UC San Diego
UC San Diego Electronic Theses and Dissertations

Title
Thermoelectric and electrical characterization of Si nanowires and GaNAs

Permalink
https://escholarship.org/uc/item/2h7846vd

Author
Pichanusakorn, Paothep

Publication Date
2012

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Thermoelectric and electrical characterization of Si nanowires and GaNAs

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Materials Science and Engineering

by

Paothep Pichanusakorn

Committee in charge:

Professor Prabhakar R. Bandaru, Chair
Professor Renkun Chen
Professor Massimiliano Di Ventra
Professor Charles W. Tu
Professor Jie Xiang

2012

Copyright
Paothep Pichanusakorn, 2012
All rights reserved.

The dissertation of Paothep Pichanusakorn is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

____

____

____

____

____

Chair

University of California, San Diego

2012

DEDICATION

To my mother, whom I love more than anyone in the world.
To my late brother and father, whom I missed more than anyone in the world.

TABLE OF CONTENTS

Signature Page . . . iii

Dedication . . . iv

Table of Contents . . . v

List of Figures . . . viii

List of Tables . . . xiii

Acknowledgements . . . xiv

Vita and Publications . . . xvi

Abstract of the Dissertation . . . xvii

1 Introduction to Thermoelectric phenomena and theory . . . 1
 1.1 Thermoelectric Effects . . . 1
  1.1.1 Seebeck effect . . . 1
  1.1.2 Peltier and Thompson effects . . . 4
 1.2 Efficiency and ZT . . . 4
 1.3 Transport coefficients . . . 7
  1.3.1 Density of states . . . 8
  1.3.2 Boltzmann transport equation . . . 9
  1.3.3 Carrier velocity . . . 11
  1.3.4 Relaxation time . . . 12
  1.3.5 The χ_i integral . . . 14
  1.3.6 Summary of equations for S, σ, and κ_e . . . 15
 1.4 Transport coefficients as function of reduced Fermi level . . . 16
  1.4.1 Carrier concentration . . . 16
  1.4.2 Electrical conductivity and Mobility . . . 19
  1.4.3 Thermal conductivity . . . 22
  1.4.4 Seebeck coefficient . . . 24

2 Power Factor Enhancement . . . 28
 2.1 Maximization of Z . . . 28
 2.2 The Optimal Reduced Fermi Level . . . 29
  2.2.1 Survey of materials . . . 31
 2.3 Density of States Enhancement . . . 37
  2.3.1 Confinement Effect . . . 38
  2.3.2 Minimum length . . . 41
  2.3.3 Effective Mass enhancement . . . 43

3 Measurement Methods 46
3.1 Measurement device layout 46
3.1.1 Device fabrication 48
3.2 External equipments 51
3.3 Measurement procedure 54
3.3.1 Van der Pauw and Hall method: ρ, n, and μ 54
3.3.2 FET method: ρ, μ and n 56
3.3.3 RTD Calibration 67
3.3.4 Seebeck measurement 70

4 Thermoelectric characterization of Si nanowires 78
4.1 Quantum confinement effect on S of Si NWs 80
4.2 Doping of SOI substrate 81
4.2.1 Diffusion doping 83
4.2.2 Ion implantation 88
4.3 Electron Beam Lithography 95
4.3.1 e-beam settings 96
4.3.2 Dosage and e-beam resist 97
4.3.3 Pattern design 104
4.4 Oxidation of Si nanowires 106
4.4.1 Self-limiting oxidation effect 106
4.4.2 Oxidation rate of planar Si and Si nanowires 111
4.4.3 Dopant segregation 114
4.5 Measurement results 119
4.5.1 FET measurement of thin film 119
4.5.2 FET measurement of Si Nanowires 126
4.5.3 Effect of oxidation on Si Thin film and Nanowires 131
4.6 Conclusion 137

5 Influence of N on effective mass and thermoelectric property of GaNAs . 139
5.1 Introduction and Motivation 139
5.1.1 Thermoelectric application for highly-mismatch alloy 139
5.1.2 Existing work on Dilute Nitride 140
5.2 Sample preparation 143
5.2.1 Growth and sample specification 143
5.2.2 Plasma etch and Ohmic contact for GaAs/GaN_xAs_{1-x} . . . 146
5.3 Electrical and Thermoelectric characterization of GaAs and GaN_xAs_{1-x} 147
5.3.1 Compensation and Passivation effects 147
5.3.2 Mobility 152
5.3.3 Seebeck Coefficient 154
5.4 Discussion 156
5.4.1 Effective Mass 156

A Symbols and Acronyms 163
Bibliography 168

LIST OF FIGURES

Figure 1.1: Seebeck effect gives rise to voltage when there is temperature difference between junctions of metals A and B. 2
Figure 1.2: Diffusion (solid arrow) and drift (dashed arrow) current of electrons with (a) zero-temperature gradient, and (b) non-zero temperature gradient. (c) shows additional effect of phonon-drag. . 3
Figure 1.3: Thermoelectric module as (a) electricity generator and (b) Peltier cooler. 5
Figure 1.4: Density of states for 3-, 2-, and 1-dimensional material. 9
Figure 1.5: Normalized carrier concentration of a *single* band/subband as a function of the reduced Fermi level. Top axis indicates $E_F = \eta k_B T$ for $T = 300$K. 19
Figure 1.6: τ_A as a function of the reduced Fermi level. Top axis indicates $E_F = \eta k_B T$ for $T = 300$K. 21
Figure 1.7: Lorentz number as a function of the reduced Fermi level. Top axis indicates $E_F = \eta k_B T$ for $T = 300$K. 23
Figure 1.8: Seebeck coefficient as a function of the reduced Fermi level. Figure is split into 2 panes for clarity; Y-axis units are same for both panes. 26

Figure 2.1: Maximization of Z. 29
Figure 2.2: Normalized $P(\eta)$ function showing optimal reduced Fermi level. 31
Figure 2.3: Power factor of PbTe versus carrier concentration [1]. 33
Figure 2.4: Power factor of PbTe [1], and La- and Nb-doped $SrTiO_3$ [2, 3]. Data label indicates n in cm^{-3}. 34
Figure 2.5: Seebeck coefficient and power factor of multiple p-type SiGe samples [4] 35
Figure 2.6: Power factor versus Seebeck coefficient of Bi_2Te_3 [5]. Each series comprises measurement from the same sample measured over temperature range of approximately 80-350K. 36
Figure 2.7: Electron density in bulk Si conduction band when the Fermi level is (a) -0.05eV, and (b) $+0.1$eV. The density of states is shown as a solid black line, while the electron density is shown by the grayed area. 39
Figure 2.8: Electron density in conduction band of Si nanowires with diameter of (a) 2 nm, and (b) 5 nm. DOS of bulk Si also plotted for comparison. 40
Figure 2.9: The minimum confinement length of quantum well and nanowire required to achieve (a) gain factor of 1 at various temperature and effective mass, and (b) gain factor between 0.5 to 4 at 300K. 42
Figure 2.10: Distortion of PbTe density of states around E_R due to Tl-resonant states. Adapted from [6]. 44

Figure 2.11: Effective mass and Fermi level comparison of Tl-PbTe and standard Na-PbTe. Adapted from [6]. 44

Figure 3.1: Various stages in the development of a measurement device. . . 47
Figure 3.2: Measurement device layout for measurement of ρ, μ, n, and S using (a) van der Pauw and (b) field-effect transistor configurations. (c) shows the micro-circuit as part of a complete device featuring large gold pads for probes and wire-bonds. 49
Figure 3.3: Photolithography process flow for VDP and FET device fabrication. 52
Figure 3.4: Continuing process flow for gate definition of FET device. Refer to Fig. 3.3 for legend. 53
Figure 3.5: The multi-probe card and sample loading stage. 54
Figure 3.6: System block diagram 55
Figure 3.7: Van der Pauw measurement configuration. 57
Figure 3.8: (a) Drain current versus gate voltage curve of n-channel bulk Si MOSFET taken at 50 mV drain voltage. (b) Equivalent total conductance (G_{tot}) and channel conductance (G_{ch}) versus gate voltage of the device. 60
Figure 3.9: SOI-MOSFET structure and equivalent circuit diagram. 62
Figure 3.10: (a) Measured capacitance of device with and without channel. (b) Oxide capacitance. 66
Figure 3.11: CV curve of n-type MOS 67
Figure 3.12: Plot of RTD resistance versus temperature profile. 69
Figure 3.13: Percentage increase in TCR and $|S|$ as a result of placing thermocouple directly on the sample (relative to placing thermocouple on a nearby chip). Results of 6 samples are shown. . . . 70
Figure 3.14: Time chart showing evolution of the temperature at drain- and source-side RTDs. The Seebeck voltage across drain and source is also showed on a bottom graph. Data taken from the same sample in Fig. 3.12. 72
Figure 3.15: Plot of the Seebeck voltage across drain and source contacts, $V = V_{ds}$, versus drain-source temperature difference, $\Delta T = T_d - T_s$. Linear fit (dashed line) indicates that $S = -163\,\mu\mathrm{V/K}$. Data taken from the same sample (n-Si) in Fig. 3.14. 73
Figure 3.16: Goodness of fit, as represent by the R^2 value, and $|S|$ both increases with ΔT as the constant δT is rendered less significant. 74
Figure 3.17: Average temperature increases with ΔT due to greater heating power requirement. Data are shown for multiple SOI and GaAs samples. 75
Figure 3.18: Comparison between the measured and calculated value of $|S|$ as a function of average temperature for a GaAs sample with $n = 3.2 \times 10^{17}\,\mathrm{cm}^{-3}$. 76

Figure 3.19: Seebeck coefficient versus carrier concentration for n-GaAs, p-Si, and n-Si as measured in this work (filled) and from literatures (open). GaAs: [7, 8, 9, 10], Si: [11, 12]. Dashed lines show the calculation results. 77

Figure 4.1: Seebeck coefficient of Si nanowires as a function of nanowires diameter. 80
Figure 4.2: Si etching recipe and etched thickness versus etch time. Two distinct etch rate is observed for Si top layer (filled diamond) and the buried oxide (open diamond), which is exposed after all 260 nm of Si is etched away. 82
Figure 4.3: Carrier concentration versus annealing temperature of P- and B-doped Si film. Doping is via thermal diffusion. Annealing time is 2 min for $T < 1,000\,^{\circ}$C, and 1 min for $T > 1,000\,^{\circ}$C. . . 86
Figure 4.4: Diffusivity of P and B in Si as a function of dopant/carrier concentration at 700 °C, 900 °C, and 1,100 °C. 86
Figure 4.5: Mobility versus carrier concentration of P- and B-doped SOI samples. Doping is via thermal diffusion. Solid and dashed lines are empirical model of P- and B-doped Si [13]. 87
Figure 4.6: Surface defects of diffusion doped Si top layer on SOI substrate. 89
Figure 4.7: Mobility and carrier concentration of ion-implanted Si thin film. Solid and dashed lines represent empirical model [13] 91
Figure 4.8: Solid-solubility and electrically-active impurity concentration limit in Si at high temperature. Adapted from Ref. [14]. 93
Figure 4.9: Seebeck coefficient as function of carrier concentration for N34 and N35 Si thin films. Solid ($r = 0$) and dashed lines ($r = -1/2$) represent first-principle calculation. 94
Figure 4.10: Process flow for e-beam lithography using (A) positive-tone resist, (B) negative-tone resist, and (C) positive-tone resist with reverse-polarity image. 99
Figure 4.11: Dosage test on 80 nm950PMMA A2 e-beam resist. Dosage of 10 kV exposure is indicated by the top axis, while that of 30 kV is indicated by the bottom axis. 100
Figure 4.12: SEM images of nanowires patterned with ∼80 nm diluted ma-N 2403 resist over a range of dosage. 102
Figure 4.13: SEM images of Si nanowires fabricated with reverse-polarity pattern ((C) process in Fig. 4.10). 103
Figure 4.14: SEM images of Si nanowires array with (a),(b) direct contact on nanowires, and (c),(d) contact on bulk Si. 105
Figure 4.15: NWs array arrangement on a chip, and the effect of tilt on line width. 107
Figure 4.16: Oxide thickness versus oxidation time for planar Si, and Si nanowires. Adapted from [15]. 108

Figure 4.17: Limiting Si core width and oxide thickness of Si nanowires with different starting width, and oxidation temperature. Adapted from [16]. 108
Figure 4.18: Oxidation rate of planar Si and Si nanowires at 1,100 °C. Adapted from Ref. [17] 110
Figure 4.19: TEM images showing Si/SiO_2 core/shell nanowires oxidized at 1,100 °C for (a) 100 s, (b) 250 s, and (c) 600 s. Si core (bright area) is surrounded by the thermally-grown oxide shell (dark gray), and the deposited Si_2N_3 encapsulation layer (medium gray). Adapted from Ref. [18]. 110
Figure 4.20: Timing profile for oxidation in RTA furnace. 111
Figure 4.21: Oxidation of Si thin film (undoped SOI, N34, and N35) in RTA at 1,100 °C in dry oxygen. 113
Figure 4.22: Oxide thickness of planar Si and Si NWs after oxidation in RTA at 1,100 °C in dry oxygen. 115
Figure 4.23: Dopant segregation where impurity is (a) absorbed by the oxide and depleted from Si, and (b) rejected by the oxide and piled up in Si. Adapted from Ref. [19]. 116
Figure 4.24: P concentration versus Si/SiO_2 thicknesses (a) before, and (b) after oxidation in planar Si. 117
Figure 4.25: Idealized cross-section of rectangular nanowire on BOX layer (a) before, and (b) after oxidation. 118
Figure 4.26: Segregation ratio (of dopant concentration in Si after and before oxidation) for planar and nanowires geometry. 119
Figure 4.27: Channel conductance versus gate voltage of (a) N34 and (b) N35 TF-FET devices with 60 nm gate oxide. $V_d = 50$ mV. . . . 121
Figure 4.28: Electrical properties of Si TF-FET with 60 nm gate oxide. . . . 124
Figure 4.29: Electrical properties of Si TF-FET with 40 nm gate oxide. . . . 125
Figure 4.30: Resistivity and Seebeck coefficient of Si NW-FET before and after gate deposition. 127
Figure 4.31: Resistivity and Seebeck coefficient of Si TF-FET before and after gate deposition. 128
Figure 4.32: Channel conductance versus gate voltage of N34 NW-FET devices. 129
Figure 4.33: Electrical properties of Si NWs measured via FET method. . . 132
Figure 4.34: Resistivity of oxidized Si thin film. 133
Figure 4.35: Carrier concentration of oxidized Si thin film. $^{(a)(b)}$See equation 4.17 134
Figure 4.36: Mobility of oxidized Si thin film. 135
Figure 4.37: Seebeck coefficient of oxidized Si thin film. Solid and dashed lines show calculation result for $|S|$ as predicted by dopant segregation model for $^{(a)(c)}r = 0$, and $^{(b)(d)}r = -1/2$. 136

Figure 4.38: Seebeck coefficient of oxidized Si nanowires with initial width of 40 and 80 nm. Solid, dotted, and dashed lines represent calculation based on dopant segregation model. The horizontal axis was shortened for the 6 to 12 min interval to highlight the nanowires data below 6 min. 138

Figure 5.1: Band Anti-Crossing model 141
Figure 5.2: N composition and carrier concentration map of GaAs and GaN_xAs_{1-x} samples. Growth windows identified by initial sample set (gray circle); adjacent numerical label indicates T_{Si} in °C. The final set of GaAs (black triangle) and GaN_xAs_{1-x} (black diamond) was then grown to specific n and x targets. 145
Figure 5.3: Carrier concentration of GaAs (open triangle) and GaN_xAs_{1-x} (filled diamond) as a function of Si cell temperature. 149
Figure 5.4: Doping efficiency of GaN_xAs_{1-x}grown at different T_{Si}. Numerical label indicates x in %. 150
Figure 5.5: Mobility of GaAs samples grown at earlier date (black triangle) and later date (black diamond). [(a)]See Ref. [8]. [(b)]Dashed lines are calculation results replicated from Ref. [20]. 153
Figure 5.6: Mobility of GaN_xAs_{1-x} as a function of N composition. [(a)]Ref. [9]. [(b)]Ref. [21]. 153
Figure 5.7: Seebeck coefficient of n-type Si-doped GaAs as a function of carrier concentration at 300 K. [a]Ref. [7]. [b]Ref. [8]. [c]Ref. [10]. [d]Ref. [9]. 154
Figure 5.8: Seebeck coefficient of Si-doped GaN_xAs_{1-x} as a function of N composition at 300 K. 155
Figure 5.9: S^2n of GaN_xAs_{1-x} as function of N composition. 156
Figure 5.10: Density of states effective mass of GaN_xAs_{1-x} with 3 types of dopant: [(a)] Si, n of $3-10\times10^{17}\,\mathrm{cm}^{-3}$, [(b)] Te, n of $3-5\times10^{17}\,\mathrm{cm}^{-3}$ (Ref. [10]), [(c)] Se, n of $5-7\times10^{18}\,\mathrm{cm}^{-3}$ (Ref. [9]). Theoretical models reproduced from [(d)]Masia *et al.* [22] and [(e)]Shan *et al.* [23], and [(f)]calculated from data in Ref. [24]. 158

LIST OF TABLES

Table 1.1: Relaxation time constants for elastic scattering processes. *POP scattering is inelastic; listed r value an approximation only. . . . 13
Table 1.2: Density of states parameters for bulk Si and GaAs. * T is in K. . 17

Table 2.1: Optimal reduced Fermi level (first entry) and optimal Seebeck coefficient (second entry) for several combination of r and D . . 32
Table 2.2: Optimal carrier concentration of PbTe and $SrTiO_3$ 33

Table 3.1: MOSFET device's geometries and $I_d - V_g$ calculation results. . . 61
Table 3.2: Thermal conductivity of various thermal grease/paste (according to manufacturer specification) 68
Table 3.3: Experimental and literature [25] TCR values 69

Table 4.1: SOI wafer specification 81
Table 4.2: Parameters for Si mobility calculation [13] 87
Table 4.3: Ion-implantation specification and approximate resultant electrical properties. All implant performed with nominal 7° substrate tilt. 90
Table 4.4: General settings for Raith 50 e-beam writer. 96

Table 5.1: GaAs and GaN_xAs_{1-x} etch recipe. 147

Table A.1: List of physical constant. 164
Table A.2: List of Symbols. 164
Table A.3: List of Acronym. 167

ACKNOWLEDGEMENTS

I give my sincere gratitude to my advisor, Professor Prabhakar Bandaru, who has given me the chance to pursue my childhood dream of being a scientist. His faith in me and the support he provided all these years are deeply appreciated.

My thanks also go to Professor Charles W. Tu for the opportunity and resources he has given me to complete my project. His kind invitations to so many party and group lunches were also a very welcome relief from the grind of graduate school.

This work could not have been completed without the tremendous resources of the Nano3 cleanroom, and the helpful staffs: Ryan, Bernd, Larry, Maribel, and Sean. I would also like to acknowledge Charlotte Lauve, our great admin who make things happen for us students.

For my peers, I would like to thank Yanjin Kuang, who performed the MBE growth and X-Ray diffraction of GaN_xAs_{1-x} for the work in Chapter 5. Additionally, he has become a very good friend and a great colleague. His quiet strength of conviction has also taught me a lesson. I also would like to acknowledge my labmates. Calvin Patel, who assisted in the set up of measurement system in Chapter 3. Max Aubain, for the fun times he brings with his positive disposition. Mark Hoefers, for being in the trench together. His presence has given me motivation to do more.

Finally, I want to thank my mom for all her love. It has been very difficult to be away from home, especially in my last year here. Thank you for your patience.

Chapter 2, in part, is a reprint of the material as it appears in the journal articles:

- P. Pichanusakorn and P.R. Bandaru, "Minimum length scales for enhancement of the power factor in thermoelectric nanostructure", *Journal of Applied*

Physics, vol. 107(7), 074304 (2010)

- P. Pichanusakorn and P.R. Bandaru, "The optimal Seebeck coefficient for obtaining the maximum power factor in themoelectrics", *Applied Physics Letters*, vol. 94, 223108 (2009)

The dissertation author was the primary investigator and author of these papers.

Chapter 5, in part, is a reprint of the material as it appears in the journal article: P. Pichanusakorn, Y. J. Kuang, C. J. Patel, C. W. Tu, and P. R. Bandaru, "The influence of dopant type and carrier concentration on the effective mass and Seebeck coefficient of GaN_xAs_{1-x} thin films", *Applied Physics Letters 99*, 072114 (2011). The dissertation author was the primary investigator and author of this paper.

VITA

2004	B. S. in Electrical and Computer Engineering *magna cum laude*, University of California, San Diego
2007	M. S. Materials Science and Engineering, University of California, San Diego
2012	Ph. D. in Materials Science and Engineering, University of California, San Diego

PUBLICATIONS

P. Pichanusakorn, Y. J. Kuang, C. J. Patel, C. W. Tu, and P. R. Bandaru, "The influence of dopant type and carrier concentration on the effective mass and Seebeck coefficient of GaN_xAs_{1-x} thin films", *Applied Physics Letters 99*, 072114 (2011).

P. Pichanusakorn, and P. R. Bandaru, "Nanostructured Thermoelectrics", *Materials Science and Engineering: R: Reports*, 67, 19-63 (2010).

P.R. Bandaru and P. Pichanusakorn, "An outline of the synthesis and properties of silicon nanowire", *Semiconductor Science and Technology*, vol. 25, 024003 (2010).

P. Pichanusakorn and P.R. Bandaru, "Minimum length scales for enhancement of the power factor in thermoelectric nanostructure", *Journal of Applied Physics*, vol. 107(7), 074304 (2010).

P. Pichanusakorn and P.R. Bandaru, "The optimal Seebeck coefficient for obtaining the maximum power factor in themoelectrics", *Applied Physics Letters*, vol. 94, 223108 (2009).

P. Pichanusakorn, N.B. Elsner, and P.R. Bandaru, "Increasing electrical conductivity in sputter-deposited Si/SiGe multilayers through electrical pulse based annealing", *Electronics Letters*, vol. 44, no. 21, 1274 (2008).

ABSTRACT OF THE DISSERTATION

Thermoelectric and electrical characterization of Si nanowires and GaNAs

by

Paothep Pichanusakorn

Doctor of Philosophy in Materials Science and Engineering

University of California, San Diego, 2012

Professor Prabhakar R. Bandaru, Chair

Thermo-electricity offers an elegant solution to the problem of heat-to-electricity conversion. As a completely solid-state heat engine, thermoelectric materials confer many advantages in electricity generation and heat pumping. However, the efficiency of thermoelectric material, represented by the figure of merit, Z, is generally low. In order to increase Z, it is desirable to increase a property of material known as the density of states (DOS). In this work, two approaches to increase DOS were explored, one based on the physical effect of quantum confinement in Si nanowires, and another based on the chemical approach where by GaAs property are drastically changed by doping of N. While the Si nanowires fabrication have proved to be extremely challenging, and results were hard to obtain, some

interesting insight into dopant redistribution in nanowires were observed. For the GaNAs, the research goal was completed, and the hypothesized enhancement in effective mass was observed. Although the power factor enhancement was not obtained, the work shows that resonant doping could be potentially useful, provided that problem with significant alloy scattering can be overcome.

In Chapter 1, a thorough explanation of the theory behind thermoelectric property of materials is given. The discussion will go through the mathematical rigor behind the derivation of n, μ, S, and κ_e.

In Chapter 2, the theoretical basis of this work is established. The optimal reduced Fermi level is identified, and its utility is discussed in relation to the identification of the maximum power factor. The two approaches to increasing density of states are discussed.

In Chapter 3, the transport coefficient measurement system that was built for this work is described. Device fabrication and measurement procedure are demonstrated.

In Chapter 4, preparation of the substrates, electron beam lithography work, oxidation of Si nanowires is discussed as method to prepare nanowires. Result of FET and Seebeck measurement on nanowires are discussed.

In Chapter 5, the utilization of resonant states in GaNAs to increase the power factor is explored. Predicted increase in effective mass was observed, but N scattering prevents the realization of a power factor enhancement.

1 Introduction to Thermoelectric phenomena and theory

1.1 Thermoelectric Effects

1.1.1 Seebeck effect

Thermoelectric phenomena are related to the conversion of heat flow into electricity, and vice versa, in solid materials. One such phenomenon is the *Seebeck* effect. When a pair of dissimilar metals that are joined as shown in Fig. 1.1 is subjected to a temperature difference ($\Delta T = T_1 - T_2$), the Seebeck voltage (V) is developed. The Seebeck coefficient is then defined as the ratio of the electric field ($\mathcal{E}$) over temperature gradient (∇T)

$$S = \frac{\mathcal{E}}{\nabla T} = -\frac{V}{\Delta T} \tag{1.1}$$

Note that there is a negative sign in the latter expression because $\mathcal{E} \propto -V$. S is also known as the "thermopower".

Although S can be defined for a single material, it is always measured for a *pair* of dissimilar materials. For example, the total Seebeck voltage of the thermocouple in Fig. 1.1 is

$$V = V_{B,1} + V_A + V_{B,2}$$
$$V = -[S_B(T_0 - T_2) + S_A(T_2 - T_1) + S_B(T_1 - T_0)]$$
$$V = -(S_A - S_B)(T_2 - T_1)$$

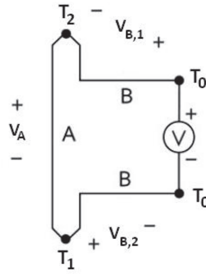


Figure 1.1: Seebeck effect gives rise to voltage when there is temperature difference between junctions of metals A and B.

,therefore

$$S = (S_A - S_B) = -\frac{V}{T_2 - T_1} \tag{1.2}$$

, where S is the Seebeck coefficient of the thermocouple, which is the difference between the individual Seebeck coefficients of metal A and metal B. Therefore, S_A could only be known if S_B is known, or if $|S_A| \gg |S_B|$, then $S \approx S_A$. In an experiment, a Pb or Pt calibration sample is sometime used to measure S_B [26]. In our work, $|S|$ of metal wirings is much smaller than that of our semiconductor samples, and the condition $|S_A| \gg |S_B|$ is often satisfied.

The Seebeck effect is due to diffusion of charge carrier, i.e. electrons or holes, under temperature gradient. When a material is under zero ∇T as in Fig. 1.2a, electrons still diffuse from one end to another, but with a net diffusion current of zero. Since there is no drift current as well, the total current is also zero. However, when one end is heated relative to another ($T_2 > T_1$) as in Fig. 1.2b, the net diffusion will be non-zero as the electrons at the hot end will have greater energy and velocity. The imbalance in carrier concentration between the two ends then generates an electric field that pushes electrons back toward the hot end. The drift current generated by this potential, i.e. the Seebeck voltage, is such that the total current remains zero. The same analysis applies to p-type material. However, electric field generated would be in opposite direction as holes carry positive charges.

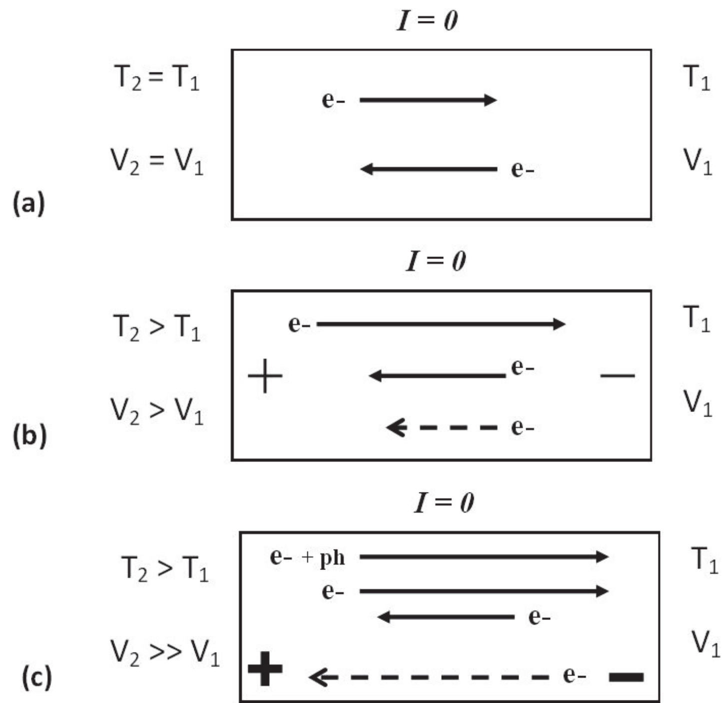


Figure 1.2: Diffusion (solid arrow) and drift (dashed arrow) current of electrons with (a) zero-temperature gradient, and (b) non-zero temperature gradient. (c) shows additional effect of phonon-drag.

Additionally, the Seebeck effect can be created by "phonon-drag" effect. Phonon represents a quantum of lattice vibrations. In electrically insulating solids, phonons exclusively provide heat flow. In conducting solids, such as metal, both electrons and phonons may conduct thermal energy. Under a temperature gradient, there is a non-zero net flow of phonons from hot end to cold end. At low temperature [27, 28], phonons may interact with and scatter electrons along the temperature gradient as well. As shown in Fig. 1.2c, this additional phonon-drag current creates even greater carrier concentration imbalance, leading to larger $|S|$.

1.1.2 Peltier and Thompson effects

The Peltier effect is observed when heat is absorbed or evolved at a junction of two dissimilar material with non-zero electrical current. This is due to the injection or withdrawal of energetic electrons from one material to another. This is different from Joule heating, where electrons energy is loss through scattering with the lattice. The Peltier coefficient (Π) is then defined as the ratio of the heat density evolved or consumed (Q) per electrical current density (J) injected, i.e.

$$\Pi = \frac{Q}{J} \tag{1.3}$$

As Q is difficult to measure experimentally, the value for Π is often not discussed in literature. Nevertheless, Π could be calculated from S through the second Thomson relation states

$$\Pi = ST \tag{1.4}$$

Seebeck and Peltier effects are different expression of the phenomena that electrons carry both charge and thermal energy, and can be induced to flow via electrical potential and/or temperature gradient.

Finally, there is a third thermoelectric effect, the Thomson effect, which describe the heating or cooling of conductor that is carrying electrical current, *and* with an imposed temperature gradient. The Thomson coefficient (μ_T) is related to the Seebeck coefficient by the first Thomson relation as

$$\mu_T = T\frac{dS}{dT} \tag{1.5}$$

1.2 Efficiency and ZT

Thermoelectric effects in materials have two widely-considered applications: the conversion of thermal energy to electrical energy via the Seebeck effect, and the pumping of heat for cooling application via the Peltier effect, as shown in Fig. 1.3. Since a thermoelectric device is devoid of any mechanical part, and does not require gas or liquid as working fluid, it is considered a solid-state device with several

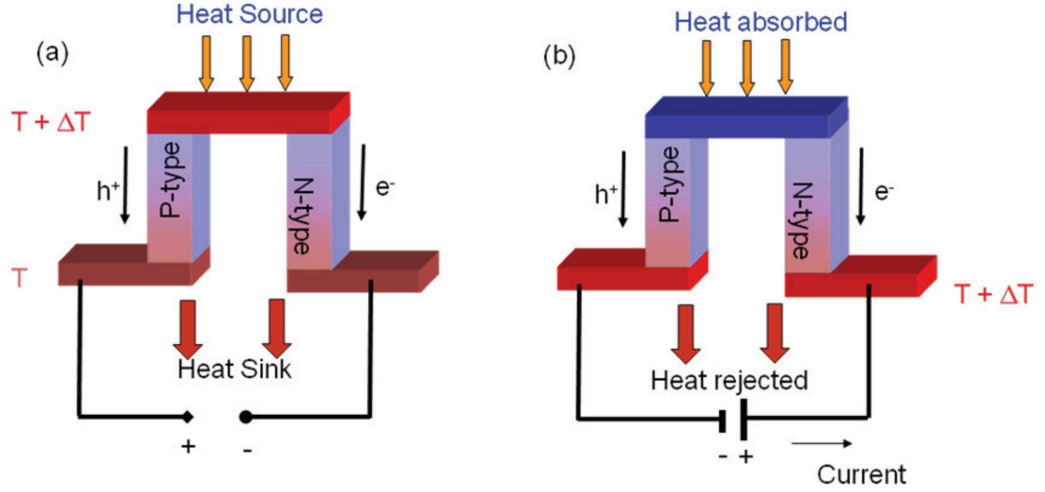


Figure 1.3: Thermoelectric module as (a) electricity generator and (b) Peltier cooler.

clear advantages. It is relatively simple and reliable, can be made very small, and is free of vibration. Thermoelectric devices are used in electronics temperature control/cooling [29], generator for vehicle [30], nuclear pacemaker, nuclear generator for deep-space satellite. Unfortunately, the efficiency of thermoelectric device has always been very low, with the best material only 5-10% efficient. The goal of thermoelectric material research is often aimed at improving the efficiency.

The efficiency of the most basic thermoelectric device, shown in Fig. 1.3, working as a heat engine is derived by accounting for electrical and thermal energy input and output of the system. A thorough derivation can be found in Ref. [28]. When operating as a Peltier cooler, the efficiency is

$$\phi_{TE} = \frac{(S_p - S_n)TI_C - K(\Delta T) - \frac{1}{2}I^2R}{(S_p - S_n)\Delta T\, I + I^2R} \tag{1.6}$$

, where S_p and S_n is the Seebeck coefficient of the p and n materials, $\Delta T (= T_H - T_C)$ is the temperature difference between the cold and hot sides, I is the electrical current, K is the total thermal conductance of both p- and n-legs *in parallel*, and R is the total electrical resistance of both p- and n-legs *in series*. The nominator of ϕ is

the sum of total heat flow from the (cold) source to (hot) sink, which is composed of the heat "pumped" via the Peltier effect from source to sink (first term), the parasitic heat due to finite K of the module which flow from hot to cold (second term), and the heat generated by Joule heating (third term). The denominator of ϕ is the total electrical power, which is spent to do the work (first term), and is dissipated through Joule heating (second term).

This efficiency can be maximized by optimizing the relative sizes of each legs, and the operating current[27]. The optimal length and cross-section area of the leg is such that

$$\frac{l_p A_p}{l_n A_n} = \left(\frac{\sigma_p \kappa_p}{\sigma_n \kappa_n}\right)^2$$

, where l_p, l_n are the length, A_p, A_n the cross-section area, σ_p, σ_n the electrical conductivity, and κ_p, κ_n the thermal conductivity of the p and n legs, respectively. The optimal I is then obtained by differentiating, and finding the maximum of ϕ. The maximum ϕ is then given by

$$\phi_{TE,max} = \frac{T_C(\sqrt{1+Z^*T}-1)}{\Delta T(\sqrt{1+Z^*T}+1)} - \frac{1}{2}$$

, where Z^* is the thermoelectric "figure of merit"for the *thermocouple*

$$Z^* = \frac{(S_p - S_n)^2}{\left(\frac{\kappa_p}{\sigma_p}^{1/2} + \frac{\kappa_n}{\sigma_n}^{1/2}\right)^2}$$

From convention, the figure of merit for a *single material* is similarly defined as

$$Z = \frac{S^2\sigma}{\kappa} = \frac{S^2\sigma}{\kappa_e + \kappa_l} \tag{1.7}$$

, where κ_e and κ_l are the components of thermal conductivity due to electrons, and the lattice vibration (i.e. phonon), respectively. Although Z only equals Z^* when the magnitude of S, σ, κ for both materials are equal, it is considered an acceptable indicator of a material's efficiency. The dimensionless figure of merit, ZT, is also commonly used. The nominator $S^2\sigma$ is known as power factor.

According to the definition of Z in equation 1.7, a good thermoelectric material should have high σ and $|S|$, but small κ, i.e. it should conduct electricity well to minimize Joule heating, generate large voltage per unit temperature difference, and suffer only small parasitic heat conduction. Unfortunately, these requirements often run in contrary to each other. For example, metals have large σ but also large κ, while insulator such as glass tend to have low κ but also low σ. In order to understand what characteristic of material yield large Z, the following sections are devoted to a thorough discussion of how S, σ, and κ are related.

1.3 Transport coefficients

The definitions of relevant electric and thermoelectric transport coefficients is shown by the set of equations below

$$\Pi = \left.\frac{Q}{J}\right|_{\frac{dT}{dx}=0}$$

$$\sigma = \left.\frac{J}{\mathcal{E}}\right|_{\frac{dT}{dx}=0} \qquad \begin{array}{c|c} J & Q \\ \hline \mathcal{E} & \frac{dT}{dx} \end{array} \qquad \kappa_e = -\left.\frac{Q}{\frac{dT}{dx}}\right|_{J=0}$$

$$S = \left.\frac{\mathcal{E}}{\frac{dT}{dx}}\right|_{J=0}$$

The heat and electrical current density is further defined as the product of (1) the number of charge/energy carriers (i.e. carrier concentration $[n]$), (2) the amount of charge/energy carried per carrier, and (3) velocity of the carrier (v). Summing over all energy and velocity, the charge and heat current density are equal to

$$J = nqv = q\int_{-\infty}^{+\infty} g(E)v(E)\big[f(E) - f_0(E)\big]dE \tag{1.8}$$

$$Q = n(E - E_F)v = \int_{-\infty}^{+\infty} g(E)(E - E_F)v(E)\big[f(E) - f_0(E)\big]dE \tag{1.9}$$

, where $q = \mp e$ is unit charge of electrons and holes, respectively, E_F is the *Fermi level*, $g(E)$ is the material's *density of states* (DOS), and $[f(E) - f_0(E)]$ is the

difference in the carrier distribution between a *non-equilibrium* and *equilibrium* states, respectively. Each term is further discussed.

1.3.1 Density of states

The density of states is the number of states that can be occupied by electrons per interval of energy. It is derived from the dispersion function of the material's band structure [31, 32]. Assuming a parabolic dispersion, the DOS *per unit volume* for a band or subband is given by the power law

$$g(E) = \frac{N_c}{d^{3-D}} \left(\frac{m_d}{\pi\hbar^2}\right)^{D/2} \left(\frac{2}{\pi}\right)^{|D/2-1|} (E - E_0)^{D/2-1} \qquad , E \geq E_0 \tag{1.10}$$

, where $\hbar$ is the reduced Planck constant, N_c is the number of conduction valley or the degeneracy factor, m_d is the density of states effective mass, E_0 is the minimum energy level of the band, D is the dimensionality factor that equal 3, 2, and 1 for bulk material, quantum wells, and nanowires, respectively, and d is the thickness of quantum well or the diameter of nanowires.Under this convention, $E < E_0$ would indicates energy level in the band gap.

DOS is strongly affected by quantum confinement effect. First of all, the energy dependence or the "shape" of the DOS is governed by D. Figure 1.4 shows the parabolic, steps, and peak profile characteristic of bulk, quantum well, and nanowires DOS. Multiple subbands are shown in the figure for quantum well and nanowires, i.e. $g(E) = \Sigma_{i=1}^{+\infty} g_i(E)$, where $g_i(E)$ is the DOS of the ith subband in quantum wells. Each subband is distinguished by its minimum energy, which depends on the principle effective mass (m_x, m_y, m_z) in the direction of confinement as

$$E_0 = E_i = \frac{\hbar^2\pi^2 i^2}{2d^2 m_z} \qquad , i = 1, 2, 3 \ldots \tag{1.11}$$

$$E_0 = E_{ij} = \frac{\hbar^2\pi^2}{2d^2}\left(\frac{i^2}{m_y} + \frac{j^2}{m_z}\right) \qquad , i = 1, 2, 3 \ldots, \; j = 1, 2, 3 \ldots \tag{1.12}$$

, where confinement in z-direction is assumed for quantum wells, and confinement in y- and z-directions is assumed for nanowires. Equations 1.10-1.12 shows that as

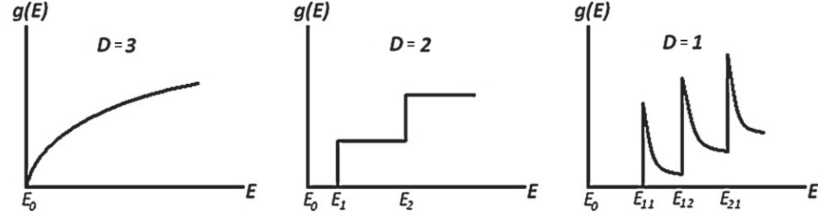


Figure 1.4: Density of states for 3-, 2-, and 1-dimensional material.

d is reduced, the magnitude DOS of subbands in quantum well and nanowires is increased, and the separation between subbands become larger. Conversely, when d is increased, each subbands get smaller and come closer together. Eventually, the summed DOS of these subbands will be the same as that of bulk material.

1.3.2 Boltzmann transport equation

A non-equilibrium distribution function ($f(E)$) for thermal/electrical transport may be determined by the Boltzmann Transport Equation (BTE). One formal expression for the BTE [27] is

$$\left(\frac{df}{dt}\right)_r = \frac{df}{dt} + \frac{d\vec{k}}{dt}\nabla_{\vec{k}} f + \frac{d\vec{r}}{dt}\nabla_{\vec{r}} f \tag{1.13}$$

, where t is time, $\vec{k}$ and $\vec{r}$ are the wave and position vectors of electrons, and f is the non-equilibrium distribution function. The BTE may describe the distribution of particles in a system under a perturbation, which redistributes the particle's momentum and position.

In regard to thermoelectric phenomenon, we consider two perturbation: $\mathcal{E}$, and $\frac{dT}{dx}$. An electric field (of strength $\mathcal{E}$) pointing along $\vec{i}$ exerts a force on electrons that changes their momentum, i.e.

$$\frac{d\vec{k}}{dt} = \frac{dk_x}{dt}\hat{i} = \frac{q\mathcal{E}}{\hbar}\hat{i} \tag{1.14}$$

Meanwhile, a spatial temperature gradient ($\frac{dT}{dx}$) induces a spatial gradient in the

electron distribution according to

$$\nabla_{\vec{r}} f = \frac{\partial f}{\partial T}\frac{dT}{dx}\hat{i} \tag{1.15}$$

These perturbations cause electrons to depart from their equilibrium distribution, leading to a current. Given the perturbations along $\vec{i}$, we may further write

$$\nabla_{\vec{k}} f \cdot \vec{i} = \frac{df}{dk_x} = \frac{df}{dE}\frac{dE}{dk_x} = v(E)\frac{df}{dE} \tag{1.16}$$

$$\frac{d\vec{r}}{dt}\vec{i} = \frac{dx}{dt} = v(E) \tag{1.17}$$

, where the velocity along $\vec{i}$, $v(E)$, is equal to both $\frac{dE}{dk_x}$ (rate of change of energy over momentum) and $\frac{dx}{dt}$ (rate of change of displacement over time).

However, carrier scattering tends to restore the equilibrium distribution, whereby electrons may lose the momentum to the atomic lattice (i.e. electron-phonon scattering). The *relaxation time approximation* then simplify the complex process of carrier scattering with the assumption that electrons will return from the non-equilibrium to the equilibrium states within a “relaxation” time τ, i.e.

$$\left(\frac{df}{dt}\right)_r = -\frac{f - f_0}{\tau} \tag{1.18}$$

, where $\tau = \tau(E)$ is strictly a function of electron energy. This relaxation time approximation should be valid only under “low-(electric) fields when the scattering is elastic or isotropic” [33].

Using the first-order approximation to the BTE where $f - f_0 \gg f_0$, the term f on the right hand side of equations 1.13, 1.15, and 1.16 may be replaced with f_0. However, the BTE does not describe f_0, which depends on the nature of the relevant particle. Electrons (fermion) follow the Fermi-Dirac (F-D) statistics, such that

$$f_0 = f_{FD}(E) = \frac{1}{\exp\left(\frac{E-E_F}{k_B T}\right) + 1} \tag{1.19}$$

, where E_F is the *Fermi level*, and k_B is the Boltzmann constant. It can then be shown that

$$\frac{df_{FD}}{dT} = -\left(\frac{E - E_F}{T}\right)\frac{df_{FD}}{dE} \tag{1.20}$$

When $(E - E_F) > 3k_BT$, f_0 may be approximated by the Maxwell-Boltzmann (M-B) statistics, whose distribution function is

$$f_{MB}(E) = \exp\left(\frac{E - E_F}{k_BT}\right) \tag{1.21}$$

M-B statistics is often used because it produces analytical solutions [27, 34]. Solutions based on the M-B statistics is accurate for intrinsic or lightly-doped semiconductor, where electrons energy $(E > E_0)$ is much greater than the Fermi level $(E_F < E_0)$ that is located in the band gap. However, it is inaccurate for moderately- to degenerately-doped semiconductors, where the Fermi level is located near or higher than the band edge $(E_F \gtrsim E_0)$ [27].

Finally, incorporating equations 1.14-1.20 into equation 1.13, the *steady state* $\left(\frac{df}{dt} = 0\right)$, *first order* solution to the BTE, assuming the Fermi-Dirac statistics, is

$$f(E) - f_0(E) = -\tau(E)v(E)\frac{df_{FD}}{dE}\left(q\mathcal{E} - \left(\frac{E - E_F}{T}\right)\frac{dT}{dx}\right) \tag{1.22}$$

Thus far, we have not define the function for $v(E)$ and $\tau(E)$. However, the relaxation time approximation (equation 1.18) necessitates that $v(E)$ and $\tau(E)$ should be as follow.

1.3.3 Carrier velocity

Under low-fields, electron motion is primarily diffusive and the velocity is given by

$$v(E) = \sqrt{\frac{2(E - E_0)}{Dm_\sigma}} \tag{1.23}$$

, where the dimensionality factor D relates to the equipartition of energy among the available degrees of freedom, and m_σ is the conductivity/inertial effective mass. Whereas m_d is the geometric mean of all principle effective masses, m_σ is equal to only the effective in the direction of current (i.e. along $\vec{i}$ in our case). However, when there are degenerate conduction valleys $(N_c > 1)$, m_σ must be taken

as the *harmonic* mean of principal effective masses the direction of conduction for each valley. For example, $m_\sigma = 3(\frac{2}{m_t} + \frac{1}{m_l})^{-1}$ along $\langle 100 \rangle$ direction for 6-fold degenerate Si conduction band, where m_t and m_l are principal effective mass along the transverse and longitudinal axes of the elliptical conduction valley, respectively.

1.3.4 Relaxation time

The relaxation time introduced in equation 1.18 is assumed to follow the power law function

$$\tau(E) = \tau_0 (E - E_0)^r \tag{1.24}$$

, where the exponent r is a characteristic scattering constant, and τ_0 represents collection of parameters related to the scattering processes and material properties. Generally, this assumption holds for elastic scattering such as ionized impurity scattering (II), neutral-impurity scattering (NI), alloy scattering (AL), and acoustic and optical phonons scattering by deformation potential (ADP and ODP, respectively) [33, 35]. The value of r and τ_0 is dependent on the type of scattering processes, though r tends to have similar value for many processes.

In fact, the relaxation time of all the aforementioned scattering processes, with the exception of NI scattering, is inversely proportional to the DOS, i.e. $\tau(E) \propto \frac{1}{g(E)}$, which implies that

$$r = 1 - D/2 \tag{1.25}$$

The relationship between $\tau(E)$ and $g(E)$ is reflective of the fact that the scattering rate of carrier may depends strongly on the number of available final states, which the scattered electrons can occupy. It is important to recognize that the scattering constant may not always remain the same. For example, while r is given by -1/2 for ionized impurity in bulk material ($D = 3$), but it may equal 0 in quantum well ($D = 2$) instead.

Table 1.1: Relaxation time constants for elastic scattering processes. *POP scattering is inelastic; listed r value an approximation only.

Scattering Process		r	$1/\tau_0 \propto$
Acoustic phonons, deformation potential	(ADP)	1-D/2	T, D_A
Optical phonons, deformation potential	(ODP)	”	T, D_O
Alloy scattering	(AL)	”	N_I
Ionized impurity, strongly-screened	($\mathrm{II_{str}}$)	”	N_{II}
Neutral impurity	(NI)	0	N_{NI}
Polar optical phonons	(POP)	0*	-
Piezoelectric	(PZ)	+1/2	T
Ionized impurity, weakly-screened	($\mathrm{II_{wk}}$)	+3/2	N_{II}

However, the r value of a few scattering processes does not follow equation 1.25. In a parabolic band, the NI scattering is an energy-independent process [35], i.e. $r = 0$. The II scattering may also be distinguished between strongly- and weakly-screened scattering. When N_{II} is low, electron does not scatter immediately between nearest-neighbor, scattering is considered weakly-screened and $r = +3/2$ for $\mathrm{II_{wk}}$ scattering [33]. Table 1.1 summarized these r value.

Special case then arises for inelastic scattering process such polar optical phonon (POP) scattering. Relaxation time for inelastic process does not follow equation 1.24, and therefore does not have an r value. However, it may be possible to assume $r \approx 0$ for POP scattering since its scattering rate are almost constant above and below 0.05eV [33]

While Table 1.1 shows a distinct r value of -1/2 to +3/2 for bulk material ($D = 3$), the effective r value may be different as multiple scattering processes could be relevant. r could be determined experimentally through (1) $\mu(T)$ relationship (see further discussion in section 1.4.2), (2) the $S-n$ curve, and (3) Nernst coefficient measurement[26].

Unlike r, τ_0 is not characterized by simple half-integers. τ_0 is typically a product of (1) the number of scattering sources, and (2) the nature of the scattering

source. For example, τ_0 for II scattering is inversely proportional the concentration of ionized impurity (N_{II}), the ionization number of that impurity, as well as screening length. Meanwhile, τ_0 for AL scattering is dependent on the concentration of the alloyed impurity (N_{AI}), and the perturbation potential introduced by lattice deformation due to the impurity. For ADP process, the scattering source is phonons, and the mechanism of scattering is the acoustic phonons deformation potential (D_A). Therefore, τ_0 is inversely proportional to T and D_A, instead. Table 1.1 also summarized these proportionality. The exact expression for τ_0 can be found in Ref. [33, 35].

1.3.5 The χ_i integral

Given the solution to the BTE in equation 1.22, the current and heat density in equations 1.8 and 1.9 can be express as a sum of a common integral

$$J = \left(q^2\mathcal{E} + q\frac{E_F}{T}\frac{dT}{dx}\right)\chi_0 - \left(\frac{q}{T}\frac{dT}{dx}\right)\chi_1 \tag{1.26}$$

$$Q = \left(q\mathcal{E} + \frac{E_F}{T}\frac{dT}{dx}\right)\chi_1 - \left(\frac{1}{T}\frac{dT}{dx}\right)\chi_2 \tag{1.27}$$

, where

$$\chi_i = -\int_{-\infty}^{+\infty} g(E)\tau(E)v^2(E)E^i\frac{df_{FD}}{dE}dE \tag{1.28}$$

If the material has a single, parabolic band, and is under low-fields with elastic scattering, then $g(E)$, $v(E)$, and $\tau(E)$ may be represented by equations 1.10, 1.23, and 1.24, respectively. χ_i is then equal to

$$\chi_i = \frac{-\tau_0 N_c}{m_\sigma d^{3-D}}\left(\frac{m_d}{\pi\hbar^2}\right)^{D/2}\left(\frac{2}{\pi}\right)^{|D/2-1|}\frac{2}{D}\left[\int_{E_0}^{+\infty}(E-E_0)^{r+D/2+i}\frac{df_{FD}}{dE}dE\right]$$

The term in the [ ] bracket is simplified with integration-by-parts

$$\begin{aligned}
&\int_{E_0}^{+\infty}(E-E_0)^{r+D/2+i}\frac{df_{FD}}{dE}dE \\
&\quad=\int_0^{+\infty}E^{r+D/2+i}\frac{df_{FD}}{dE}dE \\
&\quad=E^{r+D/2+i}f_{FD}(E)\Big|_0^{+\infty}+\left(r+\frac{D}{2}+i\right)\int_0^{+\infty}E^{r+D/2+i-1}f_{FD}(E)\,dE \\
&\quad=\left(r+\frac{D}{2}+i\right)\int_0^{+\infty}\frac{E^{r+D/2+i-1}}{\exp(\frac{E-E_F}{k_BT})+1}\,dE
\end{aligned}$$

Since a single band is assumed, we may arbitrarily set $E_0 = 0$ in the first step. By setting $E_0 = 0$, $E_F = 0$ would indicates that the Fermi level is at the band edge.

The final form for χ_i integral is then

$$\chi_i=\frac{\tau_0 N_c}{m_\sigma d^{3-D}}\left(\frac{m_d}{\pi\hbar^2}\right)^{D/2}\left(\frac{2}{\pi}\right)^{|D/2-1|}(k_BT)^{r+D/2+i}\left(\frac{2(r+i)}{D}+1\right)F_{r+D/2+i-1}(\eta) \tag{1.29}$$

, where $F_j(\eta)=\int_0^{+\infty}\frac{x^j}{\exp(x+\eta)+1}dx$ is the jth-order "Fermi integral", $x=\frac{E}{k_BT}$, and $\eta=\frac{E_F}{k_BT}$ is the *reduced* Fermi level. With the exception of the 0th-order Fermi integral, i.e. $F_0(\eta)=\ln(1+e\eta)$, $F_j(\eta)$ has no analytical solution, and can only be evaluated numerically. The algorithms published in Ref [36] were used in this work to evaluate equation 1.29. The Fermi integral is unit-less.

1.3.6 Summary of equations for S, σ, and κ_e

Given that J and Q may be expressed by equations 1.26 and 1.27, the transport coefficients as defined on page 7 is written as

$$S=\left.\frac{\mathcal{E}}{\frac{dT}{dx}}\right|_{J=0}=\frac{1}{qT}\left(\frac{\chi_1}{\chi_0}-E_F\right) \tag{1.30}$$

$$\sigma=\left.\frac{J}{\mathcal{E}}\right|_{\frac{dT}{dx}=0}=e^2\chi_0 \tag{1.31}$$

$$\kappa_e=-\left.\frac{Q}{\frac{dT}{dx}}\right|_{J=0}=\frac{1}{T}\left(\chi_2-\frac{\chi_1^2}{\chi_0}\right) \tag{1.32}$$

, where $\chi_i(\eta)$ may be evaluated with equation 1.29. Note that q^2 has been replaced with e^2 for σ, whereas q is preserved for S to preserve the sign. These equations are sufficient to calculate the transport coefficients accurately for various metals and semiconductors. It is pertinent to express these transport coefficients as a function of η, since it encapsulates both the change in T and E_F. In the following sections, the dependence of each transport coefficient on η, and their contributions toward Z are discussed.

1.4 Transport coefficients as function of reduced Fermi level

It was established in that one criterion for large Z is that σ should be large. If this was the sole criterion to material selection, then metals would appear to be a good thermoelectric material. As $\sigma = ne\mu$, it would appears that semiconductor should be highly doped to increase n as well. However, an increase in n can adversely reduces μ, $|S|$, and increases κ_e. In this section, the relationship between transport coefficient is explained through their dependence on η.

1.4.1 Carrier concentration

The relationship between n and η is considered first. Assuming that all electrons/holes in an electronic band contribute to conductivity, n may be defined as the product of the DOS and the Fermi-Dirac distribution function [34] as

$$n = \int_{-\infty}^{+\infty} g(E) f_{FD}(E) dE \tag{1.33}$$

For a parabolic dispersion, $g(E)$ is given by equation 1.10, therefore

$$n = N_0 \left(\frac{4}{\pi}\right)^{|D/2-1|} F_{D/2-1}(\eta) \tag{1.34}$$

$$N_0 = \frac{N_c}{2^{|D/2-1|} d^{3-D}} \left(\frac{m_d k_B T}{\pi \hbar^2}\right)^{D/2} \tag{1.35}$$

Table 1.2: Density of states parameters for bulk Si and GaAs. * T is in K.

Material	n-Si	p-Si	n-GaAs	p-GaAs
m_d/m_0	0.32	0.81	0.067	0.53
N_c	6	1	1	1
$N_0/T^{3/2}$ (cm^{-3})*	5.25×10^{15}	3.52×10^{15}	8.37×10^{13}	1.86×10^{15}
N_0 @300K (cm^{-3})	2.72×10^{19}	1.89×10^{19}	4.35×10^{17}	9.67×10^{18}

, where N_0 is known as the *effective density of states per unit volume*. Under the classical approximation where $f_{MB}(E)$ is used in placed of $f_{FD}(E)$ in equation 1.33, $n = N_0$ when the Fermi level is exactly at the band edge [34]. However, this is an overestimation as equation 1.34 shows that $n \approx 0.8 \times N_0$ in a bulk material ($D = 3$) when $E_F = 0$. Table 1.2 summarized m_d, N_c, and N_0 for Si and GaAs; N_0, calculated using equation 1.35 is comparable to that listed in Ref. [34, 37].

The result of Fig.1.5 can be discussed in term of change in E_F or T. Let's consider the case where T is constant. When $E_F \ll 0$, n/N_0 is larger for nanowires ($D = 1$) than for bulk material ($D = 3$) because only the states near the bottom of band are occupied, and according to Fig. 1.4, the DOS of nanowires near the band edge is large while that of bulk is close to zero. The situation is then reversed when $E_F \gg 0$, additional electrons must occupy higher energy states. Since the DOS of bulk material is large at high E and increasing with E, while that of nanowires is small and decreasing, n/N_0 is larger and increases more rapidly for bulk material.

In the non-degenerate and degenerate regimes, the carrier concentration is approximated by

$$n = N_0 \exp(\eta) \qquad , \eta \lesssim -3 \tag{1.36}$$

$$n = \frac{2N_0}{D}\eta^{D/2} \qquad , \eta \gtrsim +5 \tag{1.37}$$

, where equation 1.36 is valid for bulk ($D = 3$) only, and assume the classical approximation where $f_0(E) = f_{MB}(E)$. For quantum wells and nanowires, $n \propto N_0 \exp(\eta)$. In the degenerate regime, n in bulk material, quantum wells, nanowires may all be approximated by equation 1.37.

If T is variable, then two scenarios may apply. In the case of intrinsic semiconductor, where E_F is fixed in the band gap ($E_F \ll 0$), an increase in T necessarily increase η toward 0, and n/N_0 must increase as well. Equation 1.36 and equation 1.35 then suggests that $n \propto T^{D/2} \exp(-|E_F|/k_B T)$ for intrinsic semiconductor, which describes the increase in n due to thermal excitation.

In the case of extrinsic semiconductor, as T is increased, N_0 will also increased, but n is fixed to the concentration of donors (N_D) or acceptors (N_A). Therefore, both n/N_0 and η must decrease, when T is increased. When $\eta > 5$, equation 1.37 is valid, therefore

$$n = \frac{2N_0}{D} \eta^{D/2} \propto T^{D/2} \left(\frac{E_F}{T} \right)^{D/2} \propto E_F$$

Since n is fixed, E_F is fixed as well. Hence, an increase in T will decreases η toward 0, while E_F remained constant. When $\eta \lesssim -3$,

$$n \propto N_0 \exp(\eta) \propto T^{D/2} \exp(E_F/k_B T)$$

$$E_F \propto T \ln(n T^{-D/2}) \tag{1.38}$$

, which means E_F now decreases as T is increased. This describes the extrinsic to intrinsic semiconductor transition as T is increased.

Finally, the figure shows that if N_0 is increased while n and T are constant, i.e. through an increase in m_d or d, then n/N_0, η, and E_F must all decrease. This will have an important implication in the enhancement of $S^2\sigma$, as will be discussed in Chapter 2.

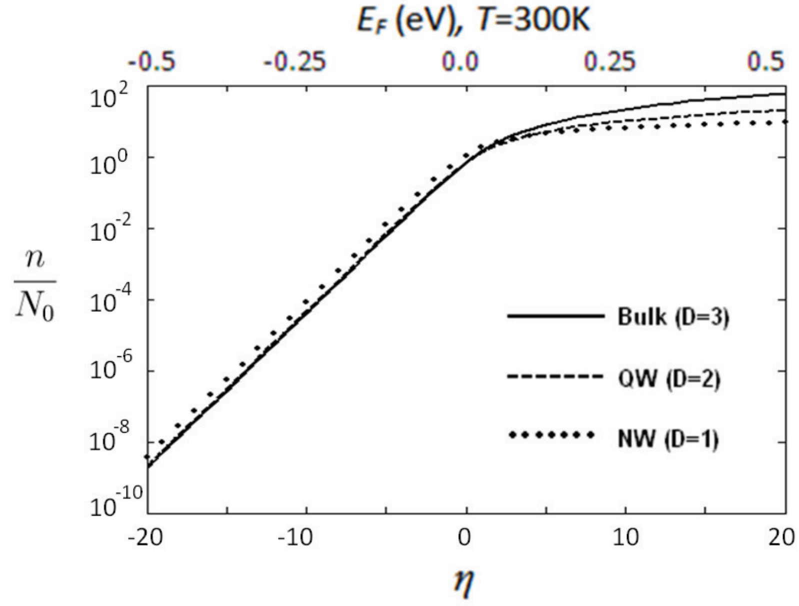


Figure 1.5: Normalized carrier concentration of a *single* band/subband as a function of the reduced Fermi level. Top axis indicates $E_F = \eta k_B T$ for $T = 300$K.

1.4.2 Electrical conductivity and Mobility

The dependent of μ on n is difficult to quantify as there are multiple factors. First, consider the electrical conductivity given by equations 1.29 and 1.31

$$\sigma = e^2\chi_0 = -e^2 \int_{-\infty}^{+\infty} g(E)\tau(E)v^2(E)\frac{df_{FD}}{dE}dE \tag{1.39}$$

$$= \frac{e^2\tau_0 N_c}{m_\sigma d^{3-D}}\Big(\frac{m_d}{\pi\hbar^2}\Big)^{D/2}\left(\frac{2}{\pi}\right)^{|D/2-1|}(k_B T)^{r+D/2}\Big(\frac{2r}{D}+1\Big)F_{r+D/2-1}(\eta) \tag{1.40}$$

Mobility may then be derived either from the definitions (equations 1.33 and 1.39), or the more explicit expression (equations 1.34 and 1.40), of n and σ. Using the

definitions, mobility is then equal to

$$\mu = \frac{\sigma}{ne} = e\frac{\int_{E_0}^{+\infty} g(E)\tau(E)v^2(E)\frac{df_{FD}}{dE}\,dE}{\int_{E_0}^{+\infty} g(E)f_{FD}(E)\,dE}$$

$$\mu = \frac{e}{m_\sigma}\left(\frac{2r}{D}+1\right)\frac{\int_{E_0}^{+\infty} g(E)\tau(E)f_0(E)\,dE}{\int_{E_0}^{+\infty} g(E)f_{FD}(E)\,dE}$$

$$\mu = \frac{e}{m_\sigma}\left(\frac{2r}{D}+1\right)\langle\tau\rangle \tag{1.41}$$

$$\langle\tau\rangle = \tau_0(k_BT)^r\frac{F_{r+D/2-1}(\eta)}{F_{D/2-1}(\eta)} \tag{1.42}$$

, where $\langle\tau\rangle$ is the equilibrium average relaxation time. The derivation above assumes $g(E)$, $v(E)$, and $\tau(E)$ from equations 1.10, 1.23, and 1.24, respectively.

μ, or more precisely, $\langle\tau\rangle$, is now shown to depends on at least three factors as emphasized by the following relation

$$(\mu \propto \langle\tau\rangle) \propto \tau_0(N,T)\cdot T^r\cdot\tau_A(\eta) \tag{1.43}$$

, where the function $\tau_A(\eta) = \left[\frac{F_{r+D/2-1}(\eta)}{F_{D/2-1}(\eta)}\right]$ is introduced for brevity. An apt description of the above equation is perhaps that τ_A describes changes in distribution of electrons, τ_0 describes changes in the environment (i.e. concentration of scattering sources), and T^r the changes in T. As shown in Table 1.1, τ_0 is mainly inversely proportional to either N or T. $\tau_A(\eta)$ is plotted in Fig. 1.6. For the purpose discussion, τ_A should be consider a pseudo-average relaxation time.

Again, two narratives based on changes in E_F and T are presented in the figure. If T is constant, then η and E_F can be used interchangeably in the discussion. When $E_F \ll 0$ and increasing, additional electrons may occupy the same energy level as the band is largely empty. Since electrons with same E should have the same $\tau(\propto E^r)$, τ_A is unchanged. However, as E_F increases beyond 0, electrons will begin to occupy states with higher E. In the case of $r > 0$, electrons with higher E will have greater τ, therefore the addition of these electrons increases τ_A. Conversely, when $r < 0$, the addition of electrons with higher E, but lower τ will decrease τ_A. If $r = 0$, then all electrons have the same τ, and the average is

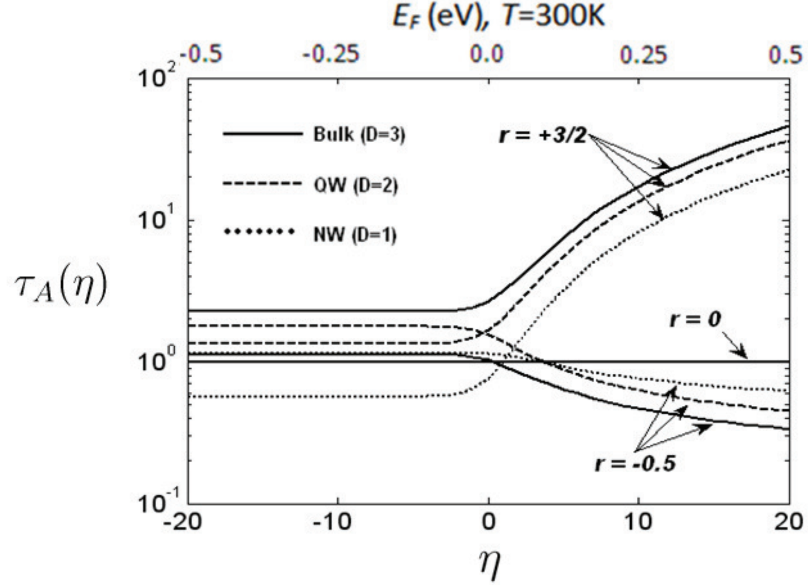


Figure 1.6: τ_A as a function of the reduced Fermi level. Top axis indicates $E_F = \eta k_B T$ for $T = 300$K.

unchanged.

This would suggests that as n is increased through doping, μ should decrease as $r = -1/2$ for $\mathrm{II}_{\mathrm{str}}$. One source of reduction is the decrease in τ_A, which describes the fact that more energetic electrons scatter more often. Another source of reduction is $\tau_0 \propto 1/N_{II}$. Doping requires addition of ionized impurity, which increases scattering centers.

If T is variable, then two cases are again considered. For intrinsic semiconductor ($E_F \ll 0$), increasing T will *increases* η toward 0, but τ_A is constant in this regime. Therefore, $\langle \tau \rangle \propto \tau_0 T^r$ for $E_F \ll 0$. Hence, $\mu \propto T^r$ for intrinsic semiconductor, which suggests that one could measure r by observing the relationship $\mu(T)$. However, if the dominant scattering process is ADP or ODP ($r = -1/2$ in bulk material), then $\tau_0 \propto T^{-1}$, and μ may vary as $T^{-1}T^{-1/2} = T^{-3/2}$ instead.

For extrinsic semiconductor (fixed n, E_F may be greater or lesser than 0), increasing T will instead *decreases* η. Therefore, when $\eta \gg 0$, τ_A would decreases with decreasing η if $r > 0$. The decrease in τ_A is balanced by increase in T^r (for $r > 0$). Therefore, without even considering τ_0, μ will not be proportional to T^r, and r could not be easily extracted from $\mu(T)$ relationship for extrinsic semiconductor. However, when $\eta < 0$, τ_A will become non-changing, and $\mu \propto \tau_0 T^r$ is true once again. This again indicates a transition from extrinsic to intrinsic semiconductor.

1.4.3 Thermal conductivity

The difference between σ and κ_e arises from the fact that while all electrons have the same amount of charge, they can have different energy. Nevertheless, both quantities are related to the concentration and mobility of carriers, such that they can related by a ratio known as the Lorentz number (L), which equals

$$L = \frac{\kappa_e}{\sigma T} = \left(\frac{k_B}{e}\right)^2 \left[\frac{(r + \frac{D}{2} + 2)F_{r+D/2+1}(\eta)}{(r + \frac{D}{2})F_{r+D/2-1}(\eta)} - \left(\frac{(r + \frac{D}{2} + 1)F_{r+D/2}(\eta)}{(r + \frac{D}{2})F_{r+D/2-1}(\eta)} \right)^2 \right] \tag{1.44}$$

L is independent of material parameters such as N_0 or τ_0, and as the Fig. 1.7 shows, is constants for the most part.

When $\eta \gg 0$, L approach a single constant value, regardless of r or D. When $E_F \gg 0$, any change in E_F will not greatly affect the overall electrons distribution as a large portion of the electronic band is already occupied. Since the material may be considered metallic in this region as the band is heavily filled, our calculation reproduced the empirical Wiedemann-Franz law, which observes that $L = \frac{\pi^2}{3}\left(\frac{k_B}{e}\right)^2 = 2.44 \times 10^{-8}$ WΩK^{-2} for metals.

When $\eta \ll 0$, L approach a constant value strongly dependent on r and D. Given the same number of electrons, σ may change with r and D, primarily through changes in μ and τ_A. The same changes would apply to κ_e, i.e. higher

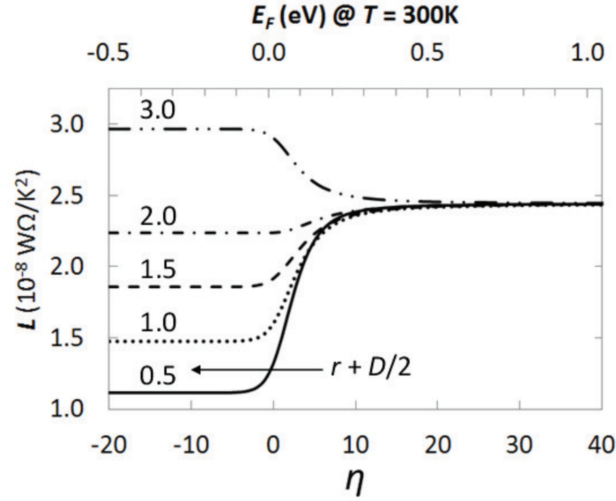


Figure 1.7: Lorentz number as a function of the reduced Fermi level. Top axis indicates $E_F = \eta k_B T$ for $T = 300$K.

velocity of electrons benefit both σ and κ_e. However, when D is large and the DOS favor high E states, more electrons can occupy higher energy states, thus increasing the amount of energy transported. Therefore, L increase with D. Likewise, when r is increased, then high energy electrons can travel faster, which benefits κ_e more than σ, and thus L increased with r as well.

κ_e cannot be modified independently of σ. An effort to increase charge conduction necessarily increase energy conduction as well. The best outcome is to ensure that charge carriers carry small amount of energy, i.e. through the use of nanowires DOS ($D = 1$) which restrict most of electrons to low E states. Additionally, scattering processes with small r are also preferred, since high energy electrons would then scatter more often, thus reducing efficiency of heat conduction.

1.4.4 Seebeck coefficient

The origin of Seebeck effect was previously explained as due to electron diffusion under temperature gradient. However, a more precise definition can be gotten through use of equations1.28 and 1.30, which shows

$$\begin{aligned} S &= \frac{1}{qT}\left(\frac{\int_{E_0}^{+\infty} g(E)\tau(E)Ev^2(E)\frac{df_{FD}}{dE}\,dE}{\int_{E_0}^{+\infty} g(E)\tau(E)v^2(E)\frac{df_{FD}}{dE}\,dE} - E_F\right) \\ &= \frac{1}{qT}\left(\frac{\langle \tau E\rangle}{\langle \tau \rangle} - E_F\right) \end{aligned} \tag{1.45}$$

, where $\frac{\langle \tau E\rangle}{\langle \tau \rangle} = \langle E\rangle_\tau$ is the relaxation time weighted-average energy of electrons. If relaxation time is constant ($r = 0$), then $\langle E\rangle_\tau = \langle E\rangle$, the average energy. Essentially, equation 1.45 shows that S is proportional to the difference between $\langle E\rangle_\tau$, and E_F. S is therefore a measure of the difference in electron energy at the current temperature relative to that at the absolute zero ($= E_F$). In other words, the diffusive Seebeck effect originates from the tendency of *heated* electrons with an average energy of $\langle E\rangle_\tau$, to return to the lowest energy level, which is E_F.

S may be calculated by using the expression for χ_i in equation 1.29, which yields

$$S = \frac{k_B}{q}\left(\frac{(r + \frac{D}{2} + 1)F_{r+D/2}(\eta)}{(r + \frac{D}{2})F_{r+D/2-1}(\eta)} - \eta\right) \tag{1.46}$$

This equation shows that S is independent of N_0 and τ_0, as these terms cancel out from the nominator and denominator. S may be calculated for a combine $r + D/2$ index, as shown in Fig. 1.8, which implies that r and D have similar influence, and are interchangeable, i.e. a reduction of D by 1 is equivalent to a reduction of r by 1/2.

The figure shows that S would varies in opposition to n. If n is increased through doping, then E_F and $|S|$ would be decreased. Conversely, if $|S|$ is increased by reducing E_F, then n would decrease. The method to find the optimal n where $S^2\sigma$ is maximized is the subject of section 2.2. S may estimated in the non-

degenerate and degenerate regimes as [27]

$$S = \frac{k_B}{q}(r + D/2 + 1 - \eta) \qquad , \eta \lesssim -3 \tag{1.47}$$

$$S = \frac{\pi^2 k_B}{Dq}\frac{(r + D/2)}{\eta} \qquad , \eta \gtrsim +5 \tag{1.48}$$

$|S|$ also increases with T for extrinsic semiconductors. As T is increased, η always decreases, but electrons are promoted to higher energy level. This thermal excitation leads to a slight increase in average energy of electrons in that band. When $\eta >> 0$, then E_F is unaffected by change in T, therefore $|S|$ increased gradually as a result of increase in $\langle E \rangle_\tau$. However, when $\eta < -3$, E_F will decrease according to equation 1.38. The decrease in E_F combines with the concurrent increase in $\langle E \rangle_\tau$ leads to a more rapid rate of change observed in the left panel of Fig. 1.8.

$|S|$ will reach a maximum as semiconductor become intrinsic. Once E_F has reached the middle of band gap and cannot be reduced further, η will begins to increases again. Although $\langle E \rangle_\tau$ will continues to increase with T, $|S|$ will decreases as the T in the reciprocal of equation 1.45 dominates. However, before that point is reached, $|S|$ is also reduced by the contribution from the opposing charge carriers. Let's assume that the material is n-type for discussion sake. As E_F is reduced into the band gap, electrons from the VB is thermally excited into the CB, leaving holes in the VB. As holes concentration increases, conductivity of VB will becomes significant. Since S for p-type material is positive, the Seebeck effect due to holes will cancel the Seebeck effect due to electrons. The extrinsic-to-intrinsic transition temperature therefore marks the maximum operating temperature for thermoelectric material.

$|S|$ also decreases as D is decreased. This dependence on the DOS shape is sometime misunderstood. It is believed that DOS of nanowires will yield large $|S|$ due to the high rate of change associated with the peak-like profile. However, as Fig. 1.8 shows that $|S|$ increases with D, it is actually the DOS of bulk material, which favors high energy states, that yield larger $|S|$ given the same η. Given the

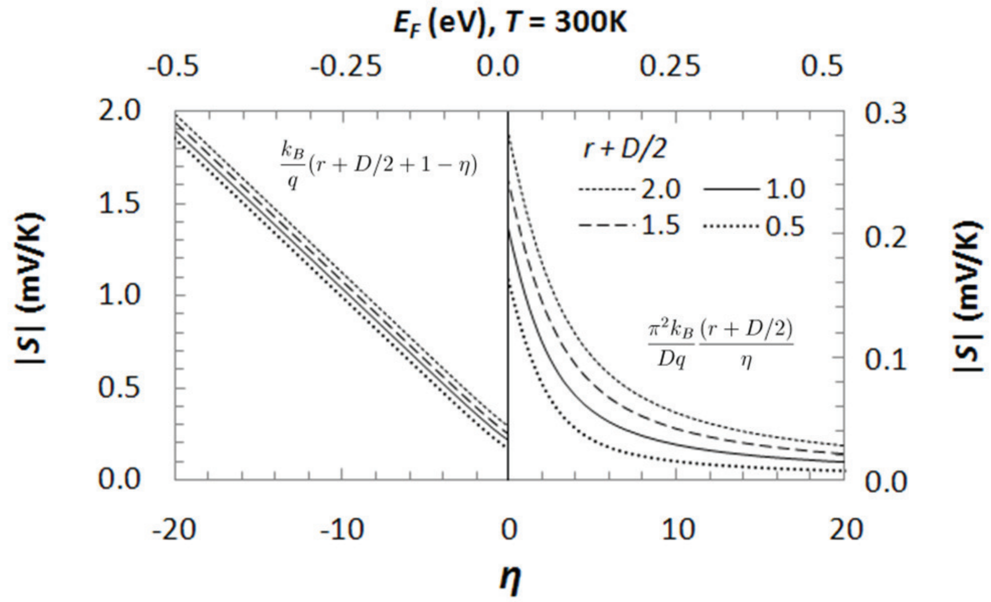


Figure 1.8: Seebeck coefficient as a function of the reduced Fermi level. Figure is split into 2 panes for clarity; Y-axis units are same for both panes.

same η, $\langle E \rangle_\tau$ would be lower in nanowires since its DOS favor low-energy states. The dependence of S on D was proposed as one method of verifying the confinement effect in two-dimensional superlattices [38].

The relaxation time parameter r also affects S in similar fashion to D, by changing the distribution of the $\tau(E)$. Just as increasing D pushes electrons up to higher energy states, and thus $\langle E \rangle_\tau$, increasing r would equate to more electrons with longer relaxation time. $\langle E \rangle_\tau$ will then increases due to the increase in weight for higher energy states. Therefore, large r and D is needed to obtain large $\langle E \rangle_\tau$ and $|S|$. However, increasing $\langle E \rangle_\tau$ will also increases κ_e. It is unavoidable that electrons should carry energy (good for κ_e) if one want to harness that energy (through S).

Part of the content for this chapter was taken from the published journal article: P. Pichanusakorn, and P. R. Bandaru, "Nanostructured Thermoelectrics", *Materials Science and Engineering: R: Reports*, 67, 19-63 (2010)

2 Power Factor Enhancement

2.1 Maximization of Z

Figure 2.1 shows the tentative trends of several transport coefficient as a function of η. Imagine that a semiconductor is being doped. As n increases, σ is increased, though $|S|$ would decrease as η is raised along with n. Therefore, $S^2\sigma$ exhibits a peak due to this competition between S and σ. Meanwhile, κ_e increases monotonically along with σ. Since κ_l is primarily a property of the lattice, it is independent of η or n, and is treated as a constant.

In the degenerate/metallic regime $\eta \gg 0$, Z is small because of (1) small $|S|$, and (2) very large κ_e due very high n. In metals, κ_e is comparable to κ_l, and as an exercise, let's consider the case where $\kappa_e \gg \kappa_l$. In such case,

$$Z = \frac{S^2\sigma}{\kappa_e} = \frac{S^2}{LT} \tag{2.1}$$

, which means Z can only decrease if η is increased, since $|S|$ would decrease, while L remains at 2.44×10^{-8} WΩK^{-2}. Therefore, a higher Z is obtained if η is reduced, i.e. if the material is made less metallic.

In the non-degenerate regime, Z is also small because while $|S|$ can be very large, σ is small. κ_e is also small, and one may consider $\kappa_l \gg \kappa_l$, such that

$$Z = \frac{S^2\sigma}{\kappa_l} \propto S^2\sigma \tag{2.2}$$

Hence in this regime, Z would increase with $S^2\sigma$. In the following section, it will be shown that $S^2\sigma$ is maximized in the moderately-doped regime. Therefore, in this

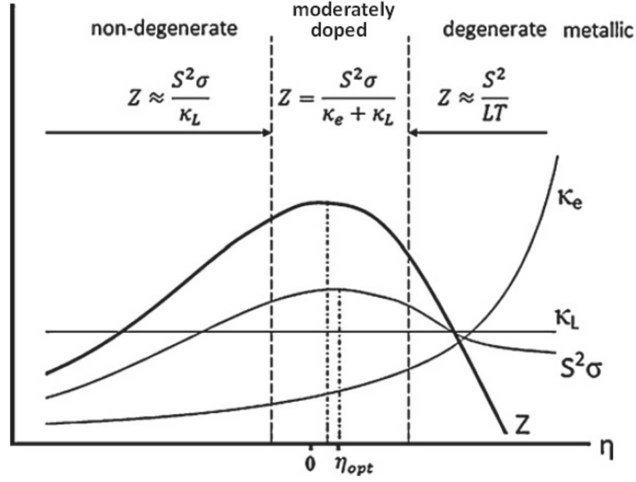


Figure 2.1: Maximization of Z.

regime, Z is increased if η is increased, i.e. if the material is made less insulating. It is not a coincidence that good thermoelectric materials such as Bi_2Te_3, PbTe, or $Si_{0.7}Ge_{0.3}$ are moderately-doped semiconductors [39, 40].

2.2 The Optimal Reduced Fermi Level

Through mathematical treatment, it can be shown that there is a narrow range of *universal* optimal reduced Fermi level (η_{opt}) where $S^2\sigma$ is maximized that extends to a large number of materials, over a wide range of working temperature [41]. From the expression of n, μ, and S in equations 1.34, 1.41, and 1.46, the power factor can be written as

$$\begin{aligned} S^2ne\mu &= \frac{N_0\tau_0}{m_\sigma}k_B^2(k_BT)^r\left(\frac{4}{\pi}\right)^{|D/2-1|}P(\eta) \\ &= \left[\frac{N_c\tau_0k_B^2}{d^{3-D}m_\sigma}\left(\frac{m_d}{\pi\hbar^2}\right)^{D/2}\left(\frac{2}{\pi}\right)^{|D/2-1|}(k_BT)^{r+D/2}\right]P(\eta) \end{aligned} \quad (2.3)$$

$$P(\eta) = \frac{2}{D}(r+D/2)F_{r+D/2-1}(\eta)\left(\frac{(r+D/2+1)F_{r+D/2}(\eta)}{(r+D/2)F_{r+D/2-1}(\eta)} - \eta\right)^2 \quad (2.4)$$

Thus, $S^2\sigma$ can be separated into a product of two parts: (1) a collection of physical constants, material parameters (such as N_c, m_d, d, etc.), and T, and (2) the $P(\eta)$ function. Since the first part is constant, it is clear that $S^2\sigma$ is maximized when $P(\eta)$ is maximized. Given that the assumptions that electron transport in the material can be described by the BTE, and that the material has a single parabolic band with elastic scattering, are broad and conventional, slight deviation, such as non-parabolicity of the band, should not have a significant impact on the results shown here.

Figure 2.2 shows the value of $P(\eta)$ function, calculated with equation 2.3, for the combined index $(r + D/2)$ of 0.5, 1.5, 1.5, and 2.0. All graphs are normalized with respect to the maximum value found in each instance. The figure then shows that $P(\eta)$ exhibits a peak at some optimal value, η_{opt}, which decreases with r and D. When $(r + D/2) = 2$, $P(\eta)$ was found to increase monotonically with no absolute maximum, and no η_{opt}. This happens as μ also increases with increasing η when $r > 0$ (minimum value of D is 1). The increase in both μ and n were then sufficient to make up for the decrease in S^2. Consequently, the $S^2\sigma$ always increases with η. This scenario is unlikely in practice since most scattering mechanism have $r = 1 - D/2$ (see Table 1.1), therefore $r + D/2 = 1$. The $\mathrm{II_{wk}}$ process with $r = +3/2$ should also converts to $\mathrm{II_{str}}$ process with r=1-D/2 when n is large.

The value of η_{opt} is listed in Table 2.1, along with the corresponding optimal Seebeck coefficient, S_{opt}. Since S also depends solely on r, D, and η according to equation 1.30, for every η_{opt} value there will be only a single corresponding S_{opt}. Identical values were also independently reached by Raseong *et al.* [42] through a different method based on Landauer formalism. Since each η_{opt} is uniquely linked to a corresponding S_{opt}, the maximum $S^2\sigma$ could be identify by measuring just S. When $S = S_{opt}$, one then know that maximum $S^2\sigma$ should have been reached. According to Table 2.1, $S^2\sigma$ of any materials will be highest when S is in range of 130-187 μV/K, depending on r and D.

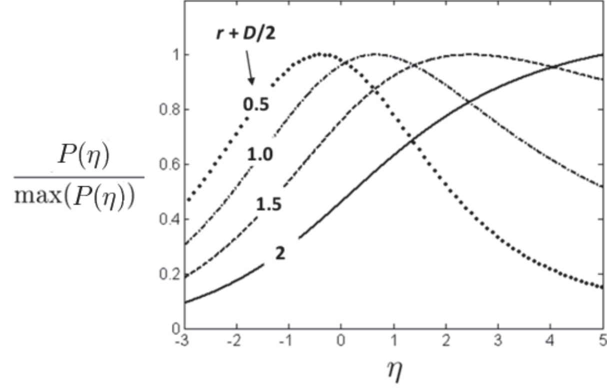


Figure 2.2: Normalized $P(\eta)$ function showing optimal reduced Fermi level.

Furthermore, the optimal carrier concentration (n_{opt}) is also easily identified with the knowledge of η_{opt}. Using equation 1.34,

$$\begin{aligned} n_{opt} &= N_0 \left(\frac{4}{\pi}\right)^{|D/2-1|} F_{D/2-1}(\eta_{opt}) \\ &= \frac{N_c}{d^{3-D}} \left(\frac{m_d k_B T}{\pi \hbar^2}\right)^{D/2} \left(\frac{2}{\pi}\right)^{|D/2-1|} F_{D/2-1}(\eta_{opt}) \end{aligned} \tag{2.5}$$

, where η_{opt} is as given in Table 2.1, and N_0 was given in Table 1.2. Note that n_{opt} increases as $T^{(}D/2)$, which means a material can only be optimally doped for a single T. If the material is to be used at higher than specified T, then its dopant must be increased again to obtain the maximum possible $S^2\sigma$.

2.2.1 Survey of materials

Literature survey of various materials confirms that for each group of material, $S^2\sigma$ is maximized for samples with $S \approx S_{opt}$. In Fig. 2.3, $S^2\sigma$ PbTe [43, 1] is shown as a function of n, where $n \approx 7 \times 10^{18}\,\mathrm{cm}^{-3}$ appears to be the optimal carrier concentration. The same data is then replotted as a function of $|S|$ in Fig. 2.4, where it is shown that $S^2\sigma$ is maximized at $S \approx 150\,\mu\mathrm{V/K}$, close to the predicted S_{opt}. In the same figure, $S^2\sigma$ of $SrTiO_3$ is also shown. The La-doped

Table 2.1: Optimal reduced Fermi level (first entry) and optimal Seebeck coefficient (second entry) for several combination of r and D

D	r = -½	0	+½	+³⁄₂
3	0.67 167	2.47 130	no limit -	no limit -
2	-0.37 187	0.67 167	2.47 130	no limit -
1	-1.14 197	-0.37 187	0.67 167	no limit -
		Neutral Impurity	ADP ODP Ionized Impurity (strongly screened)	Ionized Impurity (weakly screened)

samples [2], also seem to show a peak between 150 − 250 μV/K. Furthermore, data of Nb-doped $SrTiO_3$ [3] suggests that the maximum $S^2\sigma$ may still be found near $S \approx S_{opt}$, though there are insufficient number of samples with greater n (and lower $|S|$) to confirm the hypothesis.

n_{opt} calculations using equation 2.5 show close agreement with the measurement data as well. Table 2.2 shows the calculation results for PbTe, La-$SrTiO_3$, and Nb-$SrTiO_3$. PbTe has 4-fold degenerate band with $m_L = 0.35m_0$, and $m_t = 0.034m_0$ at 300K [44]. La−$SrTiO_3$ has 3-fold orbital degeneracy for the conducting Ti states [3]. Nb−$SrTiO_3$ have 3-fold degeneracy as well, but $N_c = 1$ was assigned since it m_d stated in the publication already included N_c in its calculation [45]. $SrTiO_3$ is known for having very large effective mass. In all cases, the range of calculated n_{opt} matches the observation in Fig. 2.3 for PbTe, and in Fig. 2.4 for both $SrTiO_3$. As the calculation and measurement data shown, S_{opt} were the same regardless of material type (PbTe vs. $SrTiO_3$) and T (300K vs. 1,000K), while n_{opt} differed by 4 order of magnitude according to material parameters and T as described in equation 2.5. Both confirmed the validity of η_{opt}.

The existence of S_{opt} that persists over a temperature range is further il-

Table 2.2: Optimal carrier concentration of PbTe and $SrTiO_3$

Material	PbTe	La-$SrTiO_3$	Nb-$SrTiO_3$
m_d/m_0	0.07	1.62	7-12
N_c	4	3	1
T (K)	300	300	1000
η_{opt}	0.67-2.47	0.67-2.47	0.67
$n_{opt}(\mathrm{cm}^{-3})$	$2.6 - 7.0 \times 10^{18}$	$2.0 - 5.5 \times 10^{20}$	$3.6 - 8.0 \times 10^{21}$

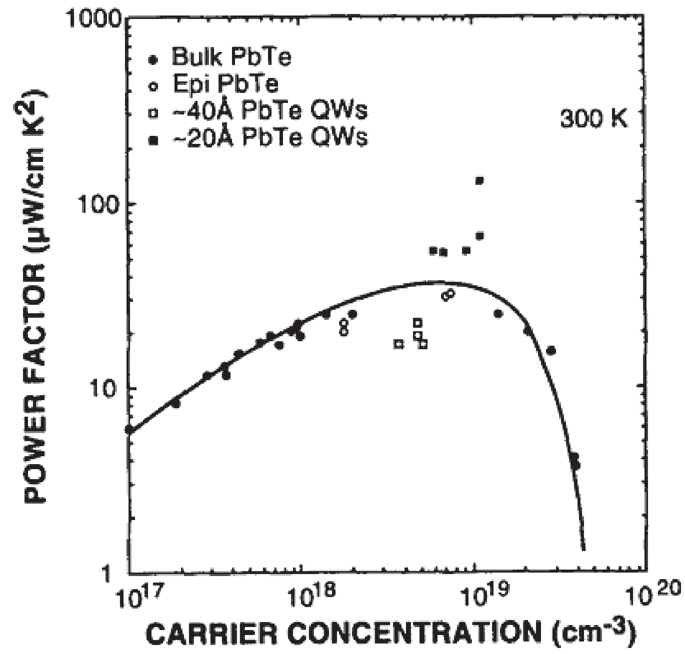


Figure 2.3: Power factor of PbTe versus carrier concentration [1].

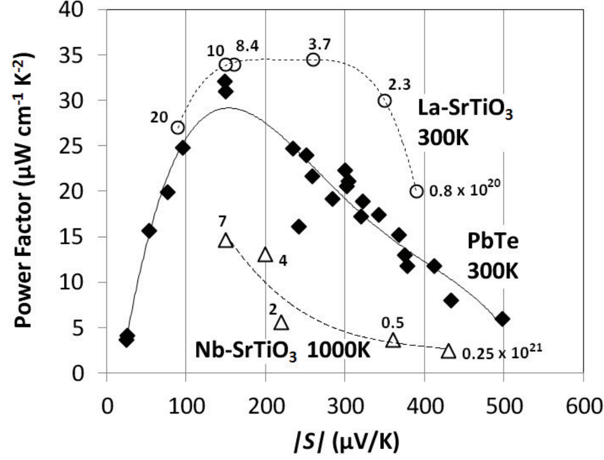
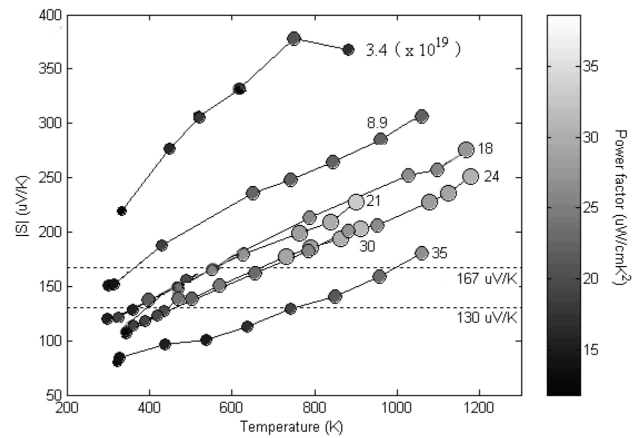


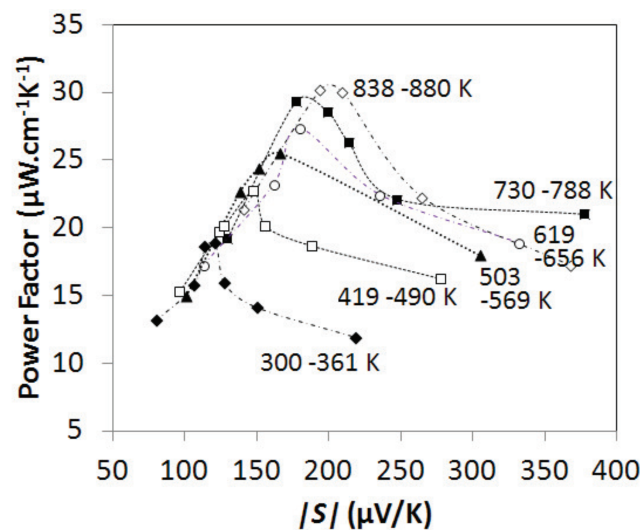
Figure 2.4: Power factor of PbTe [1], and La- and Nb-doped $SrTiO_3$ [2, 3]. Data label indicates n in cm^{-3}.

lustrated by the following examination of SiGe. In Fig. 2.5a, S of p-type SiGe is shown as a function of T for a number of samples with n of $3.4 - 35 \times 10^{20}\,cm^{-3}$. S increases with T for all samples, and is larger for sample with lower n. $S^2\sigma$ also increases with T for all samples, as indicated by the larger size and lighter tone of each data points. However, when comparing different samples over a small temperature interval, $S^2\sigma$ also appears to be largest for sample whose $S \approx S_{opt}$.

In Fig. 2.5b, $S^2\sigma$ of the same samples is shown as function of S. Each data series contains one measurement from each sample within a small T interval, i.e. the first series contain only measurements taken between 300-361 K. The peak $S^2\sigma$ is then clearly shown to be at $S \approx 120 - 194\,\mu V/K$. The figure also shows both peak $S^2\sigma$ and S_{opt} increasing with T. The increase in peak $S^2\sigma$ is explained by equation 2.3, where it is shown that $S^2\sigma \propto T^{D/2}$. Meanwhile, the shift in S_{opt} may be due to phonons scattering. At high T, phonons scattering dominates as $\tau_0 \propto T^{-1}$. Therefore, r value may decreases toward -0.5, causing the η_{opt} to decrease, and S_{opt} to increase.



(a) Each series represents measurements at various temperature of *one* sample. Numerical label indicates n in cm^{-3}. Power factor indicates by size and tone of the data point according to the scale bar on the right of figure.



(b) Each series represents measurements over the same temperature interval from *multiple* samples.

Figure 2.5: Seebeck coefficient and power factor of multiple p-type SiGe samples [4]

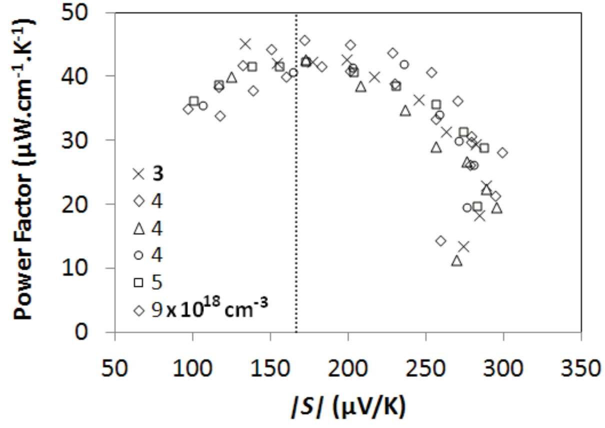


Figure 2.6: Power factor versus Seebeck coefficient of Bi_2Te_3 [5]. Each series comprises measurement from the same sample measured over temperature range of approximately 80-350K.

Similar analysis was applied to Bi_2Te_3 with data shown in Fig. 2.6. Whereas each data series in Fig. 2.5b represents measurement of multiple samples in the same T interval, the data series in Fig. 2.6 represents measurement of a single samples taken over multiple T in the range of 80-360K. Once again, the maximum $S^2\sigma$ is found at $S \approx S_{opt}$ for all samples regardless of n, which ranged from $2.5 - 90 \times 10^{18}\,\text{cm}^{-3}$. However, the temperature where $S^2\sigma$ was maximized increases with n. For example, sample with $n = 2.5 \times 10^{18}\,\text{cm}^{-3}$ was optimized for $T \approx 100\text{K}$, while sample with $n = 9 \times 10^{18}\,\text{cm}^{-3}$ was optimized for $T \approx 160\text{K}$.

Thus far, the highest $S^2\sigma$ achieved in bulk materials, as seen from Fig. 2.4-2.6, is approximately $40 - 45\,\mu\text{W/cmK}^2$. The ZT value for these materials is in the range of 0.8-1.0, with κ of 1-4 W/mK. Assuming $S_{opt} = 167\,\mu\text{V/K}$, σ at maximum $S^2\sigma$ would correspond to approximately $1,800\,\Omega^{-1}\text{cm}^{-1}$ for these materials. However, while the S and σ are similar, the constituent n and μ can vary significantly. Whereas Bi_2Te_3 and PbTe possess small n and large μ due to their small m^*, SiGe and $SrTiO_3$ have large n and small μ due to their large m^*. Larger $S^2\sigma$ exceeding $100\mu\text{W/cmK}^2$ have been reported in PbTe quantum wells [43, 1], and

Si/Ge superlattices [46], where the enhancement is due to increase in N_0 through reduction of d. The enhancement of $S^2\sigma$ through N_0 is the subject of the next section.

The identification of η_{opt} has enabled a straightforward method to find the maximum $S^2\sigma$, though comparison with S_{opt}. It also establishes S_{opt} as the reference point for what should be considered a "large" ($|S| > S_{opt}$) or "small" ($|S| < S_{opt}$) Seebeck coefficient.

2.3 Density of States Enhancement

At the optimal doping point, equation 2.3 show that the peak power factor is

$$(S^2\sigma)_{max} \propto \left[\frac{N_c \tau_0 k_B^2}{m_\sigma d^{3-D}} \left(\frac{m_d}{\pi \hbar^2} \right)^{D/2} \left(\frac{2}{\pi} \right)^{|D/2-1|} (k_B T)^{r+D/2} \right]$$

These are the same parameters that first appear in equation 1.31 for σ. Indeed, it was shown that $S = S_{opt}$ at $n = n_{opt}$, therefore it is σ that determines the maximum $S^2\sigma$. σ is increased if μ or n in enlarged. The former requires that τ_0 is increased while m_σ is reduced. Modulation doping offers one alternative to reducing τ_0 in nanostructures [47, 1]. To increase n, any of the three parameters in N_0 can be utilized. N_c could be increased by carrier pocket engineering, where confinement energy is used to bring multiple subbands into the same energy level [48, 49]. Quantum confinement can also be used to reduce the dimensionality of the electron gas. The DOS can then be increased by reducing the d, i.e. by making quantum wells thinner or nanowires smaller in diameter. Lastly, N_0 is increased if m_d is increased. Both of the two latter approaches are discussed further in the rest of this chapter, and form the basis for the experimental work covered in Chapters 4 and 5.

2.3.1 Confinement Effect

The enhancement of Z by the use of quantum confinement effect was first proposed by Hicks *et al.* [50, 51] in 1993, and experimentally demonstrated by the same group in 1996 [43, 1]. The result of the latter work was previously shown in Fig. 2.3, where it is observed that $S^2\sigma$ of 4 nm-thick PbTe quantum wells is slightly less than the maximum $S^2\sigma$ of bulk PbTe. However, the $S^2\sigma$ of 2 nm-thick sample significantly exceeded the maximum $S^2\sigma$ of bulk PbTe. This enhancement of $S^2\sigma$ is due to the increase in the DOS, through the reduction of d in the N_0 term.

To illustrate the mechanism behind the enhancement, first consider the increase in $S^2\sigma$ due to changes in n, i.e. doping. To simplify matter, μ is assumed to be constant (i.e. $r = 0$, and τ_0 is constant), such that only changes in $S^2 n$ are relevant. Figure 2.7a shows the distribution of electrons in (bulk) Si conduction band when $E_F = -0.05\,\text{eV}$ (arbitrary value). The DOS is calculated using equation 1.10 assuming $N_c = 6$, $m_d = 0.32 m_0$, and $D = 3$, i.e.

$$g(E) = 6\left(\frac{0.32 m_0}{\pi\hbar^2}\right)^{3/2}\left(\frac{2}{\pi}\right)^{1/2} E^{1/2}$$

The electron distribution is then the product of $g(E) f_{FD}(E)$. The total carrier concentration in the band is then the result of integration of the electron distribution, $n_1 = \int_0^{+\infty} g(E) f_{FD}(E)\, dE = 3.8 \times 10^{18}\,\text{cm}^{-3}$. S may also be calculated with equation 1.30, which shows that $|S_1| = 0.4\,\text{mV/K}$. According to equation 1.45, the average energy given this electron distribution must then be

$$\langle E \rangle = |S_1| eT + E_F = 0.4 \times 10^{-3} \times 300 - 0.05 \text{ eV} = 0.07 \text{ eV}$$

Finally, $S_1^2 n_1 = 6 \times 10^{23}\ \mu\text{V}^2/\text{cm}^3\text{K}^2$ for this idealized material.

Figure 2.7b then shows the case where the material is doped further, such that E_F is now arbitrarily increased to 0.10 eV. The increase in E_F caused a significant increase in occupation of the energy states. The same calculations then shows that $n_2 = 1.7 \times 10^{20}\,\text{cm}^{-3}$, $|S_2| = 0.1\,\text{mV/K}$, and

$$\langle E \rangle = |S_2| eT + E_F = 0.1 \times 10^{-3} \times 300 + 0.10 \text{ eV} = 0.13 \text{ eV}$$

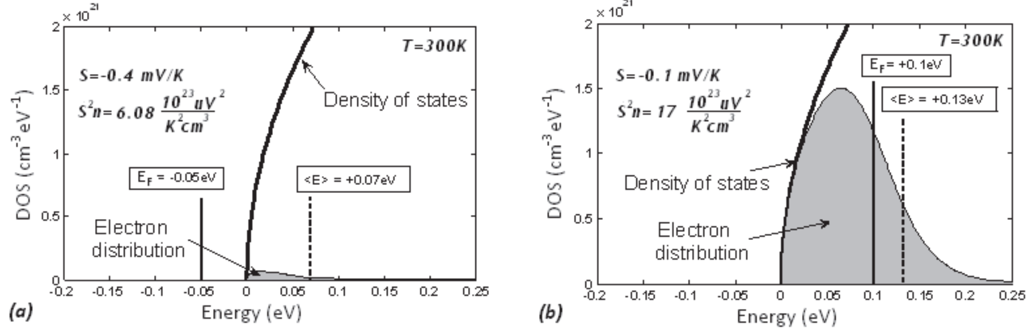


Figure 2.7: Electron density in bulk Si conduction band when the Fermi level is (a) −0.05eV, and (b) +0.1eV. The density of states is shown as a solid black line, while the electron density is shown by the grayed area.

Higher average energy is observed because higher energy states are being occupied by the additional electrons. However, the increase in $\langle E\rangle$ is smaller than that of E_F, thus the energy gap narrows and $|S|$ decreased. Despite the reduction in S, $S_2^2 n_2$ has increased by a factor of three from $S_1^2 n_1$ to $17 \times 10^{23}\,\mu\mathrm{V}^2/\mathrm{cm}^3\mathrm{K}^2$ due to increase in n. If μ is not impacted by the increase in dopant concentration, then $S^2\sigma$ would likewise increase by a factor three as well. Since $S_1 > S_{opt} > S_2$, where $S_{opt} = 0.13\,\mathrm{mV/K}$ for $r = 0$ and $D = 3$, it is predicted that $n_1 < n_{opt} < n_2$, i.e. further increase in $S^2\sigma$ can be expected if n is changed to value somewhere between $3.8 \times 10^{18}\,\mathrm{cm}^{-3}$ and $1.7 \times 10^{20}\,\mathrm{cm}^{-3}$.

The impact of a change in DOS is then illustrated. Assuming that the Si in Fig. 2.7b is now fabricated into Si nanowires with diameter of 2 nm. Figure 2.8a the show the distinctively peak-like DOS for the nanowires ($D = 1$), calculated as

$$g(E) = \frac{6}{2^2}\left(\frac{0.32m_0}{\pi\hbar^2}\right)^{1/2}\left(\frac{2}{\pi}\right)^{1/2} E^{-1/2}$$

If $n = 1.7 \times 10^{20}\,\mathrm{cm}^{-3}$ and $T = 300\,\mathrm{K}$ as before, then the interpolation of equation 1.34 shows that the E_F must decrease from +0.10eV in bulk Si to the current value of -0.01eV in Si nanowires. The decrease in E_F is due to the fact that the DOS of nanowires is significantly larger than the DOS of bulk Si for the range of energy being occupied by the electrons (approximately up to 0.15 eV as shown in

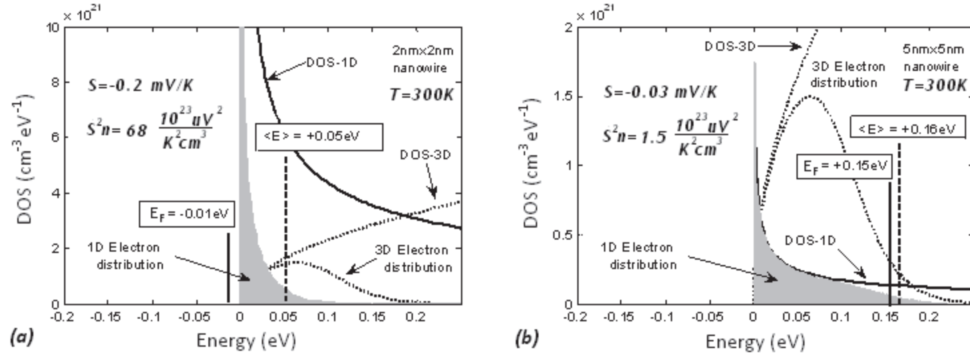


Figure 2.8: Electron density in conduction band of Si nanowires with diameter of (a) 2 nm, and (b) 5 nm. DOS of bulk Si also plotted for comparison.

the figure). Therefore, lower E_F is needed to fill the enlarged DOS of nanowires with the same amount of carriers. $S = 0.2\,\mathrm{mV/K}$ is then observed, which indicates $\langle E \rangle = 0.05$ eV. Although the amount of electrons remains the same, the average energy has decreased because electrons are now piled up in the lower energy states, instead of being spread over higher energy states as in bulk Si. Nonetheless, given the same n, $|S|$ has doubled due to large decrease in E_F. A four-fold increase in S^2n to $68 \times 10^{23}\,\mu\mathrm{V}^2/\mathrm{cm}^3\mathrm{K}^2$ was therefore observed.

However, if the nanowires diameter was 5 nm, a decrease in S^2n was observed instead due to insufficient increase in the DOS. Figure 2.8b shows that the DOS of 5 nm Si nanowires is mostly smaller than that of bulk Si over the energy range of the occupied states (up to $\sim 0.25\,\mathrm{eV}$). Given a much reduced DOS, a higher $E_F = +0.15\,\mathrm{e}$ is then needed to fill the band such that $n = 1.7 \times 10^{20}\,\mathrm{cm}^{-3}$. $|S|$ is then decreased to $0.03\,\mathrm{mV/K}$ by the combination of both smaller D, and larger E_F. $\langle E \rangle$ has increased to $0.16\,\mathrm{eV}$, as most electrons still prefers low energy states, the increase is small relative to bulk value. Overall, S^2n has reduced by an order of magnitude from bulk value in Fig. 2.7b to $1.5 \times 10^{23}\,\mu\mathrm{V}^2/\mathrm{cm}^3\mathrm{K}^2$. The implication is then obvious: there is an enhancement of S^2n only if the d is sufficiently small, such that the DOS in quantum well or nanowires may become larger than that in bulk material over the energy range of the occupied states.

2.3.2 Minimum length

The minimum d required to obtain a net increase in $S^2\sigma$ can be calculated from the ratio of the maximum $S^2\sigma$ of the nanostructures over that of its bulk value [52]. Let's define this ratio as the gain factor, i.e.

$$G = \frac{(S^2\sigma)_{D=2,1}}{(S^2\sigma)_{D=3}}$$

The identification of η_{opt} greatly simplify the problem, as the value of η to use in the equation above is known. Furthermore, S_{opt} is known as well. For clarity, the parameters belonging to the nanostructures are shown with the prime notation, while the parameters if bulk material are shown as normal. This accounts for the possibility that parameters such as N_c or m_d could change as a result of quantum confinement as well. Substituting in equation 1.40 for σ, d is related to the G and other parameters as

$$d = \sqrt{\frac{2\hbar^2}{k_B T}\left[\frac{\pi}{2G}\sqrt{\frac{{m'_d}^D}{{m_d}^3}}\left(\frac{m_\sigma \tau'_o N'_c}{m'_\sigma \tau_o N_c}\right)(k_B T)^{r'-r}\frac{3}{D}\left(\frac{2r'+D}{2r+3}\right)\right.} \\ \left.\frac{F_{r'+D/2-1}(\eta'_{opt})}{F_{r+1/2}(\eta_{opt})}\left(\frac{S'_{opt}}{S_{opt}}\right)^2\right]^{\frac{1}{3-D}} \tag{2.6}$$

The minimum confinement length d_{min}, below which there is an enhancement in $S^2\sigma$, is then defined by the condition $G = 1$.

Let's consider the simplest case where the relaxation time is constant ($r = 0$ and $\tau'_o = \tau_o$), the degeneracy of conduction valley is preserved ($N'_c = N_c$), and the effective mass is isotropic ($m'_\sigma = m_\sigma = m'_d = m_d = m$). In such case,

$$d = \sqrt{\frac{2\hbar^2}{m k_B T}}\left[\frac{\pi}{2G}\frac{F_{D/2-1}(\eta'_{opt})}{F_{1/2}(2.47)}\left(\frac{S'_{opt}}{130\,\mu\text{V/K}}\right)^2\right]^{\frac{1}{3-D}} \tag{2.7}$$

, where η'_{opt} and S'_{opt} will equal 0.67 and 167 μV/K for $D = 2$, and -0.37 and 187 μV/K for $D = 1$, respectively. Conduction valley degeneracy is preserved by

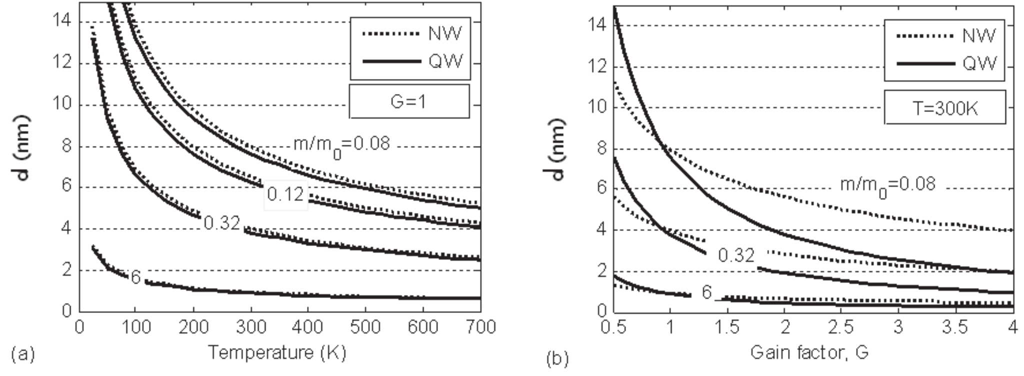


Figure 2.9: The minimum confinement length of quantum well and nanowire required to achieve (a) gain factor of 1 at various temperature and effective mass, and (b) gain factor between 0.5 to 4 at 300K.

selecting the proper confinement direction. For example, the four conduction valleys of PbTe are situated on the $\langle 111 \rangle$ axes, therefore the degeneracy of the four valleys would be broken into one longitudinal and three oblique valleys if the quantum wells were grown along the [111] direction. This would effectively reduce N_c from 4 to 1. However, if the quantum wells were grown along the $\langle 100 \rangle$ directions, the valley degeneracy would be preserved [53].

Figure 2.9a then shows the minimum quantum well thickness/nanowires diameter required obtain $G = 1$ for a number of thermoelectric materials, including Bi_2Te_3, PbTe, SiGe, and $SrTiO_3$; d_{min} was calculated using equation 2.7. The figure shows that the $S^2\sigma$ enhancement becomes increasingly difficult to obtain at higher T, or with materials with larger m, as smaller nanostructures would be needed. However, as the T is decreased, d_{min} increases, suggesting that nanostructured thermoelectrics may be suited for low-T application.

While d_{min} is approximately the same for quantum well and nanowire for $G = 1$, as shown in Fig. 2.9a, it decreases at a greater rate for quantum well at larger G value, as shown in Fig. 2.9b. This is consistent with the prediction of larger $S^2\sigma$ enhancement in nanowires [51] compared to quantum wells [50].

Finally, if ADP, ODP, II_{str}, or any other scattering processes with $r = 1-3/2 = -1/2$ for $D = 3$ is assumed (see Table 1.1), then $r' - r = 1 - D/2 - 1/2 = (1-D)/2$, $\eta'_{opt} = \eta_{opt}$, $S'_{opt} = S_{opt}$, and equation 2.6 reduces to

$$d = \sqrt{2}\hbar \left[\frac{3\pi}{2DG} \left(\frac{\tau'_0}{\tau_0} \right) \frac{1}{\sqrt{\Pi_{i=1}^{3-D} m_i}} \right]^{\frac{1}{3-D}} \tag{2.8}$$

Since a constant relaxation time is no longer assumed here, it would be necessary to know how τ_0 and τ'_0 differs. It is beyond the scope of this work to consider τ_0 in detail. Nevertheless, it is interesting to note that while a retains the same dependence on G and effective mass, it is independent of T.

This calculation provides a basis for the study of the confinement effect in n-type Si nanowires with diameter below 5 nm, where an enhancement of $S^2\sigma$ should be observed. This work is discussed in Chapter 4.

2.3.3 Effective Mass enhancement

An enhancement of the $S^2\sigma$ by increased m_d was demonstrated in bulk PbTe doped with Tl [6]. The addition of Tl significantly affects PbTe as its energy level is located within the PbTe CB. The distortion then increases the DOS near the Tl resonant energy level, E_R, as illustrated in Fig. 2.10. This increase in the DOS is equivalent to an increase in m_d as according to equation 1.35. Figure 2.11 shows that m_d is indeed increased by a factor of three over that of the PbTe control samples.

The increase in m_d has similar effect on S as the decrease in d, explained in Fig. 2.8. Namely, an increase in m_d shall increase the DOS, which then lowers E_F, but increases $|S|$, given a constant n. An enhancement of $|S|$ by a factor of 1.7 to 3.0 given the same n in Tl-PbTe as in Na-PbTe was reported. Since μ was only slightly reduced by addition of Tl and κ remained the same, an enhancement

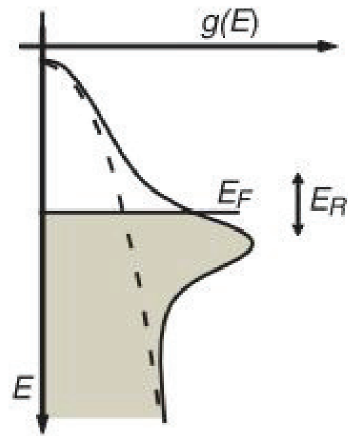


Figure 2.10: Distortion of PbTe density of states around E_R due to Tl-resonant states. Adapted from [6].

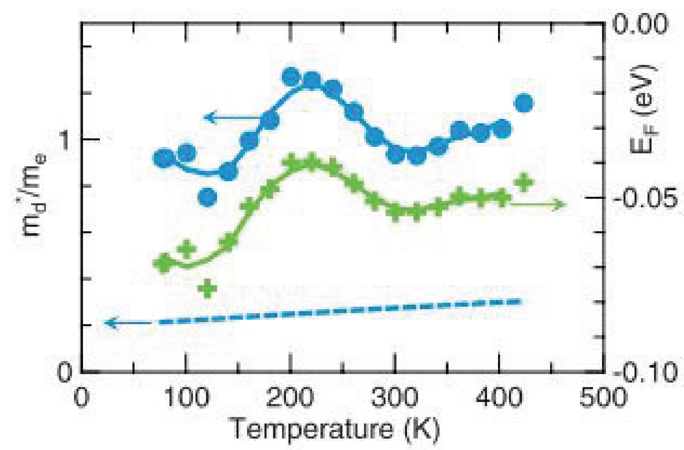


Figure 2.11: Effective mass and Fermi level comparison of Tl-PbTe and standard Na-PbTe. Adapted from [6].

of S^2n translated to an increase in ZT from 0.75 to 1.50 at 700 K.

Similar effective mass enhancement could be expected in other materials where impurity energy level is resonant with the host's conduction or valence band. One such system is the dilute Nitride, which consists of N-doped III-V semiconductor. Chapter 5 discusses this subject more extensively in lieu with our experimental results on the study of GaN_xAs_{1-x} thermoelectric properties.

Chapter 2, in part, is a reprint of the material as it appears in the journal articles:

- P. Pichanusakorn and P.R. Bandaru, "Minimum length scales for enhancement of the power factor in thermoelectric nanostructure", *Journal of Applied Physics*, vol. 107(7), 074304 (2010)
- P. Pichanusakorn and P.R. Bandaru, "The optimal Seebeck coefficient for obtaining the maximum power factor in themoelectrics", *Applied Physics Letters*, vol. 94, 223108 (2009)

The dissertation author was the primary investigator and author of these papers.

3 Measurement Methods

A major focus of this work is the accurate measurement the transport coefficient, ρ, n, μ, and S. Through the course of this work, different methodologies have been attempted. Starting with just a pair of hand-held multimeter probes used to measure resistance of a bulk material, a measurement system with the capability to measure ρ, n, μ, and S, all from a single sample, was developed from the ground up.

This chapter will first begin with the description of the sample preparation, which entails (1) the design of the testing device to be fabricated onto the material, and (2) the fabrication process of said device. The discussion will then proceed to (3) the external equipments used to take measurements of the device. Measurement procedures for (4-7) ρ, n, μ, and S are then described.

3.1 Measurement device layout

Initially, two separate paths were pursued for each types of material since the scale of the samples were very great: GaAs thin films were macroscopic, while the Si NWs could only be probed with the use of microscope and probe station. Extensive work was done to replicate the measurement system used in Ref. [26] for the thin film measurement. Eventually, design aspects from that work was folded into the system that was developed for the NWs, which is similar to design used in Ref. [54].

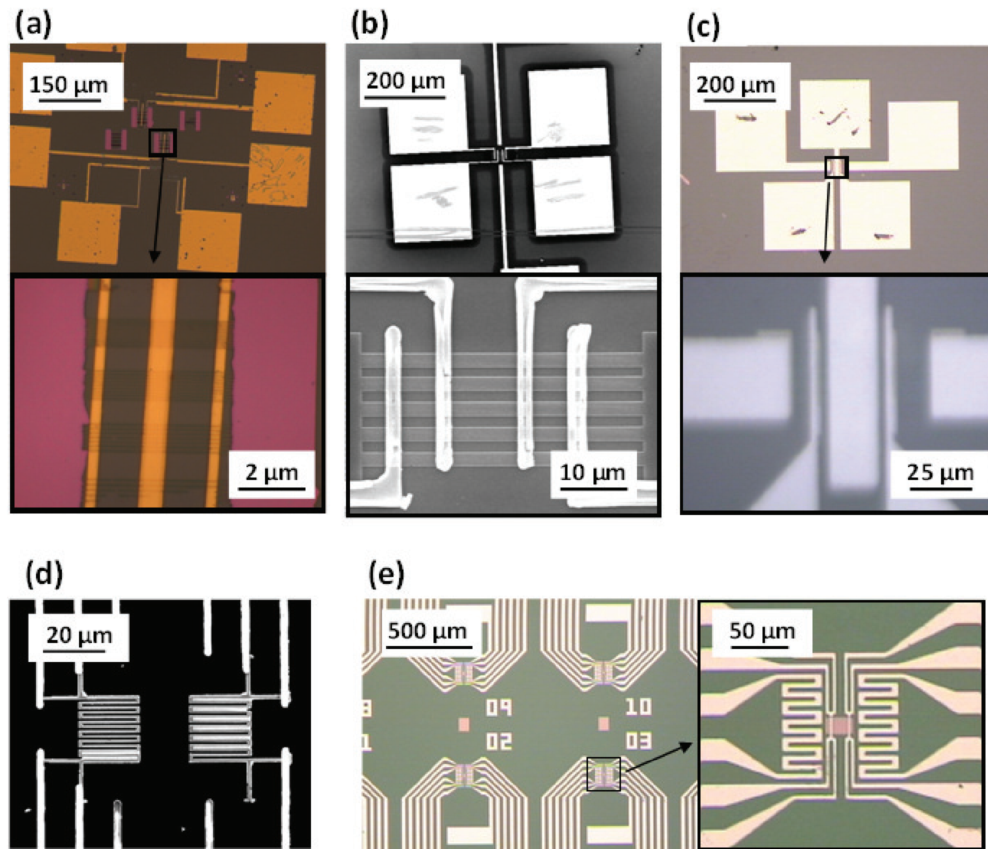


Figure 3.1: Various stages in the development of a measurement device.

In order to measure electrical properties of NWs, a measurement device must first be fabricated around the NWs. At the minimum, this device may consists of a pair of electrodes connecting the NWs to large probing pads, such as those shown in Fig. 3.1a and b. More capabilities were then added. Subfigure (c) shows the addition of a wide gate pad in between two narrow drain-source electrodes for the testing of MOSFET structure. Subfigure (d) shows an EBL-defined pair of serpentine joule heaters tested during early development of S measurement. An advance layout that integrates the heaters and allow for both resistance and S measurement then emerged in subfigure (e).

The final device layout is shown in Fig. 3.2. There are two type of device

layout: (a) van der Pauw (VDP) and (b) field-effect transistor (FET). The elements common to both design are

- Pair of 4-wires RTDs. They also function as drain and source in FET devices. Their terminals are labeled {D1, D2, D3, D4} and {S1, S2, S3, S4}, respectively.
- Pair of serpentine Joule heaters. Electrically isolated from the sample. The drain- and source-side heater terminals are labeled {HD1, HD2} and {HS1, HS2}, respectively.
- Gate contacts labeled {G1 or G2}. For the FET device, the gate is electrically insulated by gate oxide from the sample. For the VDP device, these contacts are electrically connected to the sample.
- Additional electrical contacts labeled {D0, S0} are used for 4-point resistance measurement of the FET channel.

All 16 terminals are connected to large probing/wire-bonding pads, thus forming a complete measurement device as shown in Fig. 3.2. Three VDP devices were designed to fit on a 5x5 mm^2 chip, while ten FET devices fit on a 7x7 mm^2 chip.

3.1.1 Device fabrication

The VDP and FET device is fabricated with photolithographic process shown in Fig. 3.3. The description for each steps are as follow

(*a*) Clean sample GaN_xAs_{1-x} or Si thin film by ultrasonication for 3 min each in trichloroethylene (TCE), acetone (ACE), and isopropanol (IPA) to remove organic contaminants.

(*b*) Spin-coat with negative-tone NR9 photoresist at 4,000 rpm for 40 s. Bake on a hot plate at 150 °C for 60 s. All baking in following steps are done on a hot plate.

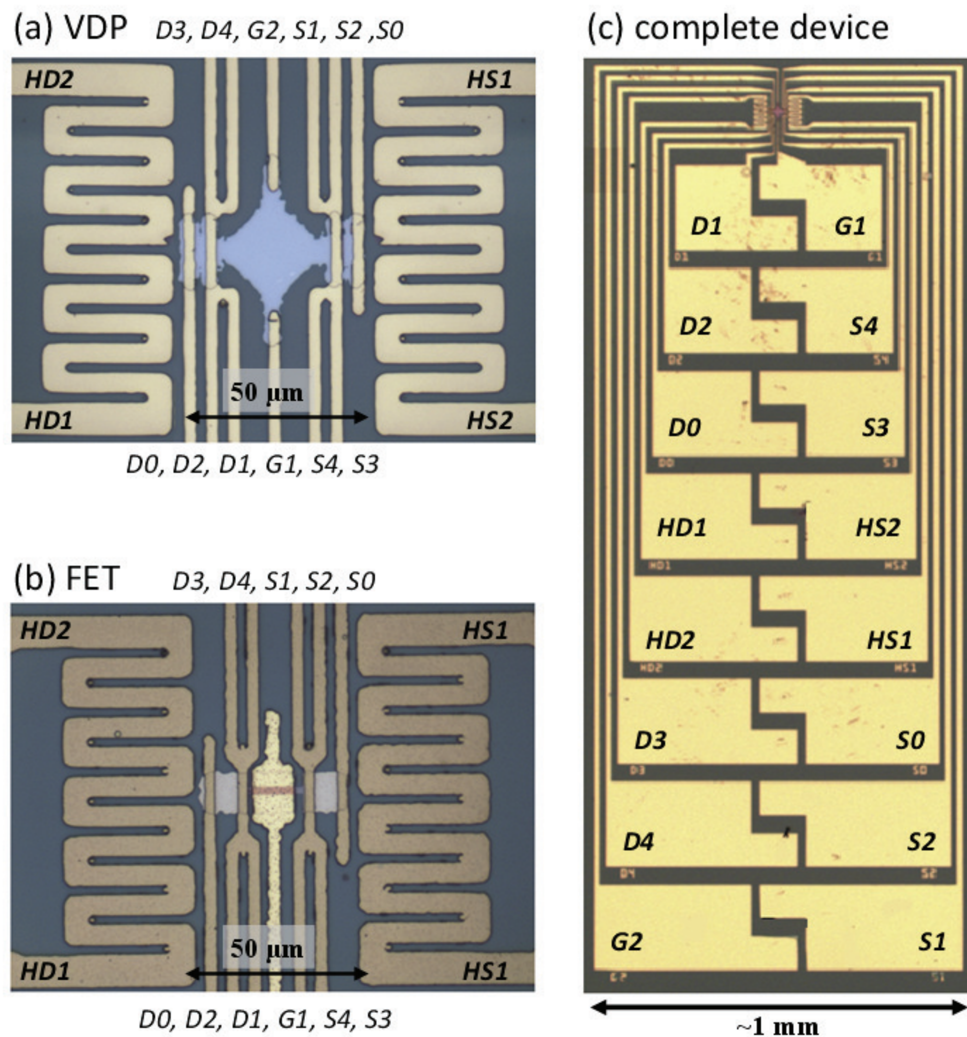


Figure 3.2: Measurement device layout for measurement of ρ, μ, n, and S using (a) van der Pauw and (b) field-effect transistor configurations. (c) shows the micro-circuit as part of a complete device featuring large gold pads for probes and wire-bonds.

(*c*) Align sample to device pattern in the MA6 mask aligner, and expose to UV light for 20 s.

(*d*) NR9 resist must be post-baked at 100 °C for 60 s before it is developed in RD6 solution for 30 s. Rinse in deionized water. The *unexposed* part of the resist is removed, leaving the resist with the shape of the device.

(*e*) Dry etch sample.

(*f*) Remove resist by ultrasonication in ACE, and then in IPA. The thin film is now etched into a cross/diamond/circle shape for VDP device, or a rectangular bar for FET device.

(*g*) Spin-coat with positive-tone S1818 resist at 4,000 rpm, for 40 s. Bake at 120 °C for 60 s. Be careful of temperature. Over bake by 5-10 °C can cause the resist to hardened, which will then fail to develop.

(*h*) Align sample to electrodes pattern, and expose for 20 s.

(*i*) Develop S1818 resist in MF-321 solution for 30-45 s. Swirl or stir the solution gently to make sure that resist is removed. Rinse in deionized water. Check with microscope that resist was properly removed. Trace of photoresist could be further removed by oxygen plasma descum in Tepla 100 etcher using power ¡150W; use of excessive power could overheat photoresist.

(*j*) Deposit metal by *e*-beam evaporator. If Ti is used as first adhesion layer, then make sure that pre-deposition vacuum is below 10^{-6}Torr as the adhesion will not be good if there are traces of moisture. In all cases, try to avoid overheating the resist by tightly affixing sample to holder, use low evaporator power, or minimize deposition time.

(*k*) Liftoff by ultrasonication in ACE at medium power. If adhesion is good, then ultrasonication will not remove the metal. 2-3 min ultrasonication should remove most of the resist. However, if resist is overheated, it may not liftoff easily. For overheated resist, the Microposit Remover 1165 can be use instead of ACE. The 1165 solution is a more potent solvent, though it does not always

work in the most severely overheated sample. After liftoff, clean sample in IPA and blow dry.

At this point, the sample may be annealed to ensure Ohmic contact if necessary. The fabrication of VDP device is completed at this point, and the device is ready for measurement. ρ and S of FET device could be measured at this point as well, but μ and n measurement will require additional fabrication of the gate. Fabrication of the gate is shown in Fig. 3.4 and is described as follow

(l) Deposit SiO_2 by PECVD, which will cover the entire sample.

(m) Spin-coat with S1818 (same condition as in step (g)). Align with pattern for probe pads and expose.

(n) Develop and remove the S1818 resist over the pads area. Then etch the oxide over the pads in BOE until the PECVD oxide is completely removed. BOE etch rate of oxide is approximately 2 nm/s. Rinse sample in flowing deionized water for 3-5 min to remove all trace of BOE. Blow dry sample.

(o) Align sample to gate pattern and expose over the same resist.

(p) Develop and remove resist over gate area. Deposit gate metal over the gate oxide, and the exposed metal pads.

(q) Liftoff.

FET device fabrication is fully completed at this point, and the device is ready for measurement.

3.2 External equipments

The device-under-test (DUT) is to be used in conjunction with the custom-built stage shown in Figure 3.5. At the bottom, there is a translation/rotation stage that is used in conjunction with an overhead optical microscope (not shown

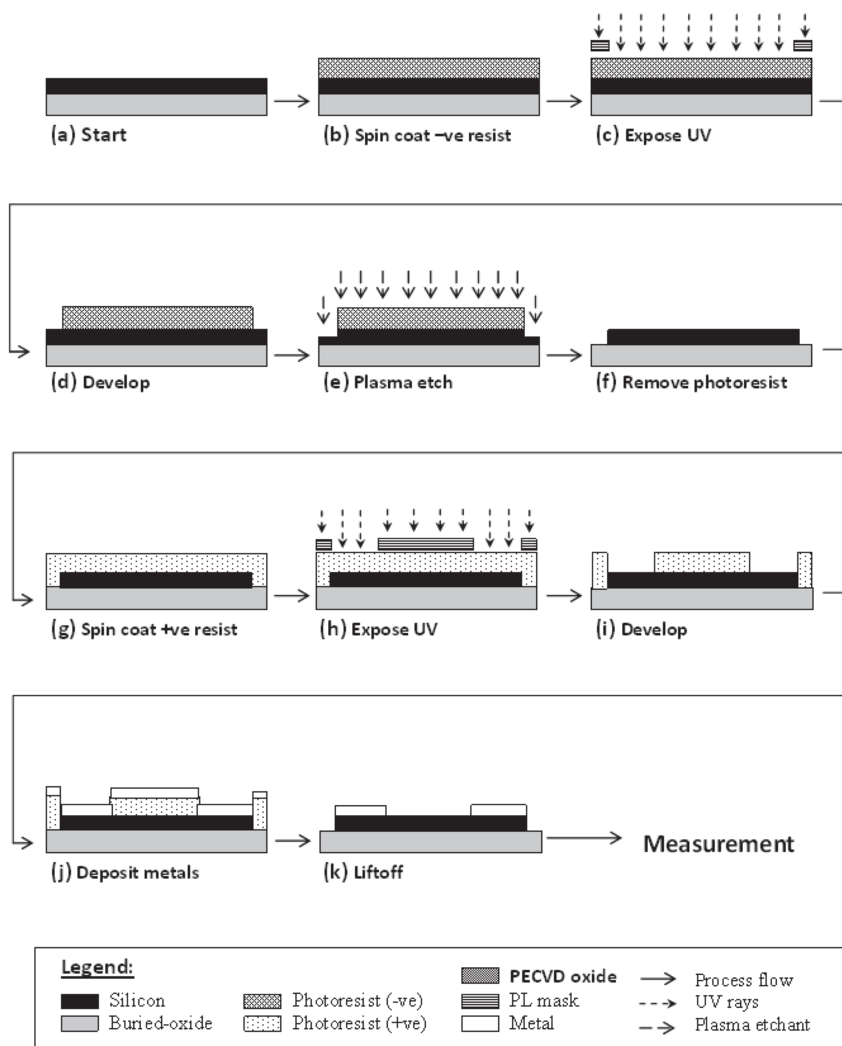


Figure 3.3: Photolithography process flow for VDP and FET device fabrication.

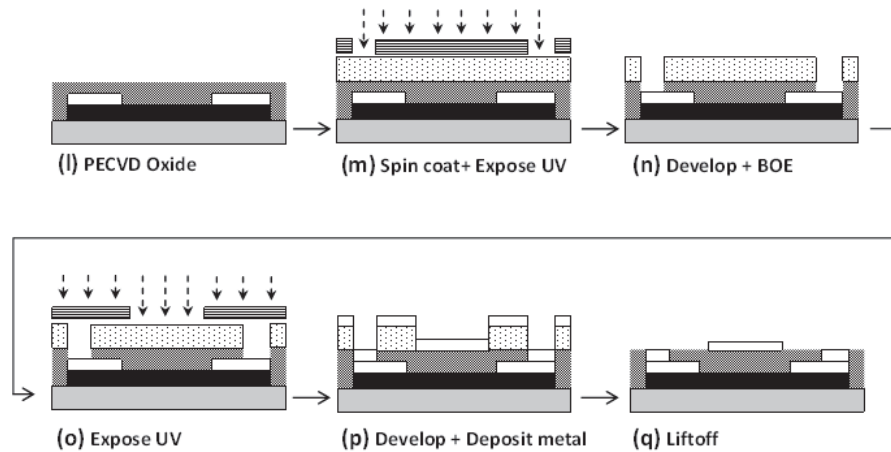


Figure 3.4: Continuing process flow for gate definition of FET device. Refer to Fig. 3.3 for legend.

in picture) for aligning the DUT to the probe. On top of the translation/rotation stages are a water-cooled block, and a thin custom-made solenoid, capable of 280G at 1A. The solenoid provides the magnetic field for Hall measurement, but also generates a lot of heat, which had to be removed by a cooling system. During measurement a sample is placed on top of the Peltier module, which is nestled in the opening at the core of the solenoid. The Peltier module is used in conjunction with a thermocouple for heating and cooling of the DUT from 10-50 °C during the calibration of the on-chip RTDs. Finally, connection between the DUT and electrical equipments is made through the multi-probe card (MCW-16 from GGB Industries, Inc.), which consists of 16 individual BeCu needles whose placement was designed to match the pads on the device.

The connection between the multi-probe card and the electrical equipments is shown by the system block diagram in Fig. 3.6. Through the probe card, the DUT is connected to the Keithley 7700 multiplexor (MUX) module, which is installed internally in the Keithley 2700 multimeter (MM). Through the MUX, the various terminals of the DUT can be connected to the MM itself, or other electrical

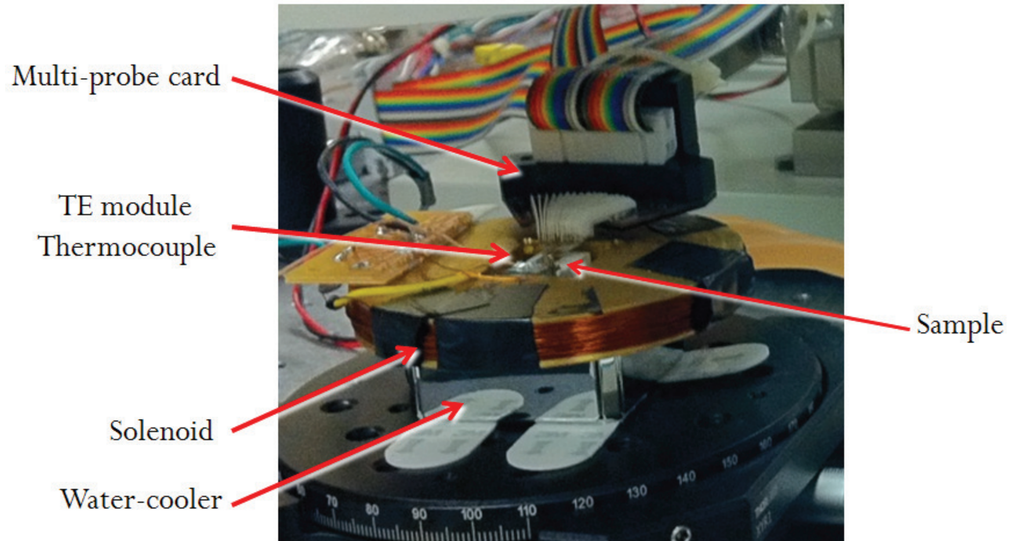


Figure 3.5: The multi-probe card and sample loading stage.

equipments such as the Keithley 2400 source-measure unit (SMU). ρ, n, μ, and S can all be measured with proper configuration of such electronic equipments. All equipments are controlled via GPIB bus by a computer running LabVIEW virtual instrument (VI) programs.

3.3 Measurement procedure

3.3.1 Van der Pauw and Hall method: ρ, n, and μ

The combination of van der Pauw and Hall techniques [55] (which will be refers together as VDP method from here on) is a simple technique for the measurement of ρ, n, and μ in a thin film sample with constant thickness. A major advantage of the VDP technique is that the sample can take any arbitrary shape as long as it is free of holes; only thickness is required for the calculation of ρ and n. A minimum of four electrical contacts are placed on the perimeter of a sample. Since the calculation assume an ideal point contact, a finite length of contact introduces error into the measurement [56, 57]. This error is reduced by making

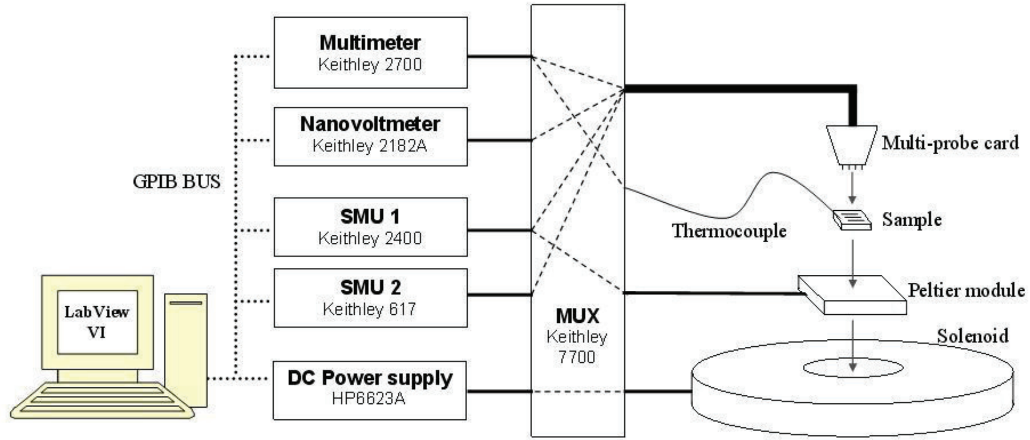


Figure 3.6: System block diagram

contact points as small as possible, or increasing the circumference of the sample, i.e. by the use of a cloverleaf shape.

In the VDP method, the sheet resistance (R_s) and Hall coefficient (R_H) are measured. The configuration required to measure R_s is shown in Fig. 3.7a. Current is passed between two *adjacent* contacts while voltage is measured from the other two contacts, giving the resistance of this configuration as

$$R_{12,43} = \frac{V_{43}}{I_{12}}$$

Similar resistance measurement is taken for all four possible current/voltage contact pairings ($R_{12,43}, R_{23,14}, R_{34,21}, R_{41,32}$), and in both forward and reverse polarity, giving a total of 8 resistance measurements. Sheet resistance is then calculated as

$$R_S = \left(\frac{\pi R_{avg}}{\ln 2}\right) \tag{3.1}$$

, where R_{avg} is the average resistance from the 8 measurements.

R_H is then measured with the configuration shown in Fig. 3.7b. While applying magnetic field perpendicular to the sample plane with the magnetic flux density of B, current is passed between two *opposing* contacts (from contact 1 to

3). The deflection of the current (i.e. $I = I_{13}$) by the applied magnetic field is then measured as the Hall voltage across the other two contacts (i.e. $V_H = V_{24}$). The Hall coefficient is then related to these experimental values, and the carrier concentration as

$$R_H = \frac{V_H d}{IB} = -\frac{r_H}{ne} \tag{3.2}$$

, where d is the sample thickness, and r_H is the Hall factor that accounts for the difference between mobility of electrons under the magnetic field (i.e. Hall mobility μ_H) and under electric field (i.e. the drift mobility μ). Hall measurement is repeated for the other configuration ($I = I_{24}, V_H = V_{13}$). For each configuration, measurement is taken for both current polarity, and field polarity. R_H is then reported as the average from all 8 measurements.

The electrical properties of the sample is now calculated as follow

$$\rho = R_s d \tag{3.3}$$

$$n = -\frac{r_H IB}{eV_H d} \tag{3.4}$$

$$\mu = \frac{1}{ne\rho} = \frac{V_H}{r_H IBR_s} \tag{3.5}$$

r_H ranges from 1 to 2, and depends on the scattering constant r [33]. For example, $r_H = 1.18$ for $r = -1/2$. Since r was not measured directly in this work, $r_H = 1$ was assumed. Note that in the final expression, μ is independent of the sample thickness d. Therefore, an error in the film thickness measurement will propagate to ρ and n, but cancel itself out for μ.

3.3.2 FET method: ρ, μ and n

Since Hall measurement could not be applied to NWs due to their shape, a FET configuration is used instead. The most common form, and the one used in this work, of FET is the metal-oxide-semiconductor field-effect transistor (MOSFET), therefore the acronym FET shall refers to MOSFET from here on. In this

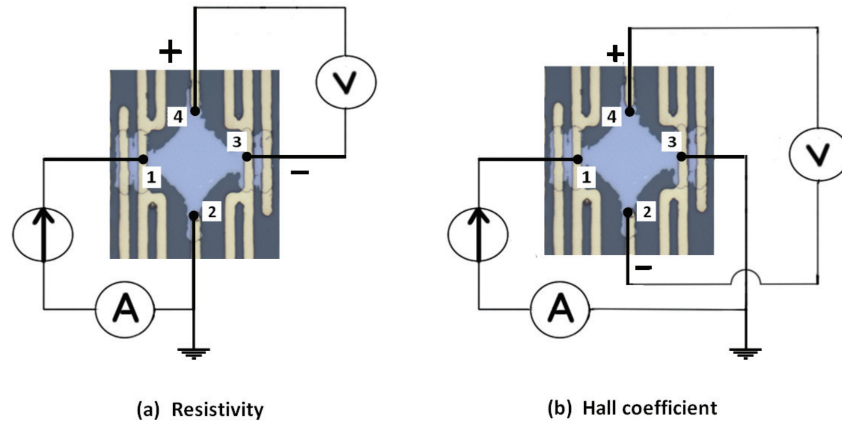


Figure 3.7: Van der Pauw measurement configuration.

section, a brief review of FET operation is discussed first, and then the final equations that take into account the particulars of our device will be given.

MOSFET equations

A typical FET consists of four terminals: drain, source, gate, and body. The gate is electrically insulated by the gate oxide from the channel, which is the region directly underneath the gate and between source and drain. The drain and source regions may be heavily doped n- or p-type, while the channel and substrate itself is oppositely doped p- or n-, respectively. In this manner, the drain-channel-source forms a current-blocking back-to-back diode. However, the current can be made to flow by inverting the channel through application of gate voltage (V_g).

There are 3 regions of operation for FET: the sub-threshold (OFF), saturation, and linear regions. In the saturation and linear region, the drain-source current for a long channel FET [58] is given by

$$I_d = \mu C_{ox} \frac{W}{L_{ch}} \left(V_g - V_t - \frac{V_d}{2} \right) V_d \tag{3.6}$$

, where W and L_{ch} is the channel width and length, V_g and V_d is the gate and drain voltages with respect to source, V_t is the threshold voltage, and C_{ox} is the

gate oxide capacitance per unit area. The capacitance may be calculated as

$$C_{ox} = \frac{\epsilon_{ox}\epsilon_0}{d_{ox}} \tag{3.7}$$

, where ϵ_{ox} is the relative permittivity of the oxide, $\epsilon_0 = 8.85 \times 10^{-12}$ F/m is the vacuum permittivity, and d_{ox} is the oxide thickness.

The MOSFET mode of operation is then understood through the relationship between I_d, and the two variable voltages, V_g and V_d. When $V_g < V_t$, $I_d \propto -V_d^2$, which is physically impossible, so equation 3.6 is invalid as the MOSFET is in the sub-threshold mode. The MOSFET is in saturation region when $V_d = V_{dsat} = V_g - V_t$, where

$$I_d = I_{dsat} = \mu C_{ox} \frac{W}{L_{ch}} \frac{(V_g - V_t)^2}{2} \tag{3.8}$$

, which is a constant depending only on V_g. If $V_d \ll V_g$, then the MOSFET is in the linear region as

$$I_d \approx \mu C_{ox} \frac{W}{L_{ch}} (V_g - V_t) V_d \tag{3.9}$$

, i.e. the MOSFET is essentially a resistor whose resistance vary proportionally to V_g.

The relationship between I_d and n in the linear region can be made more apparent by considering that the resistivity of the channel is

$$\rho = \frac{Wd}{L_{ch}} R = \frac{Wd}{L_{ch}} \left(\frac{V_d}{I_d} \right) = \frac{1}{ne\mu} \tag{3.10}$$

, where d is the thickness of the channel. Therefore, it is shown that

$$\begin{aligned} \frac{1}{ne\mu} &= \frac{d}{\mu C_{ox}(V_g - V_t)} \\ en &= \frac{C_{ox}}{d}(V_g - V_t) \end{aligned} \tag{3.11}$$

, i.e. the charge carrier density (en) is equal to capacitance per unit volume times the overdrive voltage ($V_g - V_t$).

Electrical property of the channel can be determined by the measurement of $I_d(V_g)$ curve at small V_d, where the MOSFET is in the linear region. Figure 3.8a shows the $I_d(V_g)$ curve of n-channel Si MOSFET device fabricated in this work. V_d is held at a constant 50 mV, while V_g is swept from 0 to +5 V, then to -5 V, and back to 0 V. Gate leakage current (I_g) (from gate to source) is also monitored, and is less than 200 pA. Larger $|V_g|$ is avoided since gate oxide have broken down on samples when $|V_g| > 10$ V. Gate breakdown is detected by a large I_g. The $I_d - V_g$ curve in the figure is consistent with n-type channel since a positive V_g would attracts more electrons to the channel, thus increasing I_d.

ρ, n, and μ is then extracted from the slope and y-intercept of the $I_d(V_g)$ curve. According to equation 3.9, the slope, also known as the transconductance, and the intercept are equal to

$$\frac{\Delta I_d}{\Delta V_g} = \mu C_{ox} \frac{W}{L_{ch}} V_d \tag{3.12}$$

$$I_d(0) = \mu C_{ox} \frac{W}{L_{ch}} (-V_t) V_d \tag{3.13}$$

We then derive that

$$V_t = \frac{-I_d(0)}{\frac{\Delta I_d}{\Delta V_g}} \tag{3.14}$$

$$\rho = \frac{I_d(0)}{V_d} \frac{W d}{L_{ch}} \tag{3.15}$$

$$\mu = \frac{\Delta I_d}{\Delta V_g} \left(\frac{L_{ch}}{W C_{ox} V_d} \right) \tag{3.16}$$

$$n = \frac{|V_t| C_{ox}}{e\, d} \tag{3.17}$$

Note that ρ is defined by the resistance of the channel at zero gate voltage.

For the device shown in Fig. 3.8a, whose geometry is listed in Table 3.1, it is calculated that

$$\mu = 0.13 \left[\frac{\mu\text{A}}{\text{V}} \right] \frac{11\,[\mu\text{m}] \times 40\,[\text{nm}]}{10\,[\mu\text{m}] \times 3.90 \times 8.85 \times 10^{-12}\,[\text{F/m}] \times 0.05\,[\text{V}]} = 51\,\text{cm}^2/\text{Vs}$$

, where L_{ch} is equal to the dimension of the gate defined on the PL mask (L_g), and d is equal to the thickness of Si top layer (d_{Si}). Corresponding values of ρ, n, and

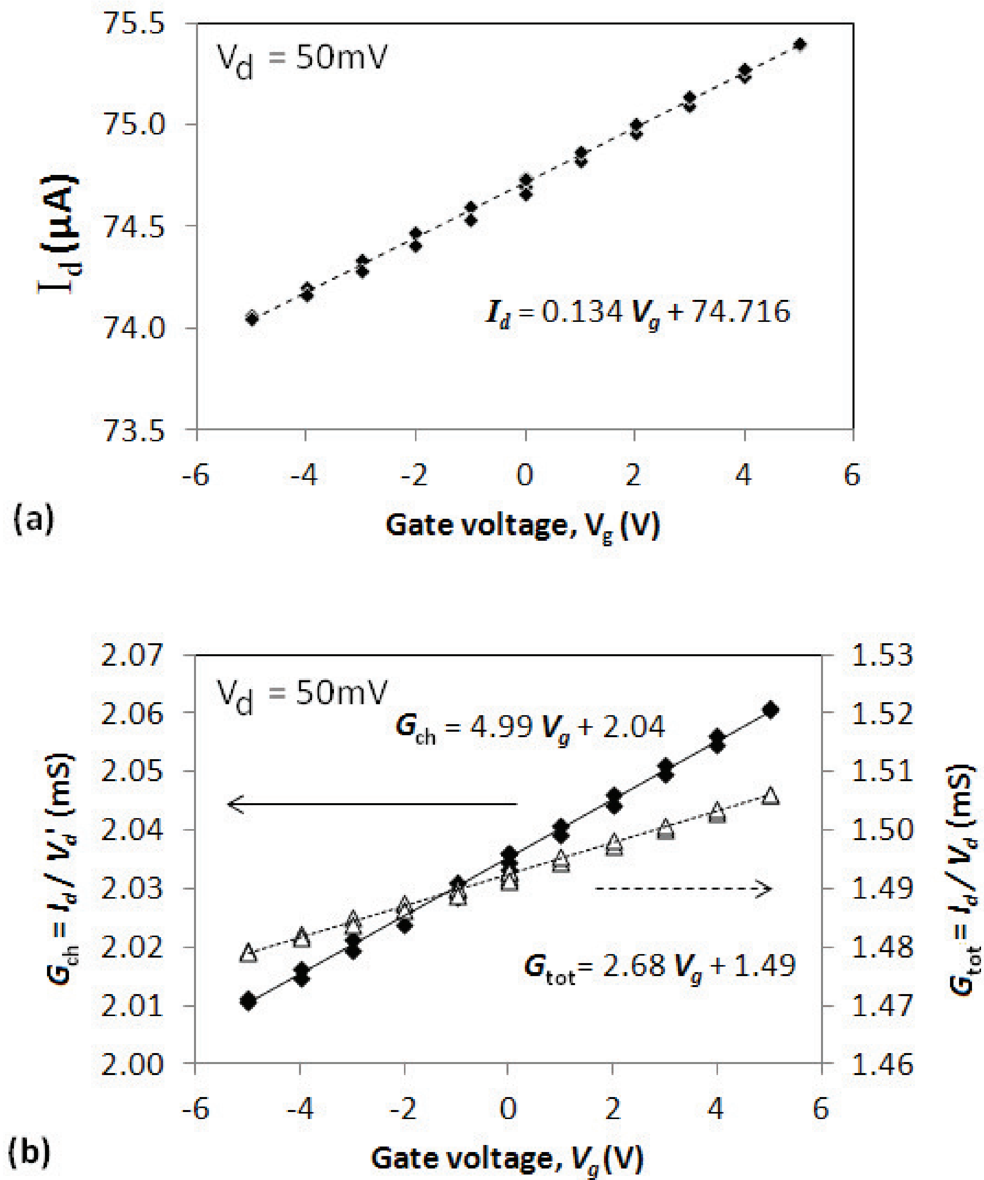


Figure 3.8: (a) Drain current versus gate voltage curve of n-channel bulk Si MOSFET taken at 50 mV drain voltage. (b) Equivalent total conductance (G_{tot}) and channel conductance (G_{ch}) versus gate voltage of the device.

Table 3.1: MOSFET device's geometries and $I_d - V_g$ calculation results.

L_g (μm)	L_{ds} (μm)	W (μm)	d_{Si} (nm)	d_{ox} (nm)
11	15	10	40	60

Case	ρ(mΩ.cm)	n(cm^{-3})	μ(cm^2/Vs)	V_t(V)	ϵ_{ox}
(a) as-measured	1.79	6.82×10^{19}	51.26	-557	3.9
(b) R_{ds}-corrected	1.79	3.67×10^{19}	95.32	-408	3.9
(c) ϵ_{ox}-corrected	1.79	6.86×10^{19}	50.93	-408	7.3

V_t are then listed under case (a) of Table 3.1. However, the calculation thus far does not account for the mismatch between L_g and the drain-source length (L_{ds}).

Drain-Source resistance correction

Figure 3.9a shows the configuration of our FET device during measurement. In this picture, Si NWs channel can be seen as the discolored bar across the gate pad. Notice that there are gaps on either side of the gate pad to drain and source contacts. This gap was intentionally designed to prevent shorting of the drain/source electrodes to the gate pad, though it is redundant with the current process, where oxide coverage is extended over drain and source electrodes as well. An equivalent side-view of the device is shown in subfigure (b), where the gap between gate-drain/source is again emphasized.

The separation between gate-drain/source will introduces error. Figure 3.9c shows that the section of the active layer not covered by the gate is represented by fixed drain and source resistances R_d and R_s, respectively. The resistance of the deposited electrodes and cabling, and the contact resistance, was ignored since 4-wires measurement was used to measure I_d and V_d. According to the model, the total resistance of the device is

$$R_{tot} = \frac{V_d}{I_d} = R_{ds} + R_{ch}(V_g) \tag{3.18}$$

, where $R_{ds} = R_d + R_s$ is the fixed drain-source resistance, and $R_{ch} = \frac{V_d'}{I_d}$ is the variable, gate-dependent channel resistance. Assuming that ρ is the same

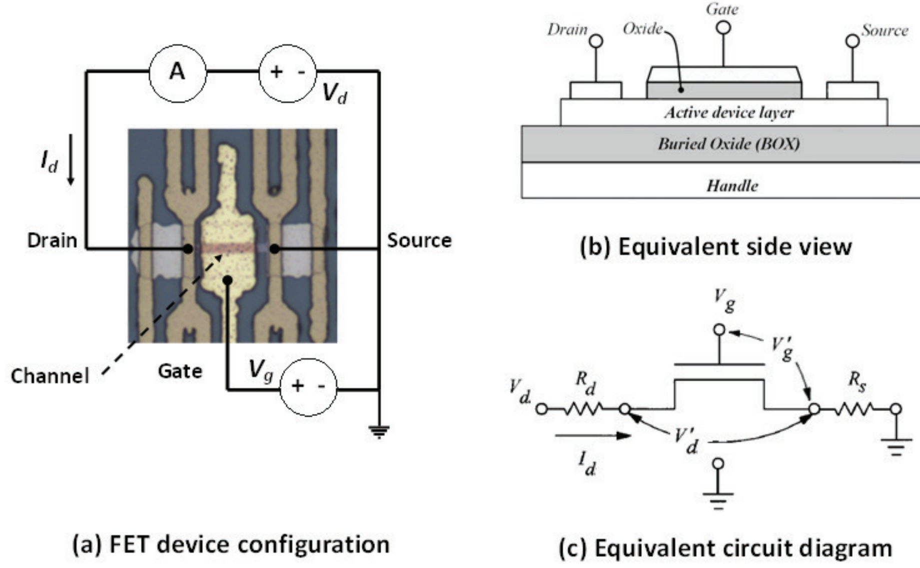


Figure 3.9: SOI-MOSFET structure and equivalent circuit diagram.

throughout the active layer, then the drain-source resistance is given by

$$R_{ds} = R_0 - R_{ch} = \frac{I_d(0)}{V_d}\left(\frac{L_{ds} - L_{ch}}{L_{ds}}\right) \tag{3.19}$$

, where $\frac{I_d(0)}{V_d} = R_0$ is the total device resistance at zero gate voltage.

Meanwhile, the current in the channel is still given by the general equation 3.6 for both linear and saturation region, but with the value of V'_d and V'_g in places of V_d and V_g. Accounting for the voltage drop over R_d and R_s, it is shown that

$$V'_d = V_d - I_d(R_d + R_s) \tag{3.20}$$

$$V'_g = V_g - I_d R_s \tag{3.21}$$

The threshold voltage is also changed to

$$V'_t = V_t + (m_b - 1)I_d R_d \ \approx V_t \tag{3.22}$$

, where m_b is the body effect coefficient. $m_b = 1$ may be assumed since the active layer is well insulated from the handle by the thick 1 µm BOX layer. If $R_d = R_s$,

then

$$V'_g - V'_t - \frac{V'_d}{2} = V_g - I_d R_s - V_t - \frac{V_d}{2} - \frac{I_d(R_d + R_s)}{2}$$
$$= V_g - V_t - \frac{V_d}{2} \tag{3.23}$$

Note that in the linear regime $V'_g = V_g - I_d R_s \approx V_g$ because $V_g \gg V_d > I_d R_s$. Substituting equation 3.23 and V'_d into equation 3.6 then yields

$$I_d = \mu C_{ox} \frac{W}{L_{ch}} \left(V_g - V_t - \frac{V_d}{2} \right) V'_d \tag{3.24}$$

$$I_d = \mu C_{ox} \frac{W}{L_{ch}} \left(V_g - V_t \right) V'_d \quad , \ V_g \gg V_d \tag{3.25}$$

However, it becomes quite difficult to express I_d as explicit function of V_g if the last V'_d term is substituted with equation 3.20.

Nevertheless, ρ, μ and n can still be determined in similar manner by considering the channel conductance, which in the linear region is defined as

$$G_{ch} = \frac{I_d}{V'_d} = \frac{I_d}{V_d - I_d R_{ds}} \tag{3.26}$$

Given the $I_d(V_g)$ curve in Fig. 3.8a, R_{ds} is calculated using equation 3.19, and G_{ch} is calculated for every pair of (I_d, V_g) data according to equation 3.26. For example, $R_{ds} = \frac{0.05\text{V}}{74\times10^{-6}\text{A}} \left(\frac{15-11}{15}\right) = 3,569\,\Omega$ for the device shown. The calculated G_{ch} is then shown in Fig. 3.8b. The total conductance ($G_{tot} = \frac{I_d}{V_d}$) is also plotted to represent the initial case where R_{ds} was not taken into account. The exclusion of R_{ds} resulted in both larger magnitude and slope for G_{ch} compared to G_{tot}.

The electrical property parameters may then be determined from the slope and intercept of the $G_{ch}(V_g)$ curve, which are given by

$$\frac{\Delta G_{ch}}{\Delta V_g} = \mu C_{ox} \frac{W}{L_{ch}} \tag{3.27}$$

$$G_{ch}(0) = \mu C_{ox} \frac{W}{L_{ch}} (-V_t) \tag{3.28}$$

Subsequently, the electrical properties of the channel is written as

$$V_t = \frac{-G_{ch}(0)}{\frac{\Delta G_{ch}}{\Delta V_g}} \tag{3.29}$$

$$\rho = \frac{1}{G_{ch}(0)} \frac{Wd}{L_{ch}} \tag{3.30}$$

$$\mu = \frac{\Delta G_{ch}}{\Delta V_g} \left(\frac{L_{ch}}{WC_{ox}} \right) \tag{3.31}$$

$$n = \frac{1}{\rho e \mu} = \frac{|V_t| C_{ox}}{e\,d} \tag{3.32}$$

These equations are of course very similar to equations 3.14-3.17, except for the substitution of I_d with G_{ch}, and the absence of V_d. Note that while n has the same expression in both equation 3.17 and 3.32, V_t as determined from $I_d(V_g)$ and $G_{ch}(V_g)$ will not be the same.

Accordingly, if correction from R_{ds} is applied, then the true mobility of the channel is

$$\mu = 4.99 \left[\frac{\mu\text{A}}{\text{V}}\right] \frac{11\,[\mu\text{m}] \times 40\,[\text{nm}]}{10\,[\mu\text{m}] \times 3.90 \times 8.85 \times 10^{-12}\,[\text{F/m}]} = 95\,\text{cm}^2/\text{Vs}$$

In comparison, the initial value may be obtained from $G_{tot}(V_g)$ as

$$\mu = 2.68 \left[\frac{\mu\text{A}}{\text{V}}\right] \frac{11\,[\mu\text{m}] \times 40\,[\text{nm}]}{10\,[\mu\text{m}] \times 3.90 \times 8.85 \times 10^{-12}\,[\text{F/m}]} = 51\,\text{cm}^2/\text{Vs}$$

Corresponding values for ρ, n, and V_t when R_{ds} is corrected for is listed as case (b) in Table 3.1. Without correction for R_{ds}, as in case (a), n and μ maybe over- and under-estimated by almost a factor of 2, given this particular geometry.

In summary, the goal of the FET method is to measure $\frac{\Delta G_{ch}}{\Delta V_g}$ and $G_{ch}(0)$, which are used to calculate V_t, ρ, μ, and n. Whereas μ is the least uncertain term (due to independence from d) in the VDP, n is perhaps the least uncertain term in the FET method since it has the least dependence on geometry, requiring only d.

C_{ox} **measurement**

Oxide capacitance measurement and analysis were performed by graduate student Hasan Faraby in the Materials Science and Engineering program of the Mechanical and Aerospace Engineering at University of California, San Diego.

C_{ox} is an important parameter required in the calculation of μ. Therefore, attempt were made to measure C_{ox} directly by treating the MOSFET device as a MOS capacitor. The drain and source pads were shorted together by wire bonding, and then connected to ground with a probe. The gate pad was then connected by a probe to the output of the Agilent 4284A Precision LCR Meter. The capacitance between gate and drain-source (C) is then measured at with gate voltage is that is composed of +2V DC and +2V AC (200 kHz) components. The parasitic capacitance in the cables and probes were compensated through calibration procedure. Figure 3.10a shows that without parasitic due to cables and probes, the measured capacitance increases from 2.4 pF at $L_g = 3\,\mu\text{m}$ to 2.7 pF at $L_g = 14\,\mu\text{m}$, which is expected since total capacitance should increase with gate area. $W = 10\,\mu\text{m}$ for all devices.

However, much of this capacitance is due to parasitic capacitance on the device itself. A test device was prepared at the same time with the same device layout and gate oxide as the actual sample, but without the Si channel, which was completely removed by dry etching prior to device fabrication. The measured capacitance of test device without the Si channel shows a near constant value of 2.2 pF regardless of gate length, which may be due to capacitance between the pads themselves. Subtracting the pads parasitic, the true capacitance between gate and drain-source is shown in Fig. 3.10b.

A linear fit of $C - L_g$ then shows that there is still an offset of 15.4 fF, which may be due to parasitic between the edges of the gate pad to the drain and source electrode. Subtracting this offset, the oxide capacitance per unit area is estimated

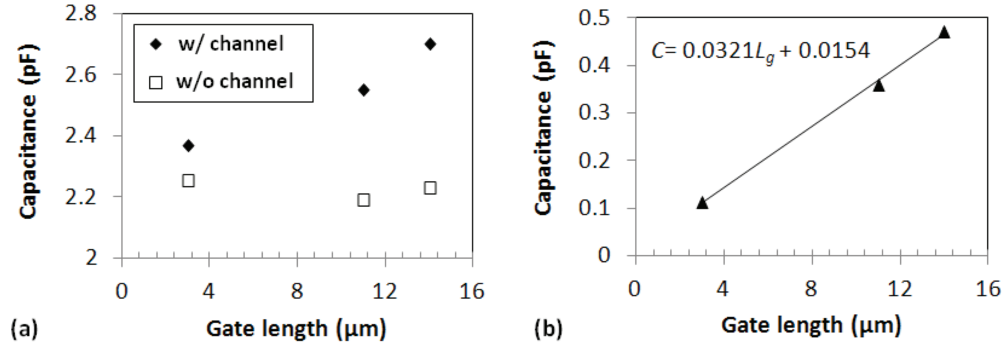


Figure 3.10: (a) Measured capacitance of device with and without channel. (b) Oxide capacitance.

to be

$$C_{ox} = \frac{0.0321 L_g}{W L_g} \frac{\text{pF}}{\mu\text{m}^2} = \frac{0.0321}{10} \frac{\text{pF}}{\mu\text{m}^2} = 0.00321 \frac{\text{pF}}{\mu\text{m}^2}$$

Given that $d_{ox} = 20\,nm$ for the devices tested here, it is estimated that

$$\epsilon_{ox} = \frac{C_{ox} d_{ox}}{\epsilon_0} = \frac{0.00321 \times 20}{8.85 \times 10^{-12}} \frac{\text{F} \cdot \text{nm}}{\text{m}^2 \cdot \text{F/m}} = 7.3 \tag{3.33}$$

, which almost twice the conventional value of $\epsilon_{ox} = 3.90$ for thermally grown oxide [58]. PECVD oxide may exhibits larger ϵ_{ox} if the oxide is non-stoichiometric and contain higher Si concentration, i.e if the oxide is SiO_x, where $x < 2$. High concentration of OH can also increase ϵ_{ox} [59]. Larger ϵ_{ox} up to 10 have been reported for PECVD oxide [60].

In row (c) of Table 3.1, the value of $\epsilon_{ox} = 7.3$ was applied to the calculation. n is therefore increased while μ is decreased. The values in case (a) and (c) are similar, but this is a coincidence. Note that V_t has not change since V_t is obtained directly from measurement.

Lastly, since the channel of our device is always in accumulation, the measured capacitance (sans parasitic capacitance) comes from the gate oxide only. Figure 3.11 shows the standard CV curve for a n-type MOS capacitor, showing a

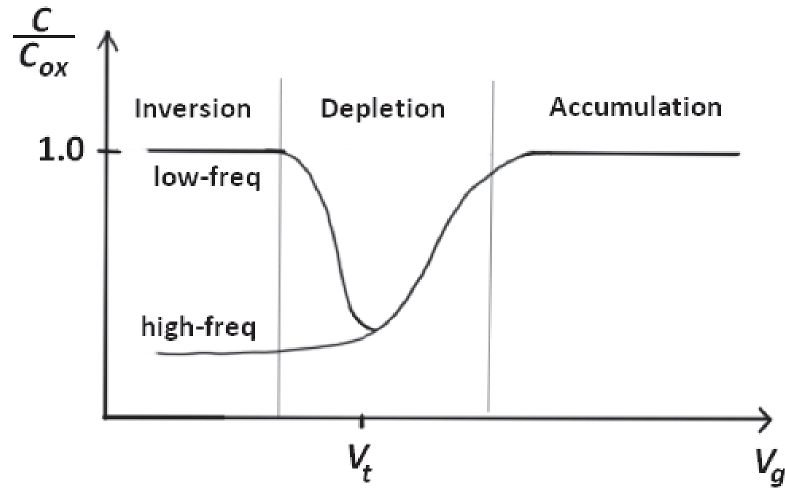


Figure 3.11: CV curve of n-type MOS

change in C/C_{ox}, where C is the total measured capacitance per unit area that is a series sum

$$\frac{1}{C} = \frac{1}{C_{ox}} + \frac{1}{C_{Si}}$$

, where C_{Si} is the capacitance of the Si channel per unit area. When the channel is in accumulation ($V_g \gg V_t$), C_{Si} is large due to heavy concentration of majority carrier, and $C = C_{ox}$. In depletion, there are no carrier, so C_{si} is small. In inversion, C_{Si} is large again, but since minority carriers are generated by thermal excitation, they cannot respond to hiqh frequency. From the $G_{ch}(V_g)$ curve in Fig. 3.8a, it is shown that $V_g = V_t = -408\,\text{V}$ would be needed to deplete the channel, which is far beyond the practical range for V_g.

3.3.3 RTD Calibration

In addition to the role of drain and source contact, the short strips of metal electrode between {D1,D2,D3,D4} and {S1,S2,S3,S4} terminals also function as resistance temperature detectors (RTDs). These on-chip RTDs are calibrated by heating or cooling the DUT using the Peltier module, while the temperature of the

Table 3.2: Thermal conductivity of various thermal grease/paste (according to manufacturer specification)

Type of thermal grease/paste	κ (W/mK)
Apiezon® "N" cryogenic grease	0.2
EPO-TEK® H70E epoxy	0.9
Omegatherm® 201 paste	2.3
SPI® Silver Paint	4.2
Arctic Silver® "Silver 5" compound	9.0
Liquid Ga-In eutectic alloy	40

chip is recorded against the resistance of the RTDs. The temperature is measured by a TC that is attached to a small area on the chip with a drop of SPI® silver paint. The mechanical bond of dried silver paint is weak but sufficient to hold down a fine-gauge TC wire, and can support the weight of a small (7x7 cm^2) chip. The silver paint is used because it has relatively high κ (see Table 3.2), and it is easily dissolved by acetone. It was assumed that the chip temperature is homogeneous, and T as measured by the TC is representative of the temperature of the RTDs. Given the high κ of GaAs ($55\,\frac{\mathrm{W}}{\mathrm{mK}}$) and Si ($149\,\frac{\mathrm{W}}{\mathrm{mK}}$) [37] substrates, and small chip size, such an assumption could hold. The resistance of the RTDs was measured with the K2700 using 4-wires resistance function with thermoelectric offset compensation ("OCOMP") function enabled.

Figure 3.12 shows the $R(T)$ profile for a pair of Ti/Ni/Au RTDs. For metals, $R(T)$ is expected to be a linear function, i.e.

$$R(T) = R_0(1 + \alpha(T - T_0))$$

, where α is the linear temperature coefficient of resistance (TCR), T_0 is an arbitrary reference temperature, and $R_0 = R(T_0)$ is the resistance at the reference temperature. The TCR is then equal to the slope of $R(T)$ divided by the resistance at the reference temperature, i.e.

$$\alpha = \frac{1}{R_0}\frac{R - R_0}{T - T_0} = \frac{1}{R_0}\frac{\Delta R}{\Delta T} \tag{3.34}$$

Therefore, if $T_0 = 0\,^\circ\mathrm{C}$, then $R_0 = 1.53\Omega$ and $\alpha = \frac{0.0042}{1.53} = 2.74\frac{10^{-3}}{^\circ\mathrm{C}}$ for the drain-

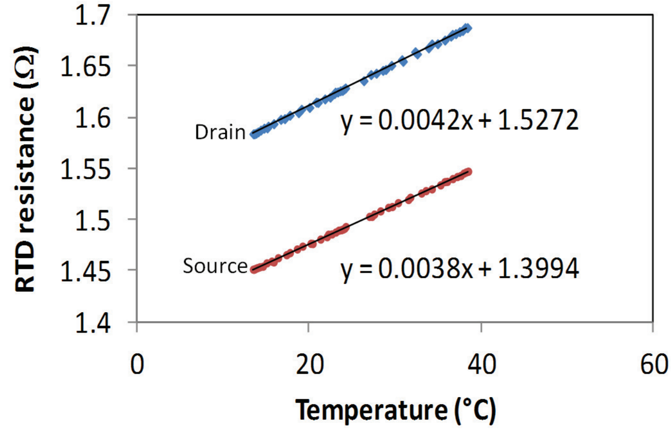


Figure 3.12: Plot of RTD resistance versus temperature profile.

Table 3.3: Experimental and literature [25] TCR values

Material	TCR ($\frac{10^{-3}}{°C}$) @ 20 °C	ρ (Ωm)
	Literature	
Al	3.9	2.8×10^{-8}
Au	3.4	2.4×10^{-8}
Ni	5.9	7.0×10^{-8}
	Experiment	
Al (p-Si)	3.12	3.3×10^{-8}
Ti/Ni/Au (5/180/30 nm) (n-Si)	2.44	3.4×10^{-7}
Ti/Ni/Au (5/100/150 nm) (n-Si)	1.86	1.2×10^{-7}

side RTD. Likewise, $R_0 = 1.40\Omega$ and $\alpha = 2.74\frac{10^{-3}}{°C}$ for the source-side RTD.

Table 3.3 compares TCR and ρ of a few metals that were used in the course of this work. Al was used for p-type Si, and Ti/Ni/Au for n-type Si. TCR and ρ were calculated at $T_0 = 20\,°C$ and compared with literature value [25]. The lower TCR and higher ρ of the RTDs is likely due to defects such as void or other impurity introduced into the metal during *e*-beam evaporation. As these source of resistance are temperature independent, they both reduce TCR and increase ρ.

As a final note, the TC should be placed directly on the chip, and attached

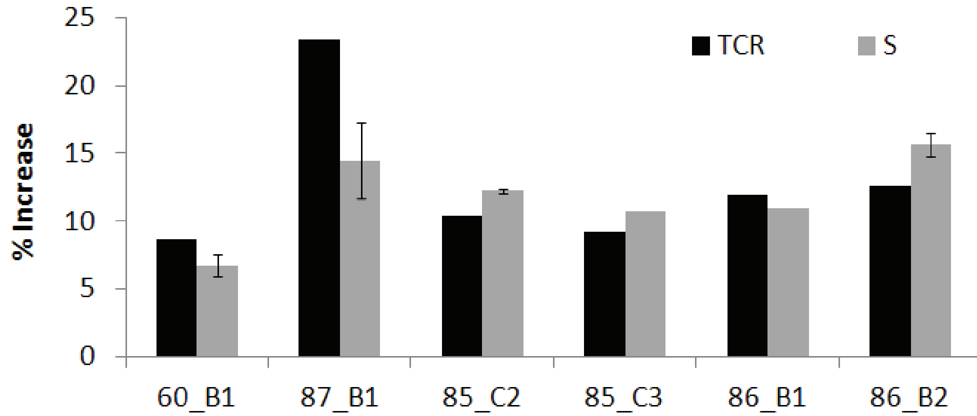


Figure 3.13: Percentage increase in TCR and $|S|$ as a result of placing thermocouple directly on the sample (relative to placing thermocouple on a nearby chip). Results of 6 samples are shown.

with thermal paste such as silver paint. Initially, the TC was affixed to the Peltier module such that it would not have to be attached and removed from the DUT. However, when TC is placed directly on the DUT, TCR increased by an average of $13 \pm 5\%$, as shown by the column chart in Fig. 3.13. Consequently, the measured $|S|$ also increased by similar amount of $12 \pm 3\%$. Therefore, it is very important that TCR be accurately measured by placing TC directly onto the sample.

3.3.4 Seebeck measurement

To measure S, one introduces and measure a temperature difference (ΔT) across a sample, and then measure the Seebeck voltage (V). Measuring temperature over a small area is the most challenging aspect of Seebeck measurement, which is overcome by the use of micro-fabricated on-chip RTDs.

Figure 3.14 shows the sequence of Seebeck measurement, which follows these steps:

1. Power the drain-side heater at a fixed voltage with the SMU to generates a

temperature difference

2. Wait 1-2 min for the temperature difference to reach steady state.

3. Measure resistances of the drain and source RTDs, and the voltage between drain and source (V_{ds}) in sequence. A minimum of 30 repeat measurements are usually required to average out the noise in RTD measurement. T_d and T_s is calculated using TCR and R_0 values measured previously.

4. Unpower drain-side heater until sample cool back down to room temperature.

5. Switch to and power the source-side heater.

Note that the average temperature, T_{AVE}, only includes temperature reading at the drain and source during the time when heater power is ON.

The thermoelectric voltage across drain and source, $V(= V_{ds})$, is then plotted versus the $\Delta T = T_d - T_s$, as shown in Fig. 3.15. Since $S = -\frac{V}{\Delta T}$, the negative slope of this plot is the measured Seebeck coefficient. The accuracy of S measurement is limited by the uncertainty in ΔT, as there is negligible uncertainty for V. The fact that V does not respond to the spread in ΔT suggests that the true temperature difference across the sample is actually stable, and suggests that the observed spread in ΔT is artifacts.

The uncertainty in ΔT is explained by the uncertainty in R measurement of the RTDs. By monitoring R of a RTD over a period of time, it was observed that $\delta R \approx 0.5\,\mathrm{m\Omega}$, which is close to the minimum of resolution of $0.1\,\mathrm{m\Omega}$ specified in the equipment manual. Given the $R(T)$ relationship in equation 3.34, i.e.

$$T(R) = \frac{R}{R_0\alpha} - \frac{1}{\alpha} + T_0$$

, the propagation of error indicates that

$$\delta T = |T'(R)|\delta R = \frac{\delta R}{R_0\alpha}$$

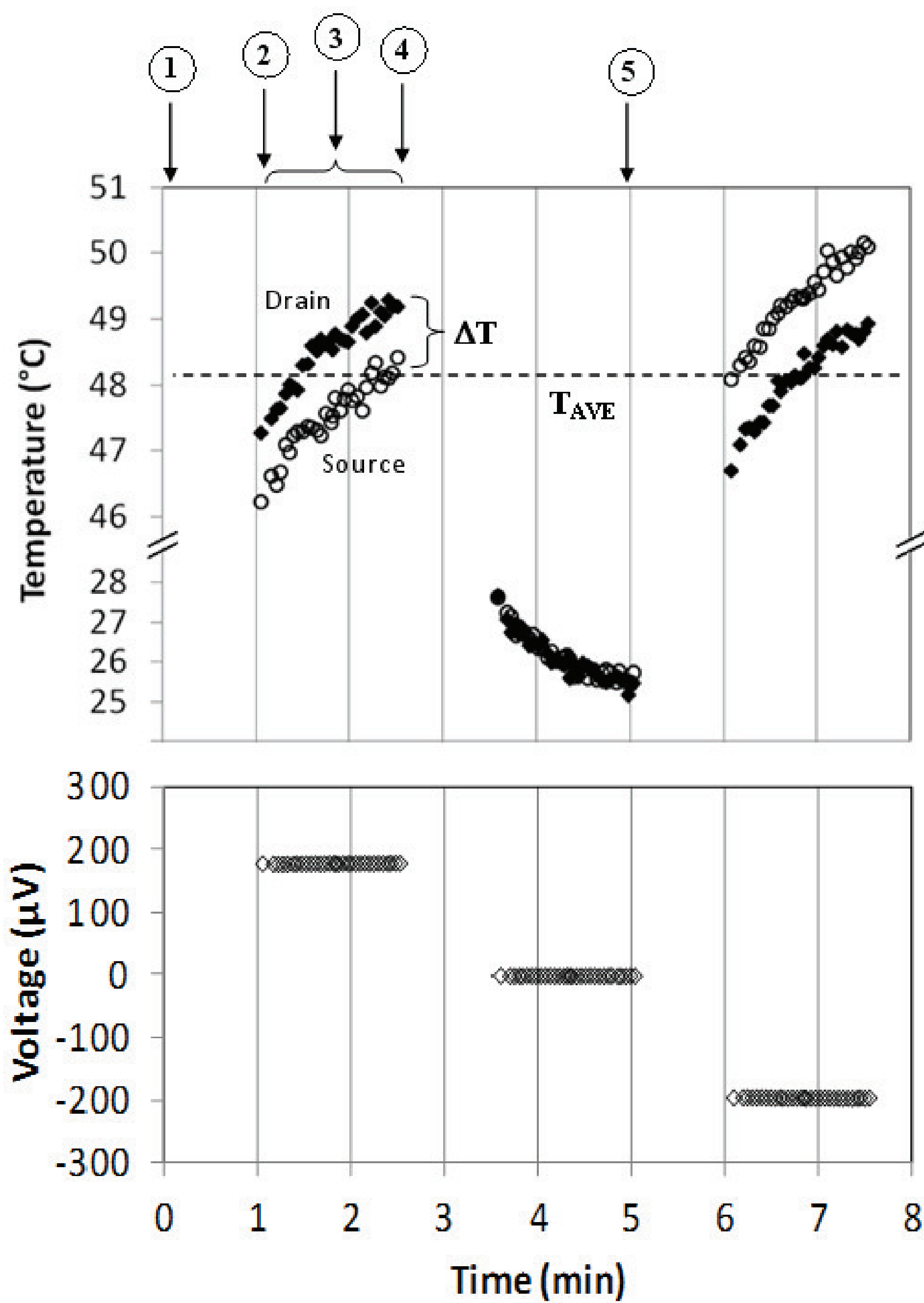


Figure 3.14: Time chart showing evolution of the temperature at drain- and source-side RTDs. The Seebeck voltage across drain and source is also showed on a bottom graph. Data taken from the same sample in Fig. 3.12.

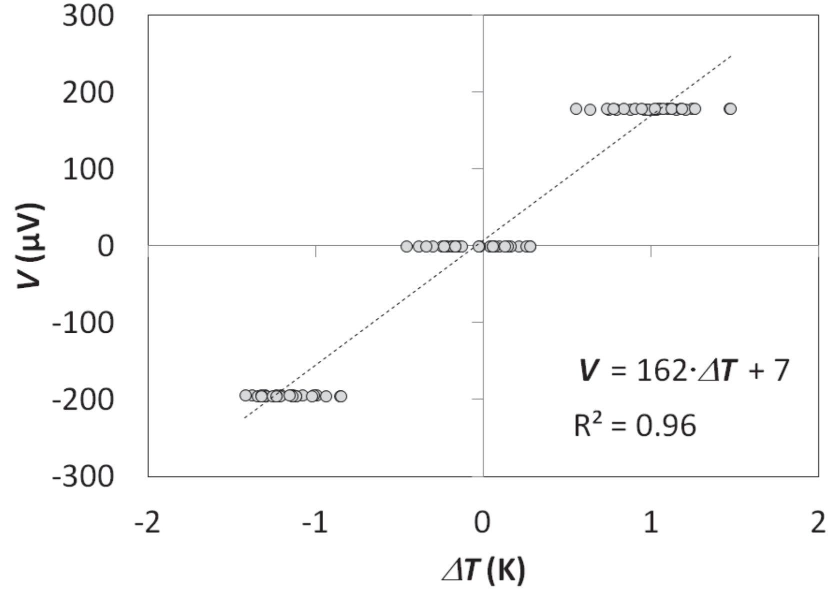


Figure 3.15: Plot of the Seebeck voltage across drain and source contacts, $V = V_{ds}$, versus drain-source temperature difference, $\Delta T = T_d - T_s$. Linear fit (dashed line) indicates that $S = -163\,\mu\text{V/K}$. Data taken from the same sample (n-Si) in Fig. 3.14.

Assuming that $\delta T_d = \delta T_s$, the uncertainty in ΔT is then

$$\delta(\Delta T) = \sqrt{(\delta T_d)^2 + (\delta T_s)^2} \approx \frac{\sqrt{2}\delta R}{R_0 \alpha} \tag{3.35}$$

Given that $R_0 \approx 1.46\,\Omega$ (average of both drain and source RTDs) and $\alpha = 2.74\frac{10^{-3}}{°\text{C}}$ for the device shown, it is estimated that

$$\delta(\Delta T) = \frac{\sqrt{2} \times 0.5 \times 10^{-3}}{1.46 \times 2.74 \times 10^{-3}} = 0.18\text{K}$$

, which is close to the actual $\delta(\Delta T) \approx 0.2\text{K}$ observed in Fig. 3.15.

According to equation 3.35, $\delta(\Delta T)$ could be minimized by increasing α or R_0 of the RTD, though each solution has its own disadvantages. For example, α could be increased by using metal with high TCR such as Ni, though the choice of metal must satisfy the requirement for Ohmic contacts. R_0 could be decrease by

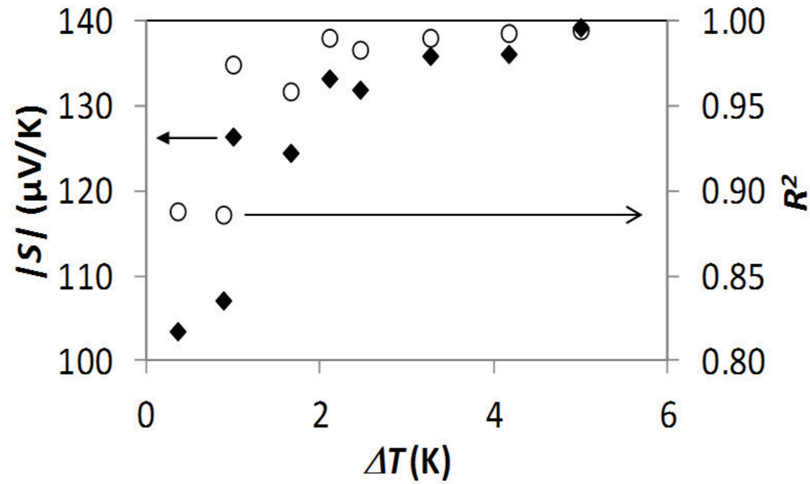


Figure 3.16: Goodness of fit, as represent by the R^2 value, and $|S|$ both increases with ΔT as the constant δT is rendered less significant.

making the RTD thinner, narrower and longer. However, reducing the thickness of metal deposition can make the device more prone to damage by electro migration. Increasing the length of the RTD may reduce accuracy of temperature detection if there is a temperature gradient along the length of the RTD. A good solution would be to reduce the width of the RTDs, though our photolithographic mask is limited to 3 μm resolution.

Uncertainty of S can also be reduced by increasing ΔT. From the propagation of error, the uncertainty for S is simply

$$\frac{\delta S}{|S|} = \frac{\delta(\Delta T)}{|\Delta T|} \tag{3.36}$$

Figure 3.16 compares $|S|$ and the coefficient of determination (R^2) of the $V - \Delta T$ linear fit to ΔT. As the figure shows, both R^2 and S increases rapidly with increased ΔT, for $\Delta T < 3$K.

Beyond $\Delta T > 3$K, $|S|$ continued to increase due to rising T_{AVE} of sample. Figure 3.17 shows that given the same T_{AVE}, the GaAs substrate can maintain a larger ΔT than the SOI substrate because κ of GaAs ($55\,\frac{\text{W}}{\text{mK}}$) is lower than that

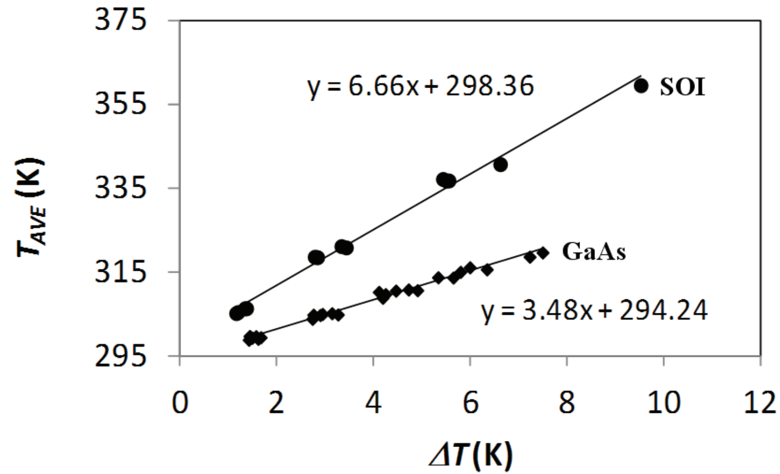


Figure 3.17: Average temperature increases with ΔT due to greater heating power requirement. Data are shown for multiple SOI and GaAs samples.

of Si (149 $\frac{\text{W}}{\text{mK}}$). Both substrates have comparable thicknesses of approximately 700 µm. Figure 3.18 shows that the observed increase in $|S|$ of GaAs with T_{AVE} is consistent with the calculation. S was calculated as a function of T using equation 1.46, where E_F was interpolated from equation 1.34 using the following parameters: $n = 3.2 \times 10^{17}\,\text{cm}^{-3}, m_d = 0.067 m_0$, and $r = 0$.

The accuracy our measurement system is confirmed in Fig. 3.19. n and S were measured using the VDP device layout for a number of doped p- and n-type Si-on-insulator, and n-type GaAs samples. Our measurements (filled data point) compare well to values found in literatures for Si [11, 12] and GaAs [7, 8, 9, 10]. Our calculations, using equations 1.34 and 1.46 assuming n-Si and n-GaAs band parameters in Table 1.2, also closely agree with measurement results. $r = -1/2$ and $r = 0$ were assumed for Si and GaAs, respectively. Calculation for p-Si underestimate S, and is not shown. The discrepancy is likely to the non-parabolicity of Si valence band.

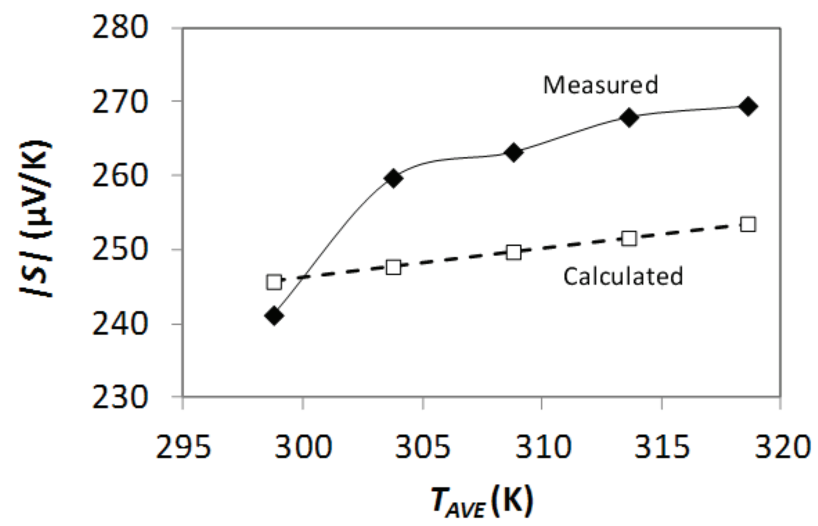


Figure 3.18: Comparison between the measured and calculated value of $|S|$ as a function of average temperature for a GaAs sample with $n = 3.2 \times 10^{17}\,\mathrm{cm}^{-3}$.

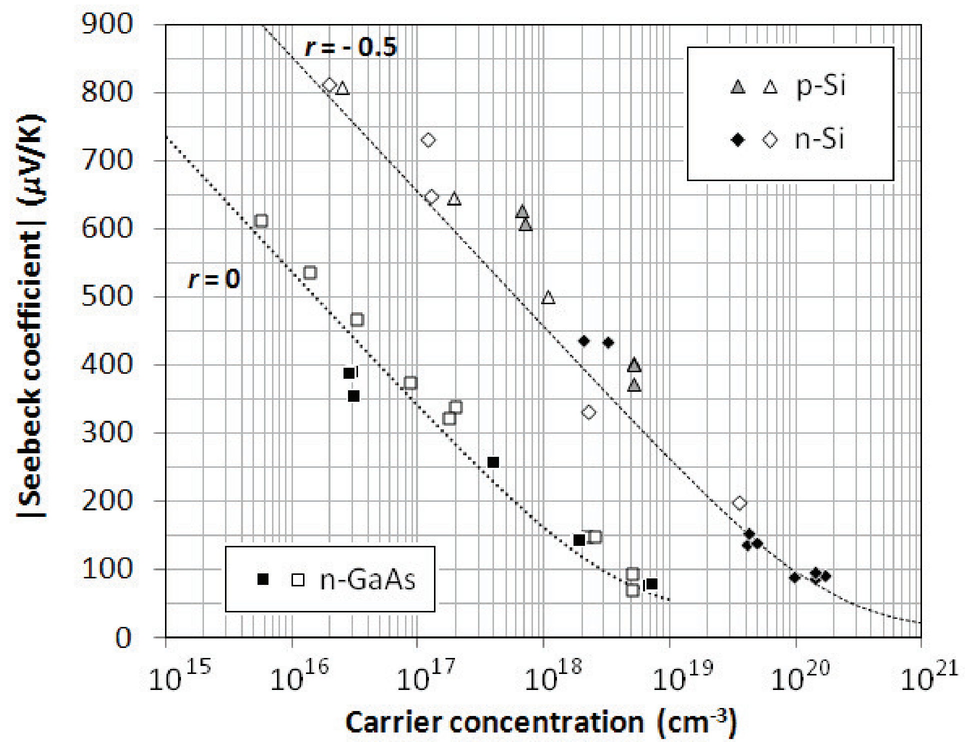


Figure 3.19: Seebeck coefficient versus carrier concentration for n-GaAs, p-Si, and n-Si as measured in this work (filled) and from literatures (open). GaAs: [7, 8, 9, 10], Si: [11, 12]. Dashed lines show the calculation results.

4 Thermoelectric characterization of Si nanowires

Silicon is an ubiquitous material for the semiconductor industry. Among its most advantageous property is the ease of which a high quality insulator, Silicon dioxide (SiO_2), can be thermally grown. In addition, Si is widely available, has high thermal conductivity that helps with heat dissipation, and can be easily doped both n- and p-type. These qualities have led to wide adoption of MOSFET as one of the most critical elements of integrated circuit technology today. However, Si is not without its disadvantages. Si is an indirect-band gap semiconductor, which generally means photons absorption or emission will be inefficient, making it an unsuitable material for devices such as light-emitting diodes or laser. Electrons mobility in Si is also not high enough for use in high-frequency microwave or radio wave circuits. These deficiencies have led to development of III-V semiconductor, most notably of which is the GaAs, which is the base material of our work in Chapter 5.

In regard to thermoelectric application, Si actually has a good balance of m_d and μ resulting in large $S^2\sigma$ [39]. Furthermore, it has a large band gap allowing for operation at very high temperature; recalling from Chapter 1, large band gap will prevent thermal excitation of minority carriers that reduces $|S|$. However, Si also has a large $\kappa \approx 150\,\frac{\mathrm{W}}{\mathrm{mK}}$ that reduces ZT. Hence, Si is often alloyed with Ge, which can reduces κ to as low as $10\,\frac{\mathrm{W}}{\mathrm{mK}}$ at 30% Ge concentration [37]. At this low-Ge composition, the electrical property of the alloy is still Si-like, thus it has the both large $S^2\sigma$ and low κ_L. $Si_{0.7}Ge_{0.3}$ is an important thermoelectric material

for operation at high temperatures up to 1,000 K.

Desirable thermoelectric and optical properties can also be obtained by modification of phonon or electrons transport at nanometer scale. For instance, it was found that κ of Si NWs with diameter of 10-100 nm, is reduced from bulk value with decreasing diameter [54, 61, 62], and increasing surface roughness [62]. The reduction is attributed to increase in phonon scattering that occurs when diameter or size of surface features approaches the phonons mean free paths, which is estimated to be 300 nm at room temperature [62]. At smaller diameter, quantum confinement effect can be utilized to affect change in electrical property. For example, band folding could induces an indirect-to-direct band gap transition, which would improve photo absorption efficiency of Si [63]. Four-time increase in μ have been observed in 3-5 nm p-type Si NWs MOSFET, which was attributed to a change in band structure due to quantum confinement effect [64].

Quantum confinement effect could also enhance $S^2\sigma$ through an increase in the density of states [50, 51, 43]. As discussed in section 2.3.1 and shown in Fig. 2.9, an enhancement of $S^2\sigma$ over bulk value could be expected at room temperature when the diameter of Si NWs is below 4 nm. To our knowledges, while there have been measurements of ρ in 6 nm doped NWs [65], S in 10 nm doped NWs [54], μ in 3-5 nm undoped NWs [64], and κ in 15 nm NWs [62], there has not been a study of thermoelectric property of doped Si NWs in the $\sim$ 4 nm region. Therefore, it was our intention to investigate the thermoelectric property of Si NWs with very small diameter close to and below 4 nm.

To study the property of such NWs, fabrication process and measurement methods (see Chapter 3) were extensively developed. In this Chapter, we will first discuss (a) expected change in $|S|$ as a function of diameter, (b) the detail of the fabrication processes, (c) comparison of FET measurement on of bulk Si and Si NWs with diameter of 40 nm, and (d) comparison of oxidized bulk Si and oxidized Si NWs exhibiting dopant segregation. Finally, future steps that may be required

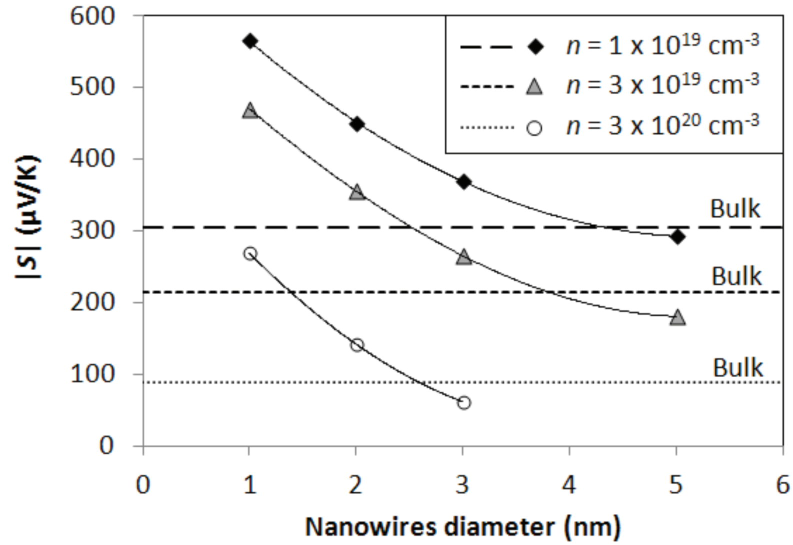


Figure 4.1: Seebeck coefficient of Si nanowires as a function of nanowires diameter.

to achieve measurement of smaller diameter NWs, and more complete analysis are discussed.

4.1 Quantum confinement effect on S of Si NWs

$|S|$ may be enhanced by quantum confinement effect, as previously discussed in section 2.3. Figure 4.1 shows the expected increase in $|S|$ as a function of d for NWs with three different n, assuming n-type Si, $T = 300K$, and $r = 0$. In all cases, $|S|$ is increased as d is decreased. However, the confinement effect can also leads to a decrease in $|S|$ compare to bulk value, if the diameter is not sufficiently small. In the figure, $|S|$ of NWs with diameter greater than 3-5 nm may be smaller than that of bulk value.

Another important feature shown in this figure is that the minimum diameter, where enhancement of $|S|$ is observed, increases with decreasing n. It was

Table 4.1: SOI wafer specification

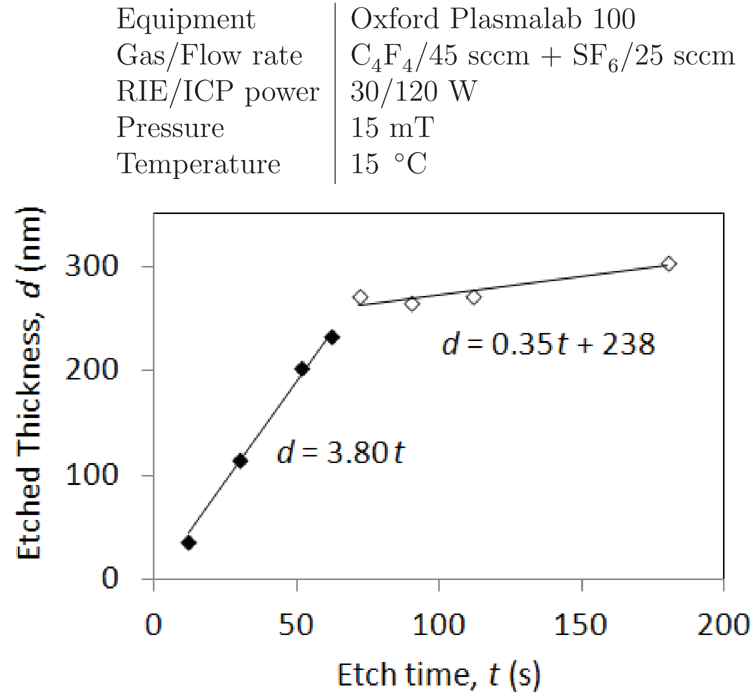
Supplier	Soitec®
Product number	G6P-029-01
SOI construction method	Wafer bonding
Top silicon layer	
- Thickness (nm)	260
- Resistivity (Ω.cm)	13.5-22.5
- Crystal orientation	⟨100⟩
Buried oxide	
- Thickness (nm)	1,000
Si handle	
- Thickness (μm)	675
- Resistivity (Ω.cm)	14-22
- Crystal orientation	⟨100⟩

previously shown that the minimum diameter needed for an enhancement of $S^2\sigma$ for Si NWs at room temperature is 4 nm, regardless of n, in section 2.3.2. In that calculation, the optimal η was assumed, and therefore n is a variable. In practice, n should be fixed by the concentration of donor, hence S was plotted with a constant n here. Figure 4.1 shows that very large increase in $|S|$ can be expected at the onset of confinement effect.

4.2 Doping of SOI substrate

The Si NWs were to be fabricated out of the Si top layer of a SOI substrate. Prime-grade SOI substrate (see. Table 4.1) made by the wafer bonding method, which produces good, well-defined interface between the Si top layer and the buried oxide (BOX) layer were used. This interface is important as it defines the bottom part of our NWs. Since the top layer's resistivity and thickness are both too high, the top layer must first be thinned and doped.

The Si top layer is thinned by plasma etch using the recipe shown in the table atop Fig 4.2. The etch rate of the Si top layer is approximately 3.80 nm/s,



Equipment	Oxford Plasmalab 100
Gas/Flow rate	C_4F_4/45 sccm + SF_6/25 sccm
RIE/ICP power	30/120 W
Pressure	15 mT
Temperature	15 °C

Figure 4.2: Si etching recipe and etched thickness versus etch time. Two distinct etch rate is observed for Si top layer (filled diamond) and the buried oxide (open diamond), which is exposed after all 260 nm of Si is etched away.

which means the top layer will be completely removed after about 68 s of etching. The oxide in the BOX layer will also be etched once it is exposed, though at a much lower rate of 0.35 nm/s. Based on this etch rate, SOI should typically be etched for approximately 58 ± 2 s to a final thickness of approximately 40 nm. The thickness of Si layer can be measured with the Filmetrics F20 thin film optical measurement system. Alternatively, a step profile is created by patterning the sample with photo-resist prior to etching. After etching, photo-resist is removed, and step height is measured with the Dektak 150 profilometer.

Si was then doped via one of the 2 methods: diffusion doping, or ion implantation. Diffusion doping were tried first since it offers the flexibility to vary dopant concentration, and was the method used in Ref. [54]. However, the dop-

ing method was switched to ion implantation after problem with surface defects in diffusion-doped sample. Doping procedure and the electrical characterization of doped Si top layer on the SOI substrate for both methods are discussed. For brevity, the doped Si top layer will be refer to as "Si thin film" for here on. Note that this thin film is still too thick to exhibit quantum confinement effect.

4.2.1 Diffusion doping

Thermal diffusion doping begins with a spin-coating of the "spin-on dopant" (SOD) solution. The SOD, also known as spin-on glass (SOG), is a solution of silicates, siloxanes, or silsesquioxanes, that contains a high concentration of dopant, such as B, P, As, or Al. Highly-concentrated Filmtronics' B155 and P509 solutions were chosen for high n. The application of the SOD follows these steps:

1. Spin-coat SOD on SOI at 4,000 rpm for 10 s.
2. Pre-bake at 100 °C to dry solvent.
3. Bake in RTA at high "drive-in" temperature of 600–1100 °C
4. Remove the SOD film with by dissolving in buffered-oxide-etch (BOE) solution.

Samples coated with P509 should be immediately loaded into the RTA furnace. Since P is very hydrophilic, moisture will condensate very quickly on the P509 film, which ruins the uniformity of the film. Sample is held for a duration of 2 min at the peak temperature in step 3, except when $T > 1,000\,^{\circ}\mathrm{C}$, where the hold time is limited to 1 min to prevent equipment damage to the RTA. Samples coated with B155 and P509 were always baked separately to prevent cross contaminant by gases phase.

n of doped samples was then measured via VDP method, which requires metallization for Ohmic contact. All metals were deposited via e-beam evaporator. Al is used for B-doped (p-type) Si, as it has good adhesion on Si and SiO_2, and is

very easy to form Ohmic contact with Si due to the low Al-Si eutectic temperature of 577 °C [14]. However, Al forms Schottky contact when used on n-type Si. Silicide contacts, such as $NiSi_2$ or PtSi, may be used instead to obtain Ohmic contact on n-type Si.

For P-doped sample, Ti/Ni/Au is sequentially deposited, starting with Ti, for contact. Thin Ti layer of 5-10 nm provides good adhesion on both Si and SiO_2. Since Ti deposition is very sensitive to moisture in the chamber, the pre-deposition vacuum should be at least below 10^{-6} Torr to ensure good adhesion. 50-100 nm thick layer of Ni is then be deposited for subsequent silicidation. Since Ni requires relatively high e-beam power to evaporate, the sample must be well affixed to the holder/heat sink to avoid overheating the resist. Thinner than 50 nm Ni layer could suffice as well given the 60% silicidation depth for NiSi [66]. Au is then deposited to provides a stable resistive element for the on-chip RTDs required for S measurement, as described in Chapter 3. Ti/Ni RTDs tend to have high resistance, and disorderly $R(T)$ relationship.

Samples were then annealed to ensure Ohmic contact. Al-Si samples were typically annealed at 400 °C, while Ti/Ni/Au-Si samples were annealed at slight higher 450 °C. Silicide formation temperature is slightly higher than that of Al-Si eutectic temperature. The best Ni-Si contact is reportedly obtained between 450-700 °C [66]. Annealing time is 1 min. Annealing is performed in the RTA furnace in flowing ($\sim$ 2L/min) N_2 atmosphere.

Figure 4.3 shows that over four order of magnitude of n, all the way up to the n solubility limit of $3-4\times10^{20}$ cm^{-3} for P and B at 1,100 °C, can be achieved for the as-received SOI sample with d_{Si} = 260 nm. The resultant $\log(n)-T$ relationship differs for each type of dopant, however. Whereas $\log(n)-T$ appears to be linear for the B155, it appears to be non-linear for the P509. These behaviors are reflective of B and P diffusivity in Si (D_B and D_P, respectively) shown in

Fig. 4.4. B and P diffusivity is calculated with the following equations [67]

$$D_B = D^0 + D^+ \left[\frac{n}{n_i}\right] \tag{4.1}$$

$$D^0 = 0.037 \exp(-3.46/k_B T)$$

$$D^+ = 0.76 \exp(-3.46/k_B T)$$

, and

$$D_P = D^0 + D^= \left[\frac{n}{n_i}\right]^2 \tag{4.2}$$

$$D^0 = 3.85 \exp(-3.66/k_B T)$$

$$D^= = 44.2 \exp(-4.37/k_B T)$$

, where n_i is the intrinsic carrier concentration for Si [34]. D_B is generally higher than D_P at all T except for when n is very high or very low. As T increases, however, the window where $D_B > D_P$ will shrinks, and both D_B and D_P will be closer in value. Thus n is similar for both type of samples when $T > 1,000\,°\mathrm{C}$, but is higher for B-doped samples at lower T.

Electron and hole mobility of diffusion-doped Si thin film were consistent with literature value [68, 13], as shown in Fig. 4.5. Standard mobility for B- and P-doped Si were calculated from an empirical model [13] using equation

$$\mu = \mu_0 + \frac{\mu_{max} - \mu_0}{1 + (n/C_r)^\alpha} - \frac{\mu_l}{1 + (C_s/n)^\beta} \tag{4.3}$$

, where the value for each parameters can be found in Table 4.2; units for μ and n in equation 4.3 are $\mathrm{cm^2/Vs}$ and $\mathrm{cm^{-3}}$, respectively. The electron mobility (in P-Si) is larger than that of hole mobility (in B-Si) due to greater holes effective mass. Note Ref. [13] specifically pointed out that μ depends largely on concentration of *ionized* impurity. Therefore, the dependence of μ on n in equation 4.3 implies that the concentration of ionized acceptor or donor (N_A or N_D, respectively) is approximately equal to concentration of free carrier, i.e. $n = N_A$ for p-type Si, and $n = N_D$ for n-type Si. Non-ionized impurity or other lattice defects may also reduces μ as well, but their contribution should be relatively insignificant.

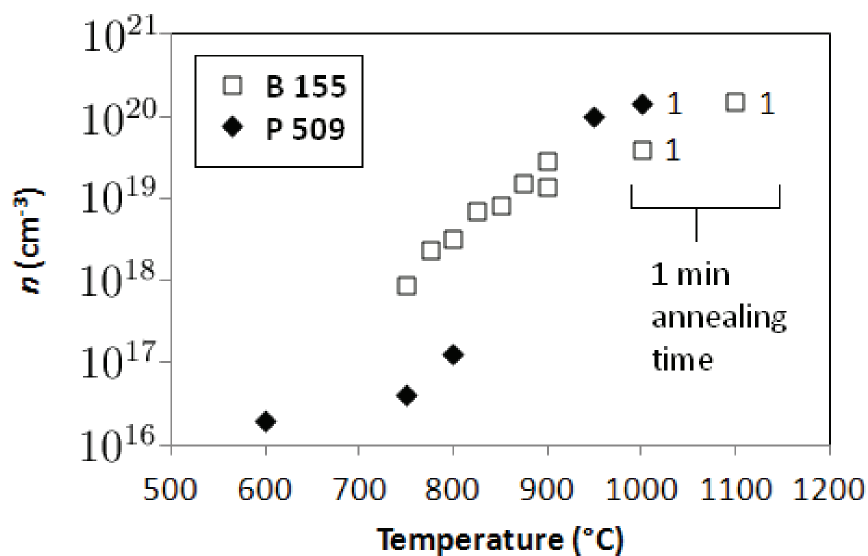


Figure 4.3: Carrier concentration versus annealing temperature of P- and B-doped Si film. Doping is via thermal diffusion. Annealing time is 2 min for $T < 1,000\,°C$, and 1 min for $T > 1,000\,°C$.

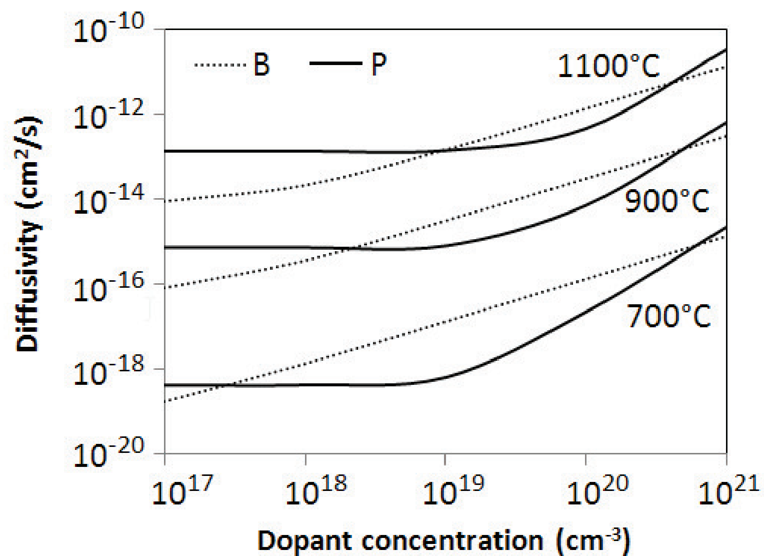


Figure 4.4: Diffusivity of P and B in Si as a function of dopant/carrier concentration at 700 °C, 900 °C, and 1,100 °C.

Table 4.2: Parameters for Si mobility calculation [13]

Dopant	P	B
μ_0	68.5	44.9
μ_{max}	1414	470.5
μ_l	56.1	29.0
C_r	9.20×10^{16}	2.23×10^{17}
C_s	3.41×10^{20}	6.10×10^{20}
α	0.711	0.705
β	1.98	2.00

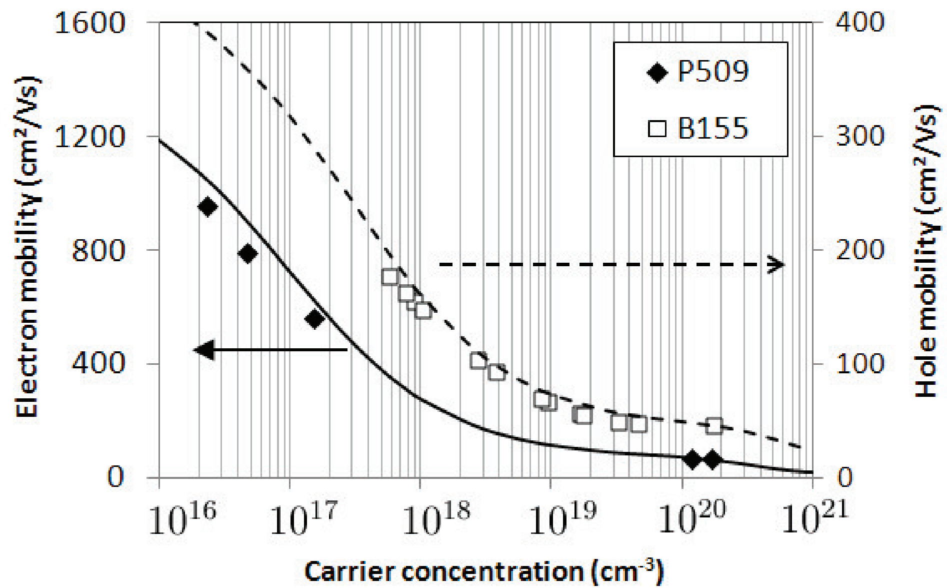


Figure 4.5: Mobility versus carrier concentration of P- and B-doped SOI samples. Doping is via thermal diffusion. Solid and dashed lines are empirical model of P- and B-doped Si [13].

However, defects were observed on many of the diffusion doped samples. Figure 4.6a shows photos of star- and round-shaped discolorations on 260 nm B-doped thin film. Surface defect can be more severe for thin film with $d_{Si} = 40\,\text{nm}$. In subfigure (b), SEM images show extensive crack and porosity. Micrometer-sized pinholes in the BOX layer is observed in subfigure (c). The porosity is very extensive for B-doped samples, as shown in subfigure (d). On the contrary, the porosity is not as severe on P-doped samples, as shown in subfigure (e). Less concentrated SOD should probably be used to avoid surface defect according to the SOD manufacturer. However, given the problem with surface defect, and the process time required by diffusion doping, the SOI doping method was replaced by ion implantation-based method instead.

4.2.2 Ion implantation

Ion implantation involves bombardment of the film with ionized species accelerated to high velocity in an electric field. The ion-implantation specification, and the resultant electrical property, of our samples are summarized in Table 4.3. The initial P14 and P15 batches were doped p-type with B, and was not thinned ($d_{Si} = 260\,\text{nm}$). In the subsequent N34 and N35 batches, the top layer was etched until $d_{Si} = 40\,\text{nm}$ prior to ion implantation, and dopant was switched to P. The accelerating voltage was selected such that the peak of the dopant concentration, whose profile is approximately Gaussian, should be at the middle of the Si top layer. Higher voltage of 100 keV was thus needed for the P14 and P15 batches that have thicker top layer. The ion-implanted dopant concentration, N_I, is then selected by varying the area dosage. The peak location and peak N_I can be calculated with SRIM simulation tool [69]. Average N_I was calculated by a simple Riemann sum of the N_I profile over the total top layer thickness. The ion implantation was performed by Leonard Kroko Inc.

The electrical property of ion-implanted substrate further depends on the activation annealing process. Annealing at high temperature (800-1,100 °C) allow

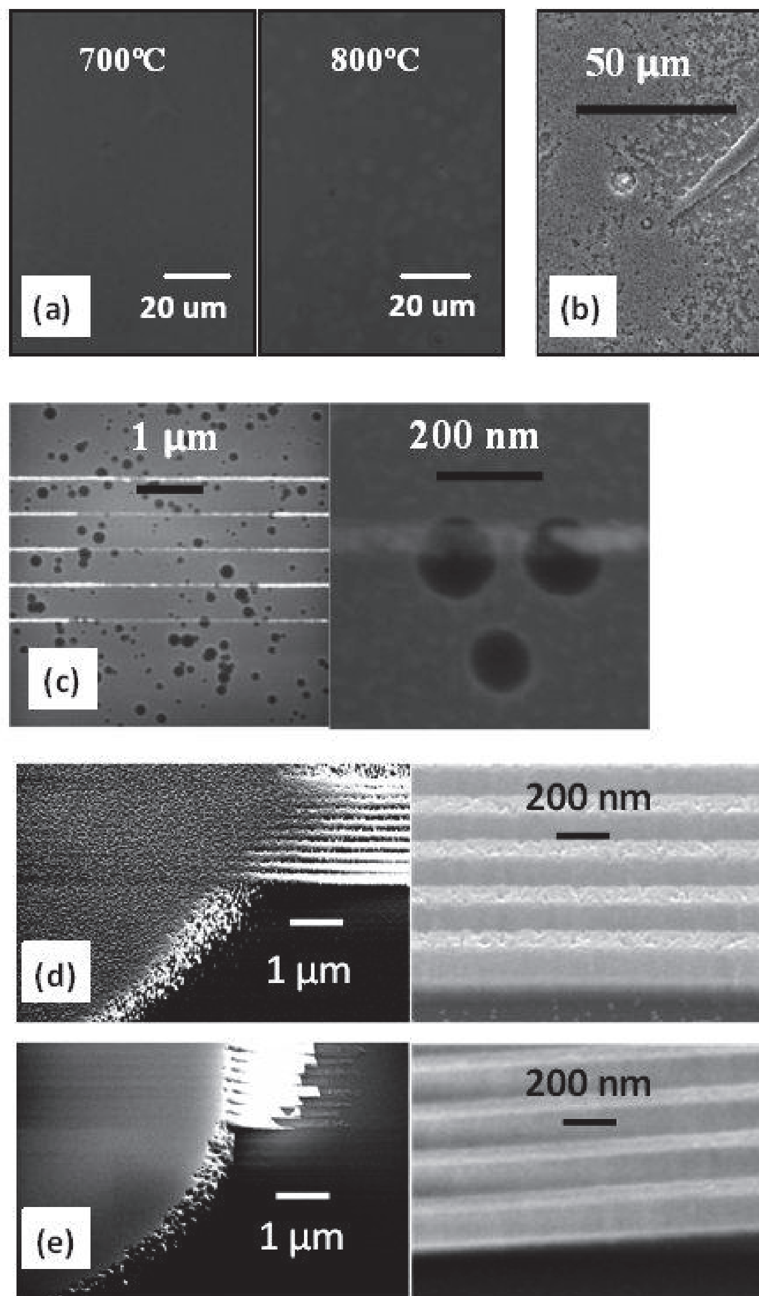


Figure 4.6: Surface defects of diffusion doped Si top layer on SOI substrate.

Table 4.3: Ion-implantation specification and approximate resultant electrical properties. All implant performed with nominal 7° substrate tilt.

Batch ID	P14	P15	N34	N35
Implant specification:				
Dopant Species	B	B	P	P
d_{Si} (nm)	260	260	40	40
Energy (keV)	100	100	15	15
Dosage (cm^{-2})	1×10^{14}	1×10^{15}	3×10^{14}	3×10^{15}
SRIM:				
Peak N_I (cm^{-3})	6.0×10^{18}	6.0×10^{19}	9.3×10^{19}	9.3×10^{20}
Average N_I (cm^{-3})	1.4×10^{18}	1.4×10^{19}	7.1×10^{19}	7.1×10^{20}
Measured value:				
n (cm^{-3})	8.0×10^{17}	6.0×10^{18}	4.0×10^{19}	2.7×10^{20}
$n/(\text{Ave.}\, N_I)$	57%	43%	56%	38%
ρ (mΩ.cm)	55.7	14.6	1.89	0.9
μ (cm^2/Vs)	140	71	83	26

both the implanted dopant and the displaced (by impact with incident ions) Si atoms in the interstitial sites to diffuse back to substitutional sites. Dopant atoms must be placed in the substitutional site to donate or accept electrons, while Si must be removed from interstitial sites where they act as defect that can trap electrons or reduce mobility. Without activation process, N_A and N_D may be less than N_I, and μ may be sub-optimal.

Figure 4.7 shows μ as a function of n for the ion-implanted samples. P14 and P15 samples were annealed at 800 °C, and of N34 and N35 samples were annealed at 800 °C and 1,100 °C. n and μ were measured via VDP method. The average n and μ for each batch are also listed in Table 4.3. With the exception of N35 samples, μ of most samples are close to the expected value according to equation 4.3. If N_I was correctly calculated, then these measurement appears to indicates that only 40-60% of dopant were activated.

The effect of activation temperature can be observed in N34 and N35 samples. The N34 samples are likely to be completely activated since μ and n does not change noticeably with annealing temperature, and μ for all samples are in

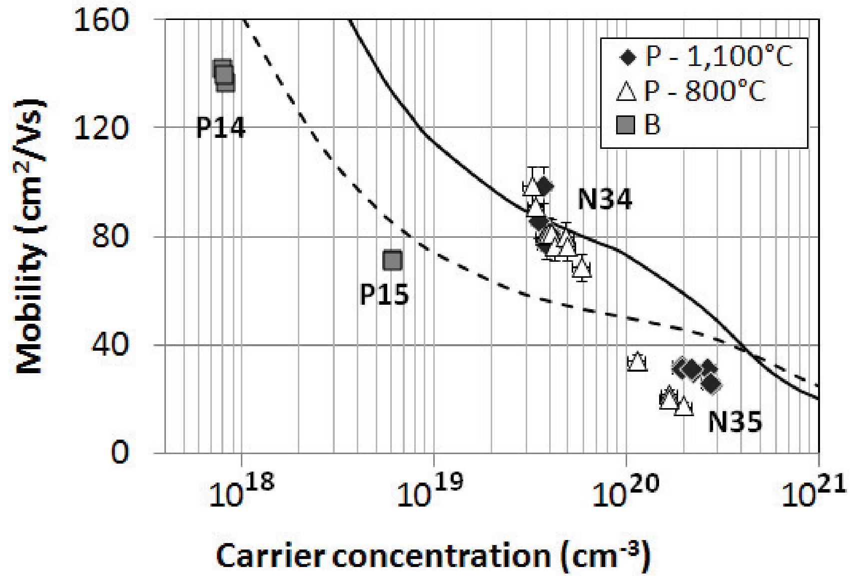


Figure 4.7: Mobility and carrier concentration of ion-implanted Si thin film. Solid and dashed lines represent empirical model [13]

reasonable agreement with the empirical model calculated with equation 4.3[13]. Since N34 sample appear fully activated, the average N_I in Table 4.3 may have been overestimated. On the contrary, N35 samples have shown that n increases when annealing temperature is increased from 800 °C to 1,100 °C. Since samples with higher N_I will requires higher annealing temperature or longer annealing time (all samples were annealed for 1 min), it is reasonable to assume that while 800 °C is sufficient for N34 samples, it may be insufficient for N35 samples. Likewise, even $1,100$ °C anneal for 1 min may also be insufficient for N35 since μ and n are still below the reference value.

However, further annealing may not increase n for N35 samples due to the solubility limit of P in Si. Figure. 4.8 shows that while solubility limit of P in Si is approximately 10^{21} cm^{-3} at 1,100 °C, only $3 - 4 \times 10^{20}$ cm^{-3} these P dopant may be electrically active and contribute to n. The inactive dopant would form non-ionized precipitates, which should contribute only negligibly to carrier scat-

tering [13].

Differences in n is also distinguished with S measurement. Figure 4.9 shows that $|S|$ of N35 samples annealed at 800 °C is larger than that of N35 samples annealed at 1,100 °C. Similarly, N34 samples that have lower μ and higher n, also have lower $|S|$. All measurement agrees reasonably well with result first-principle calculation, which was calculated as follow

1. Assume a range of E_F from approximately -0.2 to 4.0 eV.

2. Calculate n for each E_F according to equation 1.34

3. Calculate S for each E_F according to equation 1.46

$T = 300\,K$ is assummed. In step 2, standard Si conduction band parameters ($m_d = 0.32m_0, N_c = 6$), and $D = 3$ are assumed. In step 3, r of -1/2 and 0 were assumed. Calculated n and S from the same E_F value are then linked together. Measurements fall in between the two lines, but are closer to $r = -1/2$ line. $r = -1/2$ is associated with ADP and $\mathrm{II_{str}}$ scattering, both of which are relevant for highly-doped Si at room temperature. S for P14 and P15 can be found in Fig. 3.19.

Measurement uncertainty for VDP method is small since there few parameters needed. An uncertainty of ±3 nm is estimated for the Dektak profilometer and the Filmetrics F20, which equates to $3/40 = 7.5\%$ error for $d_{Si} = 40\,\mathrm{nm}$ for both N34 and N35 samples. The uncertainty in R measurement is estimated to be 0.5Ω in section 3.3.4, which is negligible compare to resistance of the sample that may be hundreds of Ohms at the least. Therefore, the uncertainty in ρ depends only on d_{Si} as

$$\frac{\delta\rho}{\rho} = \frac{\delta d_{Si}}{d_{Si}} = 7.5\%$$

On the other hand, V_H measurement was more susceptible to electronic noise and sample inhomogeneity. The uncertainty is also greater for N35 than N34 samples since the former has greater n, and therefore smaller $|V_H|$. Meanwhile, the value of B is calculated from coil current based on earlier calibration. B may differs

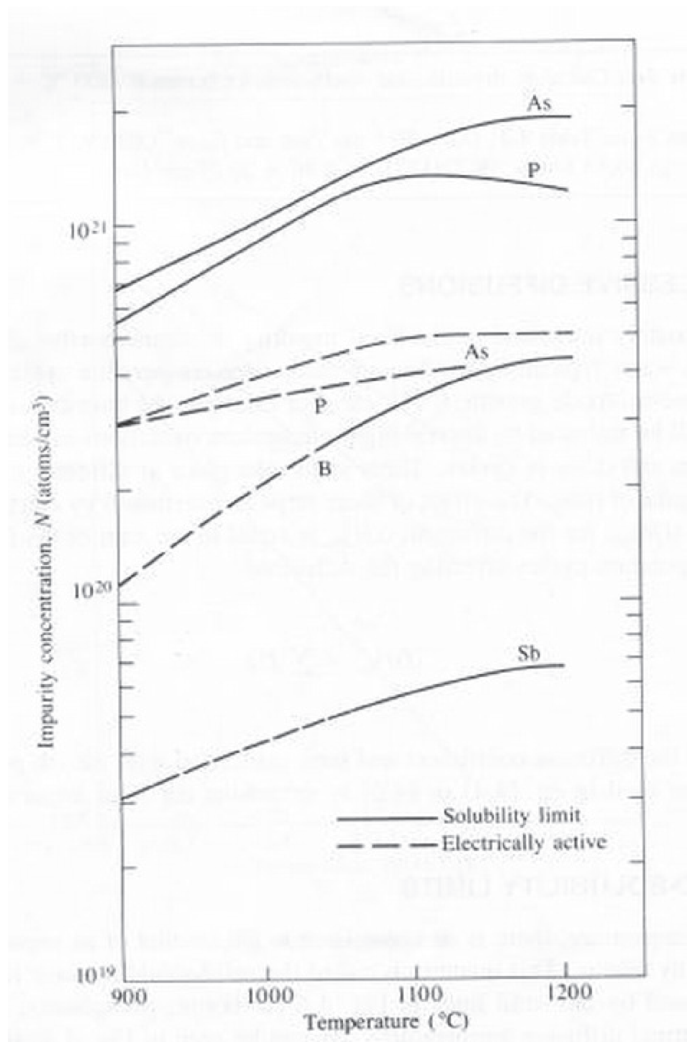


Figure 4.8: Solid-solubility and electrically-active impurity concentration limit in Si at high temperature. Adapted from Ref. [14].

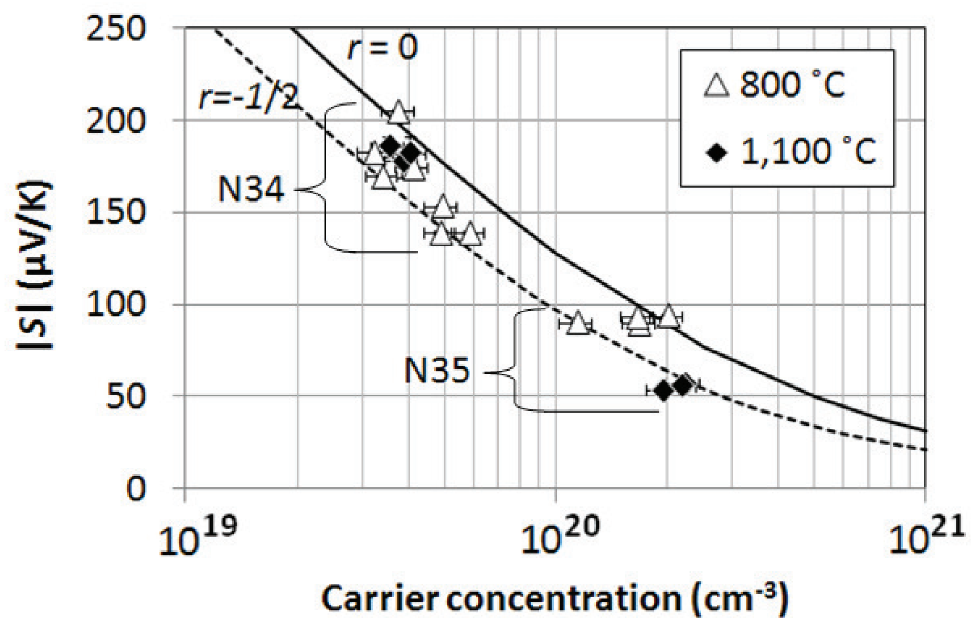


Figure 4.9: Seebeck coefficient as function of carrier concentration for N34 and N35 Si thin films. Solid ($r = 0$) and dashed lines ($r = -1/2$) represent first-principle calculation.

depending on the sample position relative to the true center of the solenoid. Based on our experience, errors for both V_H and B should be less than 5%. According to equations 3.4 and 3.5, the uncertainty for n and μ are estimated to be

$$\frac{\delta n}{n} = \sqrt{\left(\frac{\delta t_{Si}}{t_{Si}}\right)^2 + \left(\frac{\delta V_H}{V_H}\right)^2 + \left(\frac{\delta B}{B}\right)^2} \approx 10\%$$

$$\frac{\delta \mu}{\mu} = \sqrt{\left(\frac{\delta V_H}{V_H}\right)^2 + \left(\frac{\delta B}{B}\right)^2} \approx 7\%$$

Activation annealing at high T should also make dopant distribution more uniform throughout the depth of the Si layer. For example, the diffusion length of P in Si after 1 min bake at 1,100 °C is

$$L_D = \sqrt{D_P t} = \sqrt{11.45 \times 10^{-13}[\mathrm{cm}^2/\mathrm{s}] \times 60[\mathrm{s}]} = 30\,\mathrm{nm}$$

, where D_P value is taken from Fig. 4.4 for $N_D = 10^{19}\,\mathrm{cm}^{-3}$. Since $L_D \approx d_{Si}$, P should be uniformly distributed throughout the thickness of the film (i.e. N_D is a constant). Diffusion of P into the BOX layer should be negligible since the diffusivity of P in SiO_2 at 1,100 °C is $3 \times 10^{-18}\,\mathrm{cm}^2/\mathrm{s}$ [70], which is 5 order of magnitude lower than D_P in Si. Since D_P increased with N_D, all N34 and N35 samples that were annealed at 1,100 °C should have uniform P distribution in the vertical direction.

As the diffusion doped, P14, and P15 substrates were spent during the early development of this work, the FET and NWs measurements reported from here on come from samples fabricated on N34 and N35 only.

4.3 Electron Beam Lithography

Following from the preparation of a doped Si thin film, Si NWs are fabricated by patterning the substrate with electron beam lithography (EBL). To ensure the smallest line resolution affordable, the *e*-beam must be well adjusted, and the beam dosage must be optimized to match the e-beam resist. Lastly, a

Table 4.4: General settings for Raith 50 e-beam writer.

Setting	Value
Accelerating voltage (V_e)	30 kV
Probe current (I_P)	~ 100 pA
Stage height	25.5 mm
Working distance (w_d)	~ 10.6 mm
Stigmation (I_x, I_y)	$\sim -1.0\,\%, -14.7\,\%$
Beam current (I_B)	40 µA
Aperture	20 µm
Mode	OptiRes
Write field	100×100 µm^2
Magnification	1k

good design rule is needed to alleviate the tendency for *e*-beam to lose focus.

4.3.1 *e*-beam settings

EBL pattern is composed of individual pixels that were sequentially exposed to a focused electron-beam. A line, for example, is a row of such pixels. The challenge of EBL is in obtaining and maintaining a well focused *e*-beam during the course of the exposure. The Raith50 *e*-beam writer was used in this work.

In the Raith50, the *e*-beam is generated by hot-emission of electrons from the LaB_6 filament with a total beam current of I_B, and accelerated down the column with high voltage (V_e). A series of electromagnetic lenses then shape and filter out portions of the beam as it travels down the column. The *e*-beam then leaves the column through an aperture. The filtering by the lenses and the aperture will leave the beam with final probe current of I_P, which may be fractions of I_B. The settings for these parameters are listed in Table 4.4 in order of importance.

To achieve the smallest beam diameter, V_e should be high and I_P should be low. The beam diameter is limited by the repulsion of electrons that get strong as I_P increased. However, the beam divergent is lessened by increasing V_e. There-

fore, the maximum $V_e = 30\,\mathrm{kV}$, and the manufacturer's recommended value of $I_P \approx 100\,\mathrm{pA}$ were used. I_P is always measured with a Faraday cup before pattern exposure.

Next, the beam must be focused properly onto the sample. The sample stage should be raised to the maximum height of 25.5 mm, where the sample would as close as possible to the aperture. The working distance (w_d), measured from the aperture, can then be adjusted accordingly to obtain focus. Procedure in the user manual was always followed to obtains the best focus for every EBL session, though the values for these settings tend to remain the same as listed in Table 4.4.

Although the beam diameter may be only a few nm, the size of an exposed pixel will be much larger because of forward electrons scattering. Forward scattering is reduced when V_e is increased, though backward scatter off of substrate may increase instead. Backward-scattered electrons are responsible for the proximity effect, which limits how close features may be drawn next to each other. These inadvertent exposures must be minimized by optimizing beam dosage.

4.3.2 Dosage and *e*-beam resist

The resist should be as thin as possible, but still thick enough for other post-processes. Thin resist will require the least dosage, and therefore limits the extent of electron scatterings that increases feature size. In this work, two type of e-beam resist were tried: the positive-tone PMMA, and the negative-tone ma-N2403. In the positive-tone resist, polymers in the area exposed to *e*-beam is broken down, and removed during developing process. On the opposite, polymers in a negative-tone resist is further polymerize by exposure to *e*-beam, and become more resistant to developer agent. The unexposed area is then removed instead. The criteria for selecting positive- or negative-tone resist mainly depends on the process flow.

The process flow for the positive-tone PMMA is shown in left-most column of Fig. 4.10. In step (1), the Microchem® 950PMMA-A2 (“A2”) resist is applied as followed:

- Spin coat at 4,000 rpm for 60 s
- Bake on hotplate at 180 °C for 60 s

The thickness of resist may be decreased by increasing spin speed, though it is mainly dependent on resist concentration. The A2 has the lowest concentration, yielding thickness of approximately 80 nm, as measured by the Filmetrics F20. In (A.2), the resist is patterned such that exposed area will eventually define the NWs, and then developed by

- Immerse in MIBK:IPA(3:1) solution for 30 s
- Rinse in IPA for 60 s

In (A.3), 5 nm of Cr or Ni is deposited via *e*-beam evaporator to serve as etch mask. Cr or Ni have good adhesion to Si, and is easy to remove with liquid etchants. As a rule of thumb, the thickness of metal deposit should not exceed a third of the resist height, or else the subsequent liftoff process may be difficult. If thicker deposition is required such as when patterning electrodes, then more concentrated and thicker resist such as the 450PMMA-C4 can be used instead. After the liftoff process via ultra-sonic agitation in step (A.4), the exposed Si is then etched in step (A.5) using the recipe in Fig. 4.2. Care should be taken to ensure that the mask is completely removed after Si etching. Early on, there were instances where our measurements were that of the remnant metal mask, rather than the Si NWs.

The width of the NWs is then checked with the SEM to determine the optimal dosage. Figure. 4.11 shows the line width versus dosage, V_e, and I_P. The dosage is increased by increasing the exposure time. The dosage requirement is much lower for $V_e = 10\,\mathrm{kV}$ because less energetic electrons are more easily absorbed by the resist. At $V_e = 30\,\mathrm{kV}$, the same minimum line width of 40 nm is attained, but line width sensibility to dosage is lessened. Decreasing I_P to 50 pA

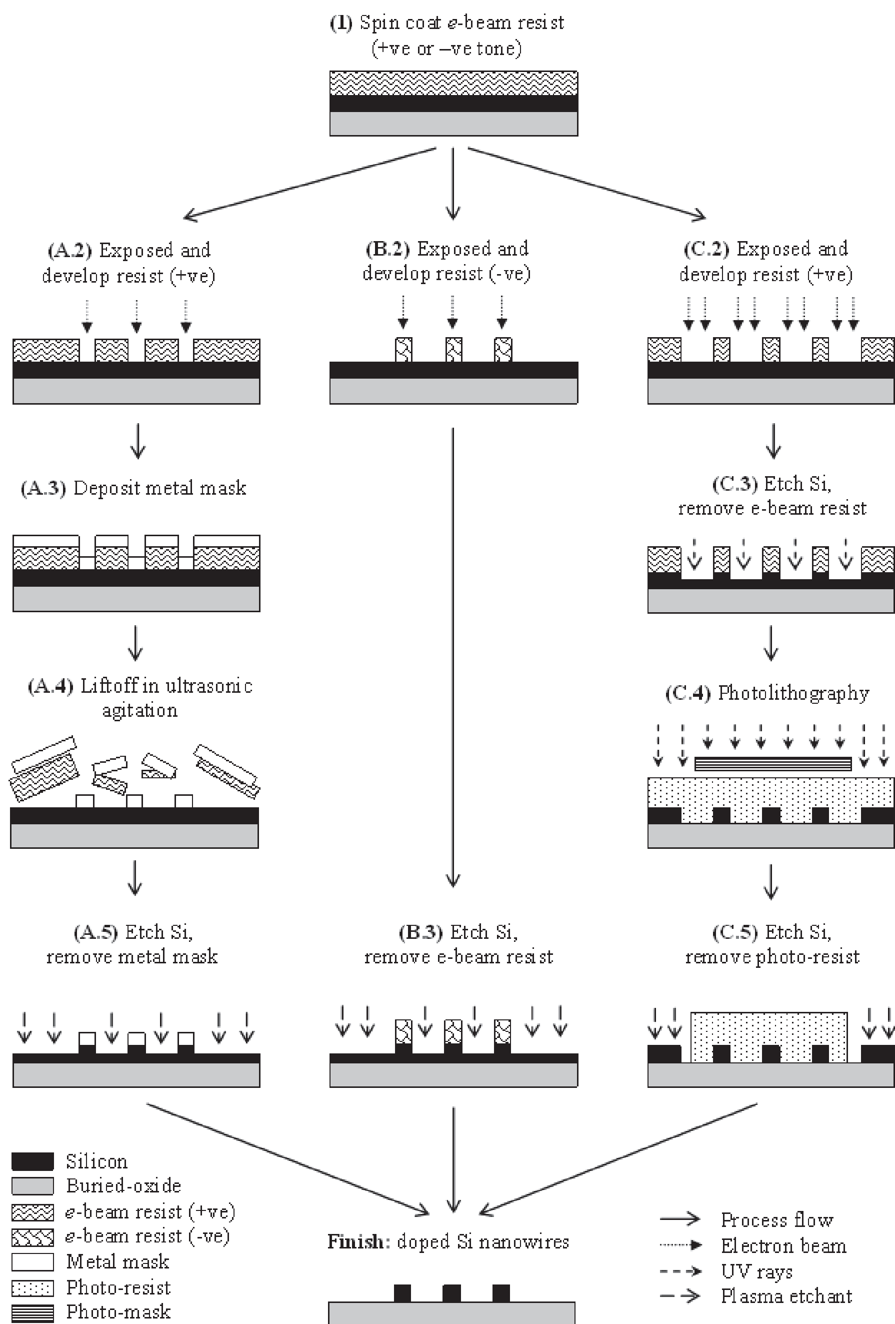


Figure 4.10: Process flow for *e*-beam lithography using (A) positive-tone resist, (B) negative-tone resist, and (C) positive-tone resist with reverse-polarity image.

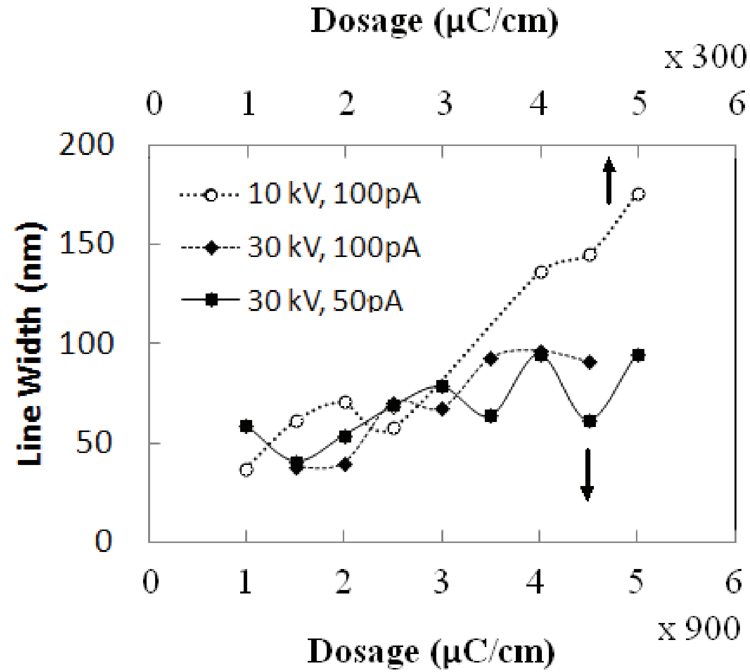


Figure 4.11: Dosage test on 80 nm950PMMA A2 e-beam resist. Dosage of 10 kV exposure is indicated by the top axis, while that of 30 kV is indicated by the bottom axis.

increase exposure time by twice as much but did not reduces line width, . Although not shown in the figure, line width does increase when $I_P > 100\,\text{pA}$. Therefore, the most optimal trade-off between exposure time and feature size is achieved at $I_P = 100\,\text{pA}$. A minimum of 900 μC/cm is required for exposure at 30 kV, though it is more reliable to use larger dosage of $\sim 1,350$ μC/cm to ensure resist development.

Metal mask deposition can be problematic. Besides increased processing, metal deposition lowers the yield as poor metal adhesion and/or liftoff can results in broken or shorted NWs. Since the Si layer to be etched is thin ($d_{Si} = 40\,\text{nm}$), the resist itself would be sufficient as etch mask. Measurement shows that the etch rate for *e*-beam- and photo-resist is about a third of the Si etch rate for the recipe

in Fig. 4.2.

The process flow for negative-tone resist is shown in middle column of Fig. 4.10. Te sample is coated with the Micro Resist Technology "ma-N 2403" resist in step (1). The minimum thickness of the ma-N 2403 resist is approximately 200 nm, so the resist is first diluted with ma-T 1090 solvent. The diluted resist (1 part ma-N 2403: 2 part ma-T 1090) then have a thickness of 80 nm when spun at 4,000 rpm. The coated sample is baked at 85 °C on the hot plate for 2 min before *e*-beam exposure in step (B.2). The same exposure pattern is used where the exposed area defines the NWs. The resist is developed in ma-D 525 solution for 30 s, and rinsed with deionized water. The sample can then be etched right away using the resist as etch in step (B.3). Subsequently, the resist is removed in acetone.

Figure 4.12 show that a minimum line width of 40 nm is likewise obtained at the optimal dosage of 360-450 μC/cm with the diluted ma-N 2403 resist. Negative-tone resist generally have lower dosage requirement than positive-tone resist. SEM images show the Si NWs etched with the diluted ma-N 2403. Unfortunately, the diluted ma-N2403 was not very reliable. Over time the dosage requirement appears to increase, which may indicates that the solvent is evaporating. Thus, the film thickness would increase, and along with it, the dosage requirement.

A good solution is to use the A2 resist with a reversed-polarity pattern instead. In step (C.2) of Fig. 4.10, the exposed area of the resist form the negative-space between and around the NWs. After etching in step (C.3), Si NWs is then formed. However, since it is impractical to use *e*-beam to expose the entire area of the chip, a large amount of Si will still be present. (C.4) The sample is then coated again with photo-resist, and patterned such that all area surrounding the NWs are exposed, and etched in step (C.5), leaving isolated NWs array. The NWs may be checked with SEM between steps (C.3) and (C.4). If the NWs pattern is not as expected, the substrate can often be reused since the majority of top layer Si has not been etched, unlike in process (A) and (B), where the top layer is removed

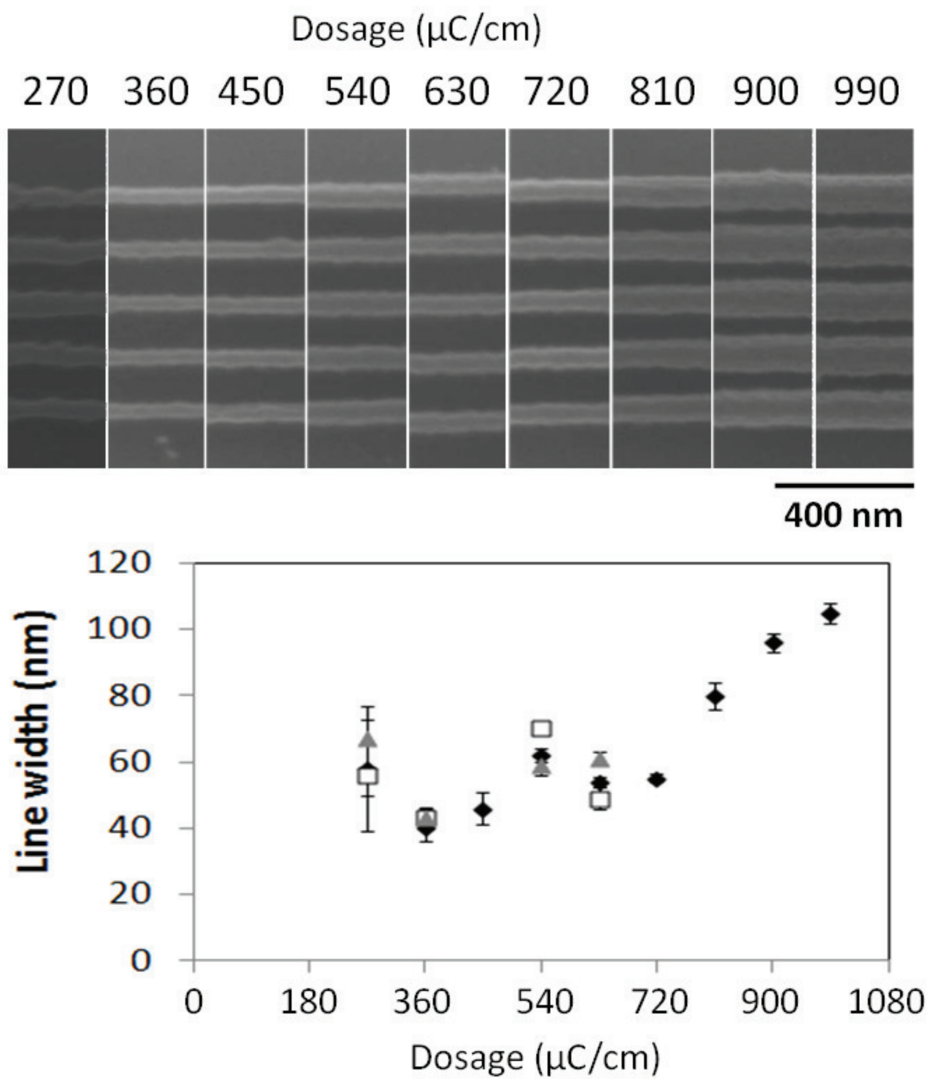


Figure 4.12: SEM images of nanowires patterned with ∼80 nm diluted ma-N 2403 resist over a range of dosage.

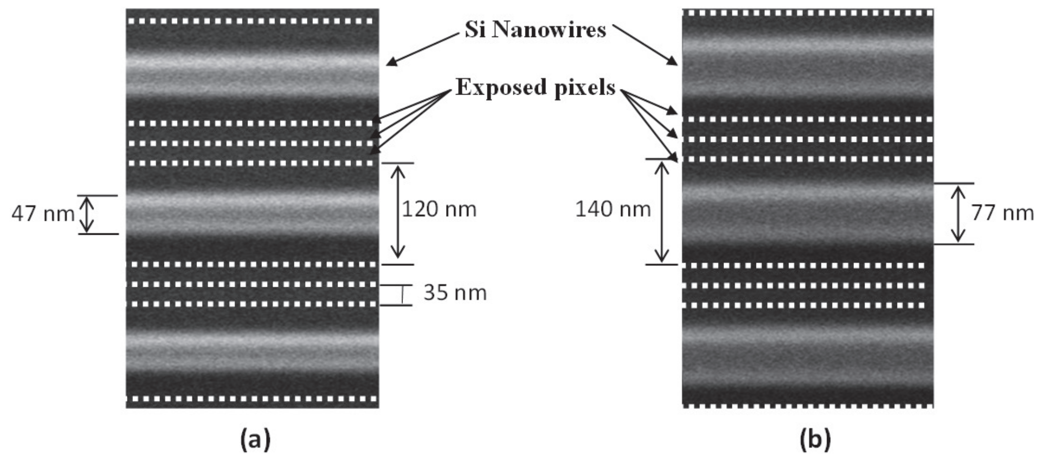


Figure 4.13: SEM images of Si nanowires fabricated with reverse-polarity pattern ((C) process in Fig. 4.10).

after step (A.5) and (B.3).

For the (C) process, the width of NWs is controlled by both the separation between the line of exposure that define the negative space, and the dosage that control the width of the line, as shown in Fig. 4.13. In subfigure (a), the negative space between NWs are defined by a group of three exposure lines spaced 35 nm apart. Each group is then separated by 120 nm pitch, leaving spaces for 47 nm-wide NWs. When the group pitch is increased by 20 nm to 140 nm, the space also increases, resulting in 77 nm wide NWs shown in subfigure (b). However, the increase in NWs width does not always match the increase group pitch. The line dosage for both figures is 1,350 μC/cm.

While all 3 processes were used to fabricate Si NWs at some time during the course of this work, all measurement data presented in the rest of this dissertation came from samples fabricated with the latest (C) process.

4.3.3 Pattern design

In most cases, the minimum width of 40 nm is targeted. However, deviation of tens of nanometers can result from loss of focus during exposure. The length of the NWs always ranges from 9 to 21 µm to match the separation of the drain and source (L_{ds}) electrodes. The number of wires are typically from 1 to 25. Having a high number of NWs increases the likelihood that the sample will be conducting, but at the same time requires longer SEM inspection time.

The type of electrical contact can also be specified. Figure. 4.14a shows an SEM image of an array of 12 NWs with metal contact deposited directly on top. The metal contact conforms around the NWs as shown in (a) and (b). This type of contact (deposited directly on NWs) mostly yield non-Ohmic IV curve, or have very large contact resistance in the range of MΩ. Despite annealing in the range of 300-650 °C, only three out of close to a hundred of such devices have yielded reliable ρ and S measurement.

Later samples were then patterned such that the Si NWs were left attached to bulk Si. Contacts can then be deposited onto this bulk contact area as shown in (c) and (d). The contact was slightly misaligned in these images, revealing that the NWs are indeed contiguous with the bulk contact area. Reliable Ohmic contact with negligible contact resistance is obtained with these bulk contacts. Ti/Ni/Au is used for these NWs samples as well. All measurement data for Si NWs shown in this dissertation come from samples with bulk contacts.

Ten sets of NWs arrays are typically patterned on a 7x7 mm^2 chip as two rows and five columns, as shown in Fig. 4.15a, to places each array as close to each other as possible while still leaving enough room for the various electrodes shown in Fig. 3.2. In practice, the sample is not parallel to the plane of focus placed at the center of the sample where w_d is adjusted, as shown by the drawing in Fig. 4.15b. Based on measurement of w_d over multiple EBL sessions, the average slope in the x-direction (for sample placed at the lower-right (LR) position of the

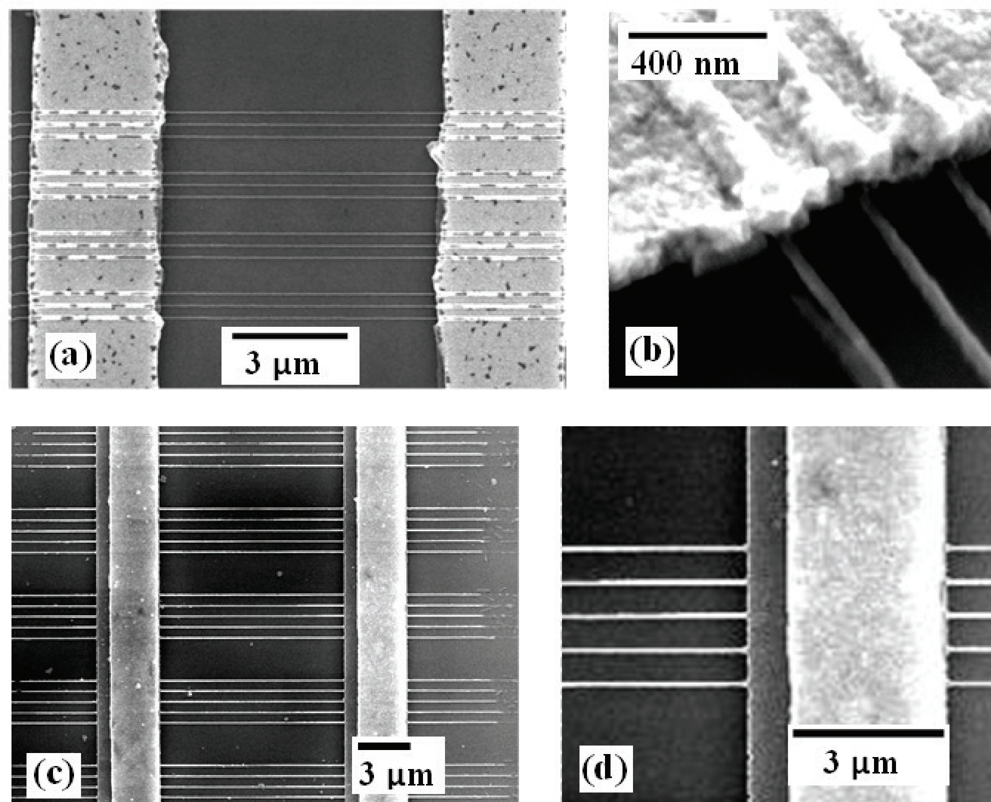


Figure 4.14: SEM images of Si nanowires array with (a),(b) direct contact on nanowires, and (c),(d) contact on bulk Si.

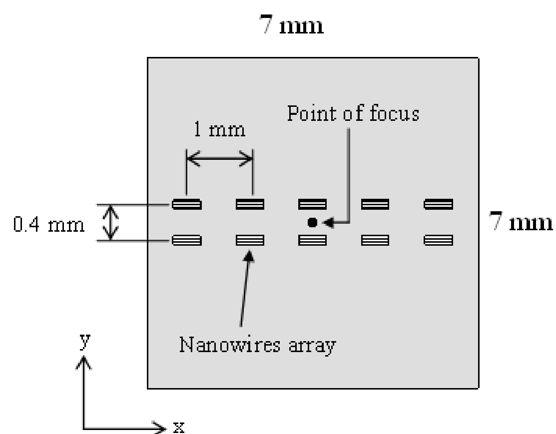
Raith50 sample holder) is +1.6 µm/mm. Therefore, the left (/right)-most NWs arrays would have been located 3.2 µm below (/above) the focal plane.

Figure 4.15c then shows variation in the average line width from 4 exposure at dosage of 1,350-2,070 µC/cm versus the approximate distance the point of focus at the center of the chip. The change in line width is non-symmetric, with loss of focus more severe when the sample is below the focal plane. This may be because electrons most strongly diverted right after they passed through the focal point, where they are most concentrated. It may help to lower the focal plane by focusing the beam on the lowest point of sample. The deviation between the top and bottom row is minimal since the travel in this direction is limited to 0.4 mm. The tilt in y-direction results in slope of approximately +0.4 µm/mm. It may also help to rearrange the NWs arrays along the y-axis instead, where the slope is smaller.

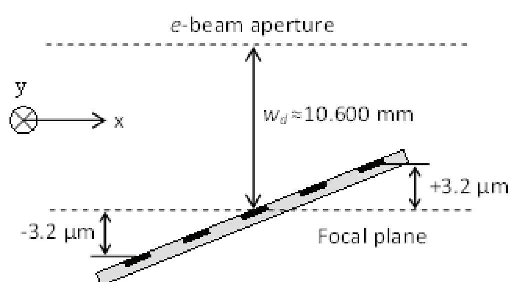
4.4 Oxidation of Si nanowires

4.4.1 Self-limiting oxidation effect

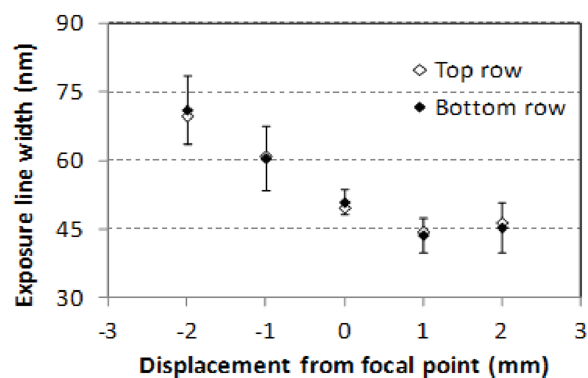
The minimum line width/nanowire diameter that could be achieved with EBL is 40 nm, which is still much larger than the 4-5 nm diameter required to observe quantum confinement effect. However, NWs fabricated by EBL can be shrunk to diameter below 5 nm with the use of "self-limiting oxidation" effect [15, 16, 71, 18]. Figure 4.16 shows the oxide thickness grown on planar Si, and vertical Si NWs with diameter of 30 nm, in dry oxidation at $T_{ox} = 850\,°C$ [15]. Compared to planar Si, the oxidation rate of Si NWs decreases rapidly to virtually zero after 5 h of oxidation, which then leaves a narrower Si core surrounded by oxide shell. The final core diameter depends on both the starting NWs diameter and the oxidation temperature [16], as shown in Fig. 4.17. The self-limiting oxidation effect offers a highly-controllable method to shrink the diameter of Si NWs, since the only parameter to control is the oxidation temperature.



(a) Top-down view of a single chip with 10 set of nanowires array



(b) Side-way view showing sample tilt during EBL



(c) Average line width for exposure with dosage of 1,350-2,070 µC/cm

Figure 4.15: NWs array arrangement on a chip, and the effect of tilt on line width.

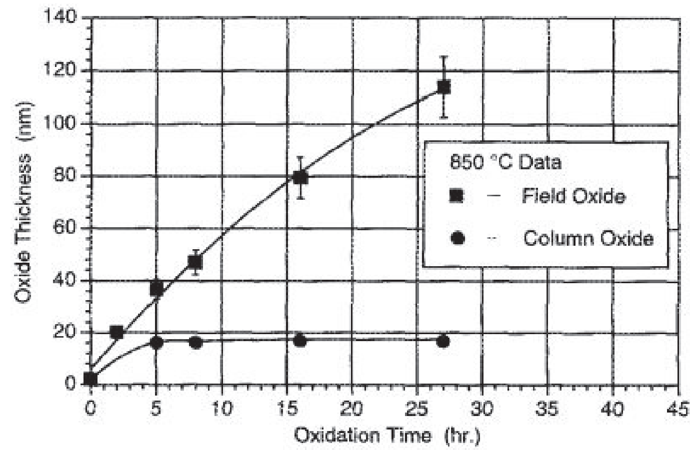


Figure 4.16: Oxide thickness versus oxidation time for planar Si, and Si nanowires. Adapted from [15].

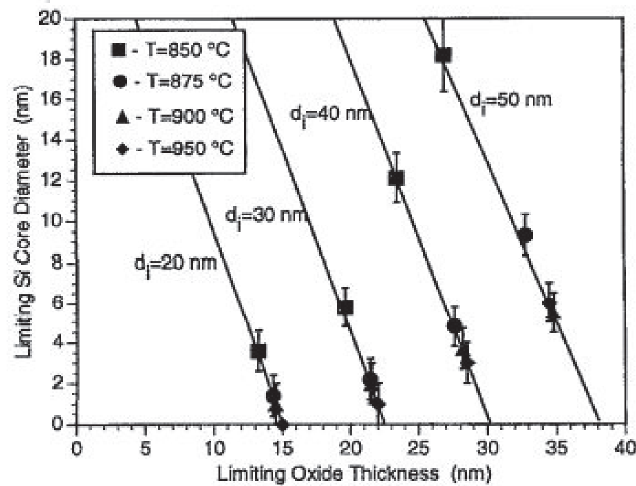


Figure 4.17: Limiting Si core width and oxide thickness of Si nanowires with different starting width, and oxidation temperature. Adapted from [16].

The self-limiting effect was explained as a result of increase in stress at the Si-SiO_2 interface [16]. Since the volume of SiO_2 is approximately 2.3 times larger than that of Si, as the new oxide grow at the Si-SiO_2 interface, the old oxide must expand. The expansion per unit volume of new oxide grown is greater as the curvature of the oxidizing surface increases from planar Si to Si NWs. The stress is therefore greater in Si NWs as well. As the stress increases, the diffusivity of oxygen is reduced and the oxidation rate is decreased. The stress is alleviated if T is increased, which allows oxide to flow better. All NWs were completely oxidized at $T > 950\,^{\circ}$C. Since $950\,^{\circ}$C is coincident with the viscoelastic flow transition temperature in planar oxidation, it was theorized that a change in oxide property may allow for continuation of the oxidation process [16].

However, subsequent work have observed a much reduced oxidation rate at higher T, and in wet oxygen as well [18]. Figure 4.18 shows a similar reduction in oxidation rate for NWs 50 nm wide square-cross section oxidized at 1,100 °C in dry oxygen for a short period of time in RTA [17]. This appears to contradicts the earlier observation that Si core always vanished for $T > 950\,^{\circ}$C. However, the oxidation time here is less than 20 min, whereas it was over several hours in the earlier work. It is likely that while the oxidation rate is drastically reduced, it simply has not truly reached zero. Therefore, the Si core would eventually be completely oxidized over a long period of time.

Interestingly, the limiting mechanism for the high-T oxidation was attributed to a reduced reaction rate [17], whereas the earlier work for low-T oxidation was attributed to a reduced diffusion rate [16]. Figure 4.19a shows that the core of NWs still remain rectangular after 100 s of oxidation at 1,100 °C. The reaction rate at these surfaces remains similar to that in bulk Si. However, as the NWs become more rounded (subfigure (b) and (c)) due to oxide flow, the reactive rate is decreased due to a combination of increased stress at the surface, and a change in crytal orientation at the oxidizing surface [17].

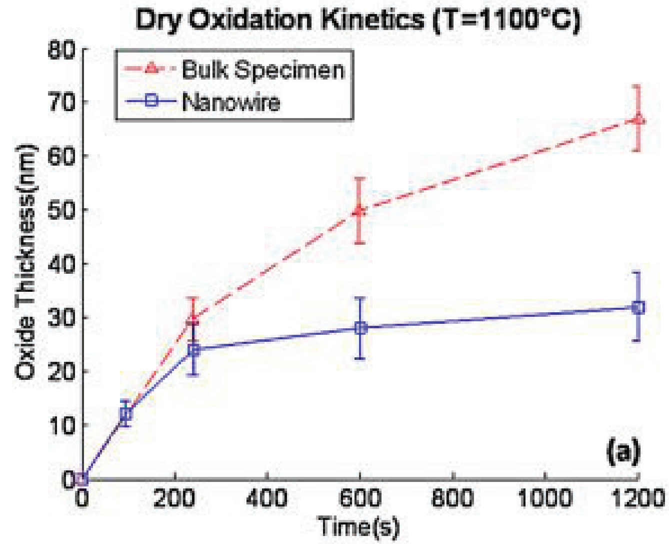


Figure 4.18: Oxidation rate of planar Si and Si nanowires at 1,100 °C. Adapted from Ref. [17]

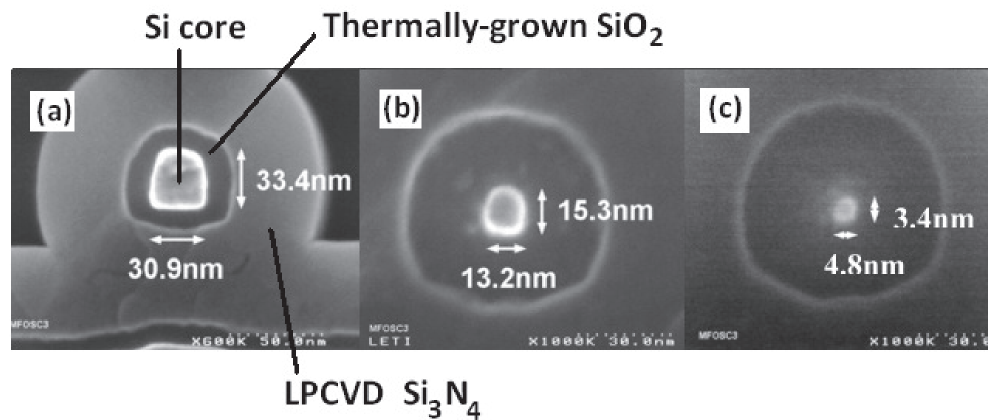


Figure 4.19: TEM images showing Si/SiO_2 core/shell nanowires oxidized at 1,100 °C for (a) 100 s, (b) 250 s, and (c) 600 s. Si core (bright area) is surrounded by the thermally-grown oxide shell (dark gray), and the deposited Si_2N_3 encapsulation layer (medium gray). Adapted from Ref. [18].

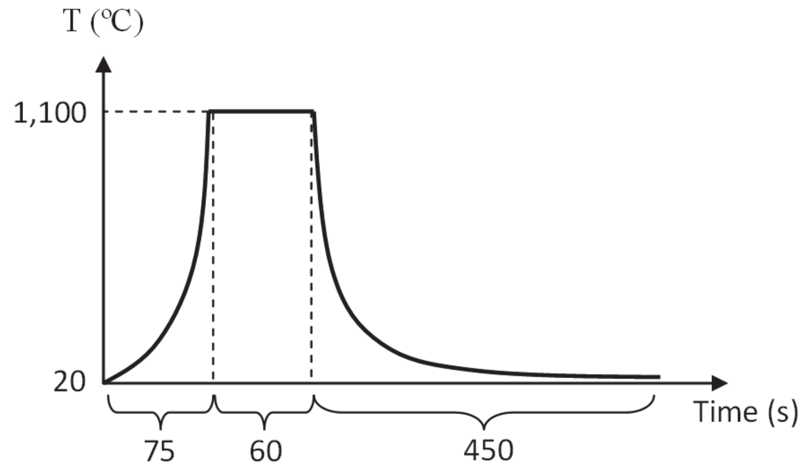


Figure 4.20: Timing profile for oxidation in RTA furnace.

4.4.2 Oxidation rate of planar Si and Si nanowires

The oxidation rate of planar Si in RTA oven is measured first. Since the RTA is limited to a maximum of 1 min hold time at 1,100 °C, samples were oxidized over multiple cycles, i.e. "12 min" oxidation is equivalent to oxidation over twelve 1-minute cycles. Figure 4.20 shows the profile of a standard heating cycle with approximately 75 s ramp up time from room temperature to 1,100 °C, 60 s holding period, and then a gradual cool down over the next 450 s. Dry oxygen is supplied at 1.75 L/min for the entire cycle, though oxidation during ramping times should be insignificant as oxidation rate drops exponentially with temperature. The thicknesses of the oxide and the remaining Si is measured with either the Filmetrics F20, or the Dektak 150 profilometer.

Figure 4.21 shows the oxide growth at 1,100 °C for 3 types of Si thin film: the as-received undoped SOI ($d_{Si} = 260\,\text{nm}$), and doped N34 and N35 substrates ($d_{Si} = 40\,\text{nm}$). The oxidation rate for the undoped SOI and the doped N34 substrate are similar. Comparison to the standard Deal-Grove model [72] would suggest, however, that the oxidation temperature is 1,200 °C rather than the nominal value of 1,100 °C. A discrepancy of 100 °C may be possible since the thermocouple

is not cemented to the sacrificial Si wafer that hold the sample, and it is placed underneath the wafer rather than on top with the sample. Additionally, although the thermocouple operating range extends beyond 1,100 °C, a pyrometer is the recommended temperature sensor for $T > 400\,^{\circ}$C. Since oxidation does not depends on flow rate, and the ratio of consumed Si over the grown oxide is close to the conventional value of 0.44, which suggests normal, stoichiometric oxide formation, the error between the true and the nominal temperature best explained the observed oxidation rate. The oxidation rate of bulk Si in RTA in Fig. 4.18 is also very close to our data, despite the nominal oxidation temperature of 1,100 °C there as well.

Oxidation appears to be reaction rate-controlled for our samples. At temperature of 1,100 and 1,200 °C, the Deal-Grove model predicts a transition to the parabolic diffusion rate-controlled regime after approximately 10 and 2 min, respectively; calculated as $t = \frac{A^2}{4B} + \tau$, where the values for parabolic rate constant B, the linear rate constant B/A, and τ can be found in Ref. [72]. Nevertheless, oxide growth on all samples still appears to be linear, indicating reaction rate-controlled regime. Greater oxidation rate observed for N35 samples can also be explained if oxidation is reaction rate-controlled. During oxidation of P-doped Si, P atoms are rejected from the growing oxide since diffusivity of P in SiO_2 is five order of magnitude lower than that in Si [70, 67]. Reaction rate may then increased once [P] at the oxidizing interface exceeds $10^{20}\,\mathrm{cm}^{-3}$ [73]. According to Table 4.3, n in N35 indicates that [P]$\geq 2.7\times10^{20}\,\mathrm{cm}^{-3}$, which is high enough for the reaction rate to be increased. On the other hand, [P] in N34 is an order of magnitude lower, and so its oxidation rate should be indifferent from that of undoped Si. If oxidation was diffusion-controlled, then the oxidation rate should be the same for all samples regardless of [P]. Dopant segregation is further discussed in the next section.

Oxidation of Si NWs were also studied with SEM. Figure 4.22a and b shows SEM images of NWs taken before and after 9 min dry oxidation at 1,100 °C in the RTA. A thin layer of Au, estimated at 5 nm, was deposited by sputtering to prevent charging effect during SEM imaging. The NWs were fabricated from

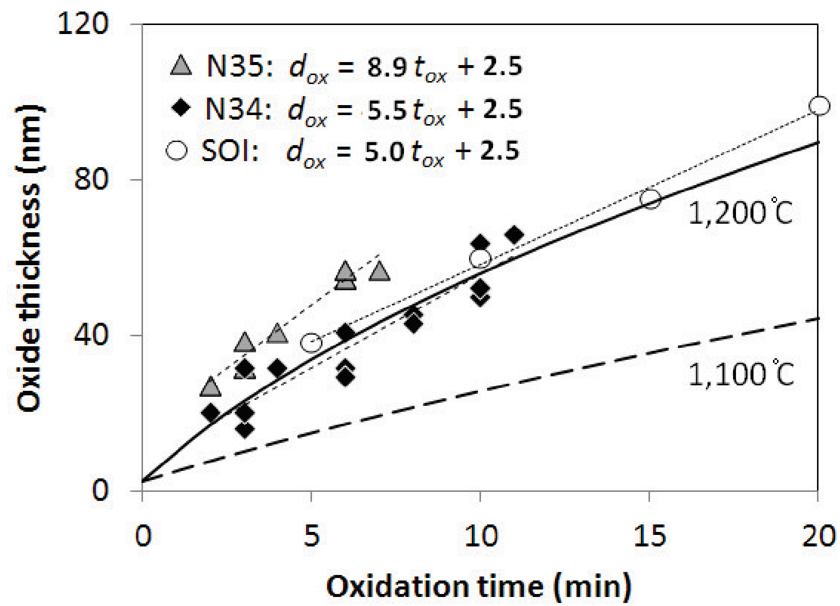


Figure 4.21: Oxidation of Si thin film (undoped SOI, N34, and N35) in RTA at 1,100 °C in dry oxygen.

thinned ($d_{Si} = 40\,nm$), but undoped SOI using EBL (C) process (see section 4.3.2). Oxide thickness is then calculated from the difference in width as

$$d_{ox} = \frac{w' - w_0 - 5}{2}\left(\frac{1}{1 - 0.44}\right) \tag{4.4}$$

, where w_0 and w' is the width of the NWs before and after oxidation, respectively. Calculation result for four set of NWs with w_0 of 62-77 nm are then shown in Fig. 4.22c, where oxidation rate appears to be lower for the NWs compare to planar Si. Additionally, it appears that oxidation rate decreases with w_0, which may indicates a self-limiting effect, though a more accurate measurement such as cross-section TEM would be needed to conclusively establish the dimension of the oxide shell and Si core.

4.4.3 Dopant segregation

The effect of dopant segregation *during thermal oxidation* is well known in bulk Si [19, 73, 14, 34], but has not been extensively studied in Si NWs. While the self-limiting oxidation have become an important part of the toolkit for the fabrication of ultra-thin Si NWs with diameter below 5 nm, it has mainly been studied in Si NWs with low dopant concentration [15, 16, 71, 18, 17], or used in the fabrication of FET with depleted channel [74, 75, 64]. The relevance of dopant segregation was only later considered, where one study has shown that B and P are indeed absorbed and rejected by the growing oxide during oxidation of Si NWs [76]. Since Si NWs in our work is heavily-doped, we then need to take into account how dopant segregation may affect n as oxidation proceed.

Dopant segregation during thermal oxidation is illustrated in Fig. 4.23. For impurity such as B that diffuses faster in oxide, they tend to be absorbed into the growing oxide layer. Therefore, the impurity concentration at Si-SiO_2 interface is increased in the oxide (C_{ox}), but reduced in Si (C_{Si}). On the contrary, impurity such as P or As that has low diffusivity in oxide tend to accumulate in Si, resulting in lower C_{ox} and higher C_{Si}. A segregation coefficient for thermal oxidation is

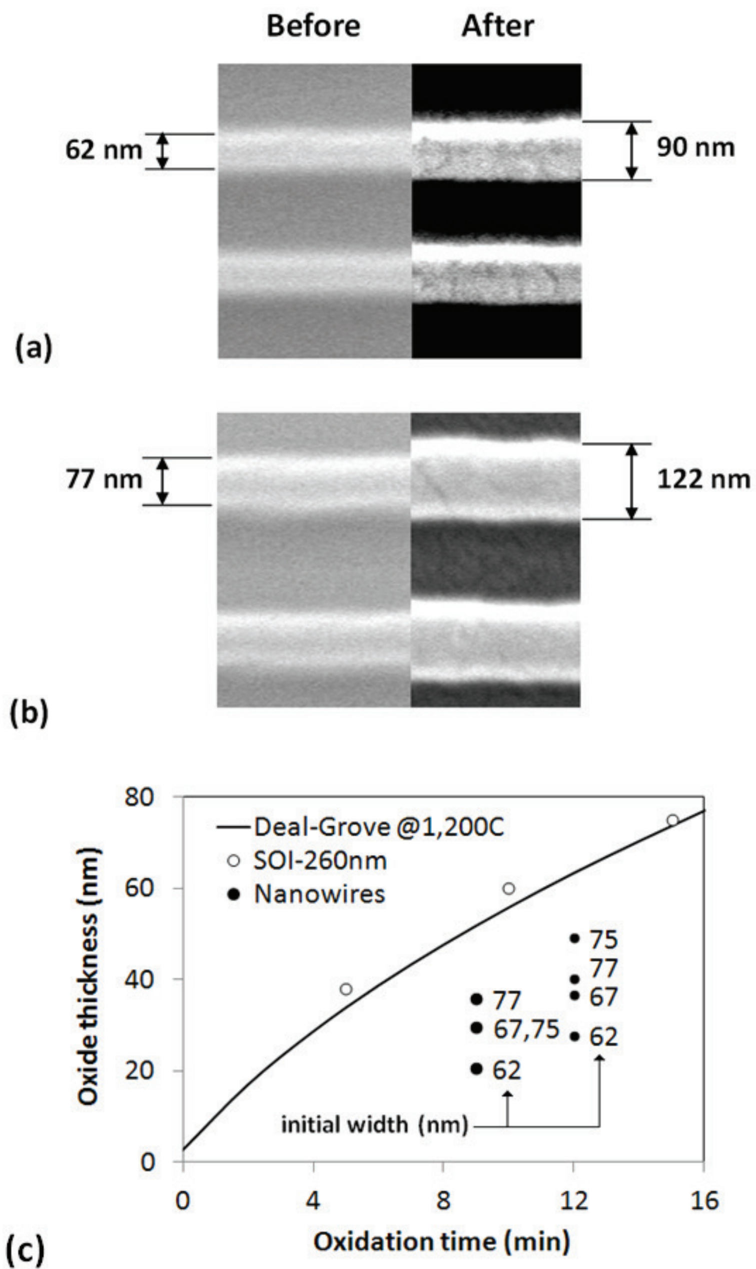


Figure 4.22: Oxide thickness of planar Si and Si NWs after oxidation in RTA at 1,100 °C in dry oxygen.

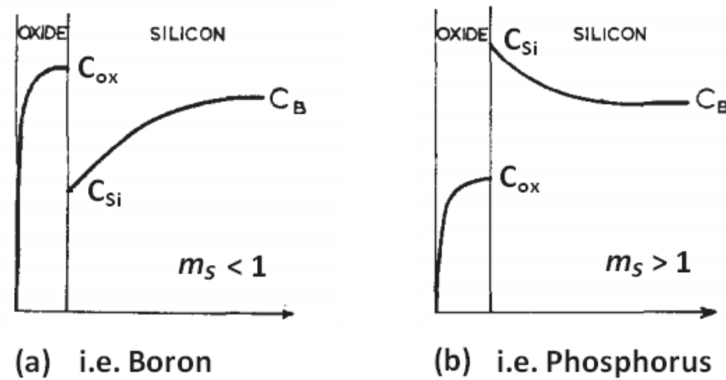


Figure 4.23: Dopant segregation where impurity is (a) absorbed by the oxide and depleted from Si, and (b) rejected by the oxide and piled up in Si. Adapted from Ref. [19].

conventionally defined as

$$m_s = \frac{C_{Si}}{C_{ox}} \tag{4.5}$$

Impurity is absorbed by the oxide when $m_s < 1$, and is rejected when $m_s > 1$. m_s is approximately 0.3 for B, and 10 for P, As, and Sb [19].

The evolution of dopant concentration during oxidation in planar Si may then be approximated with a simple conservation of the dopant area density. Figure 4.24(a) shows the situation prior to oxidation, where P is uniformly distributed with initial concentration of C_0 in Si with initial thickness d_0. After oxidation over a period of time at some temperature, as shown in subfigure (b), oxide of thickness of d_{ox} is grown, and Si thickness is reduced by the amount consumed by oxidation, i.e.

$$d_{Si} = d_0 - 0.44\, d_{ox} \tag{4.6}$$

, assuming standard volume of Si and SiO_2. C_{Si} and C_{ox} are shown to be greater and lower than C_0, though it does not matter in our calculation whether this is true. The only important relationship here is the one in equation 4.5, i.e.

$$C_{ox} = C_{Si}/m_s$$

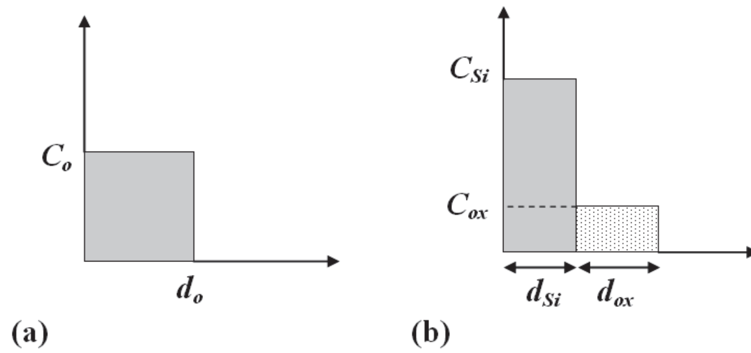


Figure 4.24: P concentration versus Si/SiO_2 thicknesses (a) before, and (b) after oxidation in planar Si.

Since it was shown in section 4.2.2 that $L_D \approx d_{Si}$ at 1,100 °C annealing for 1 min, we may assume a constant C_{Si}. C_{ox} is also assumed to be a constant for convenience. Assuming a negligible loss of P through diffusion from Si into the underlying BOX layer, and from out-gassing of P from the top oxide into the air, the conservation of mass then dictates that

$$C_0\, d_{Si,0} = C_{Si}\, dsi + C_{ox}\, d_{ox} \tag{4.7}$$

We defined a coefficient, "segregation ratio", to relate the concentration of P *in Si after and before* oxidation as

$$k_s = \frac{C_{Si}}{C_0} = \frac{0.44\, m_s\, d_0}{d_0 + (0.44 m_s - 1) d_{Si}} = \frac{d_0}{d_0 + (\frac{1}{m_s} - 0.44) d_{ox}} \tag{4.8}$$

If the oxide growth rate is known, then k_s can be expressed as a function of oxidation time as well.

The segregation ratio can also be calculated for a NWs geometry using the same conservation of mass approach. Figure 4.25(a) and (b) shows an idealized cross-section of rectangular Si NWs resting on the BOX layer before and after oxidation, respectively. Oxidation rate is assumed to be the same on all sides. After a period of oxidation, the height and width of the remaining Si core (d_{Si} and

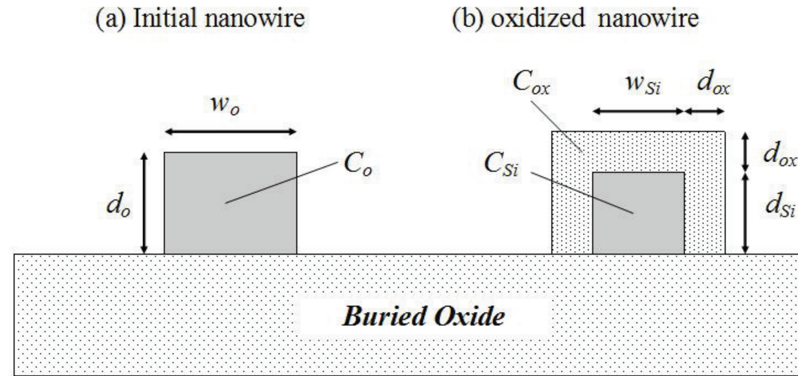


Figure 4.25: Idealized cross-section of rectangular nanowire on BOX layer (a) before, and (b) after oxidation.

w_{Si}) can be related to the oxide thickness as

$$d_{Si} = d_0 - 0.44\, d_{ox} \tag{4.9}$$

$$w_{Si} = w_0 - 2 \times 0.44\, d_{ox} \tag{4.10}$$

Assuming that all initial P atoms in Si NWs remains in either Si core or SiO_2 shell, the *line* density of P before and after oxidation are

$$C_0 d_0 w_0 = C_{Si} d_{Si} w_{Si} + C_{ox} d_{ox} (w_{Si} + 2d_{Si} + 2d_{ox}) \tag{4.11}$$

The segregation ratio for NWs geometry is then be expressed as a function of oxide thickness by substituting in equations 4.5, 4.9, and 4.10, which yield

$$k_s = \frac{C_{Si}}{C_o} = \frac{d_0}{d_0 + (1 + 2\frac{d_0}{w_0})\left(\frac{1}{m_s} - 0.44\right) d_{ox} + \frac{2}{w_0}\left(0.44^2 + \frac{1-0.88}{m_s}\right) d_{ox}^2} \tag{4.12}$$

If $w_0 = +\infty$, then this equation reverts to equation 4.8 for planar Si. Therefore, this equation is consider a more generalized equation than the previous one.

Figure 4.26 compares the segregation ratio for of planar Si, and Si NWs. First, k_s of planar Si is plotted for m_s of 0.3 (i.e. B), 1, and 10 (i.e. P). For $m_s = 0.3$, C_{Si} is expectedly reduced since B atoms is taken from Si. C_{Si} is reduced when $m_s = 1$ as well, which occurs because of an increase in total volume due

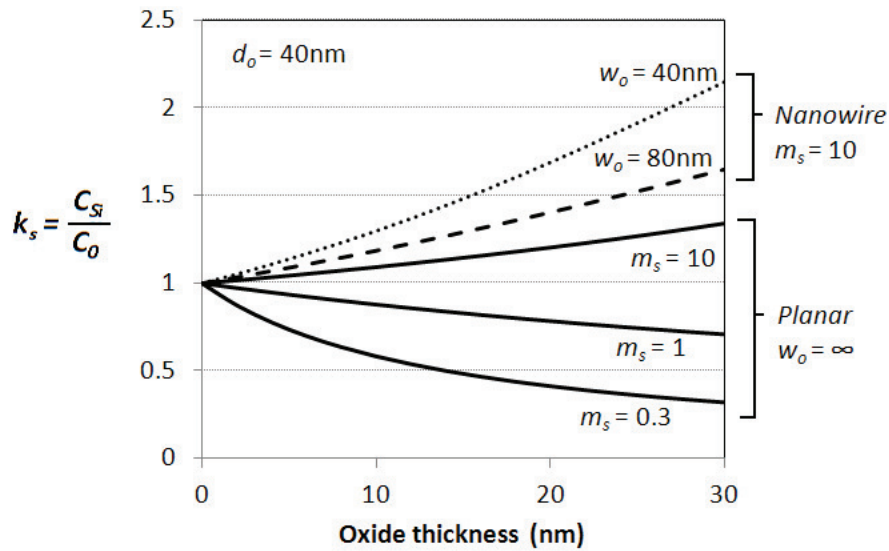


Figure 4.26: Segregation ratio (of dopant concentration in Si after and before oxidation) for planar and nanowires geometry.

to expansion of the oxide. k_s would equal 1 only when the segregation coefficient matches the volumetric ratio of Si and SiO_2, i.e $m_s = 1/0.44 = 2.\dot{2}\dot{7}$. For $m_s = 10$, C_{Si} should increase as the volume of the remaining Si is decreased. The segregation ratio is even larger in NWs than in planar geometry, and is further increased with reducing w_0. d_0 is assumed to be 40 nm in all cases. In general, the dopant segregation will become more significant as the surface-to-volume ratio increases.

4.5 Measurement results

4.5.1 FET measurement of thin film

Figure 4.27 shows the $G_c h V_g$ curves of 4 TF-FET devices, 2 from the same N34 chip ("D3" and "D4") and 2 from the same N35 chip ("d4" and "d9"). Both samples were fabricated at the same time with the same processes. Gate oxide was deposited by PECVD with thickness of d_{ox} = 60 nm. All samples behave

consistently with n-type depletion-mode MOSFET, i.e. positive slope and negative V_t. Since these devices are heavily doped n-type throughout the channel, drain, and source regions, there are no current-blocking n-p junctions and the device behaves as Ohmic resistor at $V_g = 0$. The electron concentration in the channel is nevertheless influenced by V_g. When $V_g > 0$, G_{ch} is increased as electrons are attracted to the channel. Conversely, when $V_g < 0$, G_{ch} is reduced as electrons are repelled, though it would take a very large negative $V_g = V_t$ to repel all electrons and turn off these devices. For instance, V_t of N34-D3 and N35-d4 devices are estimated from the linear fit to be

$$V_t = \text{x-intercept} = -\frac{\text{y-intercept}}{\text{slope}} = \frac{-2,091}{8.1}\frac{\text{S}}{\mu\text{S/V}} = -258\,\text{V}$$

, and $V_t = -\frac{3,335}{2.6}\text{V} = -1,283\,V$, respectively.

The hysteresis loop observed in Fig. 4.27 is found in many of our FET devices, and could be caused by presence of mobile ionic charge. Mobile ions, such as Na+ or K+, that are present in gate oxide can introduced hysteresis loop in $C_{ox}(V_g)$ curve [34]. Since $G_{ch} \propto I_d \propto C_{ox}$, the hysteresis in $G_{ch}(V_g)$ curve is likely to be of the same origin. Samples are usually taken out of the cleanroom for initial measurement before gate oxide is deposited. During this time, contaminant ions such as Na+ could be introduced from tweezers or sample boxes that were not handled with gloves. The effect of mobile ions may be mitigated by sweeping V_g at a high rate such that these ions cannot respond fast enough. A minimum sweep rate of approximately 5 V/s were used, though hysteresis was still observed. Faster sweep rate may be necessary. All hysteresis loops follows counter-clockwise sweep, where larger G_{ch} is observed when V_g is decreasing. As it stands, the effect of hysteresis on subsequent calculation of electrical property is minimized by taking a linear fit of a fully-swept $G_{ch}(V_g)$ curve, i.e. where V_g is swept from 0V to +5V, then to -5V, and back to 0V. The sweep is typically completed for 2-3 cycles, since gate oxide broke down on devices whose gate were cycled continuously for more than 5-6 cycles.

ρ, μ, and n are then calculated from the slope and intercept of the linear fit to the $G_{ch}(V_g)$ curve according to equations 3.30-3.32. For example, we calculate

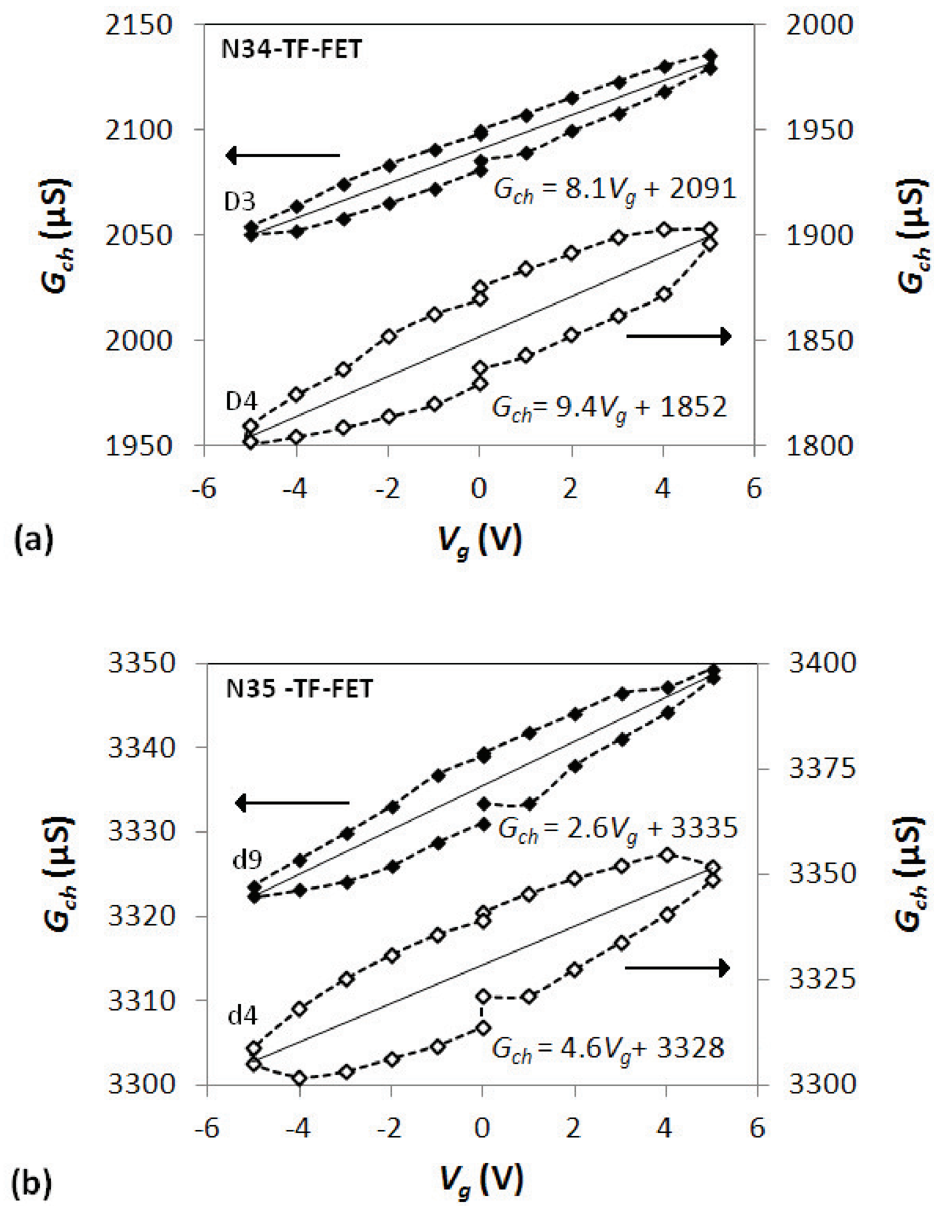


Figure 4.27: Channel conductance versus gate voltage of (a) N34 and (b) N35 TF-FET devices with 60 nm gate oxide. $V_d = 50\,\text{mV}$.

for the N34-D3 device

$$\rho = \frac{W d_{Si}}{G_{ch}(0) L_g} = \frac{10 \times 40}{2,091 \times 14} \frac{\mu\text{m} \cdot \text{nm}}{\mu\text{S} \cdot \text{nm}} = 1.37\,\text{m}\Omega.\text{cm} \tag{4.13}$$

$$\mu = \frac{\Delta G_{ch}}{\Delta V_g} \left(\frac{L_g d_{ox}}{W \epsilon_{ox} \epsilon_0} \right) = \frac{8.14 \times 14 \times 60}{10 \times 7.30 \times 8.85 \times 10^{-12}} \frac{\mu\text{S/V} \cdot \mu\text{m} \cdot \text{nm}}{\mu\text{m} \cdot \text{F/m}}$$

$$= 106\,\text{cm}^2/\text{Vs} \tag{4.14}$$

$$n = \frac{1}{\rho e \mu} = \frac{|V_t| \epsilon_{ox} \epsilon_o}{e\, d_{Si} d_{ox}} = 4.3 \times 10^{19}\,\text{cm}^{-3} \tag{4.15}$$

Note that ρ is calculated with L_g and not L_{ds} since G_{ch} is the conductance of the channel only. The results of the same calculations for other TF-FET devices are shown in Fig. 4.28a. Since all devices are from the same chips as N34-D3 and N35-d3, $d_{ox} = 60\,\text{nm}$ for all. W is also the same for all, but L_g may differs for each devices. Note that D4(/d4) and D3(/d9) devices exhibit the highest and lowest μ, respectively, among all N34(/N35) devices shown in the figure. Compared to the previous VDP measurement in Fig. 4.7, the FET method appears to systematically overestimate μ of N34 and N35 devices that are circled in the figure.

The measurement uncertainty for FET method is larger than that of VDP method since all of the FET device geometry must be specified. W and L_g were assumed from the nominal value defined in the photo-mask. However, the actual values may increase if the photo-resist was over-exposed/developed, and vice versa. Assuming the same uncertainty of 7.5% for W, L_g, and d_{ox}, the latter of which is measured in similar fashion to d_{Si}, the total measurement uncertainties are

$$\frac{\delta\rho}{\rho} = \sqrt{\left(\frac{\delta d_{Si}}{d_{Si}}\right)^2 + \left(\frac{\delta W}{W}\right)^2 + \left(\frac{\delta L}{L}\right)^2} \approx 13\%$$

$$\frac{\delta\mu}{\mu} = \sqrt{\left(\frac{\delta W}{W}\right)^2 + \left(\frac{\delta L}{L}\right)^2 + \left(\frac{\delta d_{ox}}{d_{ox}}\right)^2} \approx 13\%$$

$$\frac{\delta n}{n} = \sqrt{\left(\frac{\delta d_{Si}}{d_{Si}}\right)^2 + \left(\frac{\delta d_{ox}}{d_{ox}}\right)^2} \approx 11\%$$

The error bars then show that the discrepancy observed still cannot be explained by error in geometrical factors alone.

Considering S measurement in Fig. 4.28b, it appears that n is also underestimated. While μ and n requires assumption of many parameters, S is more likely to be accurate since it is directly measured. Furthermore, despite slightly different geometry between VDP and FET devices, S are measured in similar fashion and its accuracy should be the same. Assuming that S measurements of FET devices are accurate, then there would be a better agreement with VDP devices and calculation values if the devices circled in the figure are shifted to the right.

The concurrent overestimation of μ and underestimation of n then suggest that there is an error in the value of $\frac{\Delta G}{\Delta V_g}$ and C_{ox}, as they are the two common factors in equation 3.31, and 3.32. For example, if C_{ox} is increased, then n is increased as well, while μ is decreased. It is unlikely that ϵ_{ox} is underestimated since its current value of 7.3 is already much higher than the conventional value of 3.9. It may be possible that d_{ox} is overestimated, perhaps by uneven deposition of the gate oxide.

Conversely, if $\frac{\Delta G_{ch}}{\Delta V_g}$ is reduced, then n would increase and μ would decrease. Although measurement of I_d, V_g, and V_d should be accurate and precise, the subsequent assumption that $L_{ch} = L_g$ could be a source of error when determining the R_{ds} and $G_{ch}(V_g)$ curve. For example, if $L_{ch} > L_g$, then both R_{ds} and $\frac{\Delta G_{ch}}{\Delta V_g}$ would be decreased. R_{ds} and L_{ch} can be measured directly through "channel-resistance" or "shift-and-ratio" methods outlined in Ref. [58]. Both methods requires additional fabrication of test devices for comparison, and was not done.

Figure 4.29 shows similar electrical measurement results, but for TF-FET devices with 40 nm gate oxide. In subfigure (a), μ of N34 TF-FET devices are comparable to that of VDP devices, and follow the expected value as calculated by equation 4.3. Although these devices showed larger n than all other N34 VDP and TF-FET devices measured before, they also exhibited smaller $|S|$ in agreement with the calculation in subfigure (b). Better agreement of μ is also obtained for N35 samples, though a few samples still appear to show the same over- and

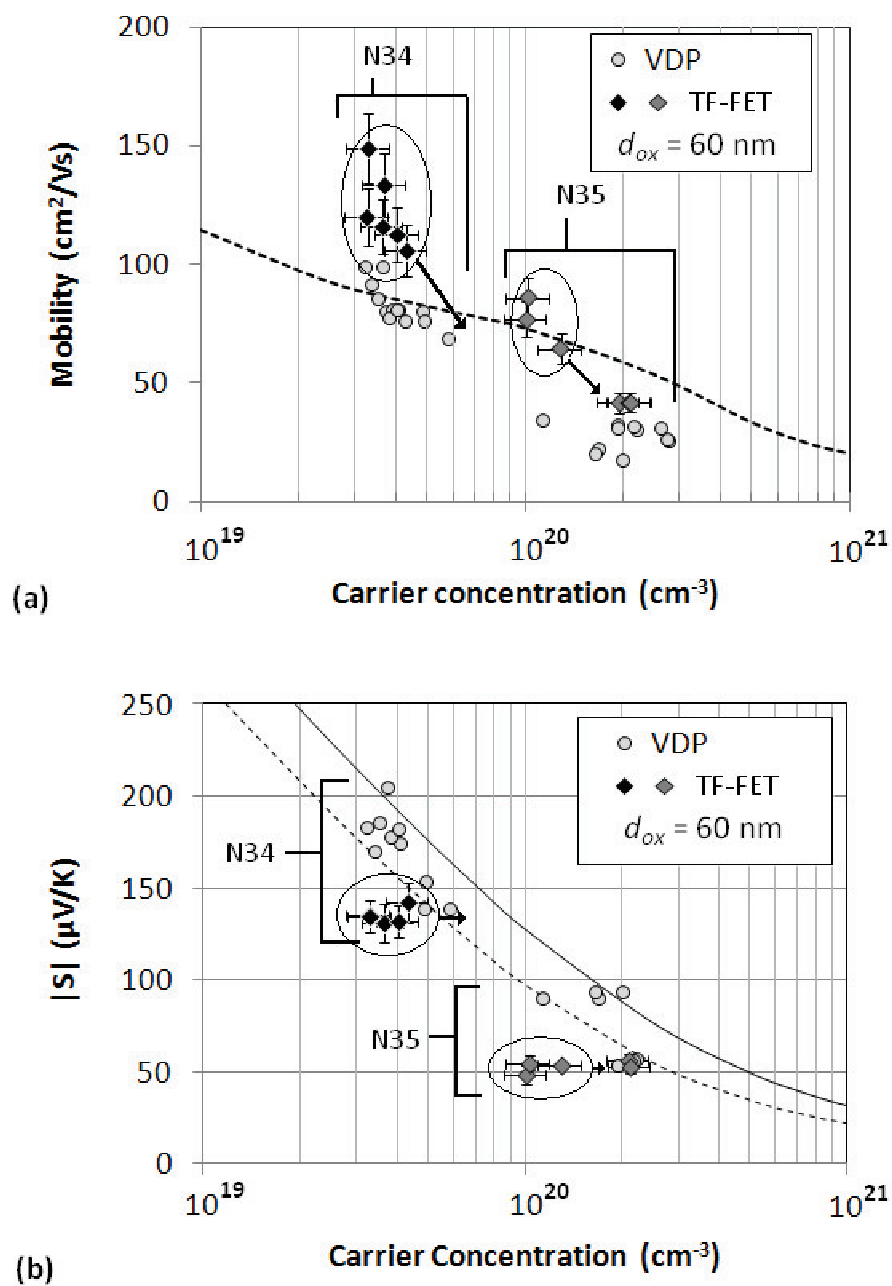


Figure 4.28: Electrical properties of Si TF-FET with 60 nm gate oxide.

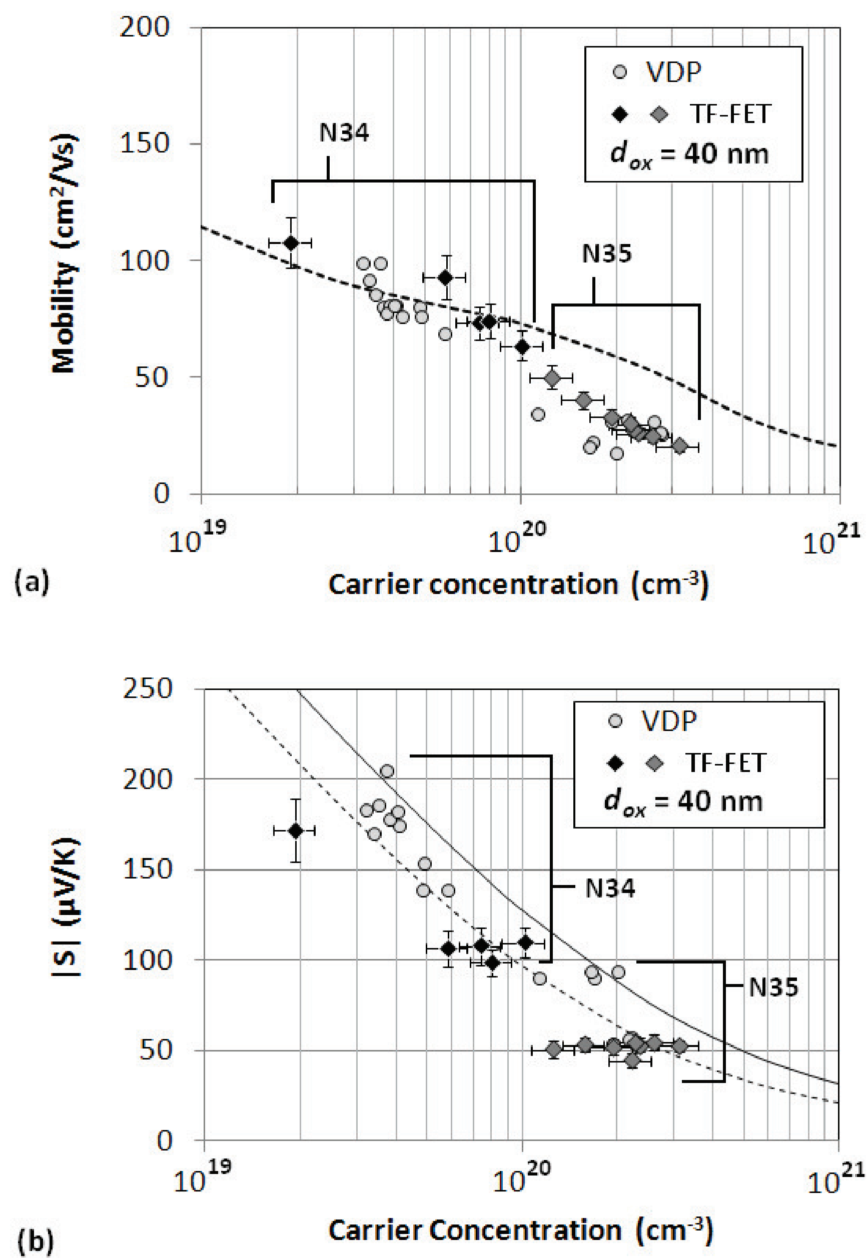


Figure 4.29: Electrical properties of Si TF-FET with 40 nm gate oxide.

under-estimation of μ and n, respectively.

4.5.2 FET measurement of Si Nanowires

After the completion of the EBL process shown in Fig. 4.10, Si NWs MOSFET (shorted to NW-FET from here on) devices were fabricated in the same manners as TF-FET devices. Right away, it is noticed that the deposition of gate metal increased ρ and decreased $|S|$ for all NW-FET devices, as indicated by measurements before and after gate deposition shown in Fig. 4.30. Prior to gate deposition, $|S|$ for both N34 and N35 NW-FET were smaller and consistent with those measured on TF-FET. An increase in S and ρ is not observed in TF-FET, e.g. Fig. 4.31. The change in n may be due to band banding by the floating gate potential. For instance, if the floating gate potential is positive, then electrons in the channel of the NW-FET would be repelled, resulting in decreased n and ρ, and increased $|S|$. The coupling between the channel and gate may be more efficient in NW-FET than in TF-FET because of increased effective gate area, which will be discussed later on.

After the completion of NW-FET devices, the same measurement method is applied to measure $I_d - V_g$. Figure 4.32 shows the $G_{ch}(V_g)$ curve for three N34 NW-FET devices with L_g of 3, 11, and 15 µm. All devices have 40 nm-thick gate oxide, and contains $N_w = 10$ NWs with an average NWs width (w) of 40, and 70 nm. The $G_{ch}(V_g)$ curve of NW-FET appears similar to that of TF-FET, with positive slope and large hysteresis loop. Electrical properties, including V_t, ρ, μ, and n, are again extracted from the best linear fit.

The slope and intercept of NW-FET $G_{ch}(V_g)$ curve reveals a significantly smaller $|V_t|$ than that of TF-FET. Whereas the average V_t of N34 and N35 TF-FET devices with 40 nm gate oxide shown in Fig. 4.29a were -364 and -858 V, respectively, V_t of these N34 and N35 NW-FET devices were approximately -25 V and -88 V, respectively. Considering that $|V_t| = \frac{ne\,d_{Si}}{C_{ox}}$ (see equation 3.32), this indicates

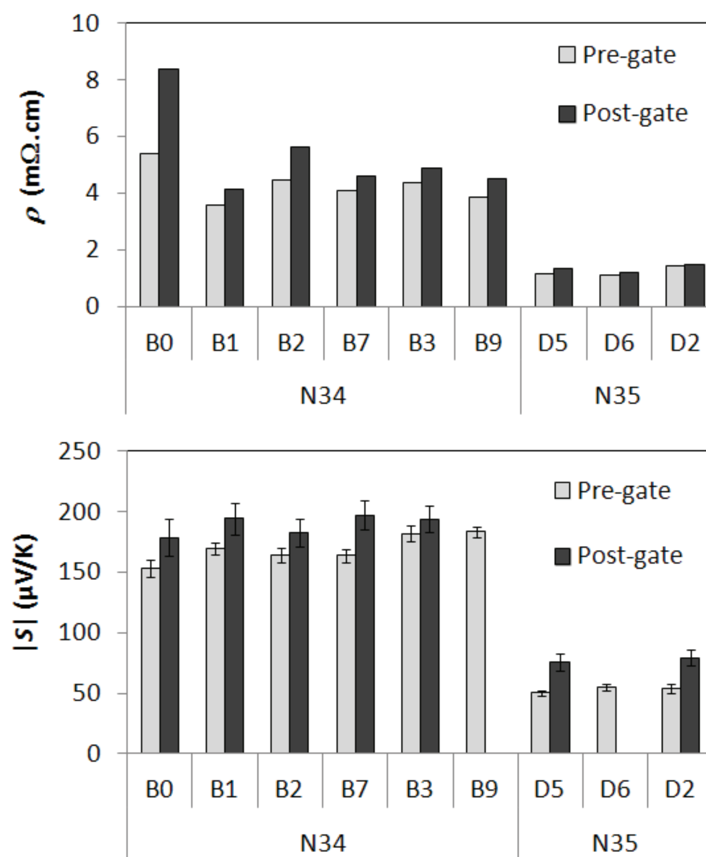


Figure 4.30: Resistivity and Seebeck coefficient of Si NW-FET before and after gate deposition.

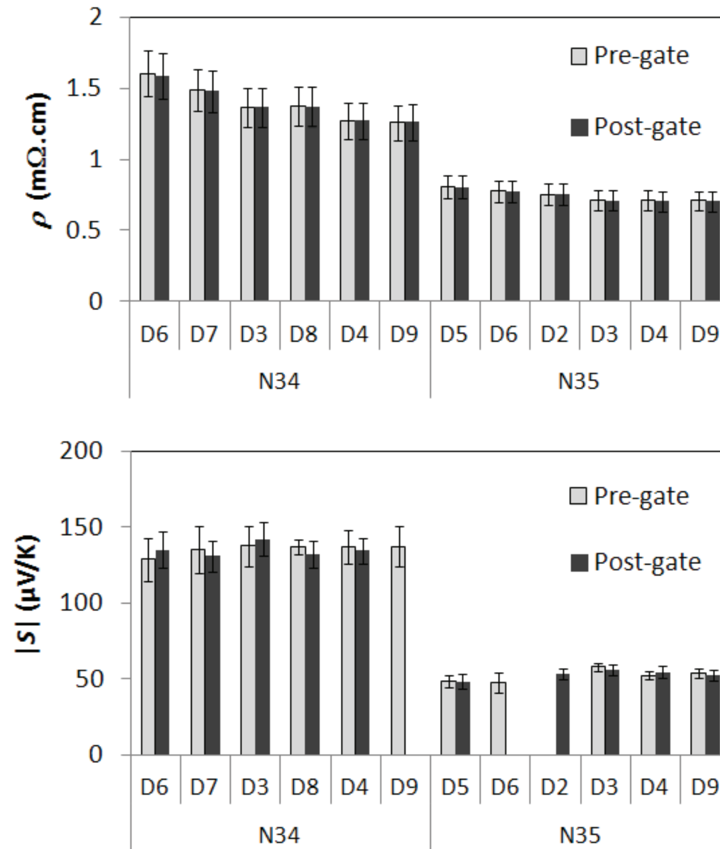


Figure 4.31: Resistivity and Seebeck coefficient of Si TF-FET before and after gate deposition.

that either that n must be smaller and/or C_{ox} must be larger in NW-FET than TF-FET. n may decrease significantly in NWs by a dopant deactivation effect, whereby charge carriers are trapped by surface states. However, this effect is only noticeable below approximately 40, 25, and 10 nm for NWs with $N_D = 9 \times 10^{18}$, 3×10^{19}, and $1.5 \times 10^{20}\,\text{cm}^{-3}$ [65], respectively. Given that $N_D > 4 \times 10^{19}\,\text{cm}^{-3}$ and $w > 40\,\text{nm}$ for our NWs, dopant deactivation should be irrelevant. Therefore, the observed increase in V_t is most likely due to increase in C_{ox}.

The increase in C_{ox} is possible with the increase in effective gate area. Unlike

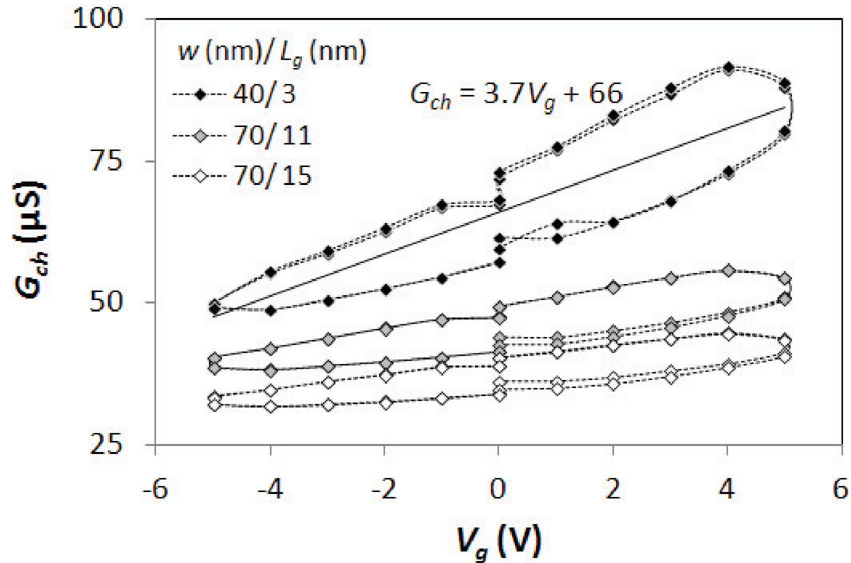


Figure 4.32: Channel conductance versus gate voltage of N34 NW-FET devices.

the TF-FET that have planar gate, the gate on the NW-FET wraps around the two sides of the NWs as well. In similar structure such as the FinFET [77] or the Gate-All-Around (GAA) FET [75], the increase in gate area is matched by an increase in C_{ox}, which may then be approximated by

$$C_{ox} = N_g \left(\frac{\epsilon_{ox}\epsilon_0}{d_{ox}} \right) \tag{4.16}$$

, where N_g is the number of effective gate. As such, a double-gate FinFET or the GAA-FET would see a 2–4 times increase in C_{ox}, though in the case of GAA-FET, a slightly different equation was used to calculate C_{ox}, which accounts for the circular structure [75]. A more precise calculation of C_{ox} may be obtained with finite-element method software such as COMSOL [78, 64].

The increase in C_{ox} is the motivation for many studies of multi-gate FETs [77, 78, 75, 64]. A large C_{ox} has always been desirable for MOSFET since it allows for high drive current I_d, or large transconductance $\frac{\Delta I_d}{\Delta V_g}$. C_{ox} can be increased by reducing d_{ox}, though this approach also leads to larger gate current leakage due to quantum tunneling. Alternatively, C_{ox} could be increases with the use of high-k

dielectric material with larger ϵ_r such as HfO_2. The use of multi-gate structure to increase the effective gate area is then another approach to increase C_{ox}. A short review of such multi-gate structure is found in Ref. [79]. In our case, we assumed $N_g = 3$ since there are gates on 3 surfaces of the NWs.

FET measurement results of a few N34 and N35 NW-FET are shown in Fig. 4.33a. It is found that unlike the VDP and TF-FET samples, μ of NW-FET appears to be fixed at approximately $50\,\mathrm{cm^2/Vs}$ for both N34 and N35 samples. This may indicates that μ of NW-FET is limited by surface defects, and is independent of bulk N_D as in the case of the thin film VDP and FET devices. The measured n also appears consistently lower than that of thin film samples as well.

The measurement uncertainty for μ in NW-FET can be larger since the width of NWs is more difficult to measure precisely. To calculate the channel width ($W = N_w \times w$), the number of wires (N_w) and their width (w) must be checked in SEM. Since these NWs have very high aspect ratio of over two hundred at the least ($L_{ds}/w = 9\,\mu\mathrm{m}/40\,\mathrm{nm} = 225$), it can be difficult to detect small break in NWs along the entire length of the NWs. w may also vary along the wires due to drift in e-beam focus or variation in beam current. Furthermore, poor contrast in SEM images can make edge detection unclear. Assuming that 1-2 NWs out of 10 may be incorrectly counted (due to break in NWs), and an uncertainty of $\pm 5\,\mathrm{nm}$ in SEM measurement, there could be a large uncertainty in total channel width of $\frac{\delta W}{W} \approx 20\%$. The total measurement uncertainty for NW-FET devices is then

$$\frac{\delta\rho}{\rho} = \sqrt{\left(\frac{\delta W}{W}\right)^2 + \left(\frac{\delta L}{L}\right)^2 + \left(\frac{\delta d_{Si}}{d_{Si}}\right)^2} = 22\%$$

$$\frac{\delta\mu}{\mu} = \sqrt{\left(\frac{\delta W}{W}\right)^2 + \left(\frac{\delta L}{L}\right)^2 + \left(\frac{\delta d_{ox}}{d_{ox}}\right)^2} = 22\%$$

$$\frac{\delta n}{n} = \sqrt{\left(\frac{\delta d_{Si}}{d_{Si}}\right)^2 + \left(\frac{\delta d_{ox}}{d_{ox}}\right)^2} = 11\%$$

, assuming the same uncertainty for L, d_{Si}, and d_{ox} as before. Note that n is

independent of W, and its uncertainty is the same as in case of TF-FET. In fact, n could be measured via FET-based method without knowing the width of the NWs, though d_{Si} and C_{ox} must still be known.

Although FET measurement of NW-FET devices have shown μ to be much different from that measured in VDP and TF-FET devices, n and S appears to be much more consistent with the expected value. Figure 4.33b shows that $|S|$ of NW-FET tend to be large, which agrees well with the observation of lower n. Since the diameter of the NWs is still much greater than the limit where modification to S by quantum confinement may be observed, these NWs are still bulk-like. Therefore, a close adherence of S to other thin film samples, and the first-principle calculation (dashed line) would be expected.

4.5.3 Effect of oxidation on Si Thin film and Nanowires

Next, we characterize the electrical property of oxidized Si thin film and NWs. We begin by checking ρ, n, and μ of oxidized Si thin film using the VDP method. Figure 4.34 shows that ρ appears to increase as d_{Si} is decreased, though the trend is not clear. ρ is difficult to determine precisely as it is difficult to measure small d_{Si} given ± 3 nm uncertainty of the Dektak profilometer. As d_{Si} is reduced, the uncertainty becomes more significant in the calculation of ρ, as reflected by the growing vertical error bar in the figure. The horizontal error bar is fixed at ± 3 nm. The thinnest sample measured is approximately 10 nm. Although about 2.5 and 3 nm of Si is oxidized during 1 min oxidation cycle for N34 and N35 substrate, respectively, further oxidation have always resulted in non-conductive samples. This likely indicates complete oxidation of the thin film, where there is no self-limiting oxidation effect.

Likewise, Fig. 4.35 shows no clear trend for n as a function of d_{Si}. Since R_H and n also required d_{Si} value, there is also large uncertainties for n, similar to ρ. However, it was previously theorized in section 4.4.3 that [P] should increase

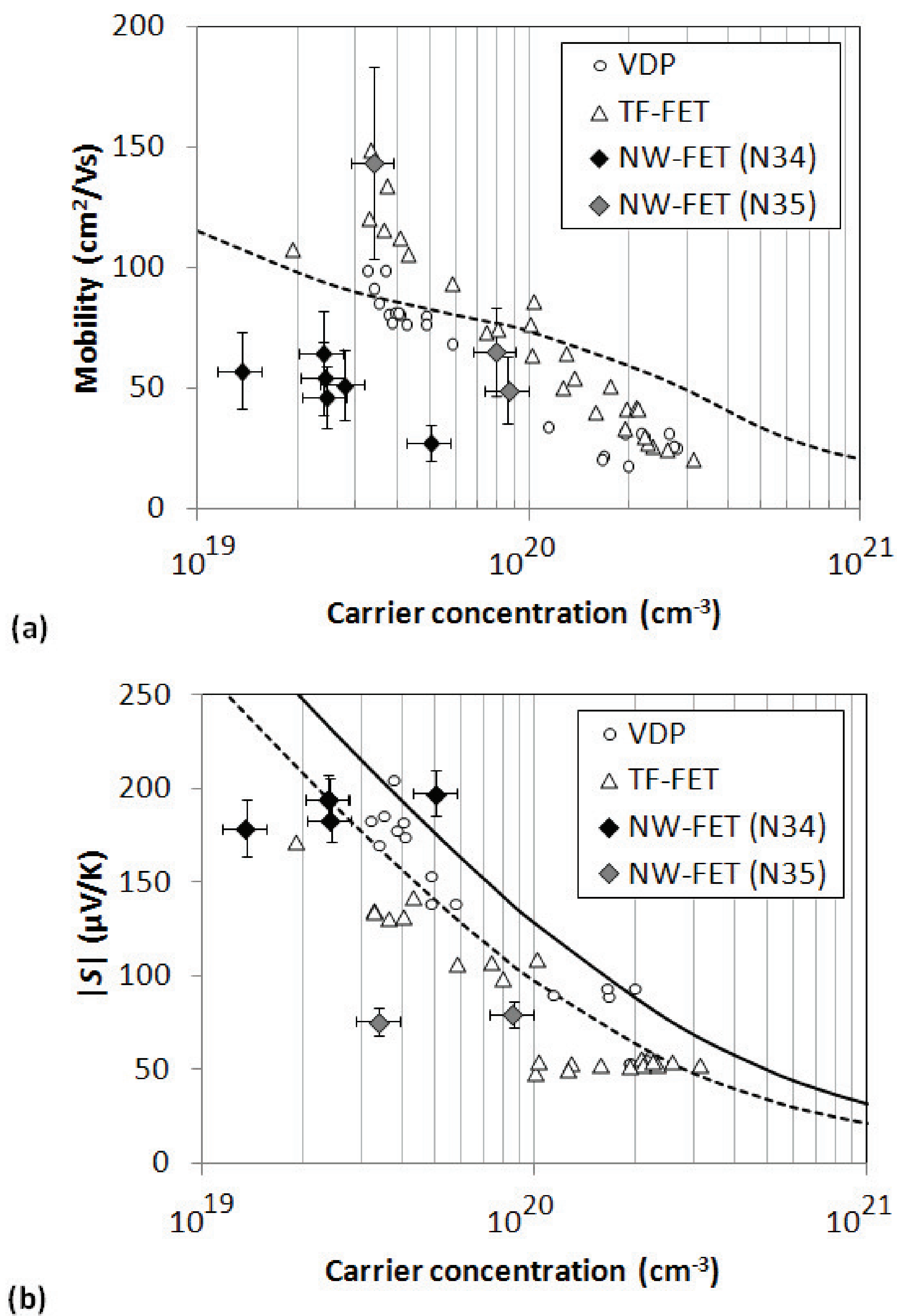


Figure 4.33: Electrical properties of Si NWs measured via FET method.

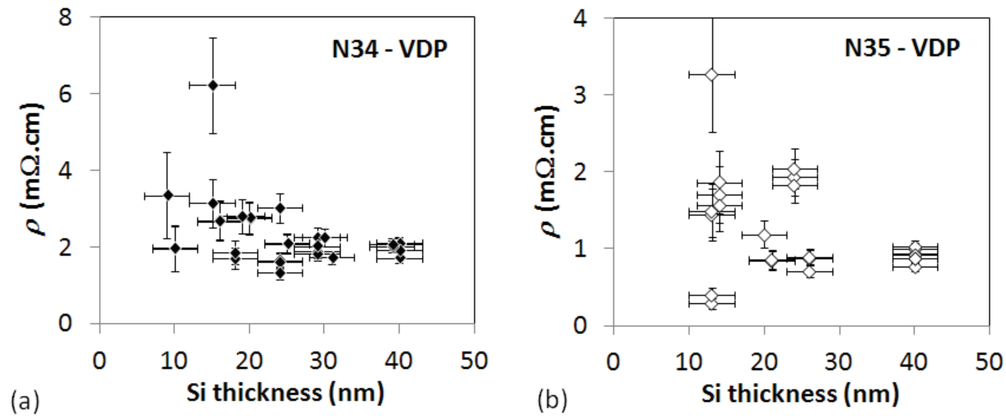


Figure 4.34: Resistivity of oxidized Si thin film.

as P atoms are rejected from the growing oxide, and accumulate in Si core. If it is assumed that all the segregated P are electrically active, then the increase in n after oxidation may be calculated as a function of remaining Si thickness as

$$n(d_{Si}) = k_s(d_{Si})n_0 = \left[\frac{0.44\, m_s\, d_0}{d_0 + (0.44m_s - 1)d_{Si}}\right] n_0 \tag{4.17}$$

, where k_s is the planar segregation ratio given in equation 4.8, $d_0 = 40\,\text{nm}$, $m_s = 10$ for P, and n_0 is the average initial carrier concentration of 3.5×10^{19} and $2.2 \times 10^{20}\,\text{cm}^{-3}$ for N34 and N35 samples shown in the figure, respectively. Calculation results of equation 4.17 is plotted as solid and dashed lines in the figure. For the N35 samples, the increase in n appears to generally agrees with the predicted increase due to dopant segregation. On the contrary, n appears to be the same regardless of d_{Si} for N34 samples. Although there is no strong agreement with the measured n value, the dopant segregation model may explain the changes in μ and S.

Figure 4.36 shows a clear trend where μ of the oxidized thin film decreases gradually with decreasing d_{Si} for both N34 and N35 samples. According to equation 3.5, $\mu = \frac{V_H}{IBR_s}$ is independent of d_{Si}, and should therefore be more accurately measured. The decrease in μ is also predicted as a result of increased [P] due to dopant segregation during oxidation. The predicted decrease in μ, shown as

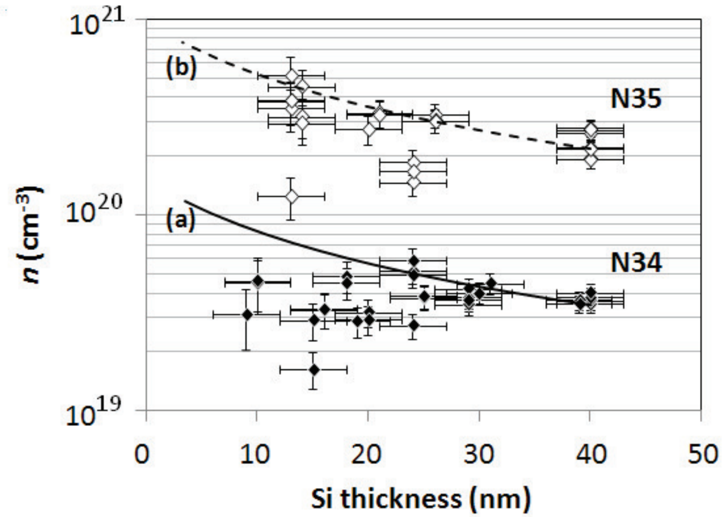


Figure 4.35: Carrier concentration of oxidized Si thin film. $^{(a)(b)}$See equation 4.17

solid and dashed lines in the figure, is calculated by applying the value of n from equation 4.17 to the empirical model [13] given by equation 4.3, i.e.

$$\mu(d_{Si}) = \mu_0 + \frac{\mu_{max} - \mu_0}{1 + (n(d_{Si})/C_r)^{\alpha}} - \frac{\mu_l}{1 + (C_s/n(d_{Si}))^{\beta}} \tag{4.18}$$

A generally close comparison between μ values as measured, and as predicted by the dopant segregation model in equation 4.18, is then observed for both N34 and N35 samples.

The decrease in $|S|$ of oxidized thin film shown in Fig. 4.37 is also in general agreement with the prediction based on the dopant segregation model. The calculation of S is performed as follows

1. Interpolate the value of E_F for each value of $n(d_{Si})$ from equation 1.34
2. Calculate S from value of E_F obtained from step 1 using equation 1.46

In step 1, the following parameters are assumed: $T = 300$K, $m_d = 0.32m_0$, $N_c = 6$, and $D = 3$. S was calculate twice for r equals to 0 and -1/2 for both N34 and N35 samples. For N34 samples, there is a good agreement between the measurement

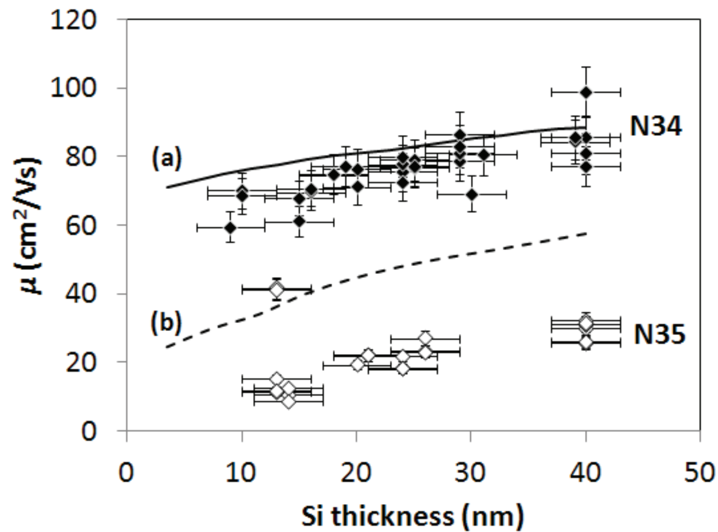


Figure 4.36: Mobility of oxidized Si thin film.

and the value predicted by the model for $r = -1/2$. However, the measured S may indicates a change to $r = 0$ at $d_{Si} = 10\,\text{nm}$. As the volume decreased, the dominant scattering mechanism may change from ionized impurity scattering in the bulk ($r = -1/2$) to (Si-SiO_2) interface defects or roughness scattering. Surface roughness scattering is largely energy-independent [33], i.e. $r = 0$. In the case of N35 sample, the measured S did not appears to change with oxidation, and is close to the value predicted by dopant segregation model assuming $r = 0$ for all the oxidized samples ($d_{Si} < 40\,\text{nm}$). This may indicates an earlier onset of significant surface defects scattering, brought on, perhaps, by precipitation of P at the Si-SiO_2 interface due to high [P] close to solubility limit in Fig. 4.8.

Alternately, the non-changing $|S|$ could indicates that n in fact did not increase with oxidation for N35. While [P] in Si may rise because of dopant segregation, n may not change due to the samples having already reached the electrically-active concentration limit of $3 - 4 \times 10^{20}\,\text{cm}^{-3}$ as shown in Fig. 4.8. In this case, the observed agreement between the measured and calculated value for μ in Fig. 4.36 for N35 samples could be merely a co-incidence. The reduc-

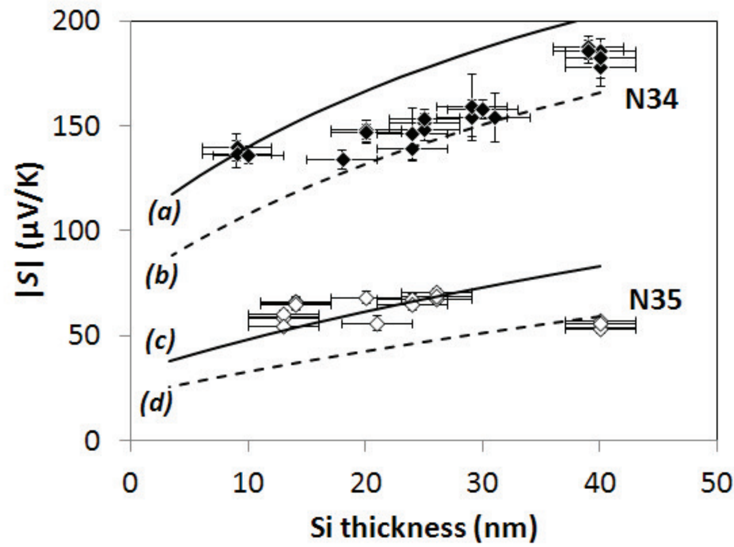


Figure 4.37: Seebeck coefficient of oxidized Si thin film. Solid and dashed lines show calculation result for $|S|$ as predicted by dopant segregation model for $^{(a)(c)}r = 0$, and $^{(b)(d)}r = -1/2$.

tion of μ could be due to increased concentration of non-ionized defects such as P precipitates or surface defect, rather than of ionized impurity as implied by the calculation.

The same reduction in $|S|$ was observed for oxidized Si NWs. Figure 4.38 shows the S of oxidized Si thin film and Si NWs fabricated on N34 substrate as a function of oxidation time, t_{ox}. The decrease in $|S|$ of Si thin film due to dopant segregation is calculated as a function of oxidation time as follow

1. Calculate $d_{ox} = 5.5t_{ox} + 2.5$ (see N34 in Fig. 4.21).
2. Calculate $n(d_{ox}) = k_s(d_{ox})n_0$ according to equation 4.8.
3. Interpolate the value of E_F for each value of $n(d_{ox})$ from equation 1.34
4. Calculate S from value of E_F obtained from step 3 using equation 1.46

The decrease in $|S|$ for Si NWs is calculated in similar manner except the segregation ratio for NWs geometry in equation 4.12 is used instead in step 2. Note that the same oxidation rate was assumed for the NWs as for bulk since t_{ox} is small. Two sets of NWs with the initial width, w_0, of 40 and 80 nm are shown. $d_0 = 40$ nm for all samples. A good fit was observed for 40 nm NWs (dotted line). However, the measured $|S|$ for the 80 nm NWs are far below the predicted value. There were no fitting parameters here since all parameters were either taken from literature ($m_s = 10$ for P, Si/SiO_2 volumetric ratio), or measured experimentally (i.e. oxidation rate, initial carrier concentration).

Only Si NWs that were oxidized for less than 4 min could be measured. Beyond this oxidation time, all NWs samples become non-conductive. Although after 4 min, about 25-30 nm of planar Si in the drain and source contact area should still be left, it may still be possible that drain and source area have been thinned sufficiently that Si core is effectively cutoff from the contacts. We were unable to measure any of the oxidized Si NWs that were fabricated from N35 substrate. Since ρ calculation is imprecise due to unknown geometry of the core, and number of conductive NWs, only S measurements were reliable. FET measurement has not been performed on oxidized NWs, so information about n and μ is also unavailable.

4.6 Conclusion

In conclusion, the entire process of Si NWs fabrication by electron-beam lithography was demonstrated. The fabricated NWs were then successfully characterized using measurement method based on the field-effect transistor, where an increase in effective gate oxide capacitance was observed through an order of magnitude reduction in the threshold voltage. Oxidation of NWs was then attempted, but device yields were low, and measurements results were limited. However, S measurements of both bulk Si and Si NWs appears to indicates an effect of dopant segregation consistent with our prediction. Future work should pay more emphasis

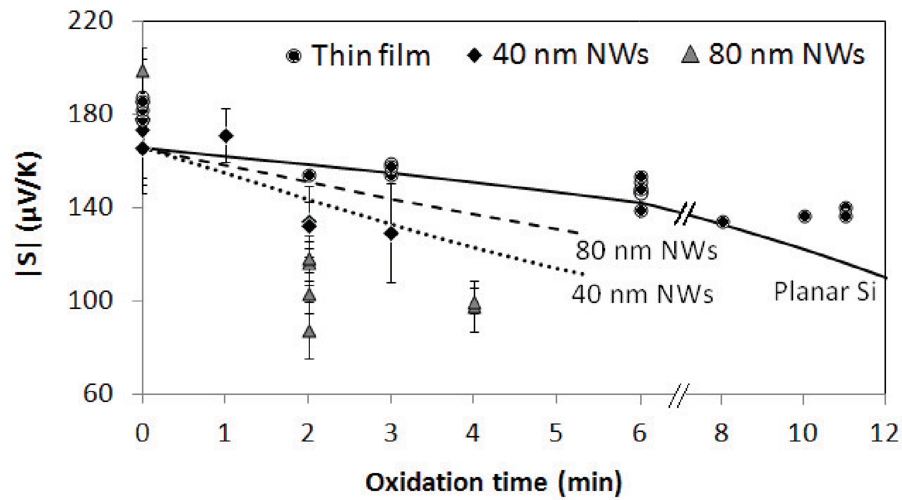


Figure 4.38: Seebeck coefficient of oxidized Si nanowires with initial width of 40 and 80 nm. Solid, dotted, and dashed lines represent calculation based on dopant segregation model. The horizontal axis was shortened for the 6 to 12 min interval to highlight the nanowires data below 6 min.

on physical characterization techniques, such as TEM, to improve the measurement of device geometry.

5 Influence of N on effective mass and thermoelectric property of GaNAs

5.1 Introduction and Motivation

5.1.1 Thermoelectric application for highly-mismatch alloy

In section 2.3.3, an enhancement of S^2n was observed for Tl-doped PbTe. This increase was explained as a result of interaction between Tl-resonant states and PbTe valence band (VB), which significantly increased m_d. Given the same n, an increase in m_d would increase $|S|$. Similar modification of the electronic band structure found in other "highly mismatched alloy" (HMA) were then touted as a way to improve $S^2\sigma$ for such material [80]. One well-known HMA material is the dilute nitride, which refers to a family of III-V semiconductor, such as GaAs or GaP, that is heavily doped with N up to typically a few atomic percent. The iso-valent N impurity that substitutes for Group V element, thus forming GaN_xAs_{1-x} or GaN_xP_{1-x} alloys, has a resonant energy level (E_N) that is inside the conduction band (CB). The interaction between E_N and the host CB then affects significant changes in the material's electrical properties, including the narrowing of band gap (E_g) due to lowering of CB edge by 0.12 eV/atomic N % [81, 24], the reduction of μ due to change in scattering process [82, 35], and an increase in m_d [83, 84, 22]. Although the reduction in μ is undesirable for improving $S^2\sigma$, it would be interesting to investigate whether the increased m_d in N-doped GaAs could at least

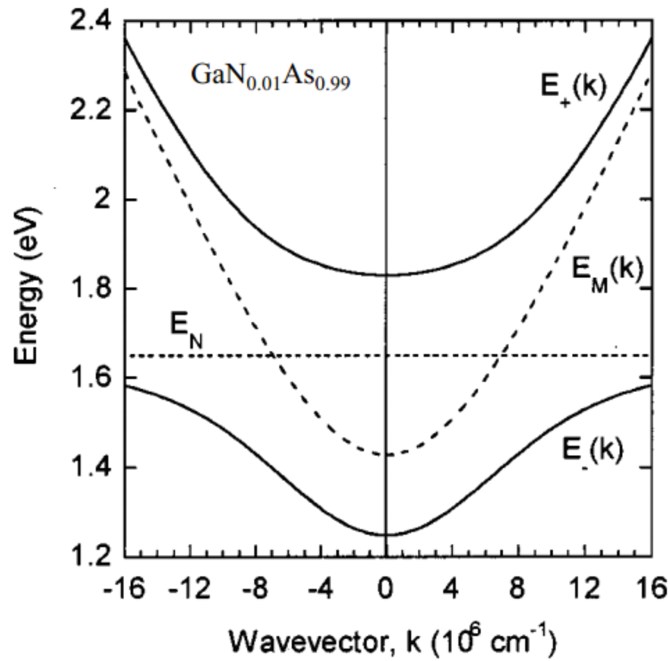
enhance S^2n in similar fashion to that observed in Tl-doped PbTe.

Another interesting and emerging HMA is the dilute bismide, which essentially swaps N for Bi. The role of Bi is similar to that of N, except that Bi energy level is resonant with GaAs VB. The addition of Bi also reduces band gap, but by raising of the VB instead of lowering of the CB [85]. Hole μ of $GaBi_xAs_{1-x}$ is also reduced by addition of Bi, but to a lesser extent [86]. Elemental Bi has high μ and low κ, which gives it a potential for high Z. However, it is a semi-metal with low $|S|$, therefore it is alloyed with Te to form the semiconductor Bi_2Te_3, which is perhaps the most common commercially available thermoelectric material. Bi nanowires and thin films were also explored with the hope of inducing semi-metal-to-semiconductor transition to enable high Z for pure Bi [87, 88]. It may then be interesting to investigate whether the innate advantages of Bi could be transferred to dilute bismide.

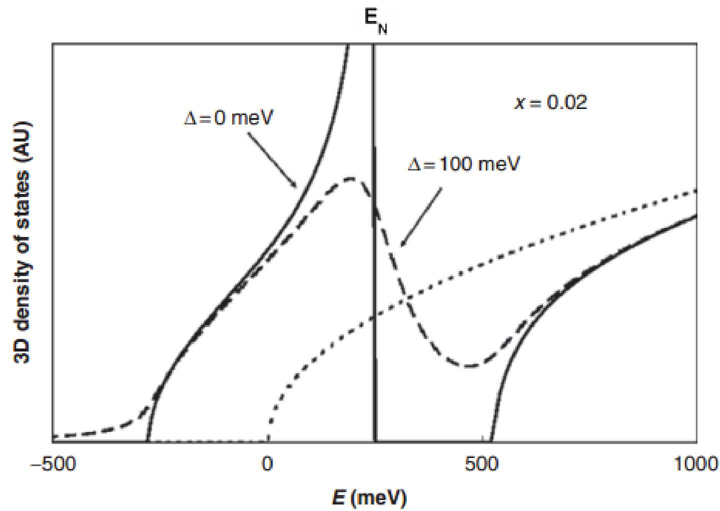
5.1.2 Existing work on Dilute Nitride

The early development of dilute nitride was for applications such as solar cell, where the ability to tune E_g in GaN_xAs_{1-x} by varying N composition, x, is important. The modification of band structure was initially described as a result of band anti-crossing (BAC) between the host GaAs CB, and E_N [23, 89]. This BAC model postulates a splitting of the CB into an upper E_+ and a lower E_- subbands, as shown in Fig. 5.1a. The electrical characteristic of GaN_xAs_{1-x} is then dominated by that of the E_- band, which has a lower CB edge relative to the original GaAs CB, and a flatter band structure that gives it a heavier m_d. As x is increased, the BAC model predicts a monotonic decrease of the CB minimum, and a monotonic increase in m_d. Calculation result in Fig. 5.1b then shows that the DOS should be significantly increases, especially near E_N.

Aggregation of N may also form "cluster" states that have localized energy level similar to E_N, which is associated with N that has substituted for As (N_{As}).



(a) Splitting of GaAs CB ($E_M(k)$) into E_+ and E_- subbands by N resonant energy level (E_N). Adapted from Ref. [90]



(b) Density of states of GaN_xAs_{1-x}(solid and dashed lines) compared to that of the unperturbed GaAs (dotted line). The Δ term in the figure represents an energy broadening term required by the calculation to prevent singularity at E_N. Adapted from Ref. [35].

Figure 5.1: Band Anti-Crossing model

Many different energy levels are possible depending on the configuration of N aggregates [91]. In the LCINS (Linear Combination of Isolated Nitrogen resonant States) framework, these cluster states further increase m_d, in addition to the increase due to E_N alone, provided that the cluster states energy levels are resonant with the E_- band [91, 92, 22]. For example, when $x = 0.5\%$, the E_- band can interact with one such cluster states, which may leads to an increase in m_d from $0.067m_0$ to as large as $0.150m_0$ [22]. However, as E_- band is shifted further with increasing x, the influence of this cluster states will reduce and m_d should reduce back to level predicted by the BAC model.

There have been a number of experiments that have measured m_d in GaN_xAs_{1-x}, and while most see an enhancement of m_d, there is a significant disagreement between the experiment results. Very large increase in m_d up to $0.19m_0$ at $x = 2.0\%$ were observed in undoped GaN_xAs_{1-x}/GaAs quantum wells, measured via cyclotron resonance technique [83]. Moderate enhancement was observed in undoped GaN_xAs_{1-x} with maximum of $m_d = 0.015m_0$ at $x = 1.8\%$, measured via magnetophotoluminescence [84, 22]. The first determination of m_d via thermoelectric measurements, to our knowledge, have instead shown a decrease in m_d to a minimum of $0.03m_0$ for Se-doped (n of $5\text{-}7\times10^{18}\,\mathrm{cm}^{-3}$) GaN_xAs_{1-x} at $x < 0.4\%$ [9]. The second thermoelectric measurement, which was published during the course of our work, then again shows an increase in m_d up to maximum of $0.16m_0$ for Te-doped (n of $3\text{-}5\times10^{17}\,\mathrm{cm}^{-3}$) GaN_xAs_{1-x} for $x \approx 1.7\%$ [10]. Thus, it is seen that greatly differing values of m_d were observed.

Whether such differences are due to the different measurement methods used, or due truly to the difference in material property is difficult to know, but it does show that there is still room for further research. For example, the differences between the various undoped, Se-doped, and Te-doped GaN_xAs_{1-x} may suggest that type of dopant plays a significant part, beyond changing n, in the property of GaN_xAs_{1-x} as well. Therefore, investigation Si-doped GaN_xAs_{1-x} could provide a useful contribution to both fields of dilute nitride, and thermoelectric material.

The aim of this work was to experimentally investigates whether an enhancement of m_d and, consequently, a net increase in S^2n could be achieved in Si-doped GaN_xAs_{1-x} over a wide range of N composition from x of 0.5% to 2.5%.

5.2 Sample preparation

5.2.1 Growth and sample specification

To observe the change due to incorporation of N, a set of GaN_xAs_{1-x} were grown with varying x but constant n. If n were constant across all samples, then S should increase with increasing m_d, and vice versa. All materials used in this work were grown by our collaborator, a fellow graduate student Yanjin Kuang from Professor Charles W. Tu's group in the Electrical and Computer Engineering department at University of California, San Diego. Description of the growth process in the next paragraph was provided by Mr. Kuang. X-ray diffraction measurement was also performed by Mr. Kuang.

GaAs and GaN_xAs_{1-x} thin films were grown on semi-insulating (100) GaAs substrate by gas source-Molecular Beam Epitaxy (GS-MBE or MBE). The substrate was initially heated to 580 °C to desorb native oxide under As overpressure, and a 200 nm buffer layer of GaAs was grown to reduce surface roughness. The substrate temperature was then lowered to 500 °C during the growth of the electrically active, doped GaAs or GaN_xAs_{1-x} layer, using 7N purity elemental Ga and thermally cracked As_2 (from AsH_3). For GaN_xAs_{1-x}, N was injected using a 13.56 MHz nitrogen radical beam source. During the growth of the active layer, Si was injected through thermal evaporation from effusion cell into the MBE chamber. The Si cell temperature is defined as T_{Si}. The active layer thickness (d) is kept below 200 nm to prevent defect formation due to lattice mismatch. The growth rate for both the buffer and active layers was approximately 0.2 nm/s. The film thickness was monitored *in situ* by RHEED (Reflection of high energy electron diffraction).

Subsequent to growth, the N composition (x) was determined from the lattice constant of the active layer measured by X-ray diffraction. x was interpolated from the lattice constant of GaAs (0.564 nm, $x = 0\%$) and GaN (0.519 nm, $x = 100\%$) through the Vegard's law [93]. It should be noted that this measurement would only account for N_{As}, since N aggregates may not contribute to the contraction of lattice constant. SIMS could be used to measure N concentration directly, though it would not distinguish the concentration of N_{As} from that of N aggregate. The best option would be to use both X-ray diffraction and SIMS, which should then yield information on $[N_{As}]$ and the total [N], respectively, and allowing for the determination of N aggregates concentration. Nevertheless, a study that used both methods have shown that N is incorporated almost entirely as N_{As} for $x < 3\%$ [94], therefore X-ray diffraction should accurately yield N composition in our samples.

An initial set of samples was grown over a range of x and T_{Si} to determine the growth window. Figure 5.2 maps x and n of a number of GaAs and GaN_xAs_{1-x} samples that were grown at various T_{Si}. Si acts as n-type dopant in GaAs, and it is expected that for the growth of GaAs

$$[\mathrm{Si}] \propto P_{Si} \propto \exp(C/T_{Si})$$
$$[\mathrm{Si}] = D \exp\left(\frac{-C}{k_B T_{Si}}\right) \tag{5.1}$$

, where P_{Si} is the Si vapor pressure, and C would be an experimentally determined constant [95]. n was measured using the VDP method, which was initially carried out in both our lab and Prof. Tu's lab to correlate both measurements. As the figure shows, n increased with T_{Si} at all x due to increased [Si]. But given the same T_{Si}, samples with higher x have lower n. This is due to passivation of Si donor by N; further discussion can be found in section 5.3.1. The range of T_{Si} is approximately 1,100-1,270 °C, and the range of x is approximately 0.5-3.0%. Typically, the lattice mismatch between GaN_xAs_{1-x} and GaAs will introduce high defect concentration when $x > 2.5\%$. Very low N concentration ($x < 0.5\%$) cannot

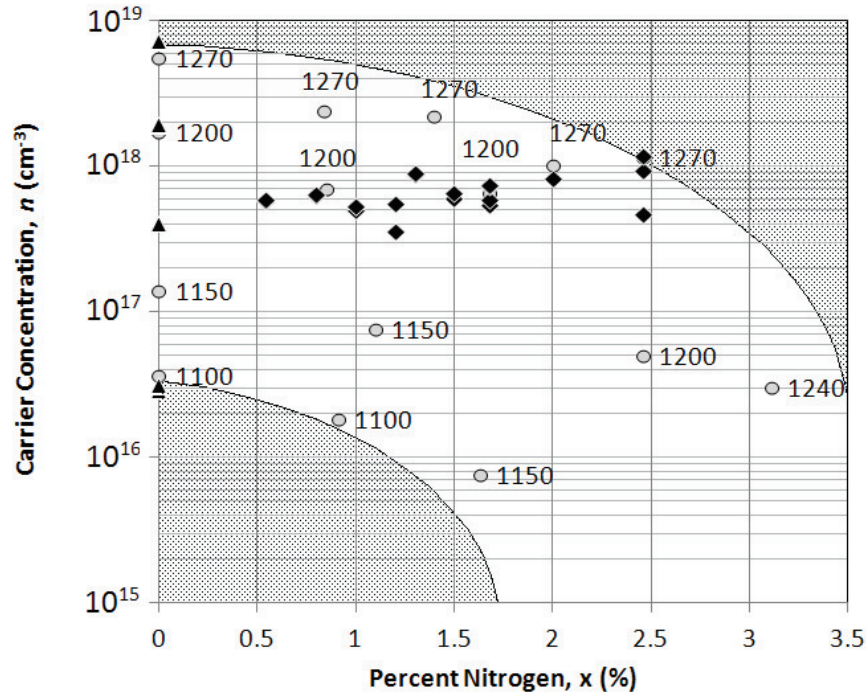


Figure 5.2: N composition and carrier concentration map of GaAs and GaN_xAs_{1-x} samples. Growth windows identified by initial sample set (gray circle); adjacent numerical label indicates T_{Si} in °C. The final set of GaAs (black triangle) and GaN_xAs_{1-x} (black diamond) was then grown to specific n and x targets.

be reliably achieved. x is also difficult to repeat precisely over different growth session. Given these constraints, the approximate growth window is shown by the unshaded area in Fig. 5.2. The measurement data from these initial samples was used to determine the dependence of n on x and T_{Si}.

To maintain the same n as x is increased, T_{Si} must be increased to compensate for the passivation of Si by N. A fitting was used to predict the required T_{Si}. First, it is assume that

$$n = A \exp\left(\frac{B}{x}\right) \tag{5.2}$$

for samples grown at the same T. Hence, three set of A, and B fitting constants were found for samples grown at $T_{Si} = 1,150$ °C, 1,200 °C, and 1,270 °C. These

fitting constants were found to vary linearly with T as

$$A = (13,444 - 12\,T_{Si}) \times 10^{15} \tag{5.3}$$

$$B = (33\,T_{Si} - 37,783) \times 10^{15} \tag{5.4}$$

Combining equations 5.2-5.4, then shows that

$$T_{Si}(n,x) = \frac{n \times 10^{-15} - 13,444x + 37,783}{33 - 12x} \tag{5.5}$$

This numerical fit was then used to determine the necessary T_{Si} that would yields the same n for a number of samples despite the variation in x.

The actual samples used in the study is then shown in Fig. 5.2 by the filled data points. The set of GaAs samples were used were used as the control samples. A new set of GaN_xAs_{1-x} samples were also grown up to a maximum of $x = 2.5\%$, but with a restricted range of n from $3 \times 10^{17}\,cm^{-3}$ to $1 \times 10^{18}\,cm^{-3}$.

5.2.2 Plasma etch and Ohmic contact for GaAs/GaN_xAs_{1-x}

Following MBE growth and X-ray diffraction measurement, the sample were prepared for VDP and S measurement as described in Chapter 3. GaAs and GaN_xAs_{1-x} may be etched with Cl-based chemistry [96, 97]; our samples were plasma etched with the recipe shown in Table 5.1. The etch rate may vary with N composition by 10-20% since GaN is more difficult to etch than GaAs [97]. Indeed, a few of the GaAs etch recipe that were tried did not etch GaN_xAs_{1-x} at all.

Measurement samples were then prepared according to Fig. 3.3. Ge/Ni/Au were sequentially deposited via electron beam evaporator for electrical contacts. Although Au on GaAs will forms Schottky contact, adding Ge will allows for Ohmic contact through tunneling once Ge has diffused into GaAs through annealing [98]. Ni is commonly added to alleviate the balling of AuGe contact, which roughen the contact surface and may affect long-term device stability [99]. The contacts were annealed at 450 °C in flowing N_2 for 5 min using the RTA. Samples were placed

Table 5.1: GaAs and GaN_xAs_{1-x} etch recipe.

Equipment	Oxford Plasmalab 80 RIE
Gas/Flow rate	BCl_3/10 sccm + Ar/5 sccm
RIE power	200 W
DC Bias	593 V
Pressure	20 mT
Temperature	15 °C
Etch rate	100-120 nm/min

with the active layer face down on a sacrificial GaAs substrate during the annealing to minimize loss of As through out-diffusion. ρ, n, μ, and S were then measured following the procedure described in Chapter 3. Indium solder can also be used to provide quick and easy contact to large-sized samples.

5.3 Electrical and Thermoelectric characterization of GaAs and GaN_xAs_{1-x}

5.3.1 Compensation and Passivation effects

From the $n-T_{Si}$ relationship of GaAs in Fig. 5.3, significant defect compensation can be observed at $T \geq 1,250\,°C$. The figure shows that while n increase with T_{Si} up to $\sim 7 \times 10^{18}\,cm^{-3}$ at $T_{Si} = 1,250\,°C$, further increase in T_{Si} to 1,270 °C reduces n. It is known that when [Si] exceed $5 \times 10^{18}\,cm^{-3}$, concentration of Si_{As} (Si incorporated in As lattice site), Ga vacancy, and Si clusters, all of which are electrons acceptor, rapidly increases [100, 101]. These defects then donates holes that compensates electrons from Si donors, which are incorporated on Ga sites (Si_{Ga}), leading to a decrease in n, which was observed at $T \geq 1,250\,°C$.

Below 1,250 °C, compensation by Ga vacancy and Si clusters become irrelevant, but a contribution from Si_{As} remains. By convention, a compensation

ratio [20] is defined as

$$\gamma = \frac{N_D + N_A}{n} \tag{5.6}$$

, where N_A=[$\mathrm{Si_{As}}$] is the acceptor concentration, and N_D=[$\mathrm{Si_{Ga}}$] is the concentration of Si incorporated on Ga sites that act as donor. It is posited that γ depends on relative availability of Ga and As vacancies, and should be independent of [Si] [102]. However, the formation energy of $\mathrm{Si_{As}}$ is lowered as [Si] is increased, therefore γ should rise with [Si] [101]. At least one estimate show that [$\mathrm{Si_{As}}$] may be as large as 30% of total [Si] at $5 \times 10^{18}\,\mathrm{cm^{-3}}$ [100], which would equate to

$$\gamma = \frac{1}{0.7 - 0.3} = 2.5$$

In the next section, it is shown that γ of 1-10 were observed in our samples.

Given the [Si]-T_{Si} relationship in equation 5.1, the carrier concentration could then be described as a function of T_{Si} as

$$n = \frac{[\mathrm{Si_{Ga}}]+[\mathrm{Si_{As}}]}{\gamma} = \frac{[\mathrm{Si}]}{\gamma} = \frac{D}{\gamma}\exp\left(\frac{C}{k_B T}\right) \tag{5.7}$$

, where the value $D/\gamma = 3.27 \times 10^{42}\,\mathrm{cm^{-3}}$ and $C = 7.09\,\mathrm{eV}$ were determined from the fitting using data of GaAs samples grown at $T_{Si} < 1,200\,^{\circ}\mathrm{C}$, where it is certain that $\mathrm{Si_{As}}$ may be the only main source of acceptor defects. The above equation is shown as solid line in Fig. 5.3.

In the case of $\mathrm{GaN}_x\mathrm{As}_{1-x}$, a further decrease in n is observed in Fig. 5.3, which is likely explained by Si-N mutual passivation [103, 104, 105, 106]. Si-N may bond with each other as a result of Coulomb interaction between the more positively-charged $\mathrm{Si_{Ga}}$ and the more negatively-charged $\mathrm{N_{As}}$ [106]. Unlike the long-range electron-hole compensation due to $\mathrm{Si_{As}}$, Si and N may only passivate each other if they are nearest-neighbor. Si-N bonding, which may result in $\mathrm{Si_{Ga}}$-$\mathrm{N_{As}}$ [103] or $\mathrm{(Si\text{-}N)_{As}}$ [104] (both Si and N share a single As site) complexes, will remove both the Si donor and N resonant energy levels. The immediate result is then a reduction of n, and if there is sufficient Si to passivate N, the BAC effect is reversed as well, which can be observed by the widening of the band gap back to

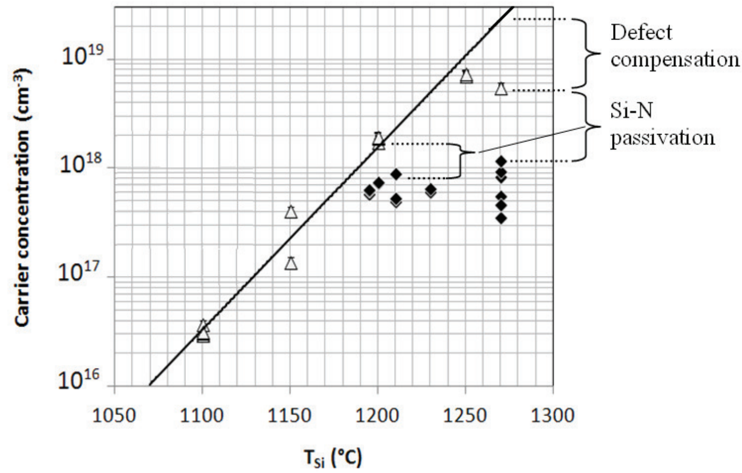


Figure 5.3: Carrier concentration of GaAs (open triangle) and GaN_xAs_{1-x} (filled diamond) as a function of Si cell temperature.

GaAs level [106].

The extent of the passivation may be represented by the doping efficiency, which is here defined as ratio of n in GaN_xAs_{1-x} and in GaAs grown at the same T_{Si}, i.e.

$$\Phi = \frac{n(x, T_{Si})}{n(0, T_{Si})} \tag{5.8}$$

For the majority of GaN_xAs_{1-x} samples grown at $T_{Si} < 1,250\,^{\circ}\mathrm{C}$, n for GaAs may be approximated with equation 5.7. For the GaN_xAs_{1-x} samples grown at $T_{Si} = 1,270\,^{\circ}\mathrm{C}$, a direct comparison is made with actual n from the GaAs samples that were grown at 1,270 °C.

Figure 5.4 shows that Φ decreases substantially with T_{Si}. However, it is unclear how T_{Si} could be physically related to Φ. Naively, one may speculate that increasing T_{Si} would impart more energy on the evaporated Si. The Si may then arrives on the substrate with more energy, allowing it to diffuse further, and increasing its chance of finding N atoms. The figure may also suggests that Φ decreases as [Si] is increased along with increasing T_{Si}. If there was comparable

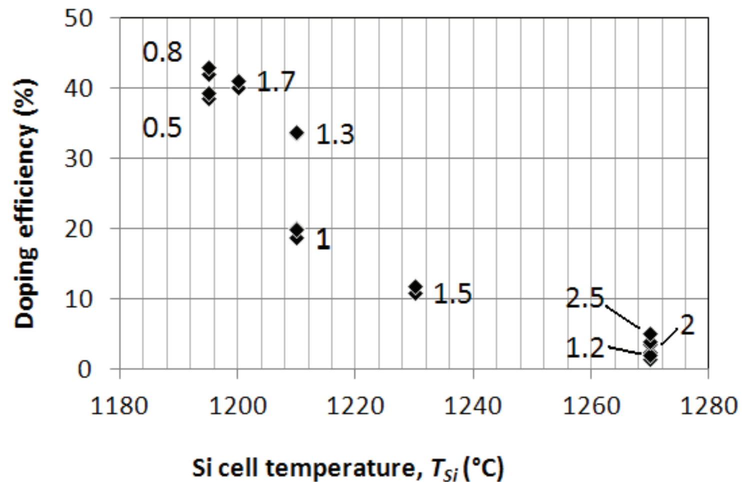


Figure 5.4: Doping efficiency of GaN_xAs_{1-x}grown at different T_{Si}. Numerical label indicates x in %.

amount of Si to N, then Φ could increases once [Si] exceeds [N]. However, [N] should be much greater than [Si]. For example, [N] $\approx 2.2 \times 10^{20}\,\text{cm}^{-3}$ for samples with the lowest $x = 0.5\%$, given the density of $4.4 \times 10^{22}\,\text{cm}^{-3}$ for GaAs [34]. Meanwhile, [Si] may be estimated from n using equation 5.7. At the maximum $T_{Si} = 1,270\,°\text{C}$, [Si] should equals $\gamma \times 2.3 \times 10^{19}\,\text{cm}^{-3}$. Even if a large $\gamma = 10$ is assumed, any addition of Si should not impact the doping efficiency there are still plenty of N for Si to bond with.

The figure also shows that at a given T_{Si}, Φ is lower for samples with higher x. Initially, this appears contradictory, as putting in more N should increases the likelihood of N finding and bonding with Si, thus reducing Φ. However, N is incorporated solely in As sites, therefore N actually competes with the formation of Si_{As}. The increase in Φ may then actually points to a decrease in [Si_{As}] and γ.

Such low Φ should have been unlikely given the low density of both [N] and [Si]. As pointed out earlier, Si and N must be nearest-neighbor in order to bonds. Since both N and Si are very diluted (with $x = 2.5\%$ at the most), the random

chance that N and Si could land in the same unit cell during deposition must be extremely low. Indeed, Yu *et al.* did not observe Si-N passivation in GaN_xAs_{1-x} when Si was doped via ion-implantation [106]. Passivation then only begins after annealing at $T > 600\,°C$ for 10 s, when the diffusion length,

$$L_D = \sqrt{D_{Si} t} \tag{5.9}$$

, where D_{Si} is diffusivity of Si in GaAs [107], were found to be comparable to the average separation between N atoms,

$$L_N = \sqrt[3]{[\mathrm{N}]} \tag{5.10}$$

Our samples were subjected to two heating cycles, i.e. (1) during growth where substrate was held at 500 °C for approximately 16 min of active layer deposition (200 nm at 0.2 nm/s), and (2) during contact annealing for 450 °C for 5 min. Using the D_{Si} value from Ref. [107], L_D was calculated to be (1) 0.07 nm, and (2) 0.01 nm, respectively. Given x of $0.5 - 2.5\%$, it is calculate $L_N = 1.0 - 1.6$ nm for our samples. Therefore, bulk diffusion should not be sufficient for extensive passivation to occurs. However, since extensive passivation did occurs during growth phase, it may indicates that surface diffusivity of Si adatoms is higher than that of Si diffusivity in bulk. Extensive Si-N passivation has also been observed in other studies when Si was incorporated during MBE growth [105].

Furthermore, the motion of Si may be more direct than Brownian as well due to Coulombic attraction. The screening length, also known as the Debye length, is given by

$$\lambda_D = \sqrt{\frac{\epsilon_r \epsilon_0 k_B T}{e^2 n}} \approx \sqrt{\frac{\epsilon_r T}{n}} \times (69\ \mathrm{m}^{-1/2}\mathrm{K}^{-1/2}) \tag{5.11}$$

, where $\epsilon_r = 13.1$ for GaAs. Given n of $3 - 10 \times 10^{17}\,\mathrm{cm}^{-3}$, it is calculate that $\lambda_D = 4.0 - 7.3$ nm. Since $\lambda_D > L_N$, it is certain that at least one N would be within a screening length of Si. Therefore, the Coulomb force is relevant and may assist in the passivation.

5.3.2 Mobility

Compensation ratio of GaAs can be inferred from μ measurement. μ of GaAs decreases with the total concentration of ionized impurity, i.e. $N_D + N_A$. Therefore, μ is strongly dependent on γ, as shown. By plotting μ versus n, as shown in Fig. 5.5, γ can be approximated. For the S085-87 samples, γ appears to steadily increases from roughly 1.4 to 4.0 with increasing n. Since $T < 1,270\,^{\circ}$C for S085-86, the increase in γ is likely to indicates a strong onset of defect compensation by Si_{As}, Ga vacancy, and Si clusters defects. Thus, this observation of increasing γ with n is well support by two independent relationship of $n - T_{Si}$ in Fig. 5.3, and $\mu - n$ in Fig. 5.5.

On the contrary, the high γ of 5 and 10 observed for S060 and 1057 likely indicates a sub-optimal growth. As discussed previously, high γ could indicates high As vacancy. Since these samples were grown at an earlier time than S085-87, it may be that the MBE chamber was not as well conditioned as for the later samples.

μ is significantly decreased by the addition of N. Figure 5.6 shows that μ is almost an order of magnitude lower than that of GaAs. Measurement results by Young *et al.* [9] shows that majority of μ reduction occur below $x = 0.1\%$. From then on, μ continues to decrease with x, but at a much reduced rate. The reduction in μ is two fold. There is the increase in m_d due to BAC with E_N, but more importantly, there is also an increase in scattering due to N impurity. The framework to describe N impurity scattering is based on alloy scattering [82], where the increase in scattering rate for GaN_xAs_{1-x} is due to relatively small radius of N compare to As that it replaces [35]. In addition to N_{As}, N complexes may also be a significant source of scattering [108]. The sharp reduction below $x = 0.1\%$ in the figure most likely reflects the sudden changes in scattering mechanism, while the gradual decrease from then forth is due to increase of both [N] and [Si]. However, since $[N] \gg [Si]$, μ should be largely independent of n or [Si]. Therefore, the different n in the samples compared in the figure should be insignificant.

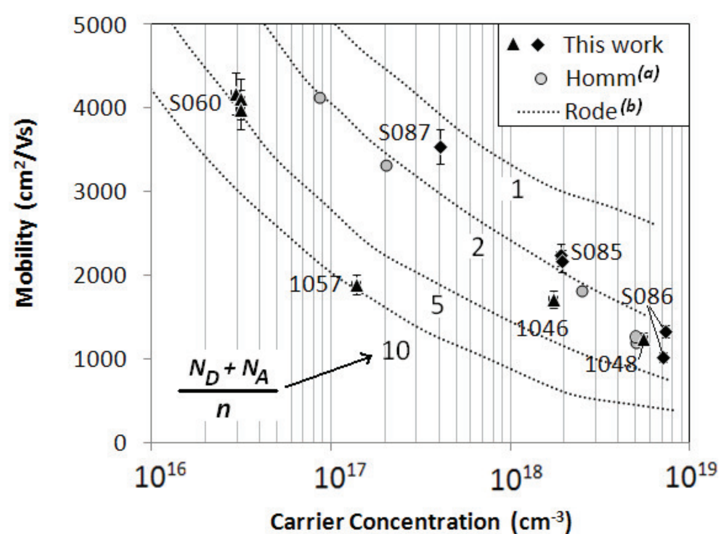


Figure 5.5: Mobility of GaAs samples grown at earlier date (black triangle) and later date (black diamond). [a]See Ref. [8]. [b]Dashed lines are calculation results replicated from Ref. [20].

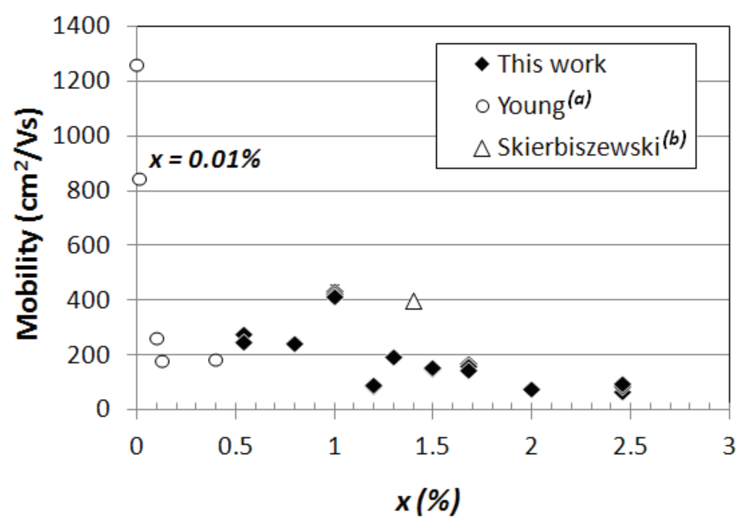


Figure 5.6: Mobility of GaN_xAs_{1-x} as a function of N composition. [a]Ref. [9]. [b]Ref. [21].

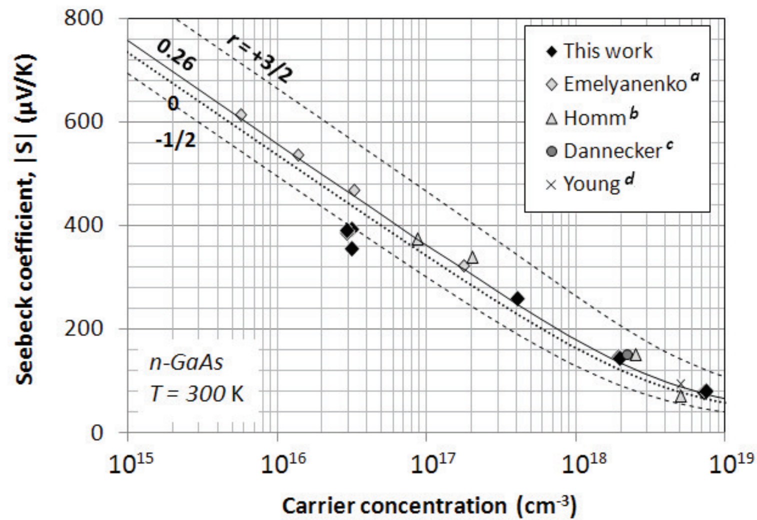


Figure 5.7: Seebeck coefficient of n-type Si-doped GaAs as a function of carrier concentration at 300 K. [a]Ref. [7]. [b]Ref. [8]. [c]Ref. [10]. [d]Ref. [9].

5.3.3 Seebeck Coefficient

Seebeck measurement of GaAs was measured and compared to literature value to verify the accuracy of our measurements. The result of the measurement and the comparison was shown in Fig. 3.19, but is reproduced again in Fig. 5.7. A very good fit between the calculation and measurement was found for $r = +0.26$, which was the value suggested through Nernst coefficient measurement [9]. Our calculation also takes into account the increase in m_d with n due to GaAs CB non-parabolicity [109]. The dominant scattering mechanism in GaAs is the Polar Optical Phonons (POP) scattering [35, 33], and cannot be accurately represented by the simple power law $\tau = \tau_0 E^r$. However, the scattering rate of POP has different constant values above and below 0.05 eV [33]. Hence, the relaxation time of POP process may be approximated with $r \approx 0$.

Having verified S measurement on GaAs, S of GaN_xAs_{1-x} were measured, and is shown in Fig. 5.8 as function of x. However, the relationship between S and x is not clear since n still varies for most samples, and S decreases with increasing

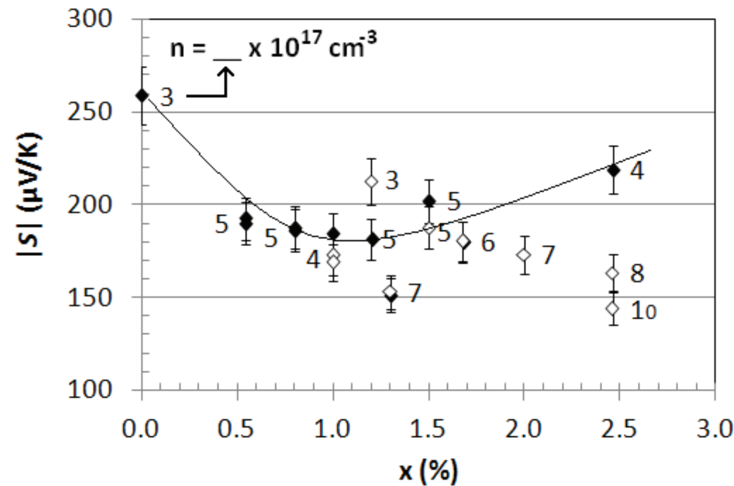


Figure 5.8: Seebeck coefficient of Si-doped GaN_xAs_{1-x} as a function of N composition at 300 K.

n. To highlight the trend, data points for samples with n closest to the average value of $4 \times 10^{17}\,cm^{-3}$ are represented in black, while the rest were unfilled. A solid line was also drawn to help guide the eye. According to the available data, it appears that $|S|$ is reduced when N is added up to $x \approx 1.0\%$, but then increases slightly with addition of N. From the figure, $|S|$ for GaN_xAs_{1-x} is lower than that of GaAs for all N composition studied.

However, the variation in n may affect the comparison. For example, the GaAs sample has the lowest $n = 3 \times 10^{17}\,cm^{-3}$. This was the situation we wanted to avoid from the beginning, where S may vary due to variation in n as well as x. Therefore, the product S^2n is compared in Fig. 5.9. The same characteristic is still observed, where a minimum is observed at $x = 1.0\%$. However, it is noticed that the spread in data at each x is removed, as the increase in n now compensate for the decrease in $|S|$. While this comparison still show no enhancement of S^2n for GaN_xAs_{1-x} over GaAs, the differences are much smaller and within the error bar. If n were actually equal, then this figure may suggests that $|S|$ of GaN_xAs_{1-x} is actually comparable to that of GaAs, and that the trends seen in the previous

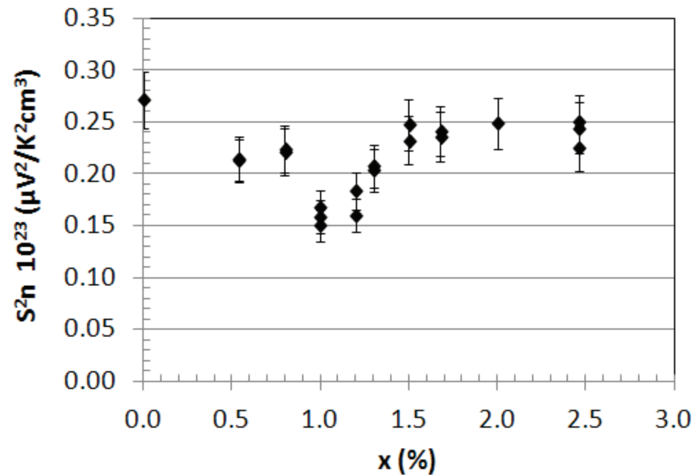


Figure 5.9: S^2n of GaN_xAs_{1-x} as function of N composition.

figure is partially due to lower n in the GaN_xAs_{1-x} samples.

The overall lack of S^2n enhancement may be due to the change in scattering mechanism. The addition of N is responsible for a large decrease in μ through a change in scattering from POP to alloy scattering. According to Table 1.1, $r = -1/2$ for alloy scattering in bulk material. A decrease of r by 1/2 can reduces $|S|$ by approximately 30-50 µV/K alone, according to Fig. 1.8. In the next section, it is shown that there may indeed be an expected enhancement of m_d.

5.4 Discussion

5.4.1 Effective Mass

The effective mass is calculated from the measured S and n to give a clear view of how GaAs CB is affected by N. The calculation process is essentially a reverse of how $S - n$ curve is calculated, and is performed as follow:

1. Given a pair of (S_i, n_i) data, η is first numerically interpolated from equa-

tion 1.46, i.e.

$$S_i = \frac{k_B}{q}\left(\frac{(r+\frac{D}{2}+1)F_{r+D/2}(\eta)}{(r+\frac{D}{2})F_{r+D/2-1}(\eta)} - \eta\right)$$

$D = 3$ is assumed for all samples. While $r = +0.26$ was assumed for GaAs, and $r = -1/2$ for GaN_xAs_{1-x}. These r values were justified based on Fig. 5.7, and the earlier discussion regarding POP and alloy scattering.

2. N_0 is then calculated from equation 1.34, i.e.

$$N_0 = \frac{n_i}{F_{D/2-1}(\eta)}\left(\frac{\pi}{4}\right)^{|D/2-1|}$$

, where η is taken from the step 1.

3. Calculate m_d according to equation 1.35, i.e.

$$m_d = \frac{\pi\hbar^2}{k_B T}\left(N_0 \frac{2^{|D/2-1|}d^{3-D}}{N_c}\right)^{2/D}$$

where N_0 is taken from step 2, $N_c = 1$ for GaAs and GaN_xAs_{1-x}, and d is irrelevant.

The the last step, there is a slight complication since the T during S measurement is typically 10 °C higher than that during n measurement. This is due to joule heating required to provide non-zero ΔT during S measurement. In our calculation, T measured during S is used, but if the lower T measured during n is used, then m_d will increase by $\sim 2-4\%$. A more significant source of uncertainty, however, is the assumption of r value during step 1, which we will discuss next. Figure 5.8 shows the values of m_d that were calculated from our measurement of S and n, in conjunction with the closely-related works by Young *et al.* [9] and Dannecker *et al.* [10].

The calculation methods used by Young *et al.*, Dannecker *et al.*, and us are largely equivalent, which means the differences observed in Fig. 5.8 originated from the samples and the measurements, and is not merely artifact. Our results were calculated in similar manner as in Dannecker *et al.* Young *et al.* calculation also share the same underlying physics, but their calculation is aided by the measurement of Nernst coefficient, which directly gives information on r value.

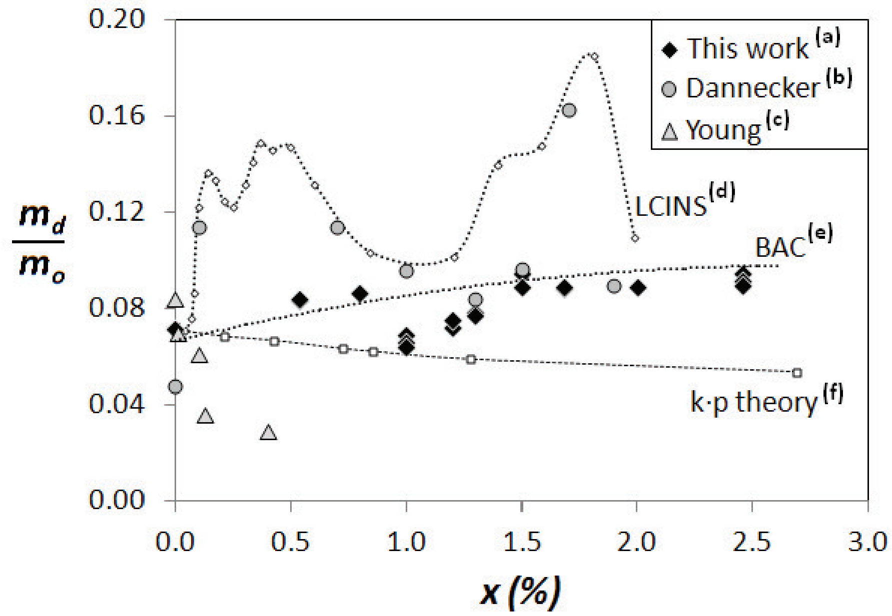


Figure 5.10: Density of states effective mass of GaN_xAs_{1-x} with 3 types of dopant: $^{(a)}$ Si, n of $3 - 10 \times 10^{17}\,\mathrm{cm}^{-3}$, $^{(b)}$ Te, n of $3 - 5 \times 10^{17}\,\mathrm{cm}^{-3}$ (Ref. [10]), $^{(c)}$ Se, n of $5 - 7 \times 10^{18}\,\mathrm{cm}^{-3}$ (Ref. [9]). Theoretical models reproduced from $^{(d)}$Masia *et al.* [22] and $^{(e)}$Shan *et al.* [23], and $^{(f)}$calculated from data in Ref. [24].

Although not stated in the publication, we were able to calculate r value from the experimental data listed in Ref. [9], using formulas from Ref. [26], which reveals that r is +0.26 for GaAs, and decreases to -0.6 for GaN_xAs_{1-x} with $x = 0.4\%$. The transition to negative r value for GaN_xAs_{1-x} is consistent the earlier discussion, and justify the use of r values in step 1 of our calculations.

Dannecker *et al*, however, chose $r = +3/2$ for GaAs instead on the basis that (weakly-screened) ionized impurity scattering should dominate for doped sample [20]. However, calculation by Vaughan *et al.* [35] suggests that POP scattering should be more dominant at room temperature, and in any case the compilation of $S - n$ data for GaAs in Fig. 5.7 does support the smaller $r = +0.26$ value. The use of larger r by Dannecker *et al.* resulted in lower evaluation for m_d of GaAs ($\approx 0.048 m_0$, compare to $0.072 m_0$ in our case) despite virtually the same S and n measurement from both work, as shown in Fig. 5.7. This is the only case where a discrepancy has arises because of assumption in the calculation. For GaN_xAs_{1-x}, Dannecker *et al.* also used the same value of $r = -1/2$.

All three experiments then show distinctly different, if not contradictory, results. Young *et al.* observed a strong decrease in m_d, which was attributed to the narrowing of band gap according to k·p theory. According to k·p theory [110], m_d for direct band gap semiconductor can be estimated as

$$\frac{m_d}{m_0} = \frac{m_0 E_g}{2P^2} \approx \frac{E_g}{20\,\text{eV}} \tag{5.12}$$

, where the matrix element P^2 is largely the same for most III-V and II-VI semiconductors [110]. The E_g value from Ref. [24] were then applied the to above equation. The calculated m_d is then plotted in Fig. 5.10 as well. The reduction of m_d according to k·p theory were found to be much less that observed by Young *et al.*

On the contrary, Dannecker *et al.* observed a drastic increase with large fluctuation in m_d that appears to follows the LCINS model [22]. As the figure shown, at least two characteristic peaks were predicted by the LCINS model in this range of x. For example, the peaks at $x = 0.3\%$ and 0.5% were attributed to

the contribution from the cluster states of N-Ga-N and N-Ga-N-Ga-N chains [22], respectively. In the absence of relevant cluster states near $x = 1.0\%$, m_d falls back to the level predicted by the BAC model.

Meanwhile, we also observed an increase in m_d, but one that is much more gradual and in closer agreement with the BAC model [23]. However, the observation of a local minimum at $x = 1.0\%$, which coincides with both Dannecker *et al.* results and the LCINS model, is puzzling. While the absence of a peak at x of 0.3-0.5% and 1.7% could be explained by the absence of N aggregates, a decrease of m_d below the level predicted by the BAC model is unexpected since E_N should always be present. It is also worth noting that the reduction is close to the level predicted by $\mathrm{k} \cdot \mathrm{p}$ theory. This may indicates that the band gap reduction can have an affect on m_d of GaN_xAs_{1-x}after all. An explanation for this minimum at $x = 1.0\%$ is not available at this time, though there may be an explanation for the absence of a peak at x of 0.3-0.5% and 1.7%

The absence of contribution from N aggregates may be related to the use of different dopant species in each experiment. While our samples were doped with Si, those of Young *et al.* and Dannecker *et al.* were doped with Se and Te, respectively. While Si may be incorporated into both Ga and As lattice site, Te and Se are incorporated solely as As substitute. Therefore, Te and Se cannot be an electron acceptor, and amphoteric defect compensation is absent in Te- or Se-doped GaAs. However, Ga vacancy defects still arise as [Te] is increased, which similarly limits n in GaAs[111]. Furthermore, Te and Se also cannot mutually passivate N, which also substitutes for As only. While Te and Se may be attracted to N by Coulomb force, they cannot form a direct bond to N as As sites are necessarily separated by Ga sites [103].

It is then possible that Si may interact and bond with N aggregate, which is the source of large increase in m_d predicted by the LCINS model, as discussed in section 5.1.2. While passivation has been described in term of bonding of single

$Si_{Ga}$to single N_{As}, it is conceivable that Si could also replace Ga in the N-Ga-N clusters. Furthermore, the high concentration of N in such aggregates could lead to stronger Coulombic force, relative to that of individual N_{As}, that disproportionately attracts Si. Considering that the N aggregates concentration"is expected to be only a small fraction of total N concentration" [91], the passivation of even a small fraction of these aggregates could prevent the large increase in m_d associated with these cluster states according to the LCINS model. But since [N] is still much greater than that of [Si], the overall increase in m_d due to BAC effect is still present.

The hypothesis that Si-N mutual passivation may counteract effect of N aggregates, could be tested by measurement of GaN_xAs_{1-x} that were Si doped by ion-implantation. Si-N mutual passivation is not observed because Si have very low chance of finding N when randomly implanted. For passivation to occur, Si must be given enough energy to diffuse to N through thermal annealing. Therefore, one could produce an electrically active Si-dope GaN_xAs_{1-x} *without* the effect of passivation through ion-implantation. Comparison of S in ion-implanted Si-GaN_xAs_{1-x}, where there is no Si-N mutual passivation, to the result shown in this work should then reveals a large increase in S as predicted by the LCINS model if the hypothesis were true.

In conclusion, it was hypothesized that the interaction between N resonant energy level and the GaAs CB could introduces a large increase in m_d, and consequently enhance $S^2\sigma$. To test this hypothesis, we have grown and measured electrical and thermoelectric property of Si-doped GaAs and GaN_xAs_{1-x} thin films with n of $3 - 10 \times 10^{17}\,\mathrm{cm}^{-3}$ and $0.5\% < x < 2.5\%$. While it was found that m_d was indeed increased in accordance with the established BAC model, an enhancement in S^2n was not found due to a change in scattering mechanism from POP in GaAs, to alloy scattering in GaN_xAs_{1-x}, which counteract the increase in m_d. Additionally, the change in scattering mechanism severely degrade μ. Therefore, GaN_xAs_{1-x} is not suitable for thermoelectric application. Nonetheless, the enhancement of m_d by chemical means such as alloying should hold a good promise

for improving ZT.

Chapter 5, in part, is a reprint of the material as it appears in the journal article: P. Pichanusakorn, Y. J. Kuang, C. J. Patel, C. W. Tu, and P. R. Bandaru, "The influence of dopant type and carrier concentration on the effective mass and Seebeck coefficient of GaN_xAs_{1-x} thin films", *Applied Physics Letters 99*, 072114 (2011). The dissertation author was the primary investigator and author of this paper.

A Symbols and Acronyms

Table A.1: List of physical constant.

e	elementary charge	$1.603 \times 10^{19}\,\mathrm{C}$
k_B	Boltzmann constant	$8.617 \times 10^{-5}\,\mathrm{eV/K}$
m_0	free electron mass	$9.109 \times 10^{-31}\,\mathrm{kg}$
$\hbar$	Reduced Planck constant	$6.583 \times 10^{-16}\,\mathrm{eV \cdot s}$
ϵ_0	Permittivity	$8.854 \times 10^{-12}\,\mathrm{F/m}$

Table A.2: List of Symbols.

$\mathcal{E}$	Electric field
η	reduced Fermi level
∇T	Temperature gradient
ΔT	Temperature difference
ϵ_{ox}	Relative permittivity
γ	Compensation factor
κ	Thermal conductivity
κ_e	Electronic thermal conductivity
κ_l	Lattice thermal conductivity
λ	Debye length
Φ	Doping efficiency
ρ	Resistivity
σ	Electrical conductivity
τ_0	scattering constant
μ	Mobility
μ_H	Hall Mobility
B	Magnetic flux density
C_{ox}	Gate oxide capacitance per unit area
C_{Si}	Dopant concentration in Si
C_0	Initial dopant concentration in Si
D	Dimensionality factor

d	Thickness/Nanowires diameter
d_{ox}	Gate oxide thickness
d_{Si}	Si (on SOI) layer thickness
E	Energy
E_F	Fermi level
I	Current
I_d	Drain-source current
I_B	EBL beam current
I_P	EBL probe current
K	Thermal conductance
k_s	Segregation ratio
G	Electrical conductance
G_{ch}	Channel conductance
J	Electrical Current density
L	Lorentz number
L_{ch}	Effective channel length
L_g	Gate length
L_{ds}	Drain-Source length
L_D	Diffusion length
L_N	Distance between N atoms
m_d	Effective mass, density of states
m_σ	Effective mass, conductivity
m_s	segregation coefficient
N_0	Effective density of states
N_c	Number of conduction valley
N_A	Acceptor concentration
N_D	Donor concentration
N_w	Number of nanowires in an array
n	Carrier concentration
P_{Si}	MBE Si vapor pressure
Q	Heat Current density

q	unit charge of electrons or holes
R	Resistance
R_s	Sheet resistance
R_H	Hall coefficient
R_{ds}	Drain-source resistance
r	scattering constant (exponent)
r_H	Hall factor
S	Seebeck coefficient
T	Temperature
T_{Si}	Si cell temperature in MBE
t	Time
V	Voltage
V_H	Hall voltage
V_e	EBL Accelerating voltage
V_g	Gate-Source voltage
V_d	Drain-Source voltage
V_t	Threshold voltage
W	FET channel width
w	Nanowires width
w_d	EBL Working distance
Z	Thermoelectric Figure of Merit
$g(E)$	Density of states
$\tau(E)$	Relaxation time
$v(E)$	Velocity
$f(E)$	Non-equilibrium Distribution function
$f_0(E)$	Equilibrium Distribution function
$f_{FD}(E)$	Fermi-Dirac Distribution function
$f_{MB}(E)$	Maxwell-Boltzmann Distribution function

Table A.3: List of Acronym.

ACE	Acetone
BOE	Buffered-oxide Etch
EBL	Electron-Beam Lithography
FET	Field-effect transistor
HMA	Highly-mismatched alloy
IPA	Isopropyl alcohol
MBE	Molecular beam epitaxy
MOSFET	Metal-oxide-semiconductor field-effect transistor
NWs	Nanowires
PECVD	Plasma-enhanced chemical vapor deposition
PL	Photo-lithography
TCE	Trichloroethylene
TF	Thin-film

Bibliography

[1] T. C. Harman, D. L. Spears, and M. J. Manfra. High thermoelectric figures of merit in pbte quantum wells. *Journal of Electronic Materials*, 25(7):1121–1127, 1996.

[2] S. Ohta, T. Nomura, H. Ohta, and K.o Koumoto. High-temperature carrier transport and thermoelectric properties of heavily La- or Nb-doped $SrTiO_3$ single crystals. *Journal of Applied Physics*, 97(3):034106, 2005.

[3] T. Okuda, K. Nakanishi, S. Miyasaka, and Y. Tokura. Large thermoelectric response of metallic perovskites: $Sr_{1-x}La_xTiO_3$. *Physical Review B*, 63(11):113104, 2001.

[4] J. P. Dismukes, L. Ekstrom, E. F. Steigmeier, I. Kudman, and D. S. Beers. Thermal and electrical properties of heavily doped Ge-Si alloys up to 1300 k. *Journal of Applied Physics*, 35(10):2899–2907, 1964.

[5] V.A. Kutasov, L.N. Lukyanova, and M.V. Vedernikov. Shifting the maximum figure-of-merit of $(Bi,Sb)_2(Te,Se)_3$ thermoelectrics to lower temperatures. In D.M. Rowe, editor, *Thermoelectrics Handbook: Macro to Nano*, pages 37–7. 2006.

[6] J. P. Heremans, V. Jovovic, E. S. Toberer, A. Saramat, K. Kurosaki, A. Charoenpakdee, S. Yamanaka, and G.J. Snyder. Enhancement of thermoelectric efficiency in PbTe by distortion of the electronic density of states. *Science*, 321:554, 2008.

[7] O.V. Emelyanenko, D.N. Nasledov, V.G. Sidorov, V.A. Skripkin, and G.N. Talalakin. Effective mass of electrons in n-GaAs. *Phys. Status Solidi*, 12(2):K93, 1965.

[8] G. Homm, P. J. Klar, J. Teubert, and W. Heimbrodt. Seebeck coefficients of n-type (Ga,In)(N,As), (B,Ga,In)As, and GaAs. *Applied Physics Letters*, 93:042107, 2008.

[9] D. L. Young, J. F. Geisz, and T. J. Coutts. Nitrogen-induced decrease of the electron effective mass in $GaAs_{1-x}N_x$ thin films measured by thermomagnetic transport phenomena. *Applied Physics Letters*, 82(8):1236–1238, 2003.

[10] T. Dannecker, Y. Jin, H. Cheng, C. F. Gorman, J. Buckeridge, C. Uher, S. Fahy, C. Kurdak, and R. S. Goldman. Nitrogen composition dependence of electron effective mass in $GaAs_{1-x}N_x$. *Physical Review B*, 82(12):125203, 2010.

[11] T. H. Geballe and G. W. Hull. Seebeck effect in silicon. *Physical Review*, 98(4):940, 1955.

[12] Y. Osamu. Effect of metal electrode on seebeck coefficient of p- and n-type Si thermoelectrics. *Journal of Applied Physics*, 95:178–183, 2004.

[13] G. Masetti, M. Severi, and S. Solmi. Modeling of carrier mobility against carrier concentration in Arsenic-, Phosphorus-, and Boron-doped Silicon. *Electron Devices, IEEE Transactions on*, 30(7):764–769, 1983.

[14] R. C. Jaeger. *Introduction to Microelectronic Fabrication*, volume V. Addison-Wesley Publishing Company, 1990.

[15] H. I. Liu, D. K. Biegelsen, N. M. Johnson, F. A. Ponce, and R. F. W. Pease. Self-limiting oxidation of Si nanowires. *Journal of vacuum Science & Technology B*, 11:2532–2537, 1993. Proceedings of the 16th international symposium on electron, ion, and photon beams 6.

[16] H. I. Liu, D. K. Biegelsen, F. A. Ponce, N. M. Johnson, and R. F. W. Pease. Self-limiting oxidation for fabricating sub-5 nm Silicon nanowires. *Applied Physics Letters*, 64(11):1383–1385, 1994.

[17] B. Pier-Francesco, F.and Caroline, C. Alain, H. Alexandre, E. Thomas, and R. Marc. Modeling stress retarded self-limiting oxidation of suspended Silicon nanowires for the development of Silicon nanowire-based nanodevices. *Journal of Applied Physics*, 110(3):033524, 2011.

[18] H. Alexandre, C. Jean-Philippe, B. Stephane, D. Cecilia, M.-A. Virginie, H. Jean-Michel, P. Sebastien, V. Christian, A. Francois, C. Catherine, D. Vincent, E. Thomas, and D. Simon. Oxidation of suspended stacked Silicon nanowire for sub-10nm cross-section shape optimization. *ECS Transactions*, 13(1):195–199, 2008.

[19] A. S. Grove, O. Leistiko, and C. T. Sah. Redistribution of acceptor and donor impurities during thermal oxidation of Silicon. *journal article*, 35:2695–2701, 1964.

[20] D. L. Rode and S. Knight. Electron transport in GaAs. *Physical Review B*, 3(8):2534, 1971.

[21] C. Skierbiszewski, P. Perlin, P. Wisniewski, T. Suski, W. Walukiewicz, W. Shan, J. W. Ager, E. E. Haller, J. F. Geisz, D. J. Friedman, J. M. Olson, and S. R. Kurtz. Effect of nitrogen-induced modification of the conduction band structure on electron transport in GaAsN alloys. *physica status solidi (b)*, 216(1):135–139, 1999.

[22] F. Masia, G. Pettinari, A. Polimeni, M. Felici, A. Miriametro, M. Capizzi, A. Lindsay, S. B. Healy, E. P. O'Reilly, A. Cristofoli, G. Bais, M. Piccin, S. Rubini, F. Martelli, A. Franciosi, P. J. Klar, K. Volz, and W. Stolz. Interaction between conduction band edge and nitrogen states probed by carrier effective-mass measurements in $GaAs_{1-x}N_x$. *Physical Review B*, 73(7):073201, 2006.

[23] W. Shan, W. Walukiewicz, J. W. Ager, E. E. Haller, J. F. Geisz, D. J. Friedman, J. M. Olson, and S. R. Kurtz. Band anticrossing in GaInNAs alloys. *Physical Review Letters*, 82(6):1221, 1999.

[24] U. Katsuhiro, M. Nobuki, and S. Ikuo. Reexamination of N composition dependence of coherently grown GaNAs band gap energy with high-resolution x-ray diffraction mapping measurements. *Applied Physics Letters*, 74(9):1254–1256, 1999.

[25] R. A. Serway. *Principles of Physics*. London: Saunders College Pub, Fort Worth, Texas, 2nd edition, 1998.

[26] D. L. Young, T. J. Coutts, V. I. Kaydanov, A. S. Gilmore, and W. P. Mulligan. Direct measurement of density-of-states effective mass and scattering parameter in transparent conducting oxides using second-order transport phenomena. *J. Vac. Sci. Technol. A.*, 18(6):2978, 2000.

[27] H. J. Goldsmid. *Applications of Thermoelectricity*. Butler & Tanner Ltd., 1960.

[28] D. K. C. MacDonald. *Thermoelectricity: An Introduction to the Principles*. Dover Publications, 2006.

[29] Tellurex Inc. Te, 2011. http://www.tellurex.com/technology/.

[30] Hi-Z Technology Inc. Thermoelectric modules, 2011. http://www.hi-z.com/products.php.

[31] R. E. Hummel. *Electronic Properties of Materials*. Springer, 2005.

[32] M. J. Kelly. *Low-Dimensional Semiconductors: Materials, Physics, Technology, Devices*. Oxford University Press, 1996.

[33] M. Lundstrom. *Fundamentals of carrier transport*, volume 2. Cambridge University Press, Cambridge, 2000.

[34] S. M. Sze and K. K. Ng. *Physics of Semiconductor Devices*. Wiley-Interscience, 2006.

[35] M. P. Vaughan and B. K. Ridley. The hall mobility in dilute nitrides. In Ayse Erol, editor, *Dilute III-V Nitride Semiconductors and Material Systems*, volume 105 of *Springer Series in Materials Science*, pages 255–281. Springer Berlin Heidelberg, 2008.

[36] N. Mohankumar and A. Natarajan. The accurate numerical evaluation of half-order fermi-dirac integrals. *Phys. stat. sol. (b)*, 188:635, 1995.

[37] F. Schaffler. *Silicon-Germanium ($Si_{1-x}Ge_x$) Material data for SiGe*, pages 149–186–. John Wiley & Sons, Inc, 2001.

[38] L. Friedman. Thermopower of superlattices as a probe of the density of states distribution. *Journal of Physics, C*, 17(22):3999, 1984.

[39] G. D. Mahan. *Good Thermoelectrics*, pages 82–157–. Academic Press, 1997.

[40] C. Wood. Materials for thermoelectric energy conversion. *Rep. Prog. Phys.*, 51:459–539, 1988.

[41] P. Pichanusakorn and P. Bandaru. The optimal seebeck coefficient for obtaining the maximum power factor in thermoelectrics. *Appl. Phys. Lett.*, 94:223108, 2009.

[42] R. Kim, S. Datta, and M. S. Lundstrom. Influence of dimenality on thermoelectric device performance. *Journal of Applied Physics*, 105:034506, 2009.

[43] L. D. Hicks, T. C. Harman, X. Sun, and M. S. Dresselhaus. Experimental study of the effect of quantum-well structures on the thermoelectric figure of merit. *Physical Review B*, 53(16):10493–10496, 1996.

[44] D. A. Broido and T. L. Reinecke. Theory of thermoelectric power factor in quantum well and quantum wire superlattices. *Physical Review B*, 64(4):45324, 2001.

[45] S. Ohta, T. Nomura, H. Ohta, H. Hirano, M.and Hosono, and K. Koumoto. Large thermoelectric performance of heavily Nb-doped $SrTiO_3$ epitaxial film at high temperature. *Applied Physics Letters*, 87(9):092108, 2005.

[46] R. Venkatasubramanian, E. Siivola, and T. S. Colpitts. In-plane thermoelectric properties of freestanding si/ge superlattice structures. In *17th International Conference on Thermoelectrics*, pages 191–197. Institute of Electrical and Electronics Engineers, 1998.

[47] X. Sun, S. B. Cronin, J. Liu, K. L. Wang, T. Koga, M. S. Dresselhaus, and G. Chen. Experimental study of the effect of the quantum well structures on the thermoelectric figure of merit in si/si1-xgex system. In *18th International Conference on Thermoelectrics*. IEEE, 1999.

[48] T. Koga, X. Sun, S. B. Cronin, and M. S. Dresselhaus. Carrier pocket engineering to design superior thermoelectric materials using GaAs/AlAs superlattices. *Applied Physics Letters*, 73:2950, 1998.

[49] X. Sun, S. B. Cronin, J. Liu, K. L. Wang, T. Koga, M. S. Dresselhaus, and G. Chen. Experimental study of the effect of the quantum well structures on the thermoelectric figure of merit in $Si/Si_{1-x}Ge_x$ system. In *18th International Conference on Thermoelectrics*. IEEE, 1999.

[50] L. D. Hicks and M. S. Dresselhaus. Effect of quantum-well structures on the thermoelectric figure of merit. *Physical Review B*, 47(19):12727–12731, 1993.

[51] L. D. Hicks and M. S. Dresselhaus. Thermoelectric figure of merit of a one-dimensional conductor. *Physical Review B*, 47(24):16631–16634, 1993.

[52] P. Pichanusakorn and P. Bandaru. Minimum length scales for enhancement of the power factor in thermoelectric nanostructures. *Journal of Applied Physics*, 107(7):074304–, 2010.

[53] T. Koga, X. Sun, S. B. Cronin, and M. S. Dresselhaus. Carrier pocket engineering applied to strained Si/Ge superlattices to design useful thermoelectric materials. *Applied Physics Letters*, 75:2438, 1999.

[54] A. I. Boukai, Y. Bunimovich, J. Tahir-Kheli, J.-K. Yu, W. A. Goddard III, and J. R. Heath. Silicon nanowires as efficient thermoelectric materials. *Nature*, 451:168–, 2008.

[55] Standard test methods for measuring resistivity and hall coefficient and determining hall mobility in single-crystal semiconductors. In *Annual Book of ASTM Standards*, pages 70–82. American Society of Testing and Materials, 1996.

[56] R. Chwang, B. J. Smith, and C. R. Crowell. Contact size effects on the van der pauw method for resistivity and hall coefficient measurement. *Solid State Electronics*, 17:1217–1227, 1974.

[57] D. W. Koon. Effect of contact size and placement, and of resistive inhomogeneities on van der Pauw measurements. *Rev. Sci. Instrum.*, 60(2):271–274, 1988.

[58] Y. Taur and T. H. Ning. *Fundamentals of Modern VLSI Devices*. Cambridge University Press, 9th edition, 2006.

[59] M. F. Ceiler, P. A. Kohl, and S. A. Bidstrup. Plasma-enhanced chemical vapor deposition of Silicon Dioxide deposited at low temperatures. *Journal of the Electrochemical Society*, 142(6):2067, 1995.

[60] K. W. Vogt, M. Houston, JR. M.F. Ceiler, C.E. Roberts, and P.A. Kohl. Improvement in dielectric properties of low temperature PECVD Silicon Dioxide by reaction with hydrazine. *Journal of Electronic Materials*, 24(6):751–755, 1995.

[61] R. Chen, A. I. Hochbaum, P. Murphy, J. Moore, P. Yang, and A. Majumdar. Thermal conductance of thin silicon nanowires. *Phys. Rev. Lett.*, 101:105501, Sep 2008.

[62] A. I. Hochbaum, R. Chen, R. D. Delgado, E.C. Liang, W.and Garnett, M. Najarian, A. Majumdar, and P. Yang. Enhanced thermoelectric performance of rough Silicon nanowires. *Nature*, 451(7175):163–167, 2008.

[63] X. Zhao, C. M. Wei, L. Yang, and M. Y. Chou. Quantum confinement and electronic properties of Silicon nanowires. *Physical Review Letters*, 92(23):236805–, 2004.

[64] K. Trivedi, H. Yuk, Herman C. F., M. J. Kim, and W. Hu. Quantum confinement induced performance enhancement in sub-5-nm lithographic Si nanowire transistors. *Nano Letters*, 11(4):1412–1417, 2011.

[65] M. T.Bjork, H. Schmid, J. Knoch, H. Riel, and W. Riess. Donor deactivation in Silicon nanostructures. *Nature Nanotechnology*, 4:103–107–, 2009.

[66] F. Deng, R. A. Johnson, P. M. Asbeck, S. S. Lau, W. B. Dubbelday, T. Hsiao, and J. Woo. Salicidation process using NiSi and its device application. *Journal of Applied Physics*, 81(12):8047–8051, 1997.

[67] R. B. Fair. Concentration profiles of diffuse dopants in Silicon. In F. F. Y. Wang, editor, *Impurity doping processes in Silicon*. North-Holland Pub. Co., New York, 1981.

[68] W. R. Thurber, R. L. Mattis, Y. M. Liu, and J. J. Filliben. The relationship between resistivity and dopant density for phosphorous and boron-doped silicon. Technical report, Department of Commerce, National Bureau of Standard, Department of Commerce, National Bureau of Standard, 1981.

[69] J. Ziegler. Stopping and range of ions in matter, 2011. www.SRIM.org.

[70] K. Shimakura, T. Suzuki, and Y. Yadoiwa. Boron and Phosphorus diffusion through an SiO2 layer from a doped polycrystalline Si source under various drive-in ambients. *Solid-State Electronics*, 18(11):991–996, 1975.

[71] J. Kedzierski and C. Bokor, J.and Kisielowski. Fabrication of planar Silicon nanowires on silicon-on-insulator using stress limited oxidation. *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures*, 15(6):2825–2828, 1997.

[72] B. E. Deal and A. S. Grove. General relationship for the thermal oxidation of Silicon. *Journal of Applied Physics*, 36(12):3770–3778, 1965.

[73] B. E. Deal and M. Sklar. Thermal oxidation of heavily doped Silicon. *Journal of The Electrochemical Society*, 112(4):430–435, 1965.

[74] J. Kedzierski, J. Bokor, and E. Anderson. Novel method for Silicon quantum wire transistor fabrication. In *Papers from the 43rd international conference on electron, ion, and photon beam technology and nanofabrication*, volume 17, pages 3244–3247, Marco Island, Florida (USA), 1999. AVS.

[75] N. Singh, A. Agarwal, L. K. Bera, T. Y. Liow, R. Yang, S. C. Rustagi, C. H. Tung, R. Kumar, G. Q. Lo, N. Balasubramanian, and D. L. Kwong. High-performance fully depleted silicon nanowire (diameter$\leq$5 nm) gate-all-around CMOS devices. *Electron Device Letters, IEEE*, 27(5):383–386, 2006.

[76] N. Fukata, S. Ishida, S. Yokono, R. Takiguchi, J. Chen, T. Sekiguchi, and K. Murakami. Segregation behaviors and radial distribution of dopant atoms in Silicon nanowires. *Nano Letters*, 11(2):651–656, 2011.

[77] X. Huang, W.-C. Lee, C. Kuo, D L. C. Hisamoto, J. Kedzierski, E. Anderson, H. Takeuchi, Y.-K. Choi, K. Asano, V. Subramanian, T.-J. King, J. Bokor, and C. Hu. Sub 50-nm FinFET: PMOS. In *Electron Devices Meeting, 1999. IEDM Technical Digest. International*, Washington, DC, 1999.

[78] X. Duan, C. Niu, V. Sahi, J. Chen, J. W. Parce, S. Empedocles, and J. L. Goldman. High-performance thin-film transistors using semiconductor nanowires and nanoribbons. *Nature*, 425(6955):274–278, 2003.

[79] V. Subramanian. Multiple gate field-effect transistors for future CMOS technologies. *IETE Tech Rev*, 27(6):446–454–, 2010.

[80] J.-H. Lee, J. Wu, and J. C. Grossman. Enhancing the thermoelectric power factor with highly mismatched isoelectronic doping. *Physical Review Letters*, 104(1):016602, 2010.

[81] M. Weyers, M. Sato, and H. Ando. Red shift of photoluminescence and absorption in dilute GaAsN alloy layers. *Jpn. J. Appl. Phys.*, 31:L 853, 1992.

[82] S. Fahy, A. Lindsay, H. Ouerdane, and E. P. O'Reilly. Alloy scattering of n -type carriers in GaN_xAs_{1-x}. *Physical Review B*, 74(3):035203, 2006.

[83] P. N. Hai, W. M. Chen, I. A. Buyanova, H. P. Xin, and C. W. Tu. Direct determination of electron effective mass in GaNAs/GaAs quantum wells. *Applied Physics Letters*, 77(12):1843–1845, 2000.

[84] F. Masia, A. Polimeni, G. Baldassarri Hoger von Hogersthal, M. Bissiri, M. Capizzi, P. J. Klar, and W. Stolz. Early manifestation of localization effects in diluted Ga(AsN). *Applied Physics Letters*, 82(25):4474–4476, 2003.

[85] S. Francoeur, M. J. Seong, A. Mascarenhas, S. Tixier, M. Adamcyk, and T. Tiedje. Band gap of $GaAs_{1-x}Bi_x$, 0¡x¡3.6%. *Applied Physics Letters*, 82(22):3874–3876, 2003.

[86] D. A. Beaton, R. B. Lewis, M. Masnadi-Shirazi, and T. Tiedje. Temperature dependence of hole mobility in $GaAs_{1-x}Bi_x$ alloys. *Journal of Applied Physics*, 108(8):083708, 2010.

[87] X. Sun, Z. Zhang, and M. S. Dresselhaus. Theoretical modeling of thermoelectricity in bi nanowires. *Applied Physics Letters*, 74(26):4005–4007, 1999.

[88] C. A. Hoffman, J. R. Meyer, F. J. Bartoli, A. Di Venere, X. J. Yi, C. L. Hou, H. C. Wang, J. B. Ketterson, and G. K. Wong. Semimetal-to-semiconductor transition in bismuth thin films. *Physical Review B*, 48(15):11431, 1993.

[89] T. Mattila, Su-Huai Wei, and Alex Zunger. Localization and anticrossing of electron levels in $GaAs_{1-x}N_x$ alloys. *Physical Review B*, 60(16):R11245, 1999.

[90] C. Skierbiszewski, P. Perlin, P. Wisniewski, W. Knap, T. Suski, W. Walukiewicz, W. Shan, K. M. Yu, J. W. Ager, E. E. Haller, J. F. Geisz, and J. M. Olson. Large, nitrogen-induced increase of the electron effective mass in $In_yGa_{1-y}N_xAs_{1-x}$. *Applied Physics Letters*, 76(17):2409–2411, 2000.

[91] P. R. C. Kent and A. Zunger. Theory of electronic structure evolution in GaAsN and GaPN alloys. *Physical Review B*, 64(11):115208, 2001.

[92] A. Lindsay and E. P. O'Reilly. Unification of the band anticrossing and cluster-state models of dilute nitride semiconductor alloys. *Physical Review Letters*, 93(19):196402, 2004.

[93] W. G. Bi, F. Deng, S. S. Lau, and C. W. Tu. High resolution x-ray diffraction studies of AlGaP grown by gas-source molecular-beam epitaxy. *Journal of Vacuum Science & Technology B*, 13:754–757, 1995.

[94] W. J. Fan, S. F. Yoon, T. K. Ng, S. Z. Wang, W. K. Loke, R. Liu, and A. Wee. Comparison of nitrogen compositions in the as-grown GaN_xAs_{1-x} on GaAs measured by high-resolution x-ray diffraction and secondary-ion mass spectroscopy. *Applied Physics Letters*, 80:4136–4138, 2002.

[95] Y. Tomooka, T.and Shoji and T. Matsui. High temperature vapor pressure of si. *Journal of the Mass Spectrometry Society of Japan*, 47(1):49–53, 1999.

[96] D. Lauvernier, S. Garidel, C. Legrand, and J.-P. Vilcot. Realization of sub-micron patterns on GaAs using a HSQ etching mask. *Microelectronic Engineering*, 77(3-4):210–216, 2005.

[97] T. Maeda, J. W. Lee, R. J. Shul, J. Han, J. Hong, E. S. Lambers, S. J. Pearton, C. R. Abernathy, and W. S. Hobson. Inductively coupled plasma etching of iii-v semiconductors in BCl_3-based chemistries: I. GaAs, GaN, GaP, GaSb and AlGaAs. *Applied Surface Science*, 143(1-4):174–182, 1999.

[98] M. Heiblum, M. I. Nathan, and C. A. Chang. Characteristics of AuGeNi ohmic contacts to GaAs. *Solid-State Electronics*, 25(3):185–195, 1982.

[99] K. Heime, U. Konig, E. Kohn, and A. Wortmann. Very low resistance Ni-AuGeNi contacts to n-GaAs. *Solid-State Electronics*, 17(8):835–837, 1974.

[100] S. Schuppler, D. L. Adler, L. N. Pfeiffer, K. W. West, E. E. Chaban, and P. H. Citrin. Can electrical deactivation of highly Si-doped GaAs be explained by autocompensation? *Applied Physics Letters*, 63(17):2357–2359, 1993.

[101] C. Domke, Ph Ebert, M. Heinrich, and K. Urban. Microscopic identification of the compensation mechanisms in Si-doped GaAs. *Physical Review B*, 54(15):10288, 1996.

[102] Y. G. Chai, R. Chow, and C. E. C. Wood. The effect of growth conditions on Si incorporation in molecular beam epitaxial GaAs. *Applied Physics Letters*, 39(10):800–803, 1981.

[103] J. Li, P. Carrier, S.-H. Wei, S.-S. Li, and J.-B. Xia. Mutual passivation of donors and isovalent nitrogen in GaAs. *Physical Review Letters*, 96(3):035505, 2006.

[104] A. Janotti, P. Reunchan, S. Limpijumnong, and C. G. Van de Walle. Mutual passivation of electrically active and isovalent impurities in dilute nitrides. *Physical Review Letters*, 100(4):045505, 2008.

[105] Y. Jin, Y. He, H. Cheng, R. M. Jock, T. Dannecker, M. Reason, A. M. Mintairov, C. Kurdak, J. L. Merz, and R. S. Goldman. Influence of Si–N complexes on the electronic properties of GaAsN alloys. *Applied Physics Letters*, 95(9):092109, 2009.

[106] K. M. Yu, W. Walukiewicz, J. Wu, D. E. Mars, D. R. Chamberlin, M. A. Scarpulla, O. D. Dubon, and J. F. Geisz. Mutual passivation of electrically active and isovalent impurities. *Nature Materials*, 1(3):185–189, 2002.

[107] E. F. Schubert, J. B. Stark, T. H. Chiu, and B. Tell. Diffusion of atomic Silicon in Gallium Arsenide. *Applied Physics Letters*, 53(4):293–295, 1988.

[108] S. Fahy and E. P. O'Reilly. Theory of electron mobility in dilute nitride semiconductors. *Physica E: Low-dimensional Systems and Nanostructures*, 21(2-4):881–885, 2004.

[109] A. Raymond, J. L. Robert, and C. Bernard. The electron effective mass in heavily doped GaAs. *Journal of Physics C: Solid State Physics*, 12(12):2289, 1979.

[110] Peter Y. Yu and Manuel Cardona. *Fundamentals of semiconductors:physics and materials properties*. Springer, 3rd edition, 2005.

[111] J. Gebauer, E. R. Weber, N. D. Jager, K. Urban, and Ebert Ph. Determination of the charge carrier compensation mechanism in Te-doped GaAs by scanning tunneling microscopy. *Applied Physics Letters*, 82(13):2059–2061, 2003.