to ground with a Keithley 614 electrometer; the error in $i_b$ will essentially the instability of the field-emission electron source.

The errors in $i_{\text{PEAK}}$, and therefore in $i_{\text{PEAK}}/i_b$ (exp.) and the collection percentage, are estimated at ±25%. As seen in Figure 101 or Figure 158, the X-EBIC contrast is inhomogeneous and varies within the plane of the film. Although the $i_{\text{PEAK}}$ along any single horizontal linescan could be measured to within a few percent of accuracy, the inhomogeneity in the sample itself leads to the $\approx\pm25\%$ error in $i_{\text{PEAK}}$. Despite these errors, it can clearly be seen in Table XVIII that:

1. The 10keV FIBed areas have the highest collection efficiency at any given EBIC beam energy, followed by the 40keV FIB and mechanically polished conditions.
2. From 3 to 6keV, the collection percent increases at any given surface condition; at 7keV, the collection takes a slight drop compared to 6keV.

Point (1) can be described in terms of the FIB milling reducing the thickness of the surface damaged layer, which allows more EHPs to enter the bulk semiconductor and therefore contribute to the diffusion and collection processes. Further, the 10keV FIB (as expected) provides a thinner dead-layer than the 40keV FIB.

Point (2) can be explained by the higher-energy SEM beams penetrating deeper into the sample and depositing more of their energy, and thus EHPs, beyond the damaged layer. The sudden but slight drop in collection at 7keV is probably due to the broadening of the interaction volume, which would result in the creation of a greater fraction of the EHPs at distances sufficiently far from the junction to prevent their diffusion to the depletion region.

An interesting experiment would be to perform these same experiments, but with a reverse bias applied across the sample in order to widen the depletion region and drift the injected minority carriers toward the junction, thus removing the diffusion aspect from the collection problem. However, this experiment is not feasible. As mentioned in Section 5.2.1, FIB-preparation of the LED results in the creation of a shunt-resistance which pushes the reverse-bias and pre-turn-on
forward-bias sections of the I-V curve to higher currents (Figure 96). Even with careful FIB cleaning, a shunt resistance $\sim 1 \Omega$ to $1 \Omega$ would be expected. A $1 \Omega$ reverse bias would be expected to produce a current of $1,000$ to $1,000,000 \text{pA}$ in this situation, which would swamp the $\sim 100$ to $1,000 \text{pA}$ EBIC signals. It is possible that improved FIB preparation procedures that result in near-zero increases to shunt resistance would make the above reverse-bias experiment possible.

Regardless, it seems clear from the EBIC data that the FIB milling and polishing results in a reduction in the surface dead-layer thickness. This is opposite to hypothesis (1) above, discussed in order to explain the reduction of CL intensity with FIB milling. There, it seems that the only way to explain the reduction of CL intensity, despite the reduction of the dead-layer thickness, is in terms of hypothesis (2), which stated that FIB must introduce near-surface non-radiative recombination sites.

The question then becomes, "how can we determine if non-radiative recombination sites have in fact been introduced by the FIB?"

Equation (3-16) is re-written here as (5-39)

$$\tau_j = \frac{1}{R_j} = \frac{1}{\sigma_j N_j v_{th}} \quad (5-39)$$

In which case $j$ is some recombination process, $\tau_j$ the lifetime of carriers associated with that process, $R_j$ the recombination rate, $\sigma_j$ is the associated cross-section of the $j$-defect-sites, $N_j$ their density, and $v_{th}$ the thermal velocity of the carriers.

Equation (3-17), for the total radiative lifetime $\tau_{RAD}$ resulting from radiative recombination processes 1, 2, 3, ... $m$ is re-written as (5-40):  

$$\frac{1}{\tau^{RAD}} = R^{RAD} = \frac{1}{\tau_1} + \frac{1}{\tau_2} + \frac{1}{\tau_3} + \cdots + \frac{1}{\tau_m} \quad (5-40)$$

Similarly, Equation (3-18) for the non-radiative lifetime $\tau_{NR}$ is re-written here as (5-41):

$$\frac{1}{\tau^{NR}} = R^{NR} = \frac{1}{\tau_1} + \frac{1}{\tau_2} + \frac{1}{\tau_3} + \cdots + \frac{1}{\tau_n} \quad (5-41)$$
From these, it is possible to find the total lifetime $\tau$:

$$\frac{1}{\tau} = R = \frac{1}{\tau_{RAD}} + \frac{1}{\tau_{NR}} = R^{RAD} + R^{NR} \quad (5-42)$$

Further, under low-injection conditions, the minority carrier diffusion length $L$ is:

$$L = \sqrt{D \tau} \quad (5-43)$$

Lastly, CL intensity $I_{CL}$ will be a strong function of:

$$I_{CL} \propto \frac{\tau_{NR}}{\tau_{RAD}} \quad (5-44)$$

Which is to say that strong CL requires that $\tau_{RAD} << \tau_{NR}$. This means that radiative channels use most of the available excess minority carriers, as opposed to non-radiative channels. This was discussed in Section 3.2.5.

From these mathematical relations, it is clear that increasing the number of sites associated with some non-radiative recombination process $j$ will result in a reduction of the non-radiative lifetime $\tau_j$. As a result, the overall non-radiative lifetime $\tau_{NR}$ will be reduced. Further, the overall minority carrier lifetime $\tau$ will thus be reduced. Lastly, and most importantly, the minority carrier diffusion length $L$ will be reduced. Because the $\tau$ values are calculated by adding in inverse, it is clear that the shortest $\tau_j$ value will dominate the overall $\tau$ value. Thus, introducing enough non-radiative recombination sites will easily result in a significant reduction in $L$. Changes in $L$ with FIB processing will indicate the presence of non-radiative recombination sites introduced by FIB.

In light of the X-EBIC results presented in Section 5.2, it is clear that we have the tools to measure the minority carrier diffusion lengths in the FIB polished areas. These values for the mechanically polished areas were found in Section 5.2, and presented in Table XIV, which indicated $L_h=92\pm21\text{nm}$, $L_e=42\pm6\text{nm}$, and $s<10^2\mu\text{m}^{-1}$.

Figure 159 shows the 10keV and 40keV FIB prepared areas of the sample, imaged in X-EBIC at 4keV and 138pA.
These were split in four 100nm tall analyses areas, similarly to that of the mechanically polished analysis in Section 5.2. Example analyses are shown in Figure 160:
Figure 160: Bonard and Ganière fits to EBIC data from FIB-milled areas. X-EBIC at 4keV, 138pA. (A) 10keV FIB; (B) 40keV FIB.
Obviously, the n-side FIB-prepared areas show a diffuse background level of current. Because the samples were n-side grounded, any sort of specimen current effect on the n-side will be suppressed. It is thus possible that the diffuse background arises from implantation of Ga ion from the FIB beam into the near-surface GaN and possibly shows a surface depletion region which collects a small fraction of carriers in the deep n-side, where carriers would otherwise recombine before reaching the bulk p-n junction. The diffuse background effect becomes less severe with increasing EBIC beam energy, indicating that whatever the cause, it is a near-surface process. However, the exact mechanism is not yet known.

This diffuse background in the FIB-prepared areas has a significant effect on the n-side behavior of the X-EBIC curves, particularly in the 40keV FIB case (Figure 160(B)). However, the major advantage of the Bonard and Ganière technique, as mentioned in Section 5.2, is that it is fits directly to the near-junction region. In this region, the signal arising from the direct collection at the junction dominates over the diffuse background. Although the diffuse background will introduce some amount of bias into the diffusion-length calculated, the fact that the theoretical curve is fit to the peak should minimize this effect. The values in Table XIX have been determined by averaging the Bonard and Ganière derived L and s values over four separate analyses areas for each surface condition (10 and 40keV FIB); the mechanically polished data is repeated from Table XIV.

Table XIX: L and s values measured at 4keV for the three different surface preparation techniques.

<table>
<thead>
<tr>
<th>Surface</th>
<th>L_h</th>
<th>L_e</th>
<th>s</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>nm</td>
<td>nm</td>
<td>μm⁻¹</td>
</tr>
<tr>
<td>Mechanically polished</td>
<td>94±21</td>
<td>42±6</td>
<td>&lt;10⁻²</td>
</tr>
<tr>
<td>10keV FIB polish</td>
<td>65±9</td>
<td>49±7</td>
<td>&lt;10⁻²</td>
</tr>
<tr>
<td>40keV FIB polish</td>
<td>58±3</td>
<td>48±6</td>
<td>&lt;10⁻²</td>
</tr>
</tbody>
</table>
The \( L_h \) values show the expected trend, and the \( L_e \) values are approximately constant. However, as discussed in detail in Section 5.2, the \( L_e \) values are estimates at best, due to the presence of the metal contact at \( x \approx 250 \text{nm} \). Clearly, however, the mechanically polished sample areas have longer hole diffusion lengths \( L_h \) in the n-side than the FIB-polished areas. The 10keV FIB-polished area has longer \( L_h \) values than the 40keV FIBed areas. Surface recombination velocities are all very low.

The apparent increase in \( L_e \) with FIB is believed to be an artifact; the FIB sample preparation required the deposition of a protective tungsten cap on top of the p-side ohmic contact to prevent the beam from sputtering away the sample area of interest. When the SEM beam tails begin to impinge on the tungsten cap, the beam electrons should suffer significant scattering, which will scatter some beam electrons back into the semiconductor, resulting in an increase in the apparent EBIC signal level.

Via equations (5-39) to (5-44), it is clear that this reduction in \( L \) likely results from an increase in concentration of some generic defect site \( j \).

5.4.3 Interpretation of CL and EBIC data of FIBed areas

The important experimental observations will be reiterated:

1. Areas polished mechanically have stronger CL signal than areas polished with 10keV FIB, which in turn have stronger CL signal the areas polished with 40keV FIB.
2. The emission mechanisms (quantified by m-exponents) do not noticeably change with the surface preparation type.
3. X-EBIC signals (quantified by collection %) are remarkably stronger in FIB-polished areas than mechanically polished areas; a 10keV FIB polish shows stronger EBIC than a 40keV FIB polish.
4. The minority carrier diffusion lengths are measurably shorter in the FIB polished areas than the mechanically polished areas.
5. These changes are more pronounced at low-keV SEM beam energies, implying that these are near-surface effects. These experimental points are consistent with the following interpretations:

1. The FIB milling reduces the surface damaged layer thickness, increasing the number of EHPs injected into the bulk semiconductor. This gives the increase in collection % in X-EBIC.

2. The FIB milling introduces non-radiative recombination sites that reduce both L and the CL intensity.

This is illustrated schematically in Figure 161.

Figure 161: Conceptual model that explains the CL and EBIC data on the differently prepared surfaces. NR=non-radiative recombination.
From this model, several points can be concluded:

1. FIB preparation of GaN samples for CL or EBIC will quantitatively change the minority carrier properties measured by CL or EBIC.

2. Use of FIB to prepare GaN samples for STEM examination will result in major changes in the sample response to electron beam excitation. Because a STEM sample is essentially all surface, the fact that the above-studied FIB-induced changes are particularly strong near the surface means that these effects will be stronger in a STEM sample than an SEM sample.

This of course begs the question, "how can this damage be ameliorated?" Lower-keV FIB polishing is one strong possibility; the ion range in a solid decreases quickly with beam energy, much the way the electron range decreases with energy. The FIB used in this work can be operated at 10keV; other ion columns typically have ultimate beam energies of 5keV, and the most modern instruments are capable of 2keV or 500eV ultimate energies. SRIM\textsuperscript{262} Monte Carlo simulations of the ion ranges of Ga\textsuperscript{+} beams with GaN substrate for 0.5, 2, 5, 10 and 40 keV are shown in Figure 162.

![Figure 162: Simulated interaction ranges of Ga\textsuperscript{+} ions of different energies with GaN. All are plotted on the same scale.](image-url)
As discussed in Section 3.4, FIB can also be used in conjunction with reactive gas species to chemically etch the surface, greatly enhancing the sputtering process. Thus, chemically-assisted low-keV FIB etching would be another good possibility for removing the damaged surface layers while introducing minimal new damage.

Electron-beam chemically-assisted etching is also possible, and because an electron beam does not cause surface amorphization or implantation like an ion beam, e-beam etching would be extremely low damage. However, electrons do not remove material as effectively as ions, and this would be a comparatively slow process. Hence, combined chemically-assisted FIB and chemically-assisted e-beam etching in a dual-column type instrument would possibly be the ideal solution.

Milling with a broad argon ion beam is used for TEM sample preparation; at low Ar⁺ energies and shallow incident angles, this produces a very thin damage layer, on the order of monolayers. If the ohmic contact could be masked to prevent sputtering it away, argon-ion-milling would also allow low-damage sample preparation. KOH etching is used to remove surface damage during TEM sample preparation of GaN, and could potentially be used to remove the surface damage induced by FIB or diamond polishing for CL and EBIC analysis.

Additionally, the conclusion is reached that combining quantitative CL and EBIC techniques is a powerful technique for elucidating minority carrier properties and the influences of defects on minority carriers.

5.5 TEM analysis of defects

This work has demonstrated proof-of-concept for combined beam-injection CL/EBIC microscopy of defect optoelectronic behavior and transmission electron microscopy of those same defect's crystallography, as facilitated by FIB.

First, defects can be mapped by either CL or EBIC imaging in plan-view in the SEM, as discussed in Section 5.3. Figure 163 shows a SE and EBIC images of an LED device; one particular dark-spot defect has been indicated on the EBIC image:
Section 3.4 discussed how FIB lift-out can be used to prepare TEM samples with good site-specificity; it has been demonstrated that combining EBIC, as in this case, or CL in principle, with FIB lift-out allows site specific sample preparation of TEM samples. Figure 164 shows the lift-out procedure for the defect marked in Figure 163(B).
Figure 164: (A) Tungsten cap deposited over site of interest and rough FIB cuts; (B) Tilting to cut sample almost free; (C) Attach needle; (D) Sample cut free and lifted out. SIM images, 40keV, <100pA.

After the sample is lifted out, it is attached to a copper or molybdenum TEM grid and thinned to 200-300nm for TEM examination. The defect marked in Figure 165(B) was prepared similarly, and is shown in TEM in Figure 166.
Figure 165: (A) SE; (B) EBIC images of LED. A defect is marked with arrow in (B). 10keV, 40pA.
Figure 166: (A) Bright field; (B) Centered dark field; (C) Zone-axis SADP; (D) Two-beam SADP; (E) Computer simulated SADP. TEM data, Topcon 002B instrument, 200keV.

In Figure 166(A), a bright field (BF) two-beam-condition image is shown; Figure 166(B) shows the associated centered dark field (CDF) image. The sample was near the (101) zone axis, as indicated by the selected area diffraction pattern in Figure 166(C). Images (A) and (B) were taken in a two-beam condition, titled slightly...
off of the (101) zone. Figure 166(D) indicates that the two-beam condition was (10 \overline{1}). Figure 166(E) is a simulation, using the software of Zuo and Mabon\textsuperscript{13}, which confirms the zone-axis and diffraction vector identities.

It can be seen in Figure 166(A) and (B) that the sample contains large numbers of threading dislocations, as is expected both from previously published works and the CL and EBIC images presented in this work. The dislocation density is estimated via TEM at \( \sim 10^9 \) to \( 10^{10} \) cm\(^{-2} \). Also visible in the BF and CDF images are the active regions, made up of multiple QWs and spacer layers. The dislocations clearly thread the active region, indicating that electronically active dislocations will affect the optoelectronic behavior of the devices, which confirms the results reported above. Visible in the CDF image (Figure 166(B)) is a mottled surface contrast; this is probably an artifact of the FIB thinning procedure, and indicates that FIB preparation of GaN samples, even for standard TEM imaging, much less CL or EBIC in STEM, will require careful optimization of parameters.

No major defects other than the dislocations are seen. Although more TEM of more samples would be needed to conclude that no extended defects are responsible for the "large-dark-spot" type contrast visible in Figure 163(B) or Figure 165(B), a tentative hypothesis would be that point defects, perhaps long-wavelength variations in In composition, are responsible.

Thus, this section has shown that combining SEM-CL/EBIC with FIB and TEM is both feasible and useful for analyzing defects both optoelectronically and crystallographically.
6. Discussion

A number of different techniques have been used in combination to attempt to explain the optoelectronic behavior of GaN-based LEDs; specifically, electroluminescence, quantum mechanical modeling of polarization fields, cathodoluminescence in cross-section and plan view, electron beam induced current in cross-section and plan view, focused ion beam microscopy and milling, and transmission electron microscopy.

The combination of EL, CL and computer modeling provided insight into the effects of polarization fields that arise from the strain in the QWs. It was concluded in Section 5.1.7 that polarization fields, dominated by piezoelectricity, are clearly present in the green-range LED devices. The effects in the blue-range devices are far milder, but blueshifts with EL excitation density were observed. Although simple blueshift with excitation density can be explained by either the indium-segregation model or the polarization field model, the blueshift-then-redshift with reverse bias can only be explained by the polarization model and its concomitant QCSE. As such, it is clear that polarization fields are present in the green LED samples, and are likely present, in much smaller magnitude, in the blue LED samples. However, the presence or absence of In-segregation can neither be ruled in nor out by these experiments. The broad low-energy tail of the QW emission in CL or EL implies the presence of In-segregation and concomitant band-tail states, but is not conclusive in and of itself.

Cross-sectional EBIC is particularly valuable for the analysis of minority carrier diffusion length. It has been shown that L values in these materials are significantly below 100nm. The best way to explain the anomalously high light-emission efficiencies of these LED devices, despite dislocation densities $\rho \sim 10^8 - 10^{10}$ cm$^{-2}$, is in terms of the short minority carrier diffusion lengths. For GaAs or similar materials, a dislocation density $\sim 10^5$ cm$^{-2}$ will usually quench luminescence. If a simple model of dislocation arrangement is taken as a regular square array, the average spacing between dislocations $d \sim \rho^{-1/2}$. Thus, for $10^6$ cm$^{-2}$, the average spacing is some tens of micrometers, but at $\rho = 10^{10}$ cm$^{-2}$, then $d \sim 100$nm. Of course,
dislocations in GaN layers are not evenly distributed, but rather clustered as subgrain boundaries.

Minority carrier diffusion lengths in device-grade silicon or III-V materials can reach many micrometers, indicating that even with a dislocation spacing on the same order, dislocations could deplete the minority carrier population in those devices. In the III-nitrides, however, despite dislocation spacing of only hundreds of nanometers, the minority carrier diffusion lengths are below 100nm, thus indicating that the areas depleted of minority carriers by the dislocations will not encompass the entire sample. Thus, measuring $L$ via X-EBIC and $\rho$ via CL, EBIC, or TEM, allows a coherent explanation for the surprisingly high light-emission efficiencies in consideration of the large dislocation densities: the carriers can, with sufficient frequency to provide good light emission, recombine radiatively before reaching a dislocation. Phillips et al.\textsuperscript{223,263} obtained very-high-resolution plan-view CL images of dislocated GaN films, and experimentally demonstrated highly dislocated regions around grain boundaries and high-quality grain interiors. Plan view CL and EBIC imaging in this work (i.e., Figure 129 and Figure 130) further confirmed that the dislocations are clustered, rather than arranged in a regular array. This leaves areas of low dislocation density and areas of high dislocation density; the low-density areas likely contribute most of the light emission in these devices. The most recent publication dealing with the anomalously high quantum efficiency in III-nitride systems is that of Chichibu et al.\textsuperscript{257} In their paper, they suggest that statistical clusters of a few ($<<10$) In atoms, referred to as "atomic condensates," rather than relatively larger quantum dots arising from InN-GaN solid-phase immiscibility, are responsible for the high internal quantum efficiency in the presence of large dislocation densities. This argument derives from very short measured positron diffusion lengths ($L_+$) in In-bearing (In,Ga,Al)N layers. Thus, they suggest that holes are captured by the In-clusters, and that this hole localization then results in exciton localization away from non-radiative sites such as dislocations. This confinement, they suggest, provides for the high quantum efficiency. Although the techniques discussed in the present work will not be sensitive to three-atom-sized clusters, the
low diffusion lengths measured are consistent with the model proposed by Chichibu et al., and is also consistent with the solid-phase immiscibility hypothesis.

Cross-sectional EBIC is the most robust means by which to measure minority carrier diffusion length. However, it is not a technique that can be applied to GaN-based devices without significant care. The literature base regarding X-EBIC was primarily developed in the late 1970's to early 1980's, and can be applied rather directly to materials that have long minority carrier diffusion lengths, such as is often the case with silicon or GaAs. It has been shown in this work that care in the acquisition of the data, and sophisticated computational treatment of the data, are necessary to extract accurate $L$ values from GaN devices. However, these $L$ values, as discussed above, help explain the other properties of the devices, such as the anomalously high quantum efficiency. Further, as GaN-based technology moves from LEDs and laser diodes into other device industries, such as power transistors or diodes, methods to measure diffusion lengths will further grow in importance as processing-structure-properties studies are undertaken in these new applications.

CL spectroscopy is a powerful technique for studying the physics of different emissions in semiconductors. Voltage-resolved CL can be used to depth-profile into a material to observe how different emission sites are located with depth. Figure 133 shows voltage resolved spectra from a GaN LED; it is clear from the spectra that deeper CL penetration (higher $E_0$) results in more YB excitation. Theory holds that the YB is associated with Ga-vacancies, and that these defects should be dominant in $n$-GaN, which will be deeper in the epilayer, which is consistent with the voltage-resolved data. Cross-sectional imaging of the YB emission (Figure 140) confirms that this is, indeed, the case. Thus, the information obtained in this study gives strong support to the theory of $V_{Ga}$-associated defects giving rise to the yellow band defect in GaN. Current resolved CL spectroscopy allows determination of the type of emission associated with each peak; the band-edge and ultraviolet peaks in these materials have been shown to be of excitonic type, and the YB has been found to be free-to-bound defect-type, which lends further weight to the defect-emission data found via voltage-resolved and cross-sectional analysis.
Further, these current- and voltage-resolved techniques can be applied to study the damage caused by FIB. As discussed above, FIB is of interest as a sample preparation tool, and combining X-EBIC with quantitative CL spectroscopy has shown that FIB reduces dead-layer thickness and is therefore tantalizing for use as a sample preparation tool, but also causes lattice damage that reduces the minority carrier lifetime, which means that the FIB preparation procedures must be further refined. As discussed in Section 3.2.6, spatial resolution of CL or EBIC imaging will be approximately limited by the interaction volume of the SEM beam, which will range from tens of nanometers (at very low keV) to several micrometers (at 30keV). However, very-high-resolution CL or EBIC is not possible in SEM, as the signal-to-noise ratio of the SEM-based signals will drop toward zero as the beam energy is reduced below 1-2keV. Thus, nanometer or sub-nanometer CL or EBIC will require scanning transmission electron microscopy, as the thin-membrane case of STEM allows the resolution to be limited to approximately the spot size of the probe. Because site-selective preparation of electron-transparent membranes of LEDs or other semiconductor devices will require FIB milling, understanding of the influence of FIB milling on GaN is a prerequisite to very-high-resolution CL or EBIC studies. The results in this study show that FIB milling will require additional damage removal before it will be suitable for CL or EBIC sample preparation.

Conventional TEM imaging has also been used. Firstly, CL and EBIC were shown to be excellent techniques for pinpointing optoelectronic defects in GaN LEDs that were not visible as defects in secondary electron imaging. FIB microsampling was then found to be ideal for preparing TEM samples of these defected regions. TEM of such a region, as discussed above in Section 5.5, shows a relatively inhomogenous dislocation distribution and a dislocation density \(\sim10^9\) cm\(^{-2}\). PV-EBIC point-counts of the small-dark-spot contrast indicate a density \(\sim10^8\) to \(10^9\) cm\(^{-2}\), which leads to the conclusion that the small-dark-spots are likely threading dislocations. Thus, it is clearly seen again that dislocations are detrimental to the optoelectronic properties of GaN LED devices. An interesting future experiment would be to use larger datasets to get counts with higher statistical accuracy, and to
use in-TEM methods to find the densities of the different dislocation types (screw, edge, mixed) in the films. It is possible that only certain dislocation types are optoelectronically active and therefore imaged by CL/EBIC, and correlation of CL/EBIC point counts with TEM-measured densities may provide insight into the varying dislocation activities.

In short, the techniques utilized in this work have very accurately measured diffusion lengths in GaN, and these measured values can help explain the high light emission efficiency of GaN layers despite high defect densities. The quantum well and GaN emissions were found to be strongly affected by extended defects, but the yellow-band emission was not. However, consistent with the theory of the yellow-band defect arising from Ga-vacancy related sites, it was found that the yellow band defect was inhomogenous, and the yellow band emission was shown by quantitative CL to be defect-related. TEM in conjunction with CL and EBIC have indicated that dislocations are strongly electrically active, which indicates that some explanation, such as short-diffusion-lengths, is necessary to understand the high light emission efficiencies.
7. Conclusions

The following major conclusions have been reached in this work:

1. Polarization fields are present, as expected, in GaN-based devices' quantum wells. An original contribution of this work has been to combine electroluminescence, and its numerical modeling, with cathodoluminescence and modeling. Polarization fields in the green-range GaN-based LEDs studied are estimated at \(-1.33\pm0.15\)MV/cm, which is a refinement over the value of \(-1.0 \pm 0.2\)MV/cm obtained by Bunker\textsuperscript{172}. Qualitative agreement has been found between the EL and CL experiments and models. Thus, it has been shown that polarization fields (primarily piezoelectric) are present in the green-range devices and profoundly affect their optical emission behavior, and the magnitude of the fields have been measured in the green-range devices. The blue-range devices were found to be minimally influenced by the polarization fields and quantum-confined Stark effect.

2. Cross-sectional electron beam induced current in GaN-based devices does not closely resemble X-EBIC in more traditional materials, such as Si or GaAs. This work presented the first comprehensive critical analysis of different literature models of X-EBIC applied to the case of GaN. This is also the first work to experimentally test the various models in comparison to this critical analysis. Further, experimental conditions for obtaining good X-EBIC data of GaN have been developed and discussed. Issues that still need improvement and artifacts that can bias quantitative results have also been enumerated. In the samples studied, \(L_h \approx 92 \pm 15\)nm in the n-side, \(L_e \approx 42 \pm 6\)nm in the p-side, and \(s<10^{-2} \ \mu m^{-1}\) for both electrons and holes. Because these diffusion lengths are very short, especially in comparison to the dislocation spacings, this indicates that a contributing factor to the high light-emission efficiencies of GaN-based LEDs is that many minority carriers or excitons can recombine radiatively before they have time to diffuse to non-radiative defect
sites, indicating that short L-values may actually benefit light emission in these highly defective GaN devices.

3. A robust, low cost, versatile, combined CL and EBIC apparatus for the examination of semiconductors was designed, developed, implemented, and refined in this work. Although CL and EBIC are established experimental techniques, the apparatus developed in this work has a number of definite advantages, in particular the ability to perform simultaneous SE, CL and EBIC imaging. Combined CL and EBIC analysis of GaN is lacking in the literature, and the correlations between CL and EBIC results presented here are original contributions. Combined CL/EBIC is an important advance, as most previous literature on GaN beam-injection microscopy used only one technique or the other. The combination of both optical and electronic assessment has allowed the demonstration of simultaneous loss of quantum efficiency and carrier lifetime around defects, which gives strong support to the theory that dislocations in GaN-based materials are strong recombination sites. Thus, combining CL and EBIC has allowed improved investigation of the recombination behavior of GaN defects.

4. This apparatus has been used to study emission mechanisms and depth distribution of emission sites in GaN devices. Although studies along these lines have been published elsewhere, it was necessary to perform these experiments on the particular materials studied in this work to find their responses in order to draw correlations between structure and properties. First, careful procedures were developed to gather quantitative data and reduce or eliminate multiscan degradation and beam damage of the materials. It has been shown in this work that the yellow-band associated defect is present in much higher densities in the deeper (n-doped) epilayers of the GaN-based devices studied here, giving strong support to the theories that gallium vacancy complexes are responsible for the yellow-band emission, and that formation of this defect will only be energetically favorable in n-doped GaN. The band-edge and near-band-edge emissions have been shown to be
free-carrier or excitonic in nature, and the yellow-band emission was shown to be free-to-bound defect-related in nature.

5. High densities of dislocations and other defects were observed in CL and EBIC imaging in plan-view. By virtue of the fact that these defects are visible in CL and EBIC, they are necessarily optoelectronically active. By combining the high density of dislocations observed with the very short minority carrier diffusion length measured in point (2) above, the anomalously high efficiency of light emission can be explained (Section 6). Thus, it is shown that defects in GaN strongly arrest light emission, and that some circumstance, such as the short minority carrier diffusion lengths discussed in point (2) above, ameliorate the defects' detrimental influence. Further, combined CL and EBIC imaging allow exploration of the defect distributions. Use of CL and EBIC imaging in conjunction with processing-properties studies to determine the structures would allow a full processing-structure-properties materials science relationships to be formulated, and possibly allow elimination or reduction of the defects present in the GaN layers leading to reduced light emission.

6. Spectroscopic CL imaging shows that the band-edge and QW emissions are very inhomogeneous and strongly affected by the structural defects within the films. The yellow band emission is far more homogenous, and showed no correlations to defects, to within the resolution achieved in this work. This indicates that the band-edge and QW emissions are controlled by dislocations or other extended defects, as is expected from current theory. This also indicates that the YB is not associated with these defects. As current theory indicates the YB emission should arise from point defects, this result is, again, consistent. This also indicates that the ~2.3eV defect states observed by valence EELS\textsuperscript{55,56} of dislocations in GaN are not associated with the YB, as has been hypothesized in the EELS literature, which is an important conclusion.
7. Evaluation of the use of FIB as a cross-sectioning technique for optoelectronic beam-injection assessment by CL and EBIC in the SEM has been performed, and previous work of this kind has not appeared in the literature. FIB-prepared cross-sectional surfaces have been found to provide a superior EBIC signal level in comparison to traditional mechanical preparation, but detailed EBIC analysis and CL spectroscopy have shown that FIB actually causes significant disruption of the minority carrier behavior in the near-surface area. As such, it is clear that further advances in FIB polishing methods will be necessary to use FIB to prepare samples for CL or EBIC, whether in the SEM or STEM. This is important as studies of GaN-based nanostructures' optoelectronic behavior will require STEM-based CL or EBIC, and site-selective STEM examination will require FIB sample preparation. Thus, improving the understanding of GaN-based nanostructures requires, as prerequisite, understanding of FIB effects on CL and EBIC.

8. A procedure has been developed and demonstrated for the combined use of SEM-CL/EBIC and TEM for both optoelectronic and crystallographic analysis of individual defects in GaN, or indeed semiconductors in general. This combined SEM-TEM technique is facilitated by the use of focused ion beam sample preparation. Densities of features hypothesized to be dislocations, measured via PV-EBIC, were \( \sim 10^8 - 10^9 \text{ cm}^{-2} \). Densities of dislocations measured via TEM were \( \sim 10^9 \text{ cm}^{-2} \). Thus, it is inferred that the small-dark-spot contrast observed in SEM-CL/EBIC is indeed due to threading dislocations, and it is further confirmed that threading dislocations in GaN are optoelectronically active.

The following thrusts for future work would likely yield valuable results:

1. Continued improvement of the CL/EBIC apparatus. A diamond turned parabolic mirror and stepper-motor-controlled 5-axis mirror positioner would allow much higher alignment accuracy and collection efficiency. A longer
monochromator (i.e., 500 to 1000mm) would allow much greater throughput
at any given bandpass. A monochromator with multiple exit ports could
mount both a PMT and a CCD simultaneously, and give the choice of parallel
or serial spectroscopy without interfering with the spectrally resolved imaging
capability of the PMT. If an SEM could be dedicated to CL microscopy, a
vacuum monochromator could be permanently mated to it, and an elliptical
mirror used for direct optical coupling to the vacuum monochromator. This
would provide the highest possible CL collection efficiency.

2. Continued EBIC apparatus improvement. Using micromanipulated probes
with coaxial shielding to contact the samples *in-situ* would allow for much less
noise pickup in the experiment. Beam blanking in conjunction with lock-in
amplification would also greatly improve signal-to-noise levels at low \( E_0 \) and
low \( i_b \). Chemically assisted FIB-etching should be explored as a means to
remove damage from FIB-cross-sectioned samples. Some means of damage
removal will be mandatory if quantitative STEM-CL or quantitative STEM-
EBIC is to be performed.

3. Correlation of SEM-CL/EBIC with TEM, in order to study defects' crystallography in conjunction with their optoelectronic behavior should
continue. This will, of course, require FIB for site-specific sample preparation.
EDS and EELS in STEM would also allow for chemical analysis of defects.
References

5. H. Morkoc, Nitride Semiconductors and Devices (Berlin: Springer-Verlag 1999)
8. Y. Wu and P. Parikh, Compound Semiconductor, Jan/Feb 2006, P.21


84 V. Fiorentini, F. Bernardini, F. D. Sala, A. D. Carlo, and P. Lugli, Phys. Rev. B. V60(19) P.8849 (1999)

305


159 M. Phillips and D. Drouin, Short Course X-11: Scanning Cathodoluminescence Microscopy and Spectroscopy, presented at Microscopy and Analysis 2006, July-August 2006, Chicago, IL.
169 D. Drouin, A. R. Coutre, R. Gauvin, P. Hovington, P. Horny and H. Demers, CASINO v 2.42 software; www.gel.usherbrooke.ca/casino/
174 J. D. Plummer, M. D. Deal, and P. B. Griffin, Silicon VLSI Technology (Upper Saddle River: Prentice Hall 2000)
175 D. Corson and P. Lorrain, Introduction to Electromagnetic Fields and Waves (San Francisco: W. H Freeman and Company 1962)
176 C. Donolato, Optik, V52(1) P.19 (1978/79)
177 C. Donolato, Appl. Phys. Lett., V34(1) P.80 (1979)
178 C. Donolato, Solid-State Electronics, V22 P.797 (1979)
207 C. van Opdorp, Phillips Research Reports, V32(3), P.192 (1977)
212 W. A. Harrison, Solid-State Theory (New York: Dover 1980)
216 K. J. Kauffmann, Photonics Spectra, July 2000, P.149

312
222 C. E. Norman, Microscopy and Analysis (Americas Edition), Issue 53, P.13, March 2002,
233 M. W. Phaneuf, Micron, V30 P.277 (1999)
243 D. K. Ferry, Quantum Mechanics: an Introduction for Device Physicists and Electrical Engineers (Bristol: IoP Publishing 2001)
246 L. Pauling and E. B. Wilson, Jr., Introduction to Quantum Mechanics (Mineola: Dover 1985)
247 R. Pozo, Template Numerical Toolkit, National Institute of Standards and Technology; math.nist.gov/tnt
250 W. Rasband, National Institutes of Health; http://rsb.info.nih.gov/ij
Appendices
A.1 Appendix 1: MathCAD 12 code for Donolato-Luke curve generation

This simulates EBIC profiles using the model of Donolato, SOLID STATE ELECTRONICS, V25(11), 1982, P1077 and then outputs the profile to a text file.

The important point to remember about the Donolato model is that it's no good for values of \( x < 1/2 \) the electron range, where the origin of \( x \) is the peak of the EBIC curve.

\[
\sigma(R) := \frac{R}{\sqrt{15}} \quad \text{This is a depth function, in microns}
\]

\[
Zo(R) := 0.3R \quad \text{Another depth function, microns}
\]

\[
\lambda(L) := \frac{1}{L} \quad \text{A simple substitution}
\]

\[
\mu(k, L) := \sqrt{k^2 + \lambda(L)^2} \quad \text{Another substitution}
\]

\[
\frac{R(E)}{\mu} = 0.0126 \cdot E^{1.57}
\]

\[
E := 7
\]

The K-O electron range of the beam for the given material

Here is the EBIC function itself:

\[
I(x, L, s, r, w) := \frac{2}{\pi} \int_{0}^{\infty} \frac{k}{\mu(k, L)} \left[ -\frac{k^2 \sigma(R)^2}{2} - \frac{\lambda(L)^2 \sigma(R)^2}{2} - \mu(k, L) \cdot Zo(R) \right] \frac{5}{\mu(k, L) + 5} \]

\[
- \frac{1}{\mu} \left[ \frac{\sigma(R)}{\sqrt{2}} \left( \frac{Zo(R)}{\sigma(R)^2} \right) \right] \sin(k \cdot x) \right] dk
\]

\[
L_{\mu} := 0.15 \quad \text{Minority carrier diffusion length, microns}
\]

\[
s_\sigma := 1 \cdot 10^{-3} \quad \text{Surface recombination, Vs/D, in inverse microns}
\]

\[
\text{increment} := 0.01
\]

\[
r_{\text{step}} := \frac{R(E)}{2} + \text{increment}
\]

These define the \( x \)-values for graphing:

\[
x := \frac{R(E)}{2}, r_{\text{step}}, 4
\]

\[
w := 500
\]
\[
\frac{dI(L, S, R, \omega)}{dx} = \frac{1}{\pi L} \int_0^\infty \frac{k^2 L}{\mu(k, L)^2} \left[ \frac{-k^2 \sigma(R)^2}{e^2} - \frac{\lambda(L)^2 \sigma(R)^2}{2} - \mu(k, L) \cdot \text{Ze}(R) \right] dk
\]

\[
\frac{S}{\mu(k, L) + S} \left[ 1 - \text{erf} \left( \frac{\sigma(R)}{\sqrt{2}} \left( \frac{\mu(k, L) - \text{Ze}(R)}{\sigma(R)} \right) \right) \cos(k \cdot x) \right] dk
\]
\( i \coloneqq 0, 1, \ldots, 40 \)

\[
\text{EBICdat}_{i, 0} \coloneqq (i \cdot \text{increment}) + \frac{R(E)}{2}
\]

\[
\text{EBICdat}_{i, 1} \coloneqq I(\text{EBICdat}_{i, 0}, L, S, R(E), w)
\]

\[
\begin{array}{c|cc}
   & 0 & 1 \\
---&---&---
0  & 0.175 & 0.317 \\
1  & 0.185 & 0.306 \\
2  & 0.195 & 0.294 \\
3  & 0.205 & 0.281 \\
4  & 0.215 & 0.267 \\
5  & 0.225 & 0.254 \\
6  & 0.235 & 0.241 \\
7  & 0.245 & 0.227 \\
8  & 0.255 & 0.214 \\
9  & 0.265 & 0.201 \\
10 & 0.275 & 0.189 \\
\end{array}
\]

\[
\text{out} = \text{WRITEPRN}("c:\text{Chad}\text{work}\text{\textbackslash ebic modelling}\text{\textbackslash ebicdatdouclate.txt}", \text{EBICdat})
\]

\[
\text{EBICdat}_{i, 2} := \text{dI}\text{dx}(\text{EBICdat}_{i, 0}, L, S, R(E), w)
\]

\[
\text{EBICdat}_{i, 3} := i
\]
A.2.1 Appendix 2.1: MathCAD 12 code for Bonard and Ganiere curve generation

This Mathcad worksheet will be an attempt to replicate the EBIC mathematical modeling of:

J.-M. Bonard and J.-D. Ganiere, "Quantitative analysis of electron-beam-induced current profiles across p-n junctions in GaAs/Al0.4Ga0.6As heterostructures," J. APPL. PHYS., V79(9), 1996, P.6897-6904

\[ \mu(k, L) := \sqrt{1 + \left( \frac{1}{L} \right)^2} \]  

This is a simple substitution

In this program, Se,h/De,h will be represented as the variables se or sh

\[ db := 30 \]

\[ \sigma_1(db) := (0.0591) \exp \left( \frac{db}{605.7} \right) \]  

These functions I obtained from using CASINO to perform Monte Carlo simulations (C.M. Parish and P. E. Russell, APPL. PHYS. LETT., in press) of GaN at different beam energies. These particular values are curve-fits to my 5keV data.

\[ \sigma_2 := 0.0140 \]

\[ E := 5 \]

Beam energy, in keV

\[ \sigma_1(db) := 0.002 \]

\[ \sigma_2 := 0.015 \]

\[ d_j := 0.050 \]

Width of the depletion region, microns

\[ \delta := 1 \times 10^{-1} \]

\[ \delta := 1 \times 10^{-1} \]  

Ratio of Vs (cm/sec) to D (cm²/sec)

\[ se := 5 \]

\[ se := 5 \]

HOLD however units here must be 1/μm

Multiply by 10⁴ to get 1/cm

\[ sh := 5 \]

\[ L_x := 0.1 \]

Minority carrier diffusion lengths, in microns

h, e represent holes, electrons if p-side is on the left.

\[ L_h := 0.1 \]

Otherwise, reversed.

The following two \( \Theta \) expressions are substituted into the final EBIC expression

For simplicity, each theta has been broken up into four parts such that \( \Theta = r t u v \)
\[ r_{e}(k, x) := \frac{1}{k^2 + \sigma_e^2} \cdot \text{erfc}\left[ \frac{\sigma_1(\text{db}) \cdot \left( \mu(k, L_e) - 2 \cdot \frac{x}{\sigma_1(\text{db})} \right)}{2} \right] \]

\[ r_{u}(k, x) := \exp\left( \frac{\sigma_1(\text{db}) \cdot \mu(k, L_e)}{2} \right) \]

\[ u_{e}(k, x) := \frac{\exp(-\mu(k, L_e) \cdot x)}{(1 + k^2 \cdot \sigma_e^2)^3} \]

\[ v_{e}(k, x) := \left[ 1 + 3 \cdot \sigma_e \cdot \sigma^2 - k^2 \cdot \left( \sigma_e \cdot \sigma^2 + 3 \sigma^2 \right) \right] \]

\[ \Theta_{e}(k, x) = r_{e}(k, x) \cdot t_{e}(k, x) \cdot u_{e}(k, x) \cdot v_{e}(k, x) \]

\[ r_{h}(k, x) := \frac{1}{k^2 + \sigma_h^2} \cdot \text{erfc}\left[ \frac{\sigma_1(\text{db}) \cdot \left( \mu(k, L_h) - 2 \cdot \frac{x}{\sigma_1(\text{db})} \right)}{2} \right] \]

\[ r_{h}(k, x) := \exp\left( \frac{\sigma_1(\text{db}) \cdot \mu(k, L_h)}{2} \right) \]

\[ u_{h}(k, x) := \frac{\exp(-\mu(k, L_h) \cdot x)}{(1 + k^2 \cdot \sigma_h^2)^3} \]

\[ v_{h}(k, x) := \left[ 1 + 3 \cdot \sigma_h \cdot \sigma^2 - k^2 \cdot \left( \sigma_h \cdot \sigma^2 + 3 \sigma^2 \right) \right] \]

\[ \Theta_{h}(k, x) = r_{h}(k, x) \cdot t_{h}(k, x) \cdot u_{h}(k, x) \cdot v_{h}(k, x) \]
Now for the actual EBIC trace function. The original paper calls it $\eta$. This terminology iritates me, but I think there's no reason to change it -- I'm trying to replicate the original work at this point.

$$
\eta(x, z, dj) := \left[ \frac{1}{\pi} \int_{0}^{2\pi} \Theta(k, x) + \Theta[-(x + dj), x] \right] \frac{dk}{k} + 0.5 \left( \text{erf} \left( \frac{x + dj}{\sigma(E)} \right) - \text{erf} \left( \frac{x}{\sigma(E)} \right) \right)
$$

start := -1.5 \quad \text{inc} := .025 \quad \text{end} := 1.5
x := start, (start + inc) .. end \quad z := 100
\[ \text{EBICdat}_{i,j} := i \times \text{inc} + \text{start} \]

\[ \text{EBICdat}_{i,1} := \eta\left[ \text{EBICdat}_{0,0}, z, 0.1 \right] \]

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-1.5</td>
</tr>
<tr>
<td>1</td>
<td>-1.475</td>
</tr>
<tr>
<td>2</td>
<td>-1.45</td>
</tr>
<tr>
<td>3</td>
<td>-1.425</td>
</tr>
<tr>
<td>4</td>
<td>-1.4</td>
</tr>
<tr>
<td>5</td>
<td>-1.375</td>
</tr>
<tr>
<td>6</td>
<td>-1.35</td>
</tr>
<tr>
<td>7</td>
<td>-1.325</td>
</tr>
<tr>
<td>8</td>
<td>-1.3</td>
</tr>
<tr>
<td>9</td>
<td>-1.275</td>
</tr>
<tr>
<td>10</td>
<td>-1.25</td>
</tr>
<tr>
<td>11</td>
<td>-1.225</td>
</tr>
<tr>
<td>12</td>
<td>-1.2</td>
</tr>
<tr>
<td>13</td>
<td>-1.175</td>
</tr>
<tr>
<td>14</td>
<td>-1.15</td>
</tr>
<tr>
<td>15</td>
<td>-1.125</td>
</tr>
</tbody>
</table>

out := WRITEPRN("c:\Chad\work\ebic\modelling\ebicdata.txt", EBICdat)
A.2.2 Appendix 2.2: Maple 9.5 code for Bonard and Ganiere curve generation

```maple
> restart;

> mu_e := sqrt(k^2 + 1/Le^2); mu_h := sqrt(k^2 + 1/Lh^2);

> re := (1/(k^2 + se^2)) * erfc(somal / 2.0 * (mu_e - 2 * (-x-dj) / somal^2));

> te := exp(somal*mu_e/2)^2;

> ue := exp(-mu_e*(-x-dj)) / (1 + k^2*sigma2^2)^3;

> ve := (1 + 3*se*sigma2 - k^2*(se*sigma2^3 + 3*sigma2^2));

> rh := (1/(k^2 + sh^2)) * erfc(somal / 2.0 * (mu_h - 2 * x / somal^2));

> th := exp(somal*mu_h/2)^2;
```
\[ \theta_h := \left( \frac{1}{2} \sigma I \sqrt{k^2 - \frac{1}{Lh^2}} \right)^2 \]

\[ u_h = e^{-\left( \sqrt{k^2 - \frac{1}{Lh^2}} x \right)} \]

\[ \frac{1}{1 + k^2 \sigma^2} \]

\[ v_h := \left( 1 + 3 st * \sigma 2 - k^2 \left( sh * \sigma 2^3 + 3 * \sigma 2^2 \right) \right) \]

\[ v_h := 1 + 3 sh \sigma 2 - k^2 \left( sh \sigma 2^3 + 3 \sigma 2^2 \right) \]

\[ \Theta e := \text{re} * \text{te} * \text{ue} * \text{ve}; \]

\[ \Theta e := \frac{1}{(k^2 + sh_0^2) (1 + k^2 \sigma^2)^2} \left\{ \text{erfc} \left( 0.5000000000 \sigma I \left( \sqrt{k^2 - \frac{1}{Le^2}} \cdot \frac{2 \left( x - d_j \right)}{\sigma I^2} \right) \right) \right\} \]

\[ e \left( \frac{1}{2} \sigma I \sqrt{k^2 - \frac{1}{Le^2}} \right)^2 e^{-\left( \sqrt{k^2 - \frac{1}{Le^2}} \left( x - d_j \right) \right)} \]

\[ \Theta h := \frac{1}{rh * \theta h * u h * v h}; \]

\[ \Theta h := \frac{1}{(k^2 + sh_0^2) (1 + k^2 \sigma^2)^2} \left\{ \text{erfc} \left( 0.5000000000 \sigma I \left( \sqrt{k^2 - \frac{1}{Le^2}} \cdot \frac{2 x}{\sigma I^2} \right) \right) \right\} \]

\[ e \left( \frac{1}{2} \sigma I \sqrt{k^2 - \frac{1}{Le^2}} \right)^2 e^{-\left( \sqrt{k^2 - \frac{1}{Le^2}} \right)} \]

\[ \Theta h := \frac{1}{1 + 3 sh \sigma 2 - k^2 \left( sh \sigma 2^3 + 3 \sigma 2^2 \right)} \]

\[ \eta := \left( \frac{1}{\pi} \right) \int (sh \Theta h + se \Theta e), k=0..Z \right) + 0.5 \left( \text{erf} \left( \frac{x + dj}{\sigma I} \right) - \text{erf} \left( \frac{x}{\sigma I} \right) \right); \]
\[
\eta := \frac{1}{\pi} \left( \int_0^z \frac{1}{(k^2 + s h^2)(k^2 + a_2^2 s \alpha^2)^3} \left( s h \text{erfc} \left( 0.5000000000 a_1 \left( \sqrt{\frac{k^2 + 1}{Le^2}} - \frac{2x}{a_1^2} \right) \right) \right) \right.

\left( e^{\left( \frac{1}{2} a_1 \sqrt{k^2 + \frac{1}{Le^2}} \right)^2} e^{-\left( \frac{x}{Le^2} \right)(k^2 s - k^2 \alpha_2^2 (k \alpha_2^2 + 3 \alpha_2^2))} \right)

\left. + \frac{1}{(k^2 + s e^2)(1 + k^2 a_2^2 s \alpha^2)^3} \left( s e \text{erfc} \left( 0.5000000000 a_1 \left( \sqrt{\frac{k^2 + 1}{Le^2}} - \frac{2(x - d_j)}{a_1^2} \right) \right) \right) \right)

\left( e^{\left( \frac{1}{2} a_1 \sqrt{k^2 + \frac{1}{Le^2}} \right)^2} e^{-\left( \frac{x}{Le^2} \right)(s e - d_j)(s e^2 - k^2 (s e a_2^2 + 3 \alpha_2^2))} \right) \right) \right)

+ 0.5 \text{erf} \left( \frac{x + d_j}{a_1} \right) - 0.5 \text{erf} \left( \frac{x}{a_1} \right)
\]

> \text{sigma1} := 1.361; \text{sigma2} := 0.015;

\>

\text{sigma1} := 1.361
\text{sigma2} := 0.015

> \text{E} := 5.0;

\>

\text{E} := 5.0

> \text{dj} := 0.030; \text{se} := 1.0e-2; \text{sh} := 1.0e-2; \text{Le} := 0.090; \text{Lh} := 0.050;

\>

\text{dj} := 0.030
\text{se} := 0.010
\text{sh} := 0.010
\text{Le} := 0.090
\[ L_h = 0.050 \]

\[ Z = 100 \]

\[ x_{\text{min}} = -2.0; \quad x_{\text{max}} = 1.0; \quad dx = 0.02; \]
\[ x_{\text{min}} = -2.0 \]
\[ x_{\text{max}} = 1.0 \]
\[ dx = 0.02 \]

\[ x_{\text{tot}} = (x_{\text{max}} - x_{\text{min}})/dx; \]
\[ x_{\text{tot}} = 150.0000000 \]

\[ \text{xlist} := \text{array}(1..150, 1..2); \]
\[ \text{xlist} = \text{array}(1..150, 1..2, []) \]

\[ \text{outfile} := \text{fopen}("ebic.txt", \text{WRITE, TEXT}); \]
\[ \text{outfile} := 0 \]

\[ \text{for } i \text{ from 1 to 150 } \text{ do} \]
\[ x := x_{\text{min}} + dx \times i; \quad y := \text{evalf}(\text{eta}); \]
\[ \text{fprintf(outfile, "\text{\textbackslash t}\text{\textbackslash e}\n", x, y);} \]
\[ \text{xlist}[i,1] := x; \quad \text{xlist}[i,2] := y; \]
\[ \text{fclose(outfile);} \]

\[ \text{print(xlist);} \]
| -1.98 | 0.0015468300 |
| -1.96 | 0.0016135254 |
| -1.94 | 0.0016823699 |
| -1.92 | 0.0017533946 |
| -1.90 | 0.0018266288 |
| -1.88 | 0.0019021000 |
| -1.86 | 0.0019798342 |
| -1.84 | 0.0020598556 |
| -1.82 | 0.0021421860 |
| -1.80 | 0.0022268449 |
| -1.78 | 0.0023138503 |
| -1.76 | 0.0024032172 |
| -1.74 | 0.0024949577 |
| -1.72 | 0.0025890822 |
| -1.70 | 0.0026855973 |
| -1.68 | 0.0027845077 |
| -1.66 | 0.0028858143 |
| -1.64 | 0.0029895153 |
| -1.62 | 0.0030956060 |
| -1.60 | 0.0032040774 |
| -1.58 | 0.0033149176 |
| -1.56 | 0.0034281116 |
| -1.54 | 0.0035436399 |
| -1.52 | 0.0036614802 |
| -1.50 | 0.0037816059 |
| -1.48 | 0.0039039020 |
A.3 Appendix 3: C++ code for polarization fields model

/***************************************************************************/

pzwell8.cpp

A program to compute the emission from an InGaN quatum well that has an arbitrary piezoelectric field across the quantum well.

This program treats the Schroedinger equation as a two-point boundary value problem and solves for the energy states using the techniques in D. Ferry's QUANTUM MECHANICS, which can be derived from the general two-point boundary value problem; see, for example, Kincaid and Cheney, NUMERICAL ANALYSIS.

Used to solve the eigenproblem are the "TNT" numerical routines developed by NIST. www.nist.gov/tnt

This is a self-consistent solution; the wavefunctions and carrier density are input into the Poisson equation to determine carrier pooling induced band bending, which is used to iterate on the Schroedinger equation.

--------------------

Written by Chad M. Parish
Analytical Instrumentation Facility
Department of Material Science and Engineering
North Carolina State University
www.ncsu.edu/aif

2004

Written and tested using Borland "Free Command Line Tools" freeware compiler 5.5.1

***************************************************************************/

/*******
Header files
**********/
#include <iostream.h>
#include <math.h>
#include <iomanip.h>
#include <stdlib.h>
#include <fstream.h>
#include <string.h>

// These headers are for the TNT and related routines
// that can be obtained from NIST.
#include <tnt.h>
#include <jama_eig.h>

// this is a personal header file that defines the
// "hist" class for histograms of data
#include "chad_hist.h"

/********************
 constants for the program
 ********************/

// constants of nature
const double hb = 1.054572e-34; // Dirac's constant; J-seconds
const double m0 = 9.109389e-31; // Electron mass; kilos
const double e = 1.602177e-19; // Electron charge; coulombs/electron
const double c = 2.998e8; // speed of light, m/sec
const double pi = 4.0 * atan( 1.0 );

// constants adjustable for the situation
const double epsilon = 8.854e-12 * 8.9; // approx dielectric constant for GaN

const double zmin = -10.0e-9; // minimum z value; meters
const double zmax = -1.0 * zmin; // maximum z value; meters
const int n = 300; // number of bins for descritization

const double a = (zmax - zmin) / ( (double)( n + 1 ) );
  // size of each array element, meters

const int num = 8; // number of energies, wavefunctions to send
to the output file
  // For wide wells, there will be more than 8 VB subband states,
  // but the difference in terms of the output spectrum is << 1%
  // (actually, about 0.001 - 0.01%). This is because the first
  // one to two subbands are by far the most significant.
  // Thus, an arbitrary cut off at 8 makes no difference.

const double LB = 3.4; // left CB barrier height, eV
const double RB = 3.85; // right CB barrier height, eV
const double LBh = 0.0; // left VB barrier height, eV
const double RBh = -0.15;  // right VB barrier height, eV
const double me = 0.2;  // effective mass of electrons, *m0
const double mh = 1.2;  // effective mass of holes, *m0
const double ratio = 0.3559; // ratio of CB to VB offsets for InGaN
           // compared to n-GaN
const double kT = 0.02586; // kT in eV at 300kelvins

/*****************
Global Variables
*****************/

    // I was reluctant to use this global variable, but I think it is
    // the proper way to handle this. E_values[0][] holds the first num
    // energies of the
    // CB subband quantized states. E_values[1][] holds the first num
    // energies of
    // the VB subband states. This is data obtained from the quantum
    // calculation
    // and utilized by the density of states functions Ze, Zh.
    double E_values[2][num];

double n_sheet;

double W;  //QW width, meters
double zmin_W;  //minimum z in well

double zmax_W;  //maximum z in well

/************************
function prototypes
***********************/

void do_calc( double bandgap, double f, hist* spec, int flag );  // for a
given bandgap and
                // pz field, it finds the
                // distribution
                // of energies
                // if flag != -1, it will
                // output the
                // wavefunctions, CB/VB, etc.
                // to the output file

double get_gap( double lambda );  // finds the bandgap under zero field
               // that corresponds to lambda
double V( double z, double f, double gap ); // return potential energy of CB at z
double Vh( double z, double f, double gap ); // return potential energy of VB at z
void Sfill( Array2D<double> &S, double f, double bandgap, double Vee[] ); // fill S with proper values
void Sfillh( Array2D<double> &S, double f, double bandgap, double Vee[] ); // fill Sh with proper values

//output to file.
void output( Array2D<double> &vec, Array1D<double> &R, Array2D<double> &vech, Array1D<double> &Rh,
            double f, double gap, char filename[], double Efcb, double Vee[] );
void integrate( Array2D<double> &CB, Array2D<double> &VB, double integrals[] ); //integrates <VB|CB>
void normalize( double dat[] ); // normalizes an n-long psi-vector for zmin->zmax

double Zh( double Eh ); // VB density of states
double Ze( double Ee ); // CB density of states
double Fermi_e( double Ee, double Ef ); // electron Fermi function
double Fermi_h( double Eh, double Efcb ); // hole Fermi function
void get_Fhvb( double &Efvb, double &Efcb );

double int_cb( int N, double Ef ); // 5 point gaussian quadrature
    // numerical integration of the number
    // of states in the CB; i.e., integral over
    Z(E)F(E)dE
    // from lowest subband to top of well.

double int_vb( int N, double Efvb ); // ditto, VB.
    // Efvb needed as input.

// do the CB calculation
void calc_cb( Array1D<double> &EigEnergiesR, Array2D<double> &psiV, double f, double bandgap );
    // ditto, VB
void calc_vb( Array1D<double> &EigEnergiesRh, Array2D<double> &psiVh, double f, double bandgap );

// The counts the number of confined subbands in VB & CB
void get_number_states( Array1D<double> &EigEnergiesR, Array1D<double> EigEnergiesRh,
                       int &CB_states, int &VB_states, double f, double gap );
// error trapping function to output a number to the screen
void tick( int c );

using namespace TNT;
using namespace JAMA;

void main( void )
{
    double f;  // PZ field, V/m
    double gap;  // peak wavelength, m
    double PZ;  // pz field

cout << "QW bandgap (eV): ";
cin >> gap;

cout << "Quantum well width (nm):";
cin >> W;
W *= 1.0e-9;

zmin_W = -1.0 * ( 0.5 * W );  //minimum X in well
zmax_W = ( 0.5 * W );    //maximum X in well

cout << "Sheet Density (EHP's/cm^2):";
cin >> n_sheet;
n_sheet *=1.0e4;

cout << "PZ field (MV/cm): ";
cin >> PZ;
PZ *= 1.0e8;

hist spectrum( 2.0, LB - LBh, 100*n );

double bias;

double V0, Wdepletion;  // built in voltage and depletion width
double Na, Nd;  // numbers of ionized acceptors, donors

Na = 1.2e19;
Nd = 9.2e17;

V0 = 3.66;
int count = 0;
ofstream biasout("biased.txt");

biasout << "Bandgap (eV):\t" << gap << endl
<< "Well Width (m):\t" << W << endl
<< "Sheet Density (m^-2):\t" << n_sheet << endl
<< "PZ field (MV/cm):\t" << ( PZ / 1.0e8 ) << endl;

cout << endl << "Bias, V\t\tEwell, V/m\t\tIterations\tPeak, nm\tDep. width, nm" << endl << setprecision(3) << fixed;
biasout << endl << "Bias, V\t\tEwell, V/m\t\tPeak, nm\tDep. width, nm" << endl << setprecision(3) << fixed;

for( bias = 0.0; bias > -31.0; bias -= 1.0 )
{
    spectrum.zero();

    Wdepletion = sqrt( 2.0 * epsilon * ( Na + Nd ) * ( V0 - bias )
    / ( e * Na * Nd * 1.0e6 ) );
    f = PZ - bias / Wdepletion;
    cout << bias << "\t" << scientific << "\t" << f;
    biasout << bias << "\t" << scientific << "\t" << f;

    do_calc( gap, f, &spectrum, -1 );

    if( spectrum.find_peak() )
    {
        cout << "\t" << fixed << hb*2*pi*c*1.0e9 / ( spectrum.find_peak() * e ) << "\t\t"
        << scientific << Wdepletion << endl << fixed;
        biasout << "\t" << fixed << hb*2*pi*c*1.0e9 / ( spectrum.find_peak() * e ) << "\t\t"
        << scientific << Wdepletion << endl << fixed;
    }
    else
    {
        cout << "Peak not found???
        biasout << "Peak not found???

    count ++;
}

biasout.close();
spectrum.zero();
/*************************************/
This function backs out the InGaN bandgap
for a given lambda at zero PZ by iterating until the
peak is near the input lambda.

It then outputs the distribution to output file
"zero_pz.txt" and returns the band gap
(in eV)
*************************************/

double get_gap( double lambda )
{
    cout << "Finding bandgap" << endl;

double f = 0.0;  // PZ field is zero for this calculation

    hist spectrum( 2.0, LB - LBh, 10*n ); // this will hold the
    predicted spectrum
    // of photon energies

    // convert lambda to E
    double Etarget = hb * 2 * pi * c / lambda / e;

    double gap = Etarget;

    double Ecurrent;

    do
    {

        do_calc( gap, f, &spectrum, -1 );

        Ecurrent = spectrum.find_peak();
        cout << "." ;

        gap -= 0.025;

    } while( Ecurrent > Etarget );

}
cout << endl << "Bandgap, eV = " << gap;
return gap;

}/**
This function takes the bandgap and the pz field f
as inputs, calculates the distribution of energies and outputs it
to spec
*/
void do_calc( double bandgap, double f, hist* spec, int flag )
{
    int count = 0; // counter for the self-consistent
    // schroedinger-poisson equation calculation
    // loops
    const int NUM_LOOPS = 100; // hold arbitray number of allowed loops.
    double Vee[n]; // Vee[i] holds the poisson-derived
    electron-electron potential // for z-position associated with i

    // zero the array
    for( int i = 0; i < n; i++ )
    {
        Vee[i]=0.0;
    }

    // OldE will be used for testing for convergence; this initializes it.
    double OldE = 20;

    // this is used in the convergence test
double error;

    // These are the data members to hold the actual Schoroedinger equation
// calculations. See the TNT documentation for explanation
Array2D<double> S(n,n);  // the S* matrix to hold the
tridiagonal for the CB
Array1D<double> EigEnergiesR(n); // This matrix holds all the
eigenenergies (real parts), joules
Array1D<double> EigEnergiesRh(n); // This matrix holds all the
eigenenergies (real parts), joules
Array1D<double> Eig_eV(n); // This matrix holds the real parts
once they are converted J -> eV
Array2D<double> psiV(n,n); // Array to hold eigenvectors
Array2D<double> Sh(n,n); // the Sh* matrix to hold the
tridiagonal for the VB
Array1D<double> EigEnergiesRh(n); // This matrix holds all the
eigenenergies (real parts), joules
Array1D<double> Eig_eVh(n); // This matrix holds the real parts
once they are converted J -> eV
Array2D<double> psiVh(n,n); // Array to hold eigenvectors

// This next code segment initializes an output file to
// hold the band shapes over the course of the iteration
// if the run is flagged to produce output to disk
ofstream fileout("temp.txt");
if( flag != -1 )
{
    fileout.close();

    // this sets up the filename to use,
    // such as "Bbands.txt"
    char fn;
    fn = (char)(flag + 65);
    char g[25] = "bands.txt\0";
    char h[25];
    h[0] = fn;
    int ii;
    for( ii = 1; g[ ii - 1 ]; ii ++ )
        h[ ii ] = g[ ii - 1 ];
    h[ii] = '\0';
    fileout.open(h);
    fileout << "PZ Field:\t" << f << "\tSheet Density:\t" << n_sheet << endl << endl;
}
// end of output file initialization

/****************
This do-while loop is the self-consistent Schoroeedinger-Poisson
 solver
****************/
do
{

/******************************
Calculate for the conduction band
*******************************/

Sfill( S, f, bandgap, Vee );  // put the proper  
tri-diagonal values into the S matrix for PZ field f

Eigenvalue<double> EigVal( S );  //initialize the  
Eigenroutines

    EigVal.getRealEigenvalues( EigEnergiesR );  // calculate 
the real eigenvalues
    EigVal.getV( psiV );  // calculate the eigenvectors

/******************************
Calculate for the valence band
*******************************/

Sfillh( Sh, f, bandgap, Vee );  // put the proper tri- 
diagonal  
// values into the Sh matrix for PZ field f

Eigenvalue<double> EigValh( Sh );  //initialize the  
Eigenroutines

    EigValh.getRealEigenvalues( EigEnergiesRh );  // calculate 
the real eigenvalues
    EigValh.getV( psiVh );  // calculate the eigenvectors

    // Because of the way I handled the holes, EigEnergiesRh will 
be negative 
their actual values
    for( int i=0; i < n; i++ )
        EigEnergiesRh[i] *= -1.0;

    // These are to hold the values of the first subbands for 
// quantum-mechanical normalization for later use in the 
// Poisson equation
    // (output from the Eigenvector routines is obviously not 
normalized

338
// in the quantum mechanical sense of the word
double psie[n], psih[n];

// copy the first subbands to the temporary array variable
for( int i = 0; i < n; i ++ )
{
    psie[i] = psiV[i][0];
    psih[i] = psiVh[i][0];
}

// normalize them in the quantum mechanically proper sense
// of $\langle i | i \rangle = 1.$
normalize( psie );
normalize( psih );

// Here comes the Poisson Equation
// I impose boundary conditions of :
// #holes=#electrons
// $E(Z_{min})=0$
// $V(Z_{min})=0$
// Integrating under all the charge density will give zero;
this results
// in a $E(Z_{max})=0$ but NOT a $V(Z_{max})=0$

double efield[n]; // hold the field E, V/m
double potential[n]; // Hold the potential V, volts
double rho[n]; // Hold the charge density, coulombs/meter$^3$

// set up charge desnity, using only lowest VB and CB subband
for( int i = 0; i < n; i ++ )
    rho[i] = n_sheet * a * a * ( psih[i] * psih[i] - psie[i] * psie[i] );

// integrate charge density to get field
efield[0] = 0.0;
for( int i = 1; i < n; i ++ )
    efield[i] = -0.5*a*( rho[i] + rho[i-1] )/epsilon + efield[i-1];

// integrate field to get potential
potential[0] = 0.0;
for( int i = 1; i < n; i ++ )
    potential[i] = 0.5*a*( efield[i] + efield[i-1] ) + potential[i-1];

// output everything to the file
for( int i = 0; i < n; i ++ )
{  
    fileout << i << "\t" << rho[i] << "\t" << -1.0*efield[i]<< "\t" << -1.0*potential[i]
    << "\t" << psie[i] << "\t" << psih[i] << "\t"
    << V( (double)( i - n/2 ) * a, f, bandgap )/e - Vee[i]
    << "\t" << Vh( (double)( i - n/2 ) * a, f, bandgap )/e - Vee[i] <<endl;
}

// configure the potential data into the Vee array
for( int i = 0; i < n; i ++ )
    Vee[i] = potential[i]; // new data

    count ++; // increment
    fileout << endl << count << endl;

    error = fabs( ( EigEnergiesR[0] - EigEnergiesRh[0] - OldE ) / OldE );

    OldE = EigEnergiesR[0] - EigEnergiesRh[0] ;

} while( ( error > 1.0e-3 ) && ( count < NUM_LOOPS ) ) ;

// 1e-3 relative error is about 1nm at 520nm.
    cout << "\t" << count;

// if 30 loops didn't give convergence, shoot the average between
the last two Vee's
if( count >= NUM_LOOPS )
{
    cout << "Didn't converge error still: " << error;
    exit(0);
}

fileout.close();
/*************
This next section gets set up for calculating DOS and using that to find the quasi-Fermi-levels
*************/

// This puts the energies in storage accessible to the DOS functions
for(int i = 0; i < num; i++) {
    E_values[0][i] = EigEnergiesR[i];
    E_values[1][i] = EigEnergiesRh[i];
}

// this code sets up the quasi fermi levels.
double Efvb, Efcb; // will hold energies of quasi-fermi levels in VB and CB
get_Fhvb( Efvb, Efcb ); // this iterates a numerical integration to find the Efvb
   // and Efcb that gives sheet carrier density numbers
   // of states in the VB and CB

/********************
Typically, there will be one or two confined CB states and several (3-4?) confined VB states.
This code finds <CBi|VBj> for the combinations and also outputs PSIcbi * PSIvbi data sets as PSI^2(z) to ints.txt,
along with the transition probability data
********************/

double integrated_values[num * num]; // this will hold the data of the transition probabilities
   // <ei|hj> squared such that the array index is [ i * num + j ]

   // this actually calculates the quantum probabilities
   integrate( psiV, psiVh, integrated_values );

// CB_states is the number of CB states confined in the well
// VB_states is the number of VB states confined in the well
int CB_states, VB_states; // number of subband states in CB, VB
get_number_states( EigEnergiesR, EigEnergiesRh, CB_states, VB_states, f, bandgap );

double delta_K = 1.0e4;
double delta_Ee, delta_Eh, P;

spec->zero();

for( int i = 0; i < CB_states; i ++ )
{
    for( int j = 0; j < VB_states; j ++ )
    {
        // gap between the bottom point of the CB subband i
        // and the top of the VB subband j
        double Ei = EigEnergiesR[ i ]/e;
        double Ej = EigEnergiesRh[ j ]/e;
        double Egap_eff = Ei - Ej;

        // momentum in the plane of the QW.
        double Kxy = 0;

        // increments resulting from Kxy
        double dEi=0.0, dEj=0.0, dEi_old, dEj_old;

        dEi_old = 0.0;
        dEj_old = 0.0;

        while(
            ( ( Ei + dEi ) < LB )
            &&
            ( ( Ej - dEj ) > LBh )
        )
        {
            dEi = hb*hb*Kxy*Kxy / ( 2.0 * me * m0 )/e;
            dEj = hb*hb*Kxy*Kxy / ( 2.0 * mh * m0 )/e;

            double Ephoton = Egap_eff + dEi + dEj;

            delta_Ee = dEi - dEi_old;
            delta_Eh = dEj - dEj_old;

            P = 2.0 * ( me * m0 ) / ( pi * hb * hb ) // parabola has 2 lobes
            density of states, CB
F for electrons
* Fermi_e( Ei + dEi, Efcb ) //

of states, VB
* ( mh * m0 ) / ( pi * hb * hb ) // density

holes
* Fermi_h( Ej - dEj, Efvb ) // F for

delta_Ee * delta_Eh );

integrated_values[ i * num + j ] * (

spec->add( EPhoton, P );

dEi_old = dEi;

dEj_old = dEj;

Kxy += delta_K;

}

}

if( flag != -1 )
{

//this sets up the filename to use,
//such as "4wavefunctions.txt"
char fn;
fn = (char)(flag + 65);
char g[25] = "wavefunctions.txt\0";
char h[25];
h[0] = fn;
int i;
for( i = 1; g[ i - 1 ]; i ++ )
h[ i ] = g[ i - 1 ];

h[i] = '\0';

output( psiV, EigEnergiesR,

psiVh, EigEnergiesRh,

f, bandgap, h, Efvb, Efcb, Vee );
g[0] = '\0'; h[0] = '\0';

{

char g[25] = "spectrum.txt";

h[0] = fn;
for( i = 1; g[ i - 1 ]; i ++ )
h[ i ] = g[ i - 1 ];
h[i] = '\0';

spec->normalize();

ofstream fout(h);

fout << "PZ field\t" << f << "\tV/m\t" << endl;

if( spec->find_peak() )
    fout << "Peak (nm)\t" << hb * 2.0 * pi * c *
1.0e9/ ( spec->find_peak() * e ) << endl;
else
    { fout << "Peak not found " << endl;
      cout << "peak error";
    }

fout << "CB states used:\t" << CB_states << endl;
fout << "VB states used:\t" << VB_states << endl;
fout << "Wavelength (nm)\tIntensity\t" << endl;

for( int i = 0; i < (10*n-1); i ++ )
{
    if( ( spec->get_E( i ) != 0.0 ) && ( spec-
>get_value_i( i ) >= 1.0e-9 ) )
        fout << hb * 2.0 * pi * c * 1.0e9/ ( spec-
>get_E( i ) * e ) << "\t" << spec->get_value_i( i ) << endl;
}

fout.close();

g[0] = '\0'; h[0] = '\0';

}
Sfill( S, f, bandgap ); // put the proper tri-diagonal values into the S matrix for PZ field f

Eigenvalue<double> EigVal( S ); // initialize the Eigenroutines

EigVal.getRealEigenvalues( EigEnergiesR ); // calculate the real eigenvalues
EigVal.getV( psiV ); // calculate the eigenvectors
return;
}

void calc_vb( Array1D<double> &EigEnergiesRh, Array2D<double> &psiVh, double f, double bandgap )
{
    Array2D<double> Sh(n,n); // the S* matrix to hold the tri-diagonal
    Sfillh( Sh, f, bandgap ); // put the proper tri-diagonal values into the Sh matrix for PZ field f
    Eigenvalue<double> EigValh( Sh ); // initialize the Eigenroutines
    EigValh.getRealEigenvalues( EigEnergiesRh ); // calculate the real eigenvalues
    EigValh.getV( psiVh ); // calculate the eigenvectors

    // Because of the way I handled the holes, EigEnergiesRh will be negative their actual values
    for( int i=0; i < n; i++ )
        EigEnergiesRh[i] *= -1.0;
}
*/

void get_number_states( Array1D<double> &EigEnergiesR, Array1D<double> EigEnergiesRh, int &CB_states, int &VB_states, double f, double gap )
{
    int count;
    count = 0;
do
    {
        count ++;
    } while(( ( EigEnergiesR[ count ]) < ( V( zmin, f, gap) ) )
            & ( ( EigEnergiesR[ count ]) < (V( zmax, f, gap) ) )
    );

    CB_states = count;
    count = 0;
do
    {
        count ++;
    } while(( ( EigEnergiesRh[ count ]) < ( V( zmin, f, gap) ) )
            & ( ( EigEnergiesRh[ count ]) < (V( zmax, f, gap) ) )
    );

    VB_states = count;
}
} while(
    ( EigEnergiesRh[ count ] > (Vh( zmin, f, gap) ) )
    && ( EigEnergiesRh[ count ] > (Vh( zmax, f, gap) ) )
);

VB_states = count;

if( CB_states > num )
    CB_states = num;

if( VB_states > num )
    VB_states = num;

return;

/*****************************/
This function returns the PE for the CB
(shape of the conduction band)
******************************/

double V( double z, double f, double gap )
{

    if( z < zmin_W )
        return LB * e + f*zmin_W*e; // returns LB if z left of the well
    if( z < zmax_W )
        return ( LB - ( ( LB - LBh - gap ) / ( 1.0 + 1.0 / ratio ) ) +
                ( z * f ) )*e;  // returns sloped value
                            // if z in the well

        return RB * e + f*zmax_W*e;  // return RB is z right of the well

}

/*****************************/
This function returns the PE for the VB
(shape of the valence band)
******************************/

double Vh( double z, double f, double gap )
{

}
// dittos for comments
if( z < zmin_W )
    return LBh * e + f*zmin_W*e;
if( z < zmax_W )
    return ( LBh + ( ( LB - LBh - gap ) / ( 1.0 + ratio ) ) + ( z
    * f ) )*e;

return RBh * e + f*zmax_W*e;

/***********
This function fills the S array with the proper values for the
computation.

Si,i = Vi + hb^2/ma^2
Si,i+1=Si,i-1=-hb^2/2ma^2
else =0
***********/
void Sfill( Array2D<double> &S, double f, double bandgap, double Vee[] )
{

    // zero out the S array; important, as only the long, sub and super
diagonals
    // have non-zero values (tri-diagonal system)
    for( int i = 0; i < n; i++ )
        for( int j=0; j<n; j++ )
            S[i][j] = 0.0;

    // now, assign proper values to the long, sub and super diagonals
    // this code SKIPS the top and bottom rows!!!
    for( int i = 1; i < (n-1); i++ )
    {
        S[i][i] = ( V( (double)( i - n/2 ) * a, f, bandgap ) -
        Vee[i]*e )
            + hb * hb / ( me * m0 * a * a );

        S[i][i-1] = - hb * hb / ( 2 * me * m0 * a * a );
        S[i][i+1] = - hb * hb / ( 2 * me * m0 * a * a );

    }
// This code gets the top and bottom diagonal terms
S[0][0] = ( V( (double)( 0 - n/2 ) * a, f, bandgap ) - Vee[0]*e ) +
  hb * hb / ( me * m0 * a * a );
S[n-1][n-1] = ( V( (double)( n - n/2 ) * a, f, bandgap ) - Vee[n-1]*e ) + hb * hb / ( me * m0 * a * a );

// this gets the top row super diagonal term and bottom row subdiagonal term
S[0][1] = - hb * hb / ( 2 * me * m0 * a * a );
S[n-1][n-2] = - hb * hb / ( 2 * me * m0 * a * a );
return;

/***********
This function fills the Sh array with the proper values for the computation.
Si,i = Vi + hb^2/ma^2
Si,i+1=Si,i-1=-hb^2/2ma^2
else =0
***********/

void Sfillh( Array2D<double> &S, double f, double bandgap, double Vee[] )
{
    // comments are dittos as for Sfill function
    for( int i = 0; i < n; i++ )
        for( int j=0; j<n; j++ )
            S[i][j] = 0.0;

    for( int i = 1; i < (n-1); i++ )
    {
        S[i][i] = -1.0*( Vh( (double)( i - n/2 ) * a, f, bandgap ) -
                      Vee[i]*e )
                    + hb * hb / ( mh * m0 * a * a );

        S[i][i-1] = - hb * hb / ( 2 * mh * m0 * a * a );
        S[i][i+1] = - hb * hb / ( 2 * mh * m0 * a * a );
    }
}
S[0][0] = -1.0*(Vh( (double)( 0 - n/2 ) * a, f, bandgap ) - Vee[0]*e ) + hb * hb / ( mh * m0 * a * a );
S[n-1][n-1] = -1.0*( Vh( (double)( n - n/2 ) * a, f, bandgap ) - Vee[n-1]*e ) + hb * hb / ( mh * m0 * a * a );
S[0][1] = - hb * hb / ( 2 * mh * m0 * a * a );
S[n-1][n-2] = - hb * hb / ( 2 * mh * m0 * a * a );
return;

}/**
This function normalizes the wavefunctions, then integrates across their products to find the transition probabilities
In other words, finds <i|j> in QM Dirac notation
Input data:
CB is the array of psi values for the electrons
VB is the array of psi values for the holes
integrals[] is a num*num long vector to hold the integrated values --> this array is basically the "return," although it is handled as pass-by-reference

*/
void integrate( Array2D<double> &CB, Array2D<double> &VB, double integrals[] )
{
    ofstream f("ints.txt"); // output file
    double cb[n], vb[n], // these hold the wavefunctions currently being worked on
        cb_dat[n][num], // these hold all the wavefunctions of interest
        vb_dat[n][num];

    // this array stores the data so it can be output to a file
    // at the end of the subroutine
    // double *store_i = new double[ n * num * num ];

    /***************/
This code normalizes the individual wavefunctions of the electrons, holes.
Uses zmin, zmax, a
**************************

for( int i = 0; i < num; i++ )
{
    // this is necessary for two reasons:
    // 1) I do not wish to modify CB (points to psiV) or VB (-> psiVh)
    // 2) I don't want to pass my normalization routine an Array2D data type,
    //    but rather a double array.
    for( int j = 0; j < (n-1); j++ )
    {
        vb[j] = VB[j][i];
        cb[j] = CB[j][i];
    }
    vb[n-1]=0.0; cb[n-1]=0.0;

    // normalize() uses the properties of the wave function to scale the cb, vb data
    // in a quantum-mechanically proper sense
    normalize( cb );
    normalize( vb );

    // this copies the vb, cb being worked with into the storage arrays cb_dat, vb_dat
    for( int l = 0; l < n; l++ )
    {
        cb_dat[l][i] = cb[l];
        vb_dat[l][i] = vb[l];
    }
}

***************************************************
Now, an array is needed to hold the different combinations of PSIel --> PSIhj. If we compute the first num electron states combining to the first num hole states, we need num^2 values.

For num=5, this is 25 combinations.

So the variable integrals[ num * num ] holds these.
store_i holds the values of the integral in order to output to a text file.
**************************/

double temp;

// this is a simple routine for integrating the "function" in the form of datapoints.
// I might modify this to a trapezoid integral later.
// But, as it is, it's quite accurate; for symmetric wells, \langle i|j \rangle
for i=j
// was nearly 1.000 and \langle i|j \rangle for i != j was <1e9.

for( int i = 0; i < num; i++ ) // i is the CB state being evaluated
{
    for( int j = 0; j < num; j ++ ) // k is the VB state being evaluated
    {
        double integral = 0.0; // this holds the running sum under the curve

        for( int k = 0; k < n; k++ ) // k is the z position being evaluated
        {
            // VB j state at point k * CB i state at point k times a 
            // (recall z = deltaz)
            temp = vb_dat[k][j] * cb_dat[k][i] * a;
            integral += temp;

            // store_i[ ( i * num + j ) * n + k ] = temp;
        }

        // this is the data desired by the program for the oscillator strength/
        // transition probability
        integrals[ i*num + j ] = integral * integral;
    }
}
f << "z (meters) \t";
for( int i = 0; i < num; i ++ )
    for( int j = 0; j < num; j ++ )
        f << i << "--->" << j << "\t";

for( int k = 0; k < n; k ++ )
{
    f << k*a + zmin << "\t";
    for( int i = 0; i < num; i ++ )
    {
        for( int j = 0; j < num; j ++ )
        {
            f << store_i[ (i * num + j) * n + k ] << "\t";
        }
    }
    f << endl;
}

// output to file
f << endl << "Integrated values " << endl << "\t";
for( int i=0; i < num; i++ )
    for( int j=0; j<num; j++ )
        f << integrals[ i*num + j ] << "\t";

// the store_i array is huge, and this takes it out of memory
delete store_i; /*

}

void normalize( double dat[] )
{

double total = 0.0;

// total is the integral under the dat^2 curve
for( int i = 0; i < n; i ++ )
{
total += dat[i] * dat[i] * a;

// if every data value is multiplied by c^2, the integral will be 1.0000000
double c = sqrt( 1.0 / total );

// do the multiplication
for( int i = 0; i < n; i ++ )
{
    dat[i] *= c;
}

/*******************
This outputs the data to a file "output.txt"
 *******************/
void output( Array2D<double> &vec, Array1D<double> &R,
              // vec are electron wavefunctions
              Array2D<double> &vech, Array1D<double> &Rh,
              // h are same for holes
              double f, double gap, char filename[],
              // f is PZ field in Volts/meter
do
tive Efvb, double Efcb,  // quasi-fermi levels
tive Vee[] )  // poisson-derived
{

/ open file stream
ofstream fn(filename);

fn << "PZ field:\t" << f << "\tV/m" << endl;
fn << "CB quasi-Fermi level:\t" << Efcb << endl << "VB quasi-Fermi level:\t" << Efvb << endl;

// print quantum numbers
fn << "Quantum state:" << endl << "CB\t" << "\t";
for( int i = 0; i < num; i++ )
    fn << i << "\t";
fn << "VB\t" ;
for( int i = 0; i < num; i++ )
    fn << i << "\t";

// Print eigenenergies in joules
fn << endl << "Energies in Joules:" << endl << "\t" << "CB\t";
for( int i=0; i < num; i++ )
{
    fn << R[i] ;
    fn << "\t";
}
fn << "VB\t";
for( int i=0; i < num; i++ )
{
    fn << Rh[i] ;
    fn << "\t";
}

fn << endl << "Energies in eV:" << endl << "\t" << "CB\t";
for( int i=0; i < num; i++ )
{
    fn << ( R[i] / e ) ;
    fn << "\t";
}
fn << "VB\t";
for( int i=0; i < num; i++ )
{
    fn << ( Rh[i] / e ) ;
    fn << "\t";
}
fn << endl << "z (m) \t V (eV) \t Wavefunctions (non-normalized): \t \t \t Vh (eV)" << endl;

for( int i=0; i < n; i++ )
{
    fn << ( ( i - n/2 ) * a ) << \t << ( V( ( i - n/2 ) * a, f,
gap ) / e + Vee[i] ) << "\t" ;

    for( int j=0; j < num; j++ )
    {
        fn << ( vec[i][j] ) << \t ;
    }

    fn << Vh( ( i - n/2 ) * a, f, gap ) / e + Vee[i] << "\t" ;

    for( int j=0; j < num; j++ )
    {
        fn << ( vech[i][j] ) << "\t" ;
    }

    fn << endl;
}

// this outputs the squares of the wave functions, multiplied by 10,
then offset by their energy;
// this is simply an aid to Excel formatting.

fn << endl << endl << "z (m) \t V (eV) \t Wavefunctions squared (non-
normalized; offset by En): \t \t \t Vh (eV)" << endl;
for( int i=0; i < n; i++ )
{
    fn << ( ( i - n/2 ) * a ) << \t << ( V( ( i - n/2 ) * a, f,
gap ) / e + Vee[i] ) << "\t" ;

    for( int j=0; j < num; j++ )
    {
        fn << ( 10.0 * vec[i][j] * vec[i][j] ) + ( R[j] / e ) << \t ;
    }

    fn << Vh( ( i - n/2 ) * a, f, gap ) / e + Vee[i] << "\t" ;

    for( int j=0; j < num; j++ )
    {
        fn << ( -10.0 * vech[i][j] * vech[i][j] ) + ( Rh[j] / e ) << \t ;
    }

    fn << endl;
}

// output energy levels
fn << endl << "Energy levels: " << endl;
for( double z = zmin_W; z < zmax_W; z += a )
{  
    fn << z << "\t";
    for( int i = 0; i < num; i ++ )
        fn << R[i] / e << "\t";
    fn << "\t";
    for( int i = 0; i < num; i ++ )
        fn << -1.0 * Rh[i] / e << "\t";
    fn << endl;
}

fn.close();

return;

} /*
 */

/**********************************************************
These next several functions are
associated with the density of
states model
**********************************************************/

// This is the CB density of states
double Ze( double Ee )
{
    int num_subbands = 0;
    if( Ee > LB )
        return 0.0;

    // E_values[0] are the energies (in J) of the different subbands.
    // this determines how many subbands are contributing to the Z at Ee
    for( int i = 0; i < num; i ++ )
    {
        if( ( E_values[0][i] / e ) > LB )
            break;
        if( ( E_values[0][i] / e ) < Ee )
            num_subbands ++;
    }
    return ( (double)num_subbands ) * me * m0 / ( pi * hb * hb);
}
// This is the VB density of states
double Zh( double Eh )
{
    int num_subbands = 0;
    if( Eh < LBh )
        return 0.0;
    for( int i = 0; i < num; i ++ )
    {
        if( ( E_values[1][i] / e ) < LBh )
            break;
        if( ( E_values[1][i] / e ) > Eh )
            num_subbands ++;
    }
    return ( (double)num_subbands ) * mh * m0 / ( pi * hb * hb);
}

//Fermi function, electrons
double Fermi_e( double Ee, double Efcb )
{
    return 1.0 / ( 1.0 + exp( -1.0 * (Efcb - Ee ) / kT ) ) ;
}

//Fermi function, holes
double Fermi_h( double Eh, double Efvb )
{
    return 1.0 / ( 1.0 + exp( -1.0 * (Eh - Efvb ) / kT ) ) ;
}

/********************
These functions integrate over Z(E)F(E)dE to find total numbers of states in the CB or VB.
Using the assumptions on the CB quasi-Fermi-level these tune the VB quasi fermi level until the numbers of states in the two are equal
********************/
// This function requires care in selecting the initial guesses for the
// Newton-Raphson
// method. If the values aren't chosen well, the iteration will bounce
// from one side
// of the root to the other and not terminate. I think the equations
// we're finding
// roots of must have an unfortunate odd-function like symmetry; I think
// they're logistic
// in shape but I've neer bothered to find out for sure.
// Moral: be careful with initial guesses if you want convergence.
void get_Fhvb( double &Efvb, double &Efcb )
{
    ********************
    This function first finds the quasi-Fermi level for the
    CB that gives a number of CB states equal to the sheet carrier
    density \texttt{n\_sheet}.

    It then uses this same carrier density as the goal of the
    iteration on \texttt{Efvb} to find the VB fermi level.

    it uses a secant method (Newton-Raphson) to
    find the \texttt{Ef}, which is then put into the pass-by-reference \texttt{Efvb}, \texttt{Efcb}
    ********************

    double num_cb, num_vb; // hold number of states in CB, VB

    int N_integral = 100; // number of boxes to use in the integration
    routines

    double E_old, E, E_new; // these hold values during iteration
    // E_old is \texttt{E(n-1)} E is \texttt{E(n)}, and \texttt{Enew} is \texttt{E(n+1)}

    double N_new, N_cb, N_old;

    // error is the difference and value the currently calculated
    integral
    double error, value;

    // here, the initial guess for \texttt{Efcb} is \texttt{E}, and is set to the first
    subband energy
    E = E_values[0][0]/e;
    E_old = 0.9*E;

    N_cb = int_cb( N_integral, E );
    N_old = int_Cb( N_integral, E_old );
// this is an error check to make sure the Newton's method will terminate
int count = 0;

// Now for the Newton's method integration to find the Efcb that makes total number
// of states in the CB equal to n_sheet
do
{
    E_new = E - ( ( E - E_old ) / ( ( N_cb - N_old ) ) * ( N_cb - n_sheet ) );

    count ++;
    N_new = int_cb( N_integral, E_new );
    error = fabs( N_new - n_sheet );
    E_old = E;
    E = E_new;
    N_old = N_cb;
    N_cb = N_new;
}
while ( ( error > 0.1 * n_sheet ) && ( count < 10000 ) );

//If the Newton's method can't get convergence this
//crashes the program
if( ( error / n_sheet ) > 1.0 )
{
    cout << "Integration to find CB quasi-Fermi level failed to converge" << endl << "Crashing program.";
    cout << endl << N_new << endl;
    exit(1);
}
else
{
    // if the integration is reasonably close (<100% error) I'll just live with it.
}
// This routine uses a Newton-like secant method to iterate
// the Efvb to convergence.

// this is an error check to make sure the Newton's method
// will terminate
count = 0;

// Newton-Raphson requires two starting values for the Efvb
// I picked these two to make sure we had values that would
// lead to convergence.
E_old = E_values[1][0]/e;
E = 1.2 * E_old;

double N_vb = int_vb( N_integral, E );
N_old = int_vb( N_integral, E_old );

// Newton's method
// See Kincaid and Cheney, 1991
do{
    E_new = E - ( ( E - E_old ) / (( N_vb - N_old ) )
                   * ( N_vb - num_cb ) );
    count ++;
    N_new = int_vb( N_integral, E_new );
    error = fabs( N_new - num_cb );
    E_old = E;
    E = E_new;
    N_old = N_vb;
    N_vb = N_new;

    if( count > 100 )
    {
        cout << "Iteration to find VB quasi-Fermi level failed to converge!";
        cout << endl << N_new ;
    }
exit(0);
}
} while ( error > (0.00001 * num_cb) );

Efvb = E_new;

return;
}

// This uses a 5-point gaussian quadrature to integrate quickly and
// accurately over the CB to
// get the total number of states.
// N is the number of boxes to use
double int_cb(int N, double Ef)
{
    const double x0 = 0.0,
                 x1 = (1.0/3.0)*sqrt( 5.0 - 2.0 * sqrt( 10.0/7.0 ) ),
                 x2 = (1.0/3.0)*sqrt( 5.0 + 2.0 * sqrt( 10.0/7.0 ) ),
                 w0 = 128.0 / 225.0,
                 w1 = 0.3 * (-0.7 + 5.0 * sqrt( 0.7 ) ) / ( -2.0 + 5.0 *
                 sqrt( 0.7 ) ),
                 w2 = 0.3 * (0.7 + 5.0 * sqrt( 0.7 ) ) / ( 2.0 + 5.0 *
                 sqrt( 0.7 ) );

    double h;
    double xl,
            xr,
            s = 0.0,
            u,
            v,
            stemp = 0.0,
            ba;

    double xmax = LB;
    double xmin = Ef - 1.0;

    h = ( xmax - xmin );
    h /= (double) N ;

    xl = xmin;
    xr = xl + h;

    do
    {
        u = ( (xr-xl) * x0 ) + xl + xr ) / (2.0);
        stemp += w0 * Ze(u) * Fermi_e(u, Ef);
    }
\[
\begin{align*}
    u &= \frac{(x_r - x_l) * x_l + x_l + x_r}{2.0}; \\
    v &= \frac{-1.0 * (x_r - x_l) * x_l + x_l + x_r}{2.0}; \\
    \text{stemp} &= (w1 * (Ze(u) * \text{Fermi}_e(u, Ef) + Ze(v) * \text{Fermi}_e(v, Ef))); \\
    u &= \frac{(x_r - x_l) * x_2 + x_l + x_r}{2.0}; \\
    v &= \frac{-1.0 * (x_r - x_l) * x_2 + x_l + x_r}{2.0}; \\
    \text{stemp} &= (w2 * (Ze(u) * \text{Fermi}_e(u, Ef) + Ze(v) * \text{Fermi}_e(v, Ef))); \\
\end{align*}
\]

\[
\text{stemp} *= \frac{(x_r - x_l)}{2.0}; \\
\text{s} += \text{stemp}; \\
\text{stemp} = 0.0; \\
x_l = x_r; \\
x_r += h; \\
\}
\]

\[
\text{while}( x_r <= x_{\text{max}} );
\]

\[
\text{stemp} = 0.0; \\
x_r = x_{\text{max}}; \\
u &= \frac{(x_r - x_l) * x_0 + x_l + x_r}{2.0}; \\
\text{stemp} += w0 * Ze(u) * \text{Fermi}_e(u, Ef); \\
u &= \frac{(x_r - x_l) * x_2 + x_l + x_r}{2.0}; \\
v &= \frac{-1.0 * (x_r - x_l) * x_2 + (x_l) + (x_r)}{2.0}; \\
\text{stemp} += (w1 * (Ze(u) * \text{Fermi}_e(u, Ef) + Ze(v) * \text{Fermi}_e(v, Ef))); \\
v &= \frac{(x_r - x_l) * x_1 + x_l + x_r}{2.0}; \\
\text{stemp} *= \frac{(x_r - x_l)}{2.0}; \\
\text{s} += \text{stemp}; \\
\text{return s;} \\
\}
\]

// This uses a 5-point gaussian quadrature to integrate quickly and accurately over the VB to get the total number of states. // N is the number of boxes to use, and Efvb the assumed quasi-Fermi level for this attempt.

double int_vb(int N, double Efvb)
{
    const double x0 = 0.0,
                x1 = (1.0/3.0)*sqrt( 5.0 - 2.0 * sqrt( 10.0/7.0 ) ),
                x2 = (1.0/3.0)*sqrt( 5.0 + 2.0 * sqrt( 10.0/7.0 ) ),
                w0 = 128.0 / 225.0,
                w1 = 0.3 * (-0.7 + 5.0 * sqrt( 0.7 ) ) / ( -2.0 + 5.0 * sqrt( 0.7 ) ),
                w2 = 0.3 * (-0.7 + 5.0 * sqrt( 0.7 ) ) / ( -2.0 + 5.0 * sqrt( 0.7 ) );
}
\[ w_2 = 0.3 \times (0.7 + 5.0 \times \sqrt{0.7}) / (2.0 + 5.0 \times \sqrt{0.7}) ; \]

double h;

double xl, xr,
s = 0.0,
u,
v,
stemp = 0.0,
ba;

double xmax = \left( E_{\text{values}[1][0]/e} + 0.1 \right) ;
double xmin = LBh;

h = (xmax - xmin);
h /= (double)N;

xl = xmin;
xr = xl + h;

do
{
    u = \left( (xr-xl) \times x0 \right) + xl + xr \right) / (2.0);
stemp += w0 * Zh(u) * Fermi_h(u, Efvb);
    u = \left( (xr-xl) \times x1 \right) + xl + xr \right) / (2.0);
v = \left( -1.0 \times (xr-xl) \times x1 \right) + xl + xr \right) / (2.0);
stemp += \left( w1 \times \right) \left( Zh(u) \times Fermi_h(u, Efvb) \right) + Zh(v) \times Fermi_h(v, Efvb) \right) ;
    u = \left( (xr-xl) \times x2 \right) + xl + xr \right) / (2.0);
v = \left( -1.0 \times (xr-xl) \times x2 \right) + xl + xr \right) / (2.0);
stemp += \left( w2 \times \right) \left( Zh(u) \times Fermi_h(u, Efvb) \right) + Zh(v) \times Fermi_h(v, Efvb) \right) ;
        stemp *= \left( (xr-xl) \right) / (2.0);
        s += stemp;
        stemp = 0.0;
        xl = xr;
xr += h;
} while(xr <= xmax);

stemp = 0.0;

xr = xmax;

u = \left( (xr-xl) \times x0 \right) + xl + xr \right) / (2.0);
stemp += w0 * Zh(u) \times Fermi_h(u, Efvb) ;
    u = \left( (xr-xl) \times x1 \right) + xl + xr \right) / (2.0);
v = \left( -1.0 \times (xr-xl) \times x1 \right) + xl + xr \right) / (2.0);
stemp += ( w1 * ( Zh( u ) * Fermi_h( u, Efvb ) + Zh( v ) * Fermi_h( v, Efvb ) ) );
    u = ( (xr-xl)* x2 + xl + xr ) / (2.0);
    v = ( -1.0*(xr-xl) * x2 + (xl) + (xr) ) / (2.0);
    stemp += ( w2 * ( Zh( u ) * Fermi_h( u, Efvb ) + Zh( v ) * Fermi_h( v, Efvb ) ) );
    stemp *= ( (xr-xl) / (2.0) );
    s += stemp;
    return s;
}

// this is a debugging routine where the program outputs the number that is the call
// to this function
void tick( int c )
{
    cout << c;
}

/********************
This is a class to act for histogram data.
Usage: include this header file
define an object of type "hist" with the minimum and maximum values of the E-axis (doubles) and the number of bins (int)

hist I( 0.0, 5.0, 100 );

For a 100-bin histogram of minimum value 0.0 and maximum value 5.0
It is important to use the (double, double, int ) constructor.

hist I;
I've got no idea how that will work, if at all. Don't do it.

Methods:
I.find_pea()` returns the double value of the peak bin
I.find_peak_index() return the int index of the peak
I.add( 3.2, 215.23 ) adds the value 215.23 to the value already in the bin to which 3.2 belongs
I.get_bin_index( 3.2 ) returns the int index of the bin to which an E of 3.2 is assigned
I.get_E( 23 ) returns the double E value associated with that bin
I.zero() zeros out the histogram
I.get_value( 3.2 ) returns the value of the histogram in the bin
associated with 3.2.
I.get_value_i( 24 ) returns the value of bin 24 (actually the 25th -- it
goes
form 0 to n-1 )

I chose not to use overloading on get_value, get_value_i, etc., to avoid
confusion.

****************************/

class hist
{
    public:
        hist( double E0, double E1, int s );        // useful
    constructor
        double find_peak();     // find E of peak value
        int find_peak_index();  // find i of peak
        void add( double E, double P );   // add a P(E) to the
    histogram
        int get_bin_index( double E );          // get bin index holding E
        double get_E( int i );                // get E of bin index i
        void zero( void );                    // set histogram to zero
        double get_value( double E );         // get P for some E
        double get_value_i( int i );
        void normalize( void );              //hist( void );

    private:
        double* I;                        // the histogram
        double Emin;                      // minimum allowed E
        double Emax;                      // max E
        double deltaE;                    // E per bin
        void set_to_zero( void );         // zeros the array
        int size;-
};
hist::hist( double E0, double E1, int s )
{
    // It is important to use this constructor.

    // These four variables are all protected data members that must be initialized
    size = s;
    Emax = E1;
    Emin = E0;
    deltaE = ( Emax - Emin ) / size;

    // This sets up an array of double in the free memory, of arbitrary size n.
    // The only way to allow the user to specify a different n for any hist object
    // was this dynamic memory handling
    I = new double[ size ];

    // Zero out the array
    set_to_zero();

    return;
}

void hist::set_to_zero( void )
{
    for( int i = 0; i < size; i++ )
        I[ i ] = 0.0;

    return;
}

void hist::add( double E, double P )
{
    int i;

    if( E < Emin ) // trap error: E below the start of the array
        return;
    if( E > Emax ) // trap error: E past end of the array
        return;

    return;
}
i = (int)( ( E - Emin ) / deltaE );
I[ i ] += P;
}

double hist::get_value( double E )
{
    int i;
    i = (int)( ( E - Emin ) / deltaE );
    return I[ i ];
}

int hist::get_bin_index( double E )
{
    int i;
    i = (int)( ( E - Emin ) / deltaE );
    return i;
}

double hist::get_E( int i )
{
    return ( (double)i * deltaE + Emin );
}

void hist::zero( void )
{
    set_to_zero();
    return;
}

double hist::find_peak( void )
{
    double a = I[ 0 ];
    int b = 0;

    for( int i = 1; i < size; i ++ )
    {
        if( fabs( I[ i ] ) > fabs( a ) )
        {
            b = i;
            a = I[ i ];
        }
    }

    return ( ( (double)b ) * deltaE + Emin );
int hist::find_peak_index( void )
{
    double a = I[0];
    int b = 0;
    for( int i = 1; i < size; i++ )
    {
        if( I[i] > a )
        {
            b = i;
            a = I[i];
        }
    }
    return b;
}

double hist::get_value_i( int i )
{
    return I[i];
}

void hist::normalize( void )
{
    double max = I[0];
    for( int i = 0; i < size; i++ )
    {
        if( I[i] > max )
            max = I[i];
    }
    if( max <= 0.0 )
        return;
    for( int i = 0; i < size; i++ )
        I[i] /= max;
    return;
}

hist::~hist( void )
{
    // This destructor is important; without it,
// the memory assigned to I is lost to the program
// (memory leak)

delete I;
A.4 Appendix 4: Imaging circuit description

Signal in
From Keithley 428 current amplifier
-10 to +10V
Coaxial BNC connector

Signal out
To SEM aux port
-10 to +10V
Coaxial BNC connector

DC power supply:
-15V
0V
+15V

Figure 167: Block diagram of the imaging circuit.
Figure 168: Circuit schematic diagram of the imaging circuit.
A.5 Appendix 5: MathCAD 12 code for Donolato-Luke curve fitting

This simulates EBIC profiles using the model of Luke, J APPL PHYS, V77(7), 1995, P.3257. This is based on the model of Donolato, SOLID STATE ELECTRONICS, V25(11), 1982, P.1077.

In the case of \( \sigma = 0 \), this reduces to Donolato’s equation.

The important point to remember about the Donolato model is that it’s no good for values of \( x < \frac{1}{2} \) the electron range, where the origin of \( x \) is the peak of the EBIC curve.

\[
\sigma(R) := \frac{R}{\sqrt{15}}
\]

This is a depth function, in microns

\[
Z_0(R) := 0.3R
\]

Another depth function, microns

\[
\lambda(L) := \frac{1}{L}
\]

A simple substitution

\[
\mu(k, L) := \sqrt{k^2 + \lambda(L)^2}
\]

Another substitution

\[
d = 0.030
\]

Electron beam FWHM

\[
\sigma_w := \frac{d}{2 \cdot \sqrt{2 \cdot \ln(2)}}
\]

Electron beam spot size, \( 1 / \sqrt{\text{e}} \) half-width, microns

\[
R_{K-O} := \frac{0.05}{\lambda} \quad \text{The K-O electron range of the beam for the given material}
\]

\[
I(\pi R, L, S, R) := \frac{2}{\pi} \left[ \frac{k}{\mu(k, L)^3} \right]^{1000} \left[ \frac{-k^2 (\sigma(R)^2 + \sigma w^2)}{2 \cdot \lambda(L)^2} \right] e^{-0.57 \cdot e - 0.5} \cdot \mu(k, L, R)
\]

\[
L) \quad \frac{Z_0(R)}{\mu(k, L)} + \left[ 1 - \frac{\sigma(R)}{\sqrt{2}} \left[ \mu(k, L) - \frac{Z_0(R)}{\sigma(R)^2} \right] \right] \sin(k \cdot x_x) \, dk
\]

EBICdat :=

<table>
<thead>
<tr>
<th>$x$</th>
<th>$y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0.997</td>
</tr>
<tr>
<td>3</td>
<td>0.999</td>
</tr>
<tr>
<td>4</td>
<td>0.982</td>
</tr>
<tr>
<td>5</td>
<td>0.984</td>
</tr>
<tr>
<td>6</td>
<td>0.970</td>
</tr>
<tr>
<td>7</td>
<td>0.956</td>
</tr>
<tr>
<td>8</td>
<td>0.950</td>
</tr>
<tr>
<td>9</td>
<td>0.940</td>
</tr>
<tr>
<td>10</td>
<td>0.925</td>
</tr>
<tr>
<td>11</td>
<td>0.915</td>
</tr>
</tbody>
</table>

| 0.040 | 0.875 |
| 0.043 | 0.859 |
| 0.048 | 0.825 |
\( \max = 137 \quad \min = 29 \quad \text{pivot} = \min \)

\( x := \text{EBIC.dat}^{(0)} \quad i := \text{EBIC.dat}^{(1)} \)

\( xx := x_{\min} \cdot (x_{\min} + 0.005) \quad x_{\max} \)

\( I_{\nu} = 0.15 \quad S_{\nu} = 10^{-2} \)

\( I_{\text{plot}}(xx, L, S, R) := \frac{I(xx, L, S, R) \cdot \text{pivot}}{I(x_{\text{pivot}}, L, S, R)} \)

\( x_{\min} = 0.126 \quad x_{\max} = 0.596 \)

Here, set up a smaller value of \( \max \) in...
\( m_{\text{max}} = 0.435 \) 

order to cut the noise out of the R^2 calculation

\( i_1 := \text{submatrix}(1, \text{min}, \text{max}, 0, 0) \)

\( \text{num} := \text{max} - \text{min} \)

\( \text{count} := 0 \ldots \text{num} \)

\( \text{Idata}_\text{count} := \text{Iplot}(\text{count} \cdot \text{min}, L, S, R) \)

\[
\text{corr} (\text{Idata}, i_1)^2 = 0.99187363
\]

This is the R^2 value of the fit

\( \text{start} := m_{\text{min}} \)

\( \text{end} := m_{\text{max}} \)

\[
\text{inc} = \frac{\text{end} - \text{start}}{\text{max} - \text{min}}
\]

\( i := 0 \ldots \frac{\text{end} - \text{start}}{\text{inc}} \)

\( \text{dat}_{i,0} := i \cdot \text{inc} + \text{start} \)

\[
\text{dat}_{i,1} := \text{Iplot}(\text{dat}_{i,0}, L, S, R)
\]

\[
\text{dat}_{i,2} := I(\text{dat}_{i,0}, L, S, R)
\]

\[
\begin{array}{ccc}
0 & 1 & 2 \\
0 & 0.125 & 0.483 & 0.444 \\
1 & 0.13 & 0.469 & 0.432 \\
2 & 0.135 & 0.456 & 0.42 \\
3 & 0.139 & 0.443 & 0.408 \\
4 & 0.143 & 0.431 & 0.396 \\
5 & 0.148 & 0.418 & 0.385 \\
6 & 0.152 & 0.406 & 0.374 \\
7 & 0.157 & 0.395 & 0.363 \\
8 & 0.161 & 0.384 & 0.353 \\
9 & 0.165 & 0.373 & 0.343 \\
10 & 0.17 & 0.362 & 0.333 \\
11 & 0.174 & 0.352 & 0.324 \\
12 & 0.178 & 0.342 & 0.314 \\
13 & 0.183 & 0.332 & 0.305 \\
14 & 0.187 & 0.322 & 0.297 \\
15 & 0.191 & 0.313 & 0.288 \\
\end{array}
\]

\( \text{dat}_{i,0} = \\
0.126 \\
0.13 \\
0.135 \\
0.139 \\
0.143 \\
0.148 \\
0.152 \\
0.157 \\
0.161 \\
0.165 \\
0.165 \\
0.17 \\
0.174 \\
0.178 \\
0.183 \\
0.187 \\
0.191 \\
\)

\( \text{dat} := \text{WRITEPRN}("c:\Chad\work\ebic modelling\ebicdata.txt", \text{dat}) \)

374
A.6 Appendix 6: MathCAD 12 code for Bonard and Ganiere curve fitting

This Mathcad worksheet will be an attempt to replicate the EBIC mathematical model of:

J.-M. Bonard and J.-D. Ganiere. "Quantitative analysis of electron-beam-induced current profiles across p-n junctions in GaAs/Al0.4Ga0.6As heterostructures." J. APPL. PHYS., V79(9), 1996.
P.8667-8664

\[
\mu(k, L) := \sqrt{k^2 + \left(\frac{1}{L}\right)^2}
\]

This is a simple substitution

In this program, Se,\,hDe,h will be represented as the variables se or sh

\[\sigma1 := .0416\]

These functions I obtained from using CASINO to perform Monte Carlo simulations of GaN at different beam energies.

\[\sigma2 := 0.0123\]

db is the spot size in nanometers (not microns)

\[\delta j := .020\]

Width of the depletion region, microns

\[\frac{A_1}{A_2} := 10^{-1}\]

Ratios of \(S\) (cm/sec) to \(D\) (cm²/sec)

\[\lambda_h := 5\]

HOWEVER units here must be 1/\(\mu\)m
Multiply by \(10^4\) to get 1/cm

\[se := 5\]

\[L_e := .1\]

Minority carrier diffusion lengths, in microns

\[L_h := .045\]

h, e represent holes, electrons if p-side is on the left
Otherwise, reversed.

The following two \(\Theta\) expressions are substituted into the final EBIC expression

For simplicity, each theta has been broken up into four parts such that \(\Theta=rtuv\)
\[ \text{re}(k, x) := \left( \frac{1}{k^2 + s e^2} \right) \text{erfc} \left[ \frac{\sigma_1 \left( \mu(k, L) - 2 - \frac{x}{\sigma_1^2} \right)}{2} \right] \]

\[ \text{te}(k, x) := \exp \left( \frac{\sigma_1 \cdot \mu(k, L)}{2} \right)^2 \]

\[ \text{ue}(k, x) := \exp \left( -\mu(k, L) \cdot x \right) \left( 1 + k \cdot \sigma_2^2 \right)^3 \]

\[ \text{ve}(k, x) := \left[ 1 + 3 \cdot s e \sigma_2 - k^2 \left( s e \cdot \sigma_2^3 + 3 \sigma_2^2 \right) \right] \]

\[ \Theta_e(k, x) := \text{re}(k, x) \cdot \text{te}(k, x) \cdot \text{ue}(k, x) \cdot \text{ve}(k, x) \]

\[ \text{rh}(k, x) := \left( \frac{1}{k^2 + s h^2} \right) \text{erfc} \left[ \frac{\sigma_1 \left( \mu(k, L) - 2 - \frac{x}{\sigma_1^2} \right)}{2} \right] \]

\[ \text{th}(k, x) := \exp \left( \frac{\sigma_1 \cdot \mu(k, L)}{2} \right)^2 \]

\[ \text{uh}(k, x) := \exp \left( -\mu(k, L) \cdot x \right) \left( 1 + k \cdot \sigma_2^2 \right)^3 \]

\[ \text{vh}(k, x) := \left[ 1 + 3 \cdot s h \sigma_2 - k^2 \left( s h \cdot \sigma_2^3 + 3 \sigma_2^2 \right) \right] \]

\[ \Theta_h(k, x) := \text{rh}(k, x) \cdot \text{th}(k, x) \cdot \text{uh}(k, x) \cdot \text{vh}(k, x) \]
Now for the actual EBIC trace function. The original paper calls it $\eta$. This terminology irritates me, but I think there's no reason to change it--I'm trying to replicate the original work at this point.

$$
\eta(x, z, dj) = \left( \frac{1}{\pi} \right) \int_0^{\infty} \left[ \text{sh} \cdot \Theta h(k, x) + \text{se} \cdot \Theta e(k, -(x + dj)) \right] \text{dk} + 0.5 \left( \text{erf} \left( \frac{x + dj}{\sigma_1} \right) - \text{erf} \left( \frac{x}{\sigma_1} \right) \right)
$$

\[
\begin{align*}
\text{start} & := -0.3 & \text{inc} & := 0.0025 & \text{end} & := 0.3 \\
x & := \text{start}, (\text{start} + \text{inc}) ... \text{end} & z & := 100
\end{align*}
\]

\[
\text{input_data} := \begin{bmatrix}
0 & 0.6 & 8.367 \times 10^{-3} \\
1 & -0.596 & 7.181 \times 10^{-3} \\
2 & -0.591 & -1.826 \times 10^{-3} \\
3 & -0.587 & -4.64 \times 10^{-4} \\
4 & -0.583 & 8.267 \times 10^{-2} \\
5 & -0.576 & 6.869 \times 10^{-2} \\
6 & -0.574 & 2.766 \times 10^{-2}
\end{bmatrix}
\]

\[
\text{xdata} := \text{input_data}_0 & \quad \text{idata} := \text{input_data}_1
\]
$\circ := .01$

Fudge factors to get peaks to line up; the $\eta$ function is not self-normalizing

$f := 1.2119$
Here, set up values of max and min in order to cut the noise out of the R^2 calculation.

\[ x_{\text{data min}} = -0.209 \quad x_{\text{data cent}} = 0 \quad x_{\text{data max}} = 0.204 \]

1: \( \text{submatrix(idata, min, cent, 0, 0)} \)
\[ \text{num} := \text{cent} - \text{min} \]
\[ \text{count} := 0 .. \text{num} \]
\[ I_{1 \text{count}} := \frac{\eta[[x_{\text{data cent}} - \text{min} - o], z, d]}{f} \]
\[ \text{corr}(I_{1}, i1)^2 = 0.99994938 \]

2: \( \text{submatrix(idata, cent, max, 0, 0)} \)
\[ \text{num} := \text{max} - \text{cent} \]
\[ \text{count} := 0 .. \text{num} \]
\[ I_{2 \text{count}} := \frac{\eta[[x_{\text{data cent}} - \text{cent} - o], z, d]}{f} \]
\[ \text{corr}(I_{2}, i2)^2 = 0.99931154 \]
\[ \text{EBICdat}_{t, 0} := i \times \text{inc} + \text{start} \]

\[ \text{EBICdat}_{t, 1} := \eta[(\text{EBICdat}_{t, 0}), z, d] \]

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-0.2</td>
<td>0.092</td>
</tr>
<tr>
<td>1</td>
<td>-0.297</td>
<td>0.098</td>
</tr>
<tr>
<td>2</td>
<td>-0.295</td>
<td>0.097</td>
</tr>
<tr>
<td>3</td>
<td>-0.292</td>
<td>0.099</td>
</tr>
<tr>
<td>4</td>
<td>-0.29</td>
<td>0.102</td>
</tr>
<tr>
<td>5</td>
<td>-0.288</td>
<td>0.104</td>
</tr>
<tr>
<td>6</td>
<td>-0.285</td>
<td>0.107</td>
</tr>
<tr>
<td>7</td>
<td>-0.282</td>
<td>0.11</td>
</tr>
<tr>
<td>8</td>
<td>-0.28</td>
<td>0.112</td>
</tr>
<tr>
<td>9</td>
<td>-0.277</td>
<td>0.115</td>
</tr>
<tr>
<td>10</td>
<td>-0.275</td>
<td>0.118</td>
</tr>
<tr>
<td>11</td>
<td>-0.272</td>
<td>0.121</td>
</tr>
<tr>
<td>12</td>
<td>-0.27</td>
<td>0.124</td>
</tr>
<tr>
<td>13</td>
<td>-0.267</td>
<td>0.127</td>
</tr>
<tr>
<td>14</td>
<td>-0.265</td>
<td>0.131</td>
</tr>
<tr>
<td>15</td>
<td>-0.263</td>
<td>0.134</td>
</tr>
</tbody>
</table>

\[ \text{EBICdat}_t = \]

\[ \text{cut} := \text{WRITEPRN}("c:\text{Class\work\ebic\modelling\ebicdata.txt"}, \text{EBICdat}) \]
A.7 Appendix 7: Maple 9.5 code for Method of Moments solver

```maple
restart;
eqn1 := mL = L^2*(l-(sL/(1+sL)))*exp(-z1/L);

eqn1 := mL = L^2 \left( 1 - \frac{sL e^{\frac{-z1}{L}}}{1 + zL} \right)

eqn2 := m2 = L^2*(1-(sL/(1+sL)))*exp(-z2/L);

eqn2 := m2 = L^2 \left( 1 - \frac{sL e^{\frac{-z2}{L}}}{1 + zL} \right)

> m1 := 335; m2 := 465; z1 := 1.3; z2 := 2.7;
ml := 335
m2 := 465
z1 := 1.3
z2 := 2.7

> fsolve( {eqn1, eqn2}, {L, sL} );
{sL = 43.24277015, L = 96.98318179}
>